# An Ensemble-based Weak-constraint 4DVar

Xiangjun Tian[1$], and Zhenghui Xie[1]

1 Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China (emails: tianxj@mail.iap.ac.cn; zxie@lasg.iap.ac.cn )

Submitted to *J. Geophys. Res.-Atmospheres*
Date: November 4, 2008

[$] *Corresponding author address*: X. Tian, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China; e-mail: tianxj@mail.iap.ac.cn

## ABSTRACT

One of the main hypothese made in variational data assimilation is to consider that the model is a strong constraint of the minimization, i.e. that the model describes exactly the behavior of the system. Obviously the hypothesis is never respected. A new approach is proposed in this paper by merging the Monte Carlo method and the proper orthogonal decomposition (POD) technique into the weak-constraint 4DVar to transform an implicit optimization problem into an explicit one, which can account for and estimate the flow-dependent model errors similar to that in the ensemble Kalman filter (EnKF). The Monte Carlo method is used to initiate an evolving forecast ensemble in a four-dimensional (4-D) space to cover the analysis state over each assimilation time window. The model errors are also represented by the evolving ensemble forecasts simultaneously. Since the 4-D analysis ensemble vectors are supposed to be in linear space, each of them can be expressed by a set of basis vectors of this space obtained through the POD technique, respectively. The 4DVar optimization problem is then resolved directly without an iterative procedure. Assimilation experiments in soil moisture assimilation show this new approach moderately outperforms another explicit strong-constraint 4DVar (referred to as ESC4DVar) method with assimilation errors can be reduced as only a fraction of the latter. Another assimilation experiment using the Lorenz model shows that it performs almost same as the ESC4DVar if the model is perfect.

## 1. Introduction

Strong constraint (perfect model assumption) 4DVar algorithms [*Johnson et al.*, 2006; *Kalnay et al.*, 2007; *Tsuyuki and Miyoshi*, 2007] are increasingly used for synoptic and global scale data assimilation at operational numerical weather prediction centers around the world. 4DVar takes into account several sources of information to produce an estimate of the forecast state at the analysis time. It is essential to transform the data assimilation problem into an optimization one, whose advantages mainly embody on the followings: 1) the physical model provides a whole dynamical constraint throughout the assimilation time window. And 2) it has the ability to assimilate the observational data at multiple times. While errors in observations and background state are accounted for, the numerical model representing the evolution of the state flow is assumed to be perfect, or at least to have errors that are negligible compared with others errors in the system.

As other aspects of the data-assimilation process have processed over the years, one might ask whether this assumption remains valid, and whether neglecting model error degrades the quality of the analysis and forecast. Is there any evidence of the presence of model error in the system, or is it still legitimate to neglect it? [*Trémolet*, 2006]

The data assimilation windows currently range from 6 hours to 12 hours at different centers. It is preferable to have as many independent observations as possible in each data assimilation window under the variational framework. Longer data assimilation windows generally increase the information content from the observations, but also make the perfect model assumption more improper [*Liang et al.*, 2007]. When model error is present, the model drifts away from the correct solution, and the discrepancy with observations increase with time, as explained, for example, by *Talagrand* [1998]. It is clear that weak constraint (imperfect model assumption) 4DVar algorithms will be required to properly combine the background forecast with high resolution observations in longer data assimilation windows in the not too distant future. There have been attempts to take model error into account in various data assimilation systems, particularly in the context of Kalman filter systems [*Dee*, 1995; *Dee and Da Silva*, 1998]. The ensemble Kalman filter (EnKF) [e.g., *Evensen*, 1994, 2003; *Kalnay et al.*, 2007; *Beezley and Mandel*, 2008] has become an increasingly popular method because of its simple conceptual formulation and relative ease of implementation. For example, it requires no derivation of a tangent linear operator or adjoint equations, and no integrations backward in time. Furthermore, by forecasting the statistical characteristics, EnKF can provide flow-dependent error estimates of the *background (model) errors* using the Monte Carlo method. Consequently, arguments on "which one is better, 4DVar or ensemble Kalman filter" [e.g., *Kalnay et al.*, 2007] appear a lot in debate due to the perfect model assumption in the strong-constraint 4DVar method, which forms a sharp contrast with that in the EnKF.

Weak-constraint 4DVar theory was firstly introduced by *Sasaki* [1970]. The main underlying idea is that, since the model's equations are not exact, it is sufficient to satisfy them only approximately: they can be imposed as a weak constraint in the optimization problem. Weak-constraint 4DVar has never been implemented fully with a realistic forecast model because of the computational cost and because of the lack of information to define the model error covariance matrix required to solve the problem [*Trémolet*, 2006, 2007]. However, even with important approximations, in the representation of model error itself, and of the model-error covariance matrix, good results have been obtained by several authors such as *Derber* [1998], *Wergen* [1992], *Zupanski* [1993] or *Bennett et al.* [1996] with atmospheric models and by *Vidard et al.* [2004] with an ocean model. *Trémolet* [2006, 2007] also discussed several formulations of weak-constraint 4DVar. These formulations were developed and evaluated at the European Centre for Medium Weather Forecasts (ECMWF).

How to represent the forecast model errors of the state flow appropriately needs to be addressed. Here we resort to the idea of the Monte Carlo method and the POD technique [*Ly and Tran*, 2001, 2002; *Volkwein*, 2008]: We merged the Monte Carlo method and the POD technique into the weak-constraint 4DVar to transform an implicit optimization problem into an explicit one. The basic idea of the POD technique is to start 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 base vectors which span the snapshot collection. The goal is to represent the ensemble of the data in terms of an *optimal* coordinate system. That is, the snapshots can be generated by a smallest possible set of base vectors. Based on this approach, an ensemble-based weak-constraint 4DVar method (referred to as EWC4DVar) is proposed in this paper: it begins with a 4-D ensemble obtained from the forecast ensembles at all times in an assimilation time window produced using the Monte Carlo method. The model errors are then represented by the evolving ensemble forecasts simultaneously. We then apply the POD technique to the 4-D forecast ensemble, so that the orthogonal base 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 base vectors obtained using the POD technique, the control variables in the weak-constraint cost function appear explicit, so that the adjoint or tangent linear model is no longer needed.

We conducted several numerical experiments using a one-dimensional (1-D) soil water equation and synthetic observations to evaluate our new method in land data assimilation. Comparisons were also made between our method and another ensemble-based strong-constraint 4DVar (referred to as ESC4DVar, [*Tian et al.*, 2008a,b]). We found that our new ensemble-based explicit weak-constraint 4DVar performs moderately better than the ESC4DVar in terms of increasing the assimilation precision. We also evaluate this approach using the Lorenz model (3D case), which shows the EWC4DVar performs almost same as the ESC4DVar if the forecast model is perfect.

## 2. Methodology

The observations of the forecast state represented by the vector $\vec{y}$ in observation space are one source of information about the state. An observation operator $H(\mathrm{x})$ represents knowledge of what the observations should be given the forecast state represented by the state variable $\vec{x}$. Errors in the observations and in the observation operator are assumed to be unbiased, Gaussian, and uncorrelated with other sources of error. They are characterized by their covariance matrix $R$.

A particular source of information available in meteorology is a prior estimate of the state of the system. In practice, in operational weather-forecasting centers, it is a forecast from the most recent analysis. This represents our prior knowledge about the state of the system without resorting to the current observations $\vec{y}$. The prior estimate of the mean of the state is represented by $\vec{x}_b$, and called the “background”. We assume that background error is unbiased and uncorrelated with other errors in the problem; it is characterized by the background-error covariance matrix $B$.

Another source of information about the system is theoretical knowledge, represented by the equation

$$F(\vec{x}) = 0\,. \tag{1}$$

In meteorological applications, $F$ can include the equations governing the evolution of the flow, as well as additional constraints, such as balance equations or prior knowledge about the state of the system. Errors in $F$ are assumed to be unbiased, Gaussian, and uncorrelated with other sources of error. They are characterized by their covariance matrix $C_f$.

