



A four-dimensional scheme based on singular value decomposition (4DSVD) for chaotic-attractor-theory-oriented data assimilation

Jincheng Wang^{1,2} and Jianping Li²

Received 31 July 2008; revised 11 November 2008; accepted 21 November 2008; published 29 January 2009.

[1] The chaotic-attractor-theory-oriented data assimilation (CDA) method is reviewed. A scheme based on the singular value decomposition (SVD) analysis, called the 4DSVD, is then updated to a real four-dimensional scheme for the CDA. This algorithm employs the SVD to extract the base vectors that span the phase spaces of both model and observation chaotic attractors. Four groups of experiments are carried out under the perfect model assumption by using the Lorenz 28-variable model to test the performance of this scheme in assimilating data. Meanwhile, this scheme is compared with the four-dimensional variational data assimilation (4DVAR). The results show that the 4DSVD is effective and efficient in assimilating “observations” to obtain analysis states that are consistent with model dynamics. The accuracy of the analysis state of the 4DSVD is similar to that of the 4DVAR; however, the required computational time of the 4DSVD is much less than that of the 4DVAR, and the 4DSVD also avoids having to estimate the background error covariance matrix. The results of observation system simulation experiments carried out using the Weather Research and Forecasting Modeling System show that the method could generate good analyses by assimilating incomplete observations. They also exhibit the ability of the 4DSVD to constrain the assimilation using the full model dynamics.

Citation: Wang, J., and J. Li (2009), A four-dimensional scheme based on singular value decomposition (4DSVD) for chaotic-attractor-theory-oriented data assimilation, *J. Geophys. Res.*, *114*, D02114, doi:10.1029/2008JD010916.

1. Introduction

[2] The purpose of data assimilation is to qualitatively obtain an initial state for the numerical model by combining the information of the observations with the model dynamics and physical properties [Daley, 1991]. Many simple data assimilation methods, such as the polynomial fitting method [Panofsky, 1949], the successive-correction method (SCM) [Bergthorsson *et al.*, 1955; Cressman, 1959] and optimal interpolation (OI) [Gandin, 1963] have been proposed since the first numerical weather prediction was successfully produced. In recent years, many advanced data assimilation methods, such as the four-dimensional variational assimilation (4DVAR) [Courtier, 1997; Zheng, 2003; Kalnay, 2005], the extended Kalman filter (EKF) [Miller *et al.*, 1994; Zheng, 2003; Kalnay, 2005], and the ensemble Kalman filter (EnKF) [Evensen, 1997; Kalnay, 2005] have been developed and applied to data assimilation in atmospheric and oceanic sciences [Pu and Braun, 2001; Zhu *et al.*, 2002; Pu and Tao, 2004; Caya *et al.*, 2005; Leeuwenburgh *et al.*,

2005; Meng and Zhang, 2007]. These data assimilation methods have been applied toward determining appropriate initial conditions for the numerical models by assimilating different observations. Theoretically, the 4DVAR data assimilation might be the best method to obtain an optimal analysis state that is consistent with model dynamics by combining different observations and states delivered from models using previous observations. Its application to practical implementation is limited by its large computation time, although many approximately algorithms [Cao *et al.*, 2006; Johnson *et al.*, 2006] were proposed to tackle this problem. Another key factor is the background error covariance matrix, on which the accuracy of the analyses of the 4DVAR depends; this factor is difficult to determine accurately in operational applications. The EnKF was proposed by Evensen [1994], and has become a popular data assimilation method in recent years [Evensen, 2006; Gao and Xue, 2007]. Its advantages include the flow-dependent background error, the use of a fully nonlinear model, and simple implementation requiring little effort or expert knowledge. Several ensemble-based Kalman filter algorithms have been proposed and successfully applied to many ideal and realistic cases [Evensen, 1997; Houtekamer and Mitchell, 2001; Ott *et al.*, 2004; Caya *et al.*, 2005; Houtekamer and Mitchell, 2005; Shu *et al.*, 2005; Anderson, 2007]. Work remains to be done in matching variational

¹College of Atmospheric Sciences, Lanzhou University, Lanzhou, China.

²State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China.

assimilation performance with ensemble methods in realistic scenarios [Houtekamer *et al.*, 2005].