Using these sources of information, four-dimensional variational data assimilation consists in minimizing the cost function:

$$\begin{aligned} J(\vec{x}) = {} & \frac{1}{2}(\vec{x}-\vec{x}_b)^T B^{-1}(\vec{x}-\vec{x}_b) + \frac{1}{2}(H(\vec{x})-\vec{y})^T R^{-1}(H(\vec{x})-\vec{y}) \\ & + \frac{1}{2}(H(\vec{x})-\vec{y}) R^{-1}(H(\vec{x})-\vec{y}) + \frac{1}{2}F(\vec{x})^T C_f^{-1} F(\vec{x})\,, \end{aligned} \tag{2}$$

The cost function can be interpreted as a weighted measure of the distance from the state $\vec{x}$ to the various available sources of information, either observational or theoretical. More details on this result are presented in, for example, *Jazwinski* [1970] or *Rodgers* [2000].

The components of $\vec{x}$ are the physical variable describing the forecast state (e.g., temperature, wind, humidity and surface pressure), discretized over the three spatial dimensions of the model's domain and the temporal dimension over the period for which observations are available. The assimilation window $[0,T]$ is discretized into $n+1$ time steps $\{t_i : i = 0, \cdots, n\}$. The state vector $\vec{x}_i$ represents the three-dimensional state of the atmosphere at time $t_i$. The observation operator will use the components of the state variable at the appropriate time to evaluate the observation term of the cost function, and will make accurate use of available observations.

In practice, approximations are necessary in order to solve the variational data assimilation problem. In operational variational data assimilation implementations, model error is assumed to be small enough to be neglected compared with initial-condition error, and the forecast model is imposed as a strong constraint. The state variable is a solution of the model equation:

$$\vec{x}_i = M_i(\vec{x}_{i-1}), \qquad (3)$$

where $M_i$ represents the model describing the evolution of the atmospheric flow from time $t_{i-1}$ to $t_i$. The evolution of the forecast state is then entirely determined by the initial condition $\vec{x}_0$, the control variable reduces to a three-dimensional state, and the constraint $F$ disappears from the cost function. This reduction of the control variable, combined with the adjoint technique to compute the gradient of the cost function (required by most minimization algorithms), was introduced by *Le Dimet and Talagrand* [1986], and is usually referred to as strong-constraint 4DVar or simple 4DVar. Although the time dimension of the information provided by the observations and the forecast model is taken into account, the control variable is defined over a three dimensional space. The size of the control variable, and the elimination of the model-error covariance matrix, make this algorithm operationally achievable with today's supercomputers.

A more general approach is to consider that the forecast model is not perfect. In such formulation, the forecast model is only imposed as a weak constraint, since the minimizing solution $\vec{x}$ does not have to be an exact solution of the model. This formulation is known as weak-constraint 4DVar. In this case, $C_f$ is the model error covariance matrix usually denoted by $Q$; the associated term in the cost function will denoted by $J_q$. A more complex introduction to the formulation of 4DVar accounting for an imperfect model is given in *Trémolet* [2006].

The weak-constraint 4DVar cost function in its most general form is defined by Eq. (2). It can be written more explicitly, as a function of the components of the control variable $\vec{x}$**,** as

$$J(\vec{x}) = (\vec{x}_0 - \vec{x}_b)^T B^{-1}(\vec{x}_0 - \vec{x}_b) + \sum_{i=0}^{m}(H_i(\vec{x}_i) - \vec{y}_i)^T R_i^{-1}(H(\vec{x}_i) - \vec{y}_i)$$

$$+ \sum_{i=0}^{m}(\vec{x}_i - M_{i,0}(\overrightarrow{x_0}))^T Q_i^{-1}(\vec{x}_i - M_{i,0}(\vec{x}_i)), \tag{6}$$

where $\vec{x}_i = M_{i,0}(\vec{x}_0)$ represents the state at time $t_i$ resulting from the forced model integrated from time $t_0$ to $t_i$, and observation and model errors are assumed uncorrelated in time. Time correlation can be taken into account, at the expense of using a non-block-diagonal model-error covariance matrix and determining the appropriate statistics.

The need of huge amount of information and then high computational costs to represent the model errors severely limits the implementation of the weak-constraint 4DVar. This problem is usually solved through some significant simplifications [*Courtier*,1997]. Obviously, such simplifications are likely subject to their poor description the evolution of the state flow. This issue is addressed in our new approach as follows.

Assuming there are $S$ time steps within the assimilation time window (*0, T*), generate $N$ random perturbation fields using the Monte-Carol method and add each perturbation field to the initial background field at $t = t_0$ to produce $N$ initial fields $\vec{x}_n(t_0), n = 1, 2, \cdots N$. Integrate the forecast model $\vec{x}_n(t_i) = M_{i,0}(\vec{x}_n(t_0))$ with the initial fields $\vec{x}_n(t_0)(n = 1, 2, \cdots N)$ throughout the assimilation time window to obtain the state series $\vec{x}_n(t_i)$ ( $i = 0, 1, \cdots S-1$) and then construct the perturbed 4-D fields (*snapshots*) $\vec{X}_n$ ( $n = 1, 2, \cdots N$ ) over the assimilation time window:

$$\vec{X}_n = (\vec{x}_n(t_0), \vec{x}_n(t_1), \cdots, \vec{x}_n(t_{S-1})), n = 1, 2, \cdots, N, \tag{7}$$

It is obvious that such vectors can capture the spatial structure of the model state and its temporal evolution. All the perturbed 4-D fields $\vec{X}_n$ ( $n = 1, 2, \cdots N$ ) can expand a finite dimensional space $\Omega(\overbrace{\vec{X}_1 \vec{X}_2 \cdots \vec{X}_N})$ . Similarly, the analysis field $\vec{x}_a(t_i)(i = 0, 1, 2, \cdots S-1)$ over the same assimilation time window can also be stored into the following vector:

$$\vec{X}_a = (\vec{x}_a(t_0), \vec{x}_a(t_1), \cdots, \vec{x}_a(t_{S-1})), n = 1, 2, \cdots, N. \tag{8}$$

When the ensemble size $N$ is increased by adding random samples, the ensemble space could cover the analysis vector $\overrightarrow{X}_a$, i.e. $\overrightarrow{X}_a$ is approximately assumed to be embedded in the linear space $\Omega(\overbrace{\overrightarrow{X}_1\overrightarrow{X}_2\cdots\overrightarrow{X}_N})$. Let $\overrightarrow{X}_{bn}$ ($n=1,2,\cdots K, K\le N$) be the base vectors of this linear space $\Omega(\overbrace{\overrightarrow{X}_1\overrightarrow{X}_2\cdots\overrightarrow{X}_N})$, the analysis vector $\overrightarrow{X}_a$ can be expressed by the linear combinations of this set of base vectors since it is in this space, i.e.

$$\overrightarrow{X}_a=\sum_{n=1}^{K}\beta_n\overrightarrow{X}_{bn}\ . \tag{9}$$

Setting $M_{i,0}(\vec{x}(t_0))=\frac{1}{N}\sum_{n=1}^{N}\vec{x}_n(t_i)$ and then substituting (8) and (9) into (6), the control variable becomes $\beta=(\beta_1\cdots\beta_K)^T$ instead of $\vec{x}(t_0)$ if the model error covariances $Q_i$ ($i=0,\cdots,m$) are known **(**This will be discussed further below**)**, so the control variable is expressed explicitly in the cost function and the computation of the gradient is simplified greatly. The tangent linear model or adjoint model is no longer required. To minimize the cost function, Eq. (6) is transformed into an explicit optimization problem with the variable vector $\beta=(\beta_1\cdots\beta_K)^T$.

*Tian and Xie* [2008a,b] proposed a concept of sample density to illustrate that the vector transformation $\delta X_n=\overrightarrow{X}_n-\overrightarrow{X}, n=1,\cdots,N$ is the optimization one in certain optimal sense, which can obtain the maximum sample density for the same ensemble forecasts and then yield the most efficient assimilation effects. That means any other vector transformation such as $\delta X_{ni}=\overrightarrow{X}_n-\overrightarrow{X}_i$, $n=1,\cdots,N$, $(\forall\overrightarrow{X}_i\in(\overrightarrow{X}_1,\overrightarrow{X}_2,\cdots,\overrightarrow{X}_N))$ can only result in some analysis vector partly close to the optimization analysis vector, whose relationship is very similar to that between the analysis ensemble and the mean analysis in the EnKF. Inspired by this similarity, we form $N$ new ensembles by focusing on deviations from the vector $\overrightarrow{X}_i, (i=1,\cdots,N)$, respectively, as follows

$$\delta X_{ni} = \overrightarrow{X}_n - \overrightarrow{X}_i, \tag{10}$$

which form the matrix $A_i$ ($M \times N$), where $M = M_g \times M_v \times S$, and $M_g, M_v$ are the number of the model spatial grid points and the number of the model variables respectively. To compute the POD modes, one must solve an $M \times M$ eigenvalue problem:

$$(A_i A_i^T)_{M\times M} V = \lambda V, \tag{11}$$

In practice, the direct solution of this eigenvalue problem is often not feasible if $M >> N$, which occurs often in numerical models. We can transform it into an $N \times N$ eigenvalue problem through the following transformations:

$$A_i^T A_i = \lambda V, \tag{12a}$$

$$A_i A_i^T A_i V = A_i \lambda V, \tag{12b}$$

$$A_i A_i^T A_i V = A_i \lambda V, \tag{12c}$$

$$A_i A_i^T (A_i V) = \lambda (A_i V), \tag{12d}$$

In the method of snapshots, one then solves the $N \times N$ eigenvalue problem.

$$TV_k = \lambda_k V_k, k = 1, \cdots N, \tag{13}$$

where $T = (A_i^T A_i)_{N\times N}$, $V_k$ is the $k$ th column vector of $V$ and $\lambda_k$ is the $k$ th row vector of $\lambda$ . The nonzero eigenvectors $\lambda_k$ ($1 \le k \le N$) may be chosen to be orthonormal, and the POD modes are given by $\phi_k = A_i V_k / \sqrt{\lambda_k}$ ,($1 \le k \le N$).

The truncated reconstruction of analysis variable in the four dimensional space $\overrightarrow{X}_a^i$ is given by

$$\overrightarrow{X}_a^i = \overrightarrow{X}_i + \sum_{j=1}^{P_i} \alpha_j^i \phi_j^i \ , \tag{14}$$

where $P_i$ (the number of the POD modes) is defined as follows

$$P_i = \min\left\{ P_i, I(P_i) = \frac{\sum_{j=1}^{P_i} \lambda_j}{\sum_{j=1}^{N} \lambda_j} : I(P_i) \geq \gamma \right\}, 0 < \gamma < 1. \tag{15}$$

Given the vector of measurements $Y = \left(\vec{y}_0, \vec{y}_1, \cdots, \vec{y}_m\right)^T$ , we can define the $N$ vectors with perturbed observations as

$$Y_i = Y + \mathrm{E}_i, \ i = 1, \cdots N \ , \tag{16}$$

where $\mathrm{E}_i = \left(\varepsilon_{i,0}, \varepsilon_{i,1}, \cdots, \varepsilon_{i,m}\right)^T$ are random real vectors. The measurement error covariance matrix can be estimated by

$$R_j = \frac{E_j E_j^T}{N-1}, \quad j = 0, \cdots m, \tag{17}$$

where $E_j = \left(\varepsilon_{1,j}, \cdots, \varepsilon_{N,j}\right)$.

Subsequently, one can construct the model error covariance $Q_i$ as follows: The ensemble matrix at time $t_i$ is constructed by

$$A_i = (\vec{x}_1(t_i), \cdots, \vec{x}_n(t_i)) \ , \tag{17}$$

The ensemble perturbation can be defined as

$$\Delta A_i = (\vec{x}_1(t_i) - \overline{x}(t_i), \cdots, \vec{x}_N - \overline{x}(t_i)) \ , \tag{18}$$

where $\overline{x}(t_i) = M_{i,0}(\vec{x}(t_0)) = \frac{1}{N}\sum_{n=1}^{N}\vec{x}_n(t_i)$.

Then the model error covariance can be represented same as that in EnKF

$$Q_i = \frac{\Delta A_i (\Delta A_i)^T}{N-1}, \tag{19}$$

The SVD of $\Delta A_i$ yields

$$\Delta A_i = U_i \Lambda_i V_i^T, \tag{20}$$

where $\Lambda_i$ is a diagonal matrix composed of the singular values of $\Delta A_i$. $U_i$ and $V_i$ are orthogonal matrices composed of the left and right singular vectors of $\Delta A_i$, respectively, then

$$Q_i = \frac{U_i \Lambda_i^2 U_i^T}{N-1}, \tag{21}$$

and

$$Q_i^{-1} = (N-1) U_i \Lambda_i^{-2} U_i^T, \tag{22}$$

Substituting (14), (17) and (22) into (6), the control variable becomes $\alpha^i = (\alpha_1^i, \cdots, \alpha_{P_i}^i)^T$ instead of $\vec{x}(t_0)$ and then the analysis vector $\overline{X}_a^i$ ($i = 1, \cdots, N$) can be easily obtained. The mean analysis state is then generated as follows:

$$\overrightarrow{X}_a = \frac{1}{N}\sum_{i=1}^{N}\overrightarrow{X}_a^i, \tag{23}$$

Similarly, the ensemble initial $A_0$ for next assimilation cycle is then constructed by

$$A_0 = (\vec{x}_a^1(t_{S-1}), \cdots, \vec{x}_a^N(t_{S-1})), \tag{24}$$

and the background error covariance $B$ can be updated by the evolving analysis ensemble forecasts (so it is flow-dependent) as follows

$$\Delta A_0 = (\vec{x}_a^{\,1}(t_{S-1}) - \bar{x}_a^{\,*}(t_{S-1}), \cdots, \vec{x}_a^{\,N} - \bar{x}_a^{\,*}(t_{S-1})), \tag{25}$$

where $\bar{x}_a^{\,*}(t_{S-1}) = \frac{1}{N}\sum_{n=1}^{N} \vec{x}_a^{\,n}(t_{S-1})$.

$$B = \frac{\Delta A_0 (\Delta A_0)^T}{N-1}, \tag{26}$$

The SVD of $\Delta A_0$ yields

$$\Delta A_0 = U_0 \Lambda_0 V_0^{\,T}, \tag{27}$$

where $\Lambda_0$ is a diagonal matrix composed of the singular values of $\Delta A_0$. $U_0$ and $V_0$ are orthogonal matrices composed of the left and right singular vectors of $\Delta A_0$, respectively, then

$$B = \frac{U_0 \Lambda_0^{\,2} U_0^{\,T}}{N-1}, \tag{28}$$

and

$$B^{-1} = (N-1) U_0 \Lambda_0^{\,-2} U_0^{\,T}. \tag{29}$$

Eqs. (24) and (29) are used to drive next assimilation cycle, which indicates that the initial condition is perturbed only once throughout the whole assimilation in this new scheme formulation.

In the above formulations, the usual optimization algorithms to find the solution of $\alpha^i = (\alpha_1^i, \cdots, \alpha_{P_i}^i)^T$ still need the iterative procedure and probably result in higher computational cost. This issue is addressed as follows:

Form the POD mode matrix

$$\Phi^i = \left(\phi_1^i, \phi_2^i, \cdots, \phi_{P_i}^i\right), \tag{30}$$

where, $\phi_j^i = \left(\phi_j^i(t_0), \phi_j^i(t_1), \cdots, \phi_j^i(t_{S-1})\right)^T$ , $j = 1, 2, \cdots, P_i$ . Transform (30) into the following format

$$\Phi^i = \left(\Phi_0^i, \Phi_1^i, \cdots, \Phi_{S-1}^i\right)^T , \tag{31}$$

where $\Phi_k^i = \left(\phi_1^i(t_k), \quad \phi_2^i(t_k), \quad \cdots, \quad \phi_{P_i}^i(t_k)\right), k = 0, 1, \cdots, S-1$.