[3] Some of the above mentioned data assimilation methods are associated with optimal estimation theory. The cost functions of three-dimensional variational assimilation (3DVAR) and 4DVAR formulations can both be derived from the Bayesian formulation [Lorenz, 1986], which is a basic principle of estimation theory used to solve inverse problems. The EKF is an important method for estimating the posterior probability distribution for a nonlinear model in an optimal estimation field [Kamen and Su, 1999]. The EnKF can also be derived from the Bayesian formulation as a suboptimal solution for a Bayesian problem; it finds the posterior probability distribution given probability densities for both model predictions and observations [Evensen and van Leeuwen, 2000].

[4] A novel chaotic-attractor-theory-oriented data assimilation (CDA) method was introduced by Qiu and Chou [2006] (hereafter QC06). This work attempted to solve the data assimilation problem in the phase space of a chaotic attractor, which was spanned by some orthogonal functions that referred to the base vectors of the atmospheric attractor in the phase space in order to reduce the degrees of the undetermined problem. Their paper suggested a CDA implementation scheme based on the singular value decomposition (SVD). One deficiency of this method is that it applies the SVD to the original fields of expanded atmosphere, and therefore it cannot deliver orthogonal base vectors that separately support the phase spaces of both the model and observation. For more details about this method, please refer to QC06.

[5] Recently, J. Wang *et al.* [2008] (hereafter WLC08) proposed a three-dimensional scheme employing the SVD technique to the covariance matrix of both model states and simulated observation states. Compared with the scheme in QC06, the scheme in WLC08 could be used to obtain orthogonal base vectors that spanned the phase space of chaotic attractors of the model and the observations. The scheme in WLC08 is also called “4DSVD”, like the name in the QC06 scheme; however, it is only a three-dimensional method. Some three-dimensional experiments were carried out to compare the performance of the WLC08 scheme with that of the QC06 scheme, and revealed that the performance of the WLC08 scheme is better than that of the QC06 scheme. The scheme in WLC08 is unable to assimilate simultaneous observations in a time window. In this paper, we upgrade the 4DSVD in WLC08 to a real four-dimensional assimilation method that is also called 4DSVD. Note that the 4DSVD scheme used here is different from the algorithm “4DSVD” in either QC06 or WLC08, although they have the same name. Unless otherwise noted, the 4DSVD referred to in this paper is the new algorithm.

[6] In this paper, the theory of the CDA is reviewed in section 2, followed by a real four-dimensional method 4DSVD in section 3. In section 4, some simple numerical experiments are described using the Lorenz 28-variable model. The results from the 4DSVD are discussed, and a comparison with the 4DVAR performance is presented. To test the performance of the 4DSVD in a more realistic model, some observation system simulation experiments (OSSEs) were designed with the Weather Research and Forecasting (WRF) Modeling System [Skamarock *et al.*,

2005], as presented in section 5. A brief summary and some discussions are presented in the last section.

2. Theoretical Description of the CDA

[7] The CDA data assimilation method based on the theory of chaotic attractor [Li and Chou, 1997; Li, 1997] was introduced by QC06; the principles are further described in this section.

[8] In addition to atmospheric models’ attractors [Lions *et al.*, 1995, 1997], there is a global atmospheric attractor [Li and Chou, 1997; Li, 1997]. The chaotic attractor of the atmosphere has a finite dimension and is much smaller than the degree of freedom of the atmospheric models’ phase space. Thus the degree of freedom, necessary for describing the atmospheric flow, is a finite number [Temam, 1991]. As a result, the underdetermined problems in atmospheric data assimilation are much less serious than previously thought; that is to say, if the data assimilation problem could be solved in the attractor phase space, the data assimilation problem may not be an underdetermined problem.

[9] The attractor (invariant point set) could be embedded into the space \mathbf{R}^{2S+1} on the basis of Whitney’s theory, where $\mathbf{R}^{2S+1} \subset \mathbf{R}^N$ and $2S+1 \leq N$, when the dimension of the attractor is S in an N -dimensional phase space [Zhang and Chou, 1992]. There is a set of orthogonal base vectors $e_i (i = 1, 2, \dots, 2S+1)$ that is not unique. Base vectors that support the attractor of the atmosphere or model may therefore be obtained from atmospheric observation samples or from numerical model solutions, through which analyses can be consistent with the model dynamics.

[10] In next section, we propose an algorithm for the CDA. The key remaining issue of the CDA is how to determine the proper base vectors spanned attractor and the analysis states from observations. In this paper, we employ the SVD analysis technique to obtain both the base vectors and the relationship between the observation space and the model space.

3. A New Scheme for the CDA

[11] The SVD analysis could be used to determine two coupled sets of orthogonal singular vectors, as well as the expansion coefficient correlations from the covariance matrix of two geophysical fields [Wallace *et al.*, 1992; Cherry, 1996]. So far, the SVD has been widely used in meteorology, especially in climatology for the detection of coupled modes of different atmospheric fields [Wang and Fu, 2002; Nan and Li, 2003; Wang *et al.*, 2003]. Our purpose is to determine analysis state that is consistent with model dynamics by mapping observations on the attractor of model phase space which is spanned by the orthogonal base vectors. The SVD analysis is an appropriate method for finding the coupled base vectors that support phase spaces of the attractors of both the model and the observations, as well as forming the correlation formulation of the expansion coefficients for the coupled base vectors. The formulation of the 4DSVD implementation algorithm for the CDA is explained in detail in the following section.

3.1. Generation of Samples

[12] We integrate a model forward over a long period of time, starting from any appropriate initial state and selecting

n samples, denoted by \mathbf{x}_i for $i = 1, 2, \dots, n$, with the dimension q from the model outputs.

3.2. Generating Expanded Simulation Observations

[13] Generating the model states $\mathbf{x}_{i,j}^f$ at observation times t_j , using each selected sample \mathbf{x}_i as the initial condition,

$$\mathbf{x}_{i,j}^f = M_{t_i, t_j}[\mathbf{x}_i], \quad (1)$$

$$t_j \in [t_i, t_i + t_m]$$

$$j = 0, 1, 2, \dots, m-1,$$

where M_{t_i, t_j} is the model-forecasting operator from time t_i to t_j , and t_j is the time when the j th observation occurs. The simulated observations $\mathbf{y}_{i,j}^{sim}$ of dimension $p(j)$ at time t_j can then be derived from $\mathbf{x}_{i,j}^f$ through

$$\mathbf{y}_{i,j}^{sim} = H_j(\mathbf{x}_{i,j}^f), \quad (2)$$

where H_j is the observation operator at time t_j . The expanded simulation observation vector \mathbf{y}_i^{sim} is obtained by

$$\mathbf{y}_i^{sim} = \left((\mathbf{y}_{i,0}^{sim})^T, (\mathbf{y}_{i,1}^{sim})^T, (\mathbf{y}_{i,2}^{sim})^T, \dots, (\mathbf{y}_{i,m-1}^{sim})^T \right)^T, \quad (3)$$

[14] Where $()^T$ is the transposed vector $()$. The dimension of the expanded simulation observation vector is then expressed by $p = \sum_{j=0}^{m-1} p(j)$.

3.3. Generation of Coupled Base Vectors

[15] The SVD technique is employed so that the covariance matrix \mathbf{C} , comprising the model state vector and the corresponding expanded simulation observation vector samples, is used to obtain base vectors. These vectors span both the phase space of the model attractors and the simulation observations in the phase space,

$$\mathbf{C} = \mathbf{X}\mathbf{Z}^T = \mathbf{U} \begin{pmatrix} \mathbf{E} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \mathbf{V}^T, \quad (4)$$

where \mathbf{X} is a $q \times n$ matrix and \mathbf{Z} is a $p \times n$ matrix, whose columns are the anomalies of model state samples and the expanded simulation observation samples, respectively; \mathbf{U} is the matrix whose columns are the singular vectors of the model sample space; \mathbf{V} is the matrix whose columns are the eigenvectors of the sample space of the expanded simulation observations; and \mathbf{E} is the diagonal matrix whose diagonal entries are singular values of the covariance matrix \mathbf{C} .

[16] The matrices \mathbf{X} and \mathbf{Z} are given by

$$\mathbf{X} = (\mathbf{x}_1 - \bar{\mathbf{x}}, \mathbf{x}_2 - \bar{\mathbf{x}}, \dots, \mathbf{x}_n - \bar{\mathbf{x}}) \quad (5a)$$

$$\mathbf{Z} = (\mathbf{y}_1^{sim} - \bar{\mathbf{y}}^{sim}, \mathbf{y}_2^{sim} - \bar{\mathbf{y}}^{sim}, \dots, \mathbf{y}_n^{sim} - \bar{\mathbf{y}}^{sim}), \quad (5b)$$

where $\bar{\mathbf{x}}_0$ and $\bar{\mathbf{y}}^{sim}$ are the sample mean or the expected value of $\mathbf{x}_{i,0}$ and \mathbf{y}_i^{sim} , respectively. The singular value diagonal matrix could be written as

$$\mathbf{E} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_r),$$

where λ_k , $k = 1, \dots, r$ are the singular values,

$$\lambda_1 > \lambda_2 > \dots > \lambda_r.$$

The singular vectors matrices \mathbf{U} and \mathbf{V} are given by

$$\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r) \quad (6a)$$

$$\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r), \quad (6b)$$

where \mathbf{u}_k and \mathbf{v}_k are the left and right singular vectors, and $r \leq \min(q, p)$; a notable feature of the singular vectors is that they are mutually orthogonal.