Eq. (14) is rewritten as follow:

$$\overrightarrow{X}_a^i = \overrightarrow{X}_i + \Phi^i \alpha^i , \tag{32}$$

where $\alpha^i = \left(\alpha_1^i, \alpha_2^i, \cdots, \alpha_{P_i}^i\right)^T$.

The cost function (14) can be transformed into the following

$$\begin{aligned} J(\alpha^i) = {} & (\vec{x}_i(t_0) - \vec{x}_b + \Phi_0^i \alpha^i) B^{-1} (\vec{x}_i(t_0) + \Phi_0^i \alpha^i - \vec{x}_b) \\ & + \sum_{j=0}^{m} \left[\vec{y}_j - \mathrm{H}\vec{x}_i(t_j) - \mathrm{H}_j \Phi_j^i \alpha^i\right]^T R_j^{-1} \left[\vec{y}_j - \mathrm{H}_j \vec{x}_i(t_j) - \mathrm{H}_j \Phi_j^i \alpha^i\right] \\ & + \sum_{j=0}^{m} \left[\overline{x}(t_j) - \vec{x}_i(t_j) - \Phi_j^i \alpha^i\right]^T \mathrm{Q}_j^{-1} \left[\overline{x}(t_j) - \vec{x}_i(t_j) - \Phi_j^i \alpha^i\right], \end{aligned} \tag{33}$$

where $\mathrm{H}_j$ is the tangent linear observation operator.

Because $\mathrm{R}_j^{-1}$ and $\mathrm{Q}_j^{-1}$ are symmetrical (see (17,22)), we can obtain the gradient of the cost function through simple calculations:

$$\begin{aligned} \nabla J(\alpha^i) = {} & (\Phi_0^i)^T B^{-1} (\vec{x}_i(t_0) - \vec{x}_b + \Phi_0^i \alpha^i) + \sum_{j=0}^{m} -\left[\mathrm{H}_j \Phi_j^i\right]^T \mathrm{R}_j^{-1} \left[\vec{y}_j - \mathrm{H}_j \vec{x}_i(t_j) - \mathrm{H}_j \Phi_j^i \alpha^i\right] \\ & + \sum_{j=0}^{m} -\left[\Phi_j^i\right]^T \mathrm{Q}_j^{-1} \left[\overline{x}(t_j) - \vec{x}_i(t_j) - \Phi_j^i \alpha^i\right], \end{aligned} \tag{34}$$

One can solve the optimization problem

$$\nabla J_i(\alpha^i) = 0 , \tag{35}$$

and

$$\left((\Phi_0^i)^T \mathrm{B}^{-1}\Phi_0^i + \sum_{j=0}^{m}\left[\mathrm{H}_j\Phi_j^i\right]^T \mathrm{R}_j^{-1}\left[\mathrm{H}_j\Phi_j^i\right] + \sum_{j=0}^{m}\left[\Phi_j^i\right]^T \mathrm{Q}_j^{-1}\left[\Phi_j^i\right]\right)\alpha^i$$

$$= \sum_{j=0}^{m}\left[\mathrm{H}_j\Phi_j^i\right]^T \mathrm{R}_j^{-1}\left[\vec{y}_j - \mathrm{H}_j\vec{x}_i(t_j)\right] + \sum_{j=0}^{m}\left[\Phi_j^i\right]^T \mathrm{Q}_j^{-1}\left[\overline{x}(t_j) - \vec{x}_i(t_j)\right] - (\Phi_0^i)^T \mathrm{B}^{-1}(\vec{x}_i(t_0) - \vec{x}_b) \tag{36}$$

Eq. (36) can be solved directly without an iterative procedure.

## 3. Evaluations in a 1-D soil water model

In this section, the applicability of this new method is evaluated through several assimilation experiments with a simple 1-D soil water equation model used in the NCAR Community Land Model (CLM) [*Oleson et al.*, 2004]. Since we have compared the ESC4DVAR with the usual strong-constraint 4DVar, the EnKF and another explicit strong-constraint 4DVar [*Qiu et al.*, 2007] in *Tian et al.* [2008b] and found that the ESC4DVar's performance is superior to the others' involved in term of increasing the assimilation precision, it is enough for us to only compare the EWC4DVar method with the ESC4DVar.

### 3.1. Set-up of experiments

The volumetric soil moisture ($\theta$) for 1-D vertical water flow in a soil column in the CLM is expressed as

$$\frac{\partial\theta}{\partial t} = -\frac{\partial q}{\partial z} - E - R_{fm}, \tag{37}$$

where $q$ is the vertical soil water flux , $E$ is the evapotranspiration rate, and $R_{fm}$ is the melting (negative) or freezing (positive) rate, 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 [*Darcy*, 1856]:

$$q = -k\frac{\partial(\varphi + z)}{\partial z}, \tag{38}$$

where $k = k_s\left(\frac{\theta}{\theta_s}\right)^{2b+3}$ is the hydraulic conductivity, and $\varphi = \varphi_s\left(\frac{\theta}{\theta_s}\right)^{-b}$ is the soil matric potential , $k_s, \varphi_s, \theta_s$ and $b$ are constants. The CLM computes soil water content in the 10 soil layers through (37-38) (see [*Oleson et al.*, 2004] for details). The upper boundary condition is

$$q_0(t) = -k\frac{\partial(\varphi + z)}{\partial z}\bigg|_{z=0}, \tag{38b}$$

where $q_0(t)$ is the water flux at the land surface (referred to as infiltration), and the lower boundary condition is $q_l = 0$. The time step $\Delta t$ is 1800 s (0.5 hour).

We took a site at (47.43$^{\circ}$N, 126.97$^{\circ}$E) as the experimental site. The soil parameters $k_s, \varphi_s, \theta_s$ and $b$ at this site were calculated by the CLM using the high-resolution soil texture data released with the CLM by NCAR: $\theta_s$ =0.46m$^3$/m$^3$, $k_s$ =2.07263E-6 m/s, $b$ =8.634, $\varphi_s$ =-3.6779m**.** We then ran the model at the site forced with observation-based 3-hourly forcing data [*Qian et al.,* 2006; *Tian et al.*, 2007] from January 1, 1992 to December 31, 1993 after ten-year spinning-up to obtain a two-year time series of simulated infiltration (i.e., the water flux *q* at the surface, c.f., Eq.(38b)) for driving the soil water hydrodynamic equation (24). We used the first year (January 1, 1992 to December 31, 1992) data of CLM-simulated infiltration as the "perfect" infiltration series, and took the second year data as the "imperfect" infiltration series (Fig. 1). In our experiments, we integrated the soil water hydrodynamic equation (37) forced by the two infiltration time series for 365 days separately:

Eq. (37) forced by the “perfect” infiltration series represents the perfect *forecast model*, whose forecast error comes only from the noise in the initial (soil moisture) field; on the contrary, Eq. (37) forced by the “imperfect” infiltration series acts as the “imperfect” forecast model, whose forecast error comes from not only the noise of the initial field but also the uncertainty in the forecast model itself.

Figure 2 shows the “imperfect” and the “perfect” initial soil moisture profiles, which were obtained by randomly taking two arbitrary CLM-simulated soil moisture profiles in the process of the infiltration series producing. These profiles represent the initial fields with and without noise. The “perfect” (or “true”) state was produced by integrating the “perfect” model with the “perfect” initial soil moisture profile for 365 days (Figure 3). The “observations” were generated by adding 3% random error perturbations to the time series of the “perfect” state (i.e., “observation” = $(1+\varepsilon)\times$“perfect”, where $\varepsilon$ is a real random number varying from -3% to 3%), and these “observations” were assimilated using the two methods in the assimilation experiments (but not in the forecast experiments). In addition, a forecast states were produced by integrating the imperfect model with the “imperfect” initial soil moisture: The forecast error comes from both the noise of the initial field and the uncertainty in the forecast model (Figure 3), which shows that the forecast model drifts seriously away from the “perfect” solution. Two observation frequencies (twelve-hourly and two-hourly) are used to test their sensitivity on the assimilation effects.