[17] Next, we assume that the eigenvectors \mathbf{u}_k and \mathbf{v}_k in equations (6a) and (6b), derived from the sample covariance matrix, could be considered the base vectors supporting the phase space of both the model and observation space attractors. Assuming that the dimension of the atmosphere attractor is S , any model sample could be expanded by using the model base vectors,

$$\mathbf{x} - \bar{\mathbf{x}} = \sum_{k=1}^{2S+1} a_k \mathbf{u}_k, \quad (k = 1, \dots, 2S+1), \quad (7a)$$

where a_k ($k = 1, \dots, 2S+1$) are the expansion coefficients. Each observation vector could be expanded using the observation attractor base vectors,

$$\mathbf{y} - \bar{\mathbf{y}} = \sum_{k=1}^{2S+1} b_k \mathbf{v}_k, \quad (k = 1, \dots, 2S+1), \quad (7b)$$

where b_k are the observation expansion coefficients. The k th pair of expansion coefficients is assumed to have the relationship

$$a_k = \rho_k b_k + \varepsilon_k, \quad (k = 1, \dots, 2S+1), \quad (8)$$

where ρ_k are coefficients of the linear fitting and ε_k are constants. Equation (8) is a general linear regression problem; ρ_k and ε_k can be derived from the n pairs of the expansion coefficients a_k and b_k via least squares estimation. The sum of squared residual of equation (8) can be expressed as $\sum_{i=1}^n (a_{k,i} - \rho_k b_{k,i} - \varepsilon_k)^2$. To minimize the sum of squared residuals, the ρ_k and ε_k can be written as

$$\begin{aligned} \rho_k &= \frac{\sum_{i=1}^n (a_{k,i} - \bar{a}_k)(b_{k,i} - \bar{b}_k)}{\sum_{i=1}^n (b_{k,i} - \bar{b}_k)^2} \\ &= \frac{\text{Cov}(a_k, b_k)}{\text{Var}(b_k)}, \quad (k = 1, \dots, 2S+1) \end{aligned} \quad (9a)$$

$$\varepsilon_k = \bar{a}_k - \rho_k \bar{b}_k, \quad (k = 1, \dots, 2S+1), \quad (9b)$$

where \bar{a}_k and \bar{b}_k are the sample mean of the expansion coefficients $a_{k,i}$ and $b_{k,i}$, respectively, and $Var(b_k)$ is the variance of b_k .

[18] From these equations, we obtain the base vectors supporting the attractors of the model and observation phase spaces. Then the relationship between the expansion coefficient and the corresponding observation expansion coefficient is described by equation (8).

3.4. Obtaining the Analysis State by Real Observations

[19] Now we compute the analysis state from the observations. To obtain the expansion coefficients, the real observations are mapped to the attractor of the expanded simulation observation phase space, yielding

$$\mathbf{y}^o = \sum_{k=1}^{2S+1} b_k^o \mathbf{v}_k + \bar{\mathbf{y}}, \quad (k = 1, \dots, 2S+1), \quad (10)$$

where

$$b_k^o = \mathbf{v}_k^T (\mathbf{y}^o - \bar{\mathbf{y}}) \quad \text{and} \quad (k = 1, \dots, 2S+1) \quad (11)$$

are the expansion coefficients of the real observations. We can then compute the model base vector expansion coefficients from equation (8).

$$a_k^o = \rho_k b_k^o + \varepsilon_k, \quad (k = 1, \dots, 2S+1). \quad (12)$$

Now the analysis state can be computed from the base vectors of the model phase space attractor and their expansion coefficients,

$$\mathbf{x}^a = \sum_{k=1}^{2S+1} a_k^o \mathbf{u}_k + \bar{\mathbf{x}}. \quad (13)$$

Submitting equations (11) and (12) to equation (13), the final analysis state can be written in the form

$$\mathbf{x}^a = \sum_{k=1}^{2S+1} [\rho_k \mathbf{v}_k^T (\mathbf{y}^o - \bar{\mathbf{y}}) \mathbf{u}_k + \varepsilon_k] + \bar{\mathbf{x}}. \quad (14)$$

[20] From the formulations stated above, we find that if the error characteristics of the observations change, it is not necessary to repeat the whole procedure of generating the SVD. The procedure of the 4DSVD does not depend on the error characteristics of the observations; thus, even if the characteristics of the observations change, none of the SVD generation procedure must be changed. If the types, locations, and times of the observations change, only the procedure of generating the simulated observation states and the SVD must be repeated. This means that once the samples are generated by integrating the model forward over enough time that the runs cover the attractor, they can be used for any situation. In contrast, in real data assimilation situations the model runs for a long time and amasses large numbers of model states. Consequently, the 4DSVD could save a lot of computation time. Even the other steps, generating the simulated observations and obtaining the

base vectors using SVD, require less computation time compared with the 4DVAR.

[21] There is one distinct difference between 4DSVD and other data assimilation methods: 4DSVD seeks to solve the data assimilation problem in the attractor space of the model and simulation observation on the basis of the chaotic attractor theory, while others solve the problem in the error subspace mainly on the basis of estimation theory. The 4DSVD method makes no explicit consideration of the background and observation uncertainties measured by error covariance. It only assumes that the column vectors of \mathbf{U} in equation (4) span the phase space of the chaotic attractors of the dynamic system. This assumption is valid when there are enough generated samples to cover the chaotic attractors of the dynamic system, so that the column vectors of \mathbf{U} can span the phase space of the chaotic attractors of the dynamic system.

[22] The 4DSVD described above can be altered with the shadowing method. The shadowing method proposed by *Judd and Smith* [2001], *Judd* [2003], and *Judd et al.* [2004] attempts to seek a model trajectory that always remains close to the original observations and the derived analysis sequences by minimizing a cost function, which is used to measure the distance a sequence of states are from being a suitable shadowing trajectory. The gradient descent methods are used to find the shadowing trajectories. These shadowing methodologies need relatively complete observations, meaning that each variable on all grids of the model should be observed; in reality, of course, observations of the atmosphere are incomplete. Therefore, a series of analyses obtained by data assimilation methods like 3DVAR are a necessary precondition for applying shadowing methodologies. As *Judd et al.* [2004] pointed out, the shadowing techniques are not necessarily intended to be a replacement for data assimilation methods such as 3DVAR, but rather to augment them so as to provide better analyses. However, the 4DSVD method is a new data assimilation method, which cannot only interpolate the incomplete observations into full model grids but also ensures that the analyses are consistent with the model dynamics.

4. Some Simple Experiments With the Lorenz 28-Variable Model

[23] In this section, some numerical experiments are designed using the Lorenz 28-variable model to examine the performance of the 4DSVD scheme (described in section 3) under the perfect model scenario.

4.1. Model and the Reference State

[24] The Lorenz 28-variable model was introduced by *Lorenz* [1965], consisting of 28 variables and some physical parameters. Some details of this model were subsequently studied by *Reinhold and Pierrehumbert* [1982]. To some extent, the dynamic behavior of this model is determined by the forcing parameter θ^* . It was proven that the chaotic attractors of the model exist at some special value of the forcing parameter θ^* [*Krishnamurthy*, 1993]. More details of the characters and behaviors of the Lorenz 28-variable model have been discussed in numerous studies [*Lorenz*, 1965; *Reinhold and Pierrehumbert*, 1982; *Krishnamurthy*, 1993].

[25] In this paper, the value of the forcing parameter θ_0^* is set to 0.15, because chaotic attractors exist at this value. The values of the other parameters are held the same as those used in Krishnamurthy's research [Krishnamurthy, 1993].

[26] We perform data assimilation experiments with the numerical version of the Lorenz 28-variable model. The fourth-order Runge-Kutta scheme is used for time integration. The time step is $\Delta t = 0.001$ in nondimensional units.

[27] Assuming the model is perfect, the model is run forward in time for a long time period starting from an initial condition. An output value is then selected as the true initial condition. The reference state is obtained by integrating the model with this initial condition.

4.2. Observations

[28] The observations are defined by perturbing the reference state, where perturbations are drawn from a white noise Gaussian distribution. The variance of the observations error is set to 10% of the variance of the reference values, where this means that the errors of different variables are different.

[29] All variables are observed in all experiments. Hence, the observation operator for this study is a linear (identity) transformation operator, defined as $H(\mathbf{x}) = \mathbf{x}$.