**3.2. Experimental results**

To evaluate the performance of the two algorithms (ESC4DVar, EWC4DVar), a relative error is defined as follows

$$E_{t_{0\to S-1}} = \frac{\sum_{i=0}^{S-1}\sum_{j=1}^{M_g \times M_v} (\vec{x}_j^{\,a}(t_i) - \vec{x}_j^{\,t}(t_i))^2}{\sum_{i=0}^{S-1}\sum_{j=1}^{M_g \times M_v} (\vec{x}_j^{\,f}(t_i) - \vec{x}_j^{\,t}(t_i))^2}, \tag{39}$$

where the index $t_{0\to S-1}$ denotes an assimilation time window (one day in our experiments), $S$ is the length of an assimilation window ($S$ =48 in our experiments), $f$ and $a$ denote the forecast state (without assimilation of the "observations") and the analysis state, respectively, $t$ represents the "true" ("perfect") state. Thus, a relative error of 1% for a given assimilation method would mean that the mean error of the analyzed soil moisture is only 1% of that in the forecast case.

Figure 4a shows that the EWC4DVar method performs moderately better than the ESC4DVar: the relative errors of the ESCW4DVar for the analyzed soil moisture are all lower than12% and most of them are even lower than 3%. However, the relative errors of the ESC4DVar for the analyzed soil moisture fluctuate between 0 and 18%, which are higher than the EWC4DVar's as a whole, especially during Day 100 to Day 200. This is expected because model error is not negligible in such data assimilation: The pure simulated (Im with Im) deviates from the true (P with P) apparently during Day 100 to Day 200 (Fig.3). The perfect model assumption in the ESC4DVar introduces larger errors and leads to sub-optimal performance. With the observation frequency being increased, there is so much observation information merged into the analyzed soil moisture that the relative errors of the

EWC4DVar's become very small (<2.0%)(Fig.4b). The relative errors of the ESC4DVar are not reduced as so much as the EWC4DVar: Some of are even up to 8%.

## 4. Evaluations within the Lorenz model

In this section, our approach (EWC4DVar) is further evaluated within the Lorenz model for investigating its wider applications. The Lorenz model is widely used to test the new proposed methods in data assimilation community: e.g. *Xiong*, *Navon and Uzunoglu* [2006] used it to test the performances of the EnKF and PF (particle filter) methods. Their results show that the PFGR (PF with Gaussian resampling) method possesses good stability and accuracy and is potentially applicable to large-scale data assimilation problems.

### 4.1. Set-up of experiments

The Lorenz system under chaotic regime is used as a test problem, which is given by equation(e.g., see http://www.taygeta.com/perturb/node2.html):

$$\frac{dx}{dt} = -s(x-y), \tag{40a}$$

$$\frac{dy}{dt} = rx - y - xz, \tag{40b}$$

$$\frac{dz}{dt} = xy - bz, \tag{40c}$$

For numerical experiment the Lorenz system with parameters $s=10, r=28, b=\frac{8}{3}$ was integrated using a second order Runge Kuatta's method, with $\Delta t = 0.1$, and initial conditions $x(0)=-1.5$, $y(0)=-1.5$, $z(0)=25$ for the true solution (observations) and $x(0)=-1.52$, $y(0)=-1.3$, $z(0)=27$ for background solution (a-priori forecast). The observation insertion is done at each 12 time-step. The length of each assimilation time window is 24 time-step.

### 4.2. Experimental results

Figure 5 shows time series of the Lorenz curve coordinates (x,y,z) from observations, the EWC4Dvar and ESC4Dvar assimilations: the forecast Lorenz curve is adjusted to approach the true curve rapidly at the end of the first assimilation cycle by the EWC4DVar method, even though only twice observations in each assimilation time window. On the contrary, the pure forecast state without assimilations begins to deviate from the true solution seriously after 60 time-step or so, even though the noise of the initial filed (x,y,z) only results in small departures from the true state in the first 48 time steps or two assimilation time windows (not shown). Figure 6 shows the root mean square (rms) errors for the EWC4Dvar assimilated Lorenz curve are mostly less than 4 in the first assimilation window, and become close to zero at the start of the second assimilation cycle. On the other hand, the rms errors for the simulated curve fluctuate drastically in the magnitude from 1 to 30 (not shown). The ESC4DVar method was also applied in the same experiments. Because the forecast model (the Lorenz model) used in this experiments is perfect and the forecast errors come only from the noise of the initial fields, the EWC4DVar method with consideration of model errors doesn't show superior performance compared with the ESC4DVar method: The two methods performs almost same during this assimilation experiments.

## 5. Summary and concluding remarks

Weak-constraint 4DVar is a generalization of the more widely developed strong-constraint 4DVar: In weak-constraint 4DVar one simplifying assumption—namely, that the forecast model is perfect—has been removed. A new approach is proposed in this paper by merging the Monte Carlo method and the POD technique into the weak-constraint 4DVar formulation to transform it into an implicit optimization problem, which can account for and estimate model error similar to that in the EnKF. The model errors are then represented by the evolving ensemble forecasts.

Assimilation experiments in soil moisture assimilation show this new approach moderately outperforms the strong-constraint 4DVar method with assimilation errors can be reduced only a fraction of the latter, which shows whether considering model error or not in data assimilation plays some role that can not be easily ignored. Another assimilation experiment conducted within the Lorenz model shows that it performs almost same as the usual strong-constraint 4DVar method if the model is perfect.

It should be pointed out that the additional computational costs resulting from representing model error in the proposed method could possibly limit its further operational applications, even though it is not very obvious our experiments. How to reduce the computational costs as much as possible is a critical step in using this method. This aspect requires more evaluations and investigations.

*Acknowledgments*. This work was supported by the National Natural Science Foundation of China under grant 40705035, the Knowledge Innovation Project of Chinese Academy of Sciences under grants KZCX2-YW-217 and KZCX2-YW-126-2, the National Basic Research Program under grant 2005CB321704, and the Chinese COPES project (GYHY200706005).

REFERENCES

Beezley, J. D., and J. Mandel (2008), Morphing ensemble Kalman filters, *Tellus A, 60(1)*,131—140

Bennett A., B. Chua and L. Leslie (1996), Generalized inversion of a global numerical weather prediction model. *Meteorol. Atmos. Phys., 60*, 165—178

Courtier P. (1997), Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc. 123*, 2449—2461

Darcy, H. (1856), *Les Fontaines Publiques de la Ville de Dijon,* Dalmont, Paris.

Dee D. (1995), On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Weather Rev. 123*, 1128—1145

Dee D. and A. Da Silva (1998), Data assimilation in the presence of forecast bias. *Q.J.R.Meteorol. Soc. 124*, 1783—1807

Derber J. (1989), A variational continuous assimilation technique. *Mon. Weather Rev. 117*, 2437—2446

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

Evensen, G. (2003), The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics, 53*, 343—367

Jazwinski A. H. (1970), *Stochastic Processes and Filtering Theory*. Academic Press.

Johnson, C., B. J. Hoskins, N. K. Nichols, and S. P. Ballard (2006), A singular vector perspective of 4DVAR: The spatial structure and evolution of baroclinic weather systems, *Mon. Wea. Rev.,134(11)*, 3436—3455

Kalnay, E., H. Li, T. Miyoshi, S. C. Yang, and J. Ballabrera-Poy (2007), 4-D-Var or ensemble Kalman filter?, *Tellus A*, *59(5)*, 758—773 .

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

Liang X., R. Thomas, G. James, and C. Boon (2007), Toward a weak constraint operational 4D-Var system: application to the Burgers' equation, *Meteorologische Zeitschrift*, *16(6)*, 741—753, doi: 10.1023/A:1011452303647

Ly, H. V., and H.T. Tran (2001), Modeling and control of physical processes using proper orthogonal decomposition, *Mathematical and Computer Modeling*, *33*, 223—236

Ly, H. V., and H.T. Tran (2002), Proper orthogonal decomposition for flow calculations and optimal control in a horizontal CVD reactor, *Quarterly of Applied Mathematics*, *60(3)*,631—656

Oleson K. W. et al. (2004), *Technical description of the community land model (CLM)*, NCAR/TN-461+STR, 186pp

Qian, T., A. Dai, K. E. Trenberth and K. W. Oleson (2006), Simulation of global land surface conditions from 1948 to 2004. Part I: Forcing data and evaluations, *J. Hydrometeor., 7*, 953—975

Qiu, C. J., L. Zhang and A. M. Shao(2007), An explicit four-dimensional variational data assimilation method, *Science in China (D)*, *50(8)*, 1232—1240

Rodgers C. (2000), *Inverse Methods for Atmospheric Sounding*. World Scientific.

Sasaki Y. (1970), Some basic formalisms in numerical variaitonal analysis. *Mon. Weather Rev. 98*, 875—883

Talagrand O. (1998), ‘A posterior evaluation and verification of analysis and assimilation algorithms’. In: *Workshop on Diagnosis of Data Assimilation Systems*, ECMWF: 407—409

Tian X. and Z. Xie (2008a), Accounting for flow-dependence in the background error covariance within an ensemble-based explicit four-dimensional variaitonal assimilation method, *Water Resour. Res.,* revised

Tian X., Z. Xie and A. Dai (2008b), An ensemble-based explicit four-dimensional variational assimilation method, *J. Geophys. Res.*, in press.

Tian, X., A. Dai, D. Yang and Z.Xie (2007), Effects of precipitation-bias corrections on surface hydrology over northern latitudes, *J. Geophys. Res.*, *112*, D14101, doi:10. 029/2007JD008420.

Trémolet Y. (2006), Accounting for an imperfect model in 4D-Var, *Q. J. R. Meteorol. Soc.*, *321*, 2483—2504

Trémolet Y.(2007), Model-error estimation in 4D-Var, *Q.J.R. Meteorol. Soc.*, *133*, 1267—1280

Tsuyuki, T., and T. Miyoshi (2007). Recent progress of data assimilation methods in meteorology, *Journal of the Meteorological Society of Japan*, *85B*, 331-361.

Vidard, P. A., A. Piacentini and F. X. Le Dimet (2004), Variational data analysis with control of the forecast bias, *Tellus, 56A*, 177—178

Volkwein, S. (2008), Model Reduction using Proper Orthogonal Decomposition., Script in English language, 41 pages, available as a PDF-file. Available from: http://www.uni-graz.at/imawww/volkwein/publist.html

Wergen, W. (1992), The effect of model errors in variational assimilation. *Tellus, 44A*, 297—313

Xiong X., I. M. Navon and B. Uzunoglo (2006), A note on the particle filter with posterior Gaussian resampling. *Tellus A*, 58A, 456—460

Zupanski, M. (1993), Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Weather Rev.*, 121, 2396—2408

FIGURE CAPTIONS:

**FIG. 1.** The “perfect” (solid line) and “imperfect” (dashed line) infiltration time series used in the assimilation experiments.

**FIG. 2**. The “perfect” (solid line) and “imperfect” (dashed line) initial soil moisture profiles used in the assimilation experiments.

**FIG. 3**. Time series of skip volumetric soil moisture simulated by the perfect model with the “perfect” initial soil moisture, and the imperfect model with the “imperfect” initial soil moisture.

**FIG. 4.** Relative error ( $E_n$ ) for analyzed soil moisture in the assimilation experiments by the imperfect model with the “imperfect” initial field.

**FIG.5**. Time series of the Lorenz curve coordinates (x,y,z) from observations (solid line), the EWC4DVar assimilations (long-dashed line) and the ESC4DVar assimilations (short-dashed line)

**FIG.6**. Root mean square error for the EWC4Dvar assimilated (solid line) or the ESC4DVar assimilated (long-dashed line) Lorenz curves.

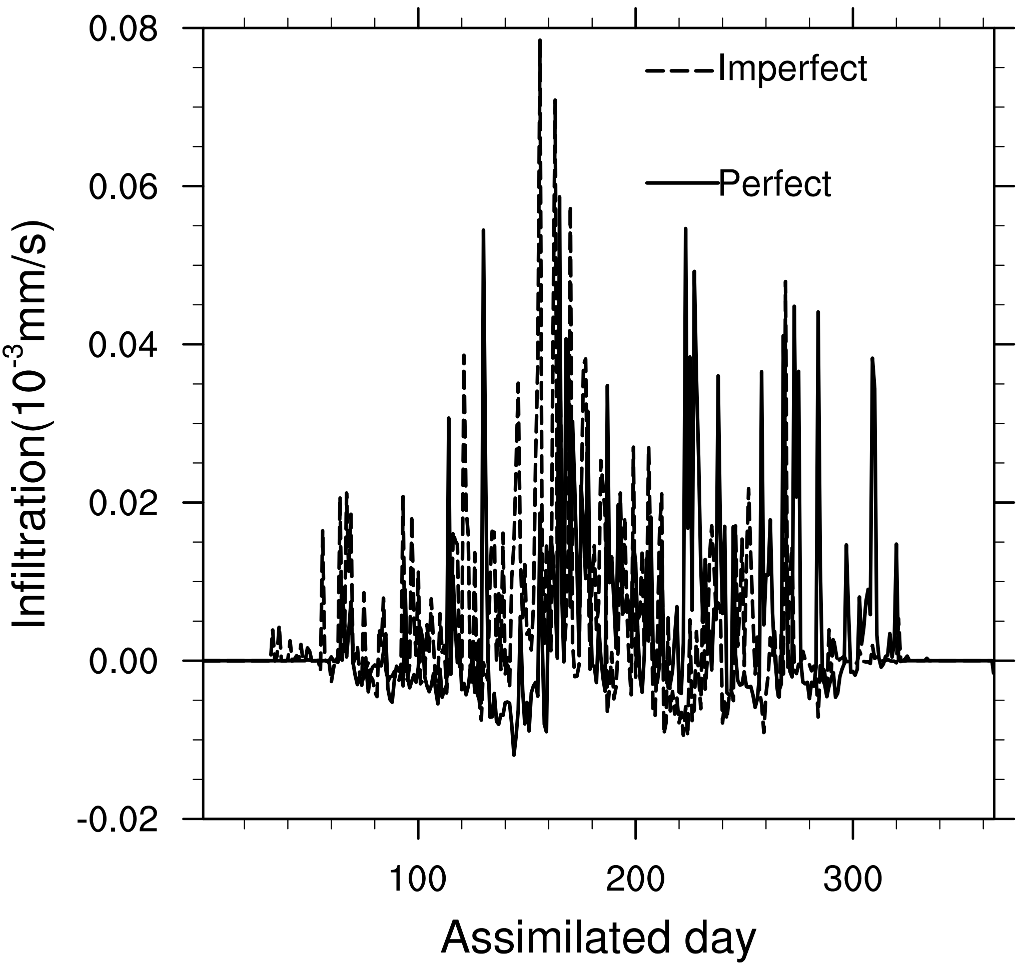


**FIG. 1.** The “perfect” (solid line) and “imperfect” (dashed line) infiltration time series used in the assimilation experiments.