4.3. Sampling Strategy and Sampling Size

[30] Samples are selected from the model output by running the model starting from a random initial condition. To ensure the samples are linearly independent, samples are selected every 100 time steps. Although this strategy may not be the most efficient, it can generate good base vectors. Using this method, 300 samples are selected for all 4DSVD experiments in this paper. This large number is chosen to ensure there are sufficient samples representing the attractor of the model to obtain good base vectors.

4.4. Experiments Setup

[31] To ensure consistency between the 4DSVD and 4DVAR experiments, we use the same reference state and observations for the assimilation experiments. The assimilation time window is set to $[0, 1.0]$. There are four experimental groups, ExpG1, ExpG2, ExpG3, and ExpG4, with observations at time intervals 0.1, 0.2, 0.5, and 1.0. In each group, 30 independent experiments are performed at different initial times using different observations with different observation errors.

[32] For the 4DVAR experiments under the perfect model scenario, the constant observation error covariance is computed using the observation error explained above. The background is randomly selected from the outputs of the model, and the background error covariance is computed using the climate covariance of "true" model states. The nonmonotone gradient-projection algorithm SPG2 [Birgin *et al.*, 2000], which has been applied widely in optimal problems, is used to optimize the cost function [Duan *et al.*, 2004; Mu and Zhang, 2006; Mu *et al.*, 2007a, 2007b].

[33] The accuracy of analysis states is measured by the analysis error of all the variables using root mean squared errors (RMS), as expressed by

$$\varepsilon^a = \sqrt{\frac{(\mathbf{x}^a - \mathbf{x}^t)^T (\mathbf{x}^a - \mathbf{x}^t)}{28}}$$

The average analysis errors of all the experiments in each group are also computed,

$$\varepsilon_{ave}^a = \frac{1}{n_e} \sum_{i=1}^{n_e} \varepsilon_i^a,$$

where n_e is the number of experiments.

4.5. Results

[34] Figures 1a and 1b show the analysis errors of different experiments at the initial assimilation time. Figure 1a presents 4DSVD, which uses 28 base vectors; it can be seen herein that almost all the analysis errors of the 4DSVD are smaller than the observation errors as calculated in an RMS sense under different observation intervals. That is, for each experiment, the initial conditions for the model improve significantly by assimilating observations using 4DSVD. The analysis errors of experiments in ExpG1 are the smallest of the four groups, while those of ExpG4 are the largest. For all these experiments, more observations in the fixed assimilating time window lead to more accurate analysis states. The analysis errors are nearly the same, although the values of observation errors in different experiments within the same experiment groups differ greatly. This may imply that analysis errors may be determined by the statistic characters of the observation errors, although the observation error covariance is not explicitly used in 4DSVD. Moreover, it implies that the 4DSVD method only needs limited observations to deliver good analyses for a specific problem and it means that additional observations may do little to improve the initial conditions. The analysis errors of 4DVAR are shown in Figure 1b. For ExpG1 and ExpG2, the analysis errors are much smaller than the observation errors, whereas the analysis errors of experiments in ExpG3 and ExpG4 are equal to or even greater than the observation errors. The analysis errors increase rapidly with decreasing numbers of observations. Figure 1c shows the difference between the 4DVAR and the 4DSVD analysis errors (the 4DVAR minus the 4DSVD). Comparing the analysis errors of the two methods shows that the analysis states generated by assimilating observations via 4DVAR are more accurate than those from 4DSVD are when there are many observations. This is the case in ExpG1, where the average analysis error difference (the 4DVAR minus the 4DSVD) is -8.9×10^{-4} . The performance of the 4DVAR and 4DSVD are comparable with sufficient numbers of observations, such as in ExpG2 and ExpG3 where the average analysis error differences (the 4DVAR minus the 4DSVD) are -4.8×10^{-4} and 6.8×10^{-4} , respectively. The performance of 4DVAR is inferior to 4DSVD in ExpG4, where there was an insufficient number of observations in the data assimilation time window; its average analysis error difference is 1.2×10^{-3} . These results suggest that given enough observations, the 4DSVD method would be capable of improving the initial conditions of all the experiments; and further, that its performance is comparable with, and sometimes superior to, 4DVAR.

[35] Moreover, the computation time required for 4DSVD is much less than that for 4DVAR. For example, the 4DSVD needs about 6277 s of CPU time to complete all 30 experi-

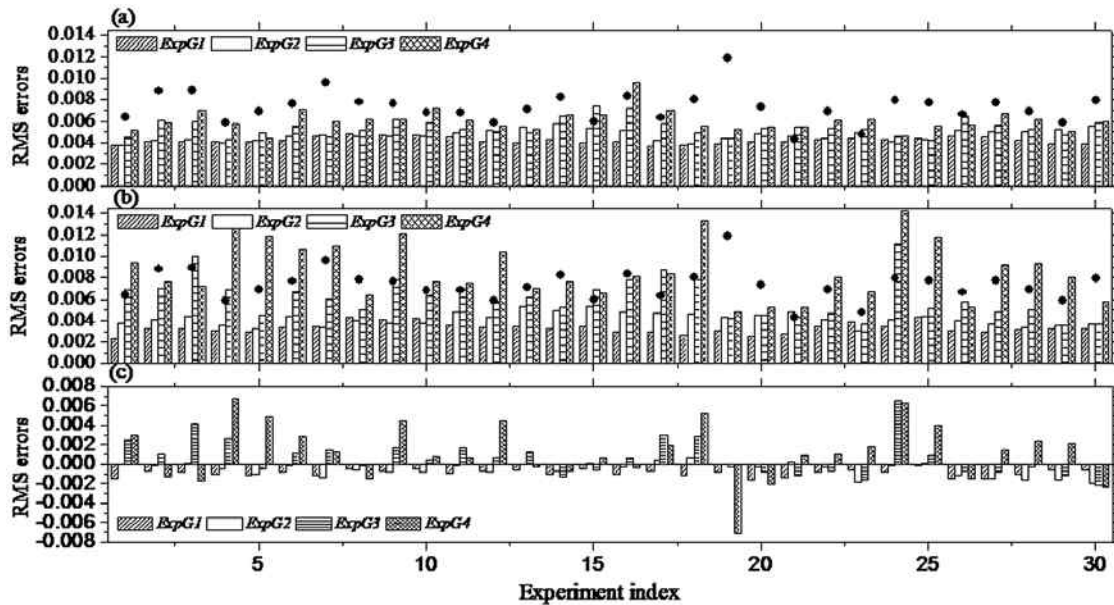


Figure 1. (a) Analysis errors of 4DSVD for the experiment groups ExpG1, ExpG2, ExpG3, and ExpG4. The observation errors are indicated by closed circles. (b) Same as Figure 1a but for 4DVAR. (c) Difference in the analysis errors of 4DVAR and 4DSVD.

ments of the group ExpG1, while the 4DVAR takes about 16 times longer at 98,000 s of CPU time. It is a major advantage of the 4DSVD method that under the same conditions, much less computation time is required than with 4DVAR to achieve similar analysis states.

[36] Figure 2 presents the average analysis errors during the assimilation time window. The analysis errors are shown to be stable during the assimilating time window for all situations. This illustrates that the analysis states of 4DSVD may be consistent with both model dynamics and the analysis states of 4DVAR. These results indicate that 4DSVD may have the ability to ensure consistency between analysis state and model dynamics, which would represent yet another advantage of this method compared with other simple ones like 3DVAR. However, more evidence would be required to prove this point. Figure 2 shows that the analysis errors from 4DSVD and 4DVAR can have quite different dependencies on the observation time interval; this

phenomenon is an interesting and important issue. Possible explanations for the different dependencies are discussed in section 6.

[37] Another important issue for 4DSVD is determining how many base vectors are needed. Figure 3 shows the average of the analysis errors as a function of the number of base vectors. For all experiments in all four groups (ExpG1, ExpG2, ExpG3, and ExpG4) the same relationship exists: the analysis error decreases with increasing number of base vector, although the data assimilation intervals and the number of observations are different. This phenomenon may reflect the nature of the model, as the dimension of its chaotic attractor is not changed. The analysis errors are the smallest when the number of the base vectors is 28; this result might be confusing since the optimal truncation number is as large as the dimension of the model. In fact, it could call into question the advantage of having large degrees of the freedom for the real atmosphere or ocean