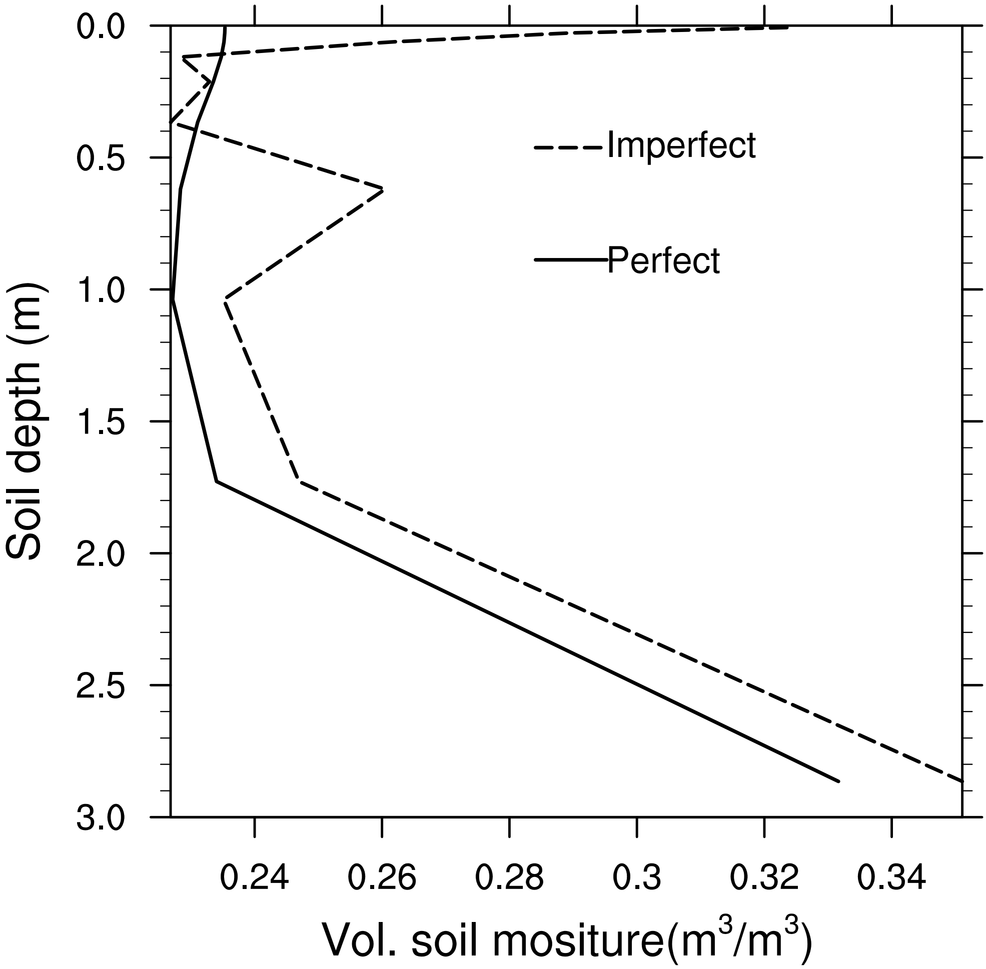


**FIG. 2**. The “perfect” (solid line) and “imperfect” (dashed line) initial soil moisture profiles used in the assimilation experiments.

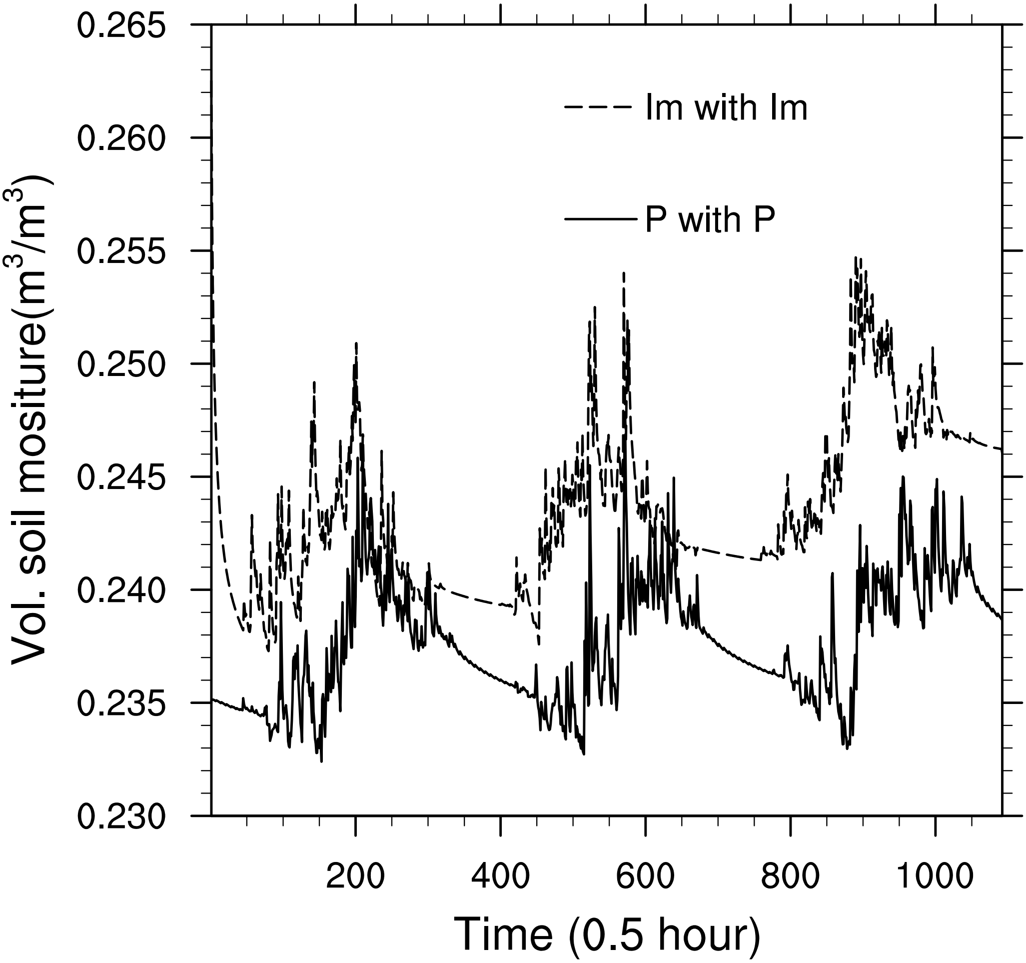


**FIG. 3**. Time series of skip volumetric soil moisture simulated by the perfect model with the "perfect" initial soil moisture, and the imperfect model with the "imperfect" initial soil moisture.

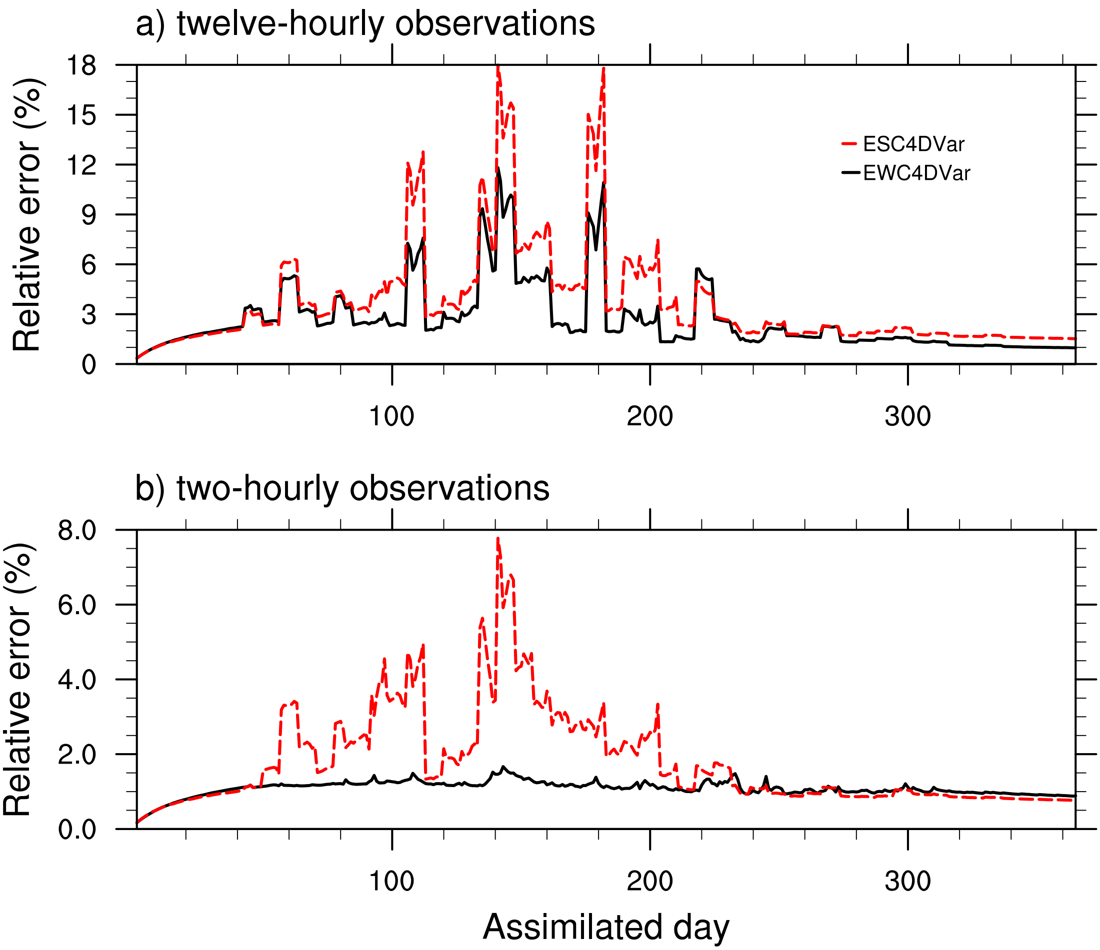


**FIG. 4.** Relative error ($E_n$) for analyzed soil moisture in the assimilation experiments by the imperfect model with the "imperfect" initial field.

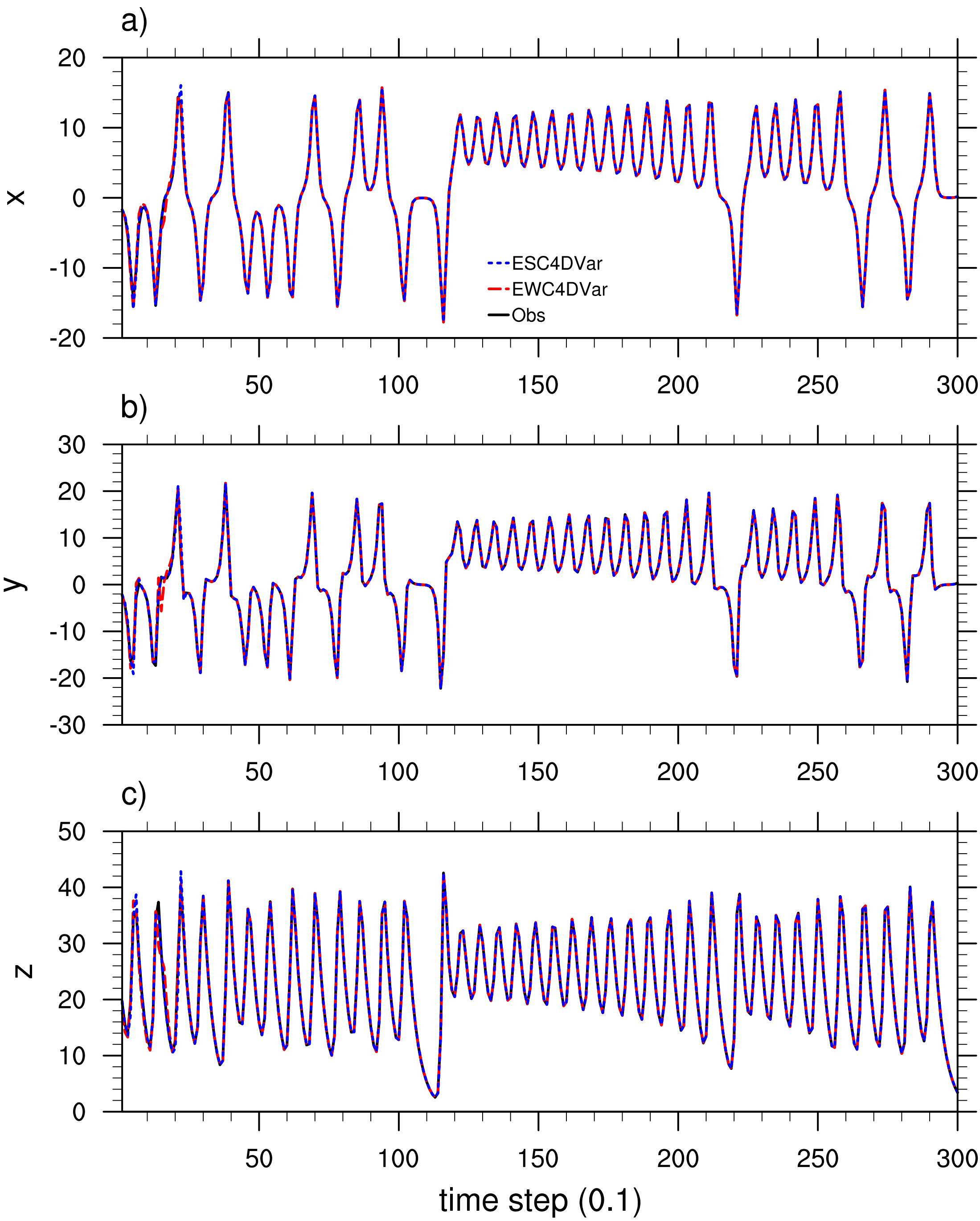


**FIG.5**. Time series of the Lorenz curve coordinates (x,y,z) from observations (solid line), the EWC4DVar assimilations (long-dashed line) and the ESC4DVar assimilations (short-dashed line)

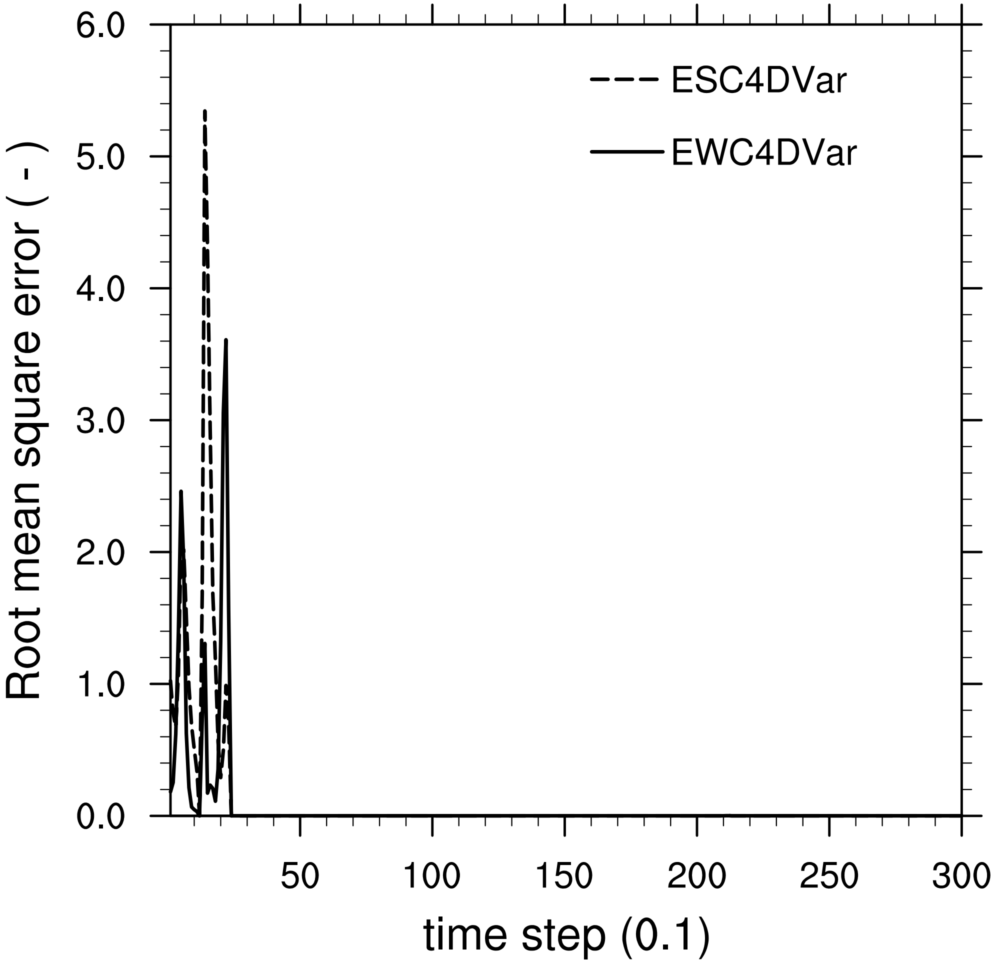


**FIG.6**. Root mean square error for the EWC4Dvar assimilated (solid line) or the ESC4DVar assimilated (long-dashed line) Lorenz curves.