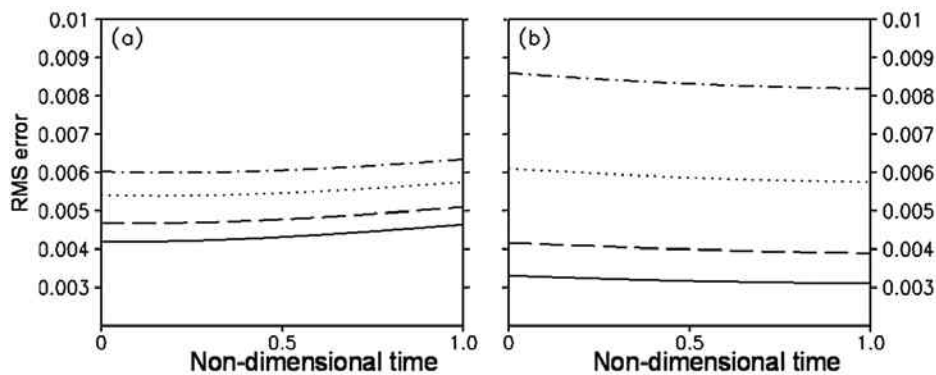


Figure 2. Averaged analysis errors of all the experiments in groups ExpG1 (solid line), ExpG2 (dashed line), ExpG3 (dotted line), and ExpG4 (dash-dotted line) in the assimilation time window. (a) 4DSVD. (b) 4DVAR.

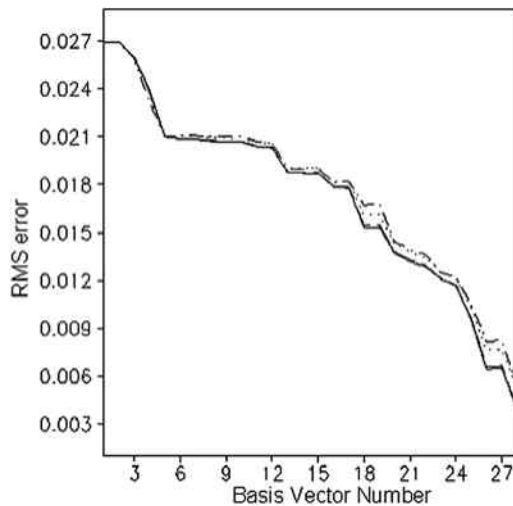


Figure 3. The 4DSVD averaged analysis errors of all the experiments in groups ExpG1 (solid line), ExpG2 (dashed line), ExpG3 (dotted line), and ExpG4 (dash-dotted line) as a function of base vector number.

model of the 4DSVD. According to the theory of the CDA, the number of the base vectors is determined by the dimension of the chaotic attractors of the model. For the Lorenz 28-variable model used here, the Lyapunov dimension, which is a good estimate of the dimension of the chaotic attractors [Farmer *et al.*, 1983], is about 13 [Krishnamurthy, 1993]. According to Whitney's theory the chaotic attractors could be embedded into space R^{28} , where the analysis errors would then reach the minimum value when using 28 base vectors in these experiments. The dimension of the atmosphere model's chaotic attractors is much smaller than the degree of the freedom of the model, as mentioned in section 3. Thus, the optimal truncation number of the base vectors is much smaller than the dimension of atmospheric model. To confirm this assertion, and to test the performance

of 4DSVD in more realistic model, more experiments are carried out using the WRF model.

5. Observation System Simulation Experiments With the WRF Model

[38] Observation system simulation experiments (OSSEs) are usually conducted to assess the potential impact of instruments or observing systems that are proposed or under construction [Miller, 1990; Atlas, 1997; Liu and Rabier, 2003; Lahoz *et al.*, 2005], and to evaluate the data assimilation methodology [Kuo and Guo, 1989; Bishop *et al.*, 2001; Xue *et al.*, 2006; Etherton, 2007; Qiu *et al.*, 2007; X. Wang *et al.*, 2008]. In this section, some OSSEs are designed using the WRF model in order to evaluate the 4DSVD scheme.

5.1. Model, the True State, and Observations

[39] The Advanced Research WRF (ARW) Modeling System [Skamarock *et al.*, 2005] version 3 is used to test the performance of the 4DSVD. As an initial attempt to test the 4DSVD in a realistic atmospheric model with limited computational resources, some relatively simple and manageable observation system simulation experiments are designed.

[40] Experiments are performed by running the WRF model on a domain whose center is located at (25°N, 112°E) in eastern Asia (Figure 4). The modeling domain is setup with 40×60 horizontal grid points at 30 km resolution and 27 vertical levels and a model top 50 hPa.

[41] The model is assumed to be perfect. The initial conditions and lateral boundary conditions (LBCs) are generated using NCEP "Final" analysis (FNL, <https://dss.ucar.edu/datazone/dsszone/ds083.2>). The nature run (the "truth") is obtained by integrating the model from time $t = -24$ h to the analysis time $t = 0$ using the above initial conditions and LBCs.

[42] The simulated observations of radiosonde temperature and u and v wind components are generated by adding Gaussian random noise to the WRF nature run (the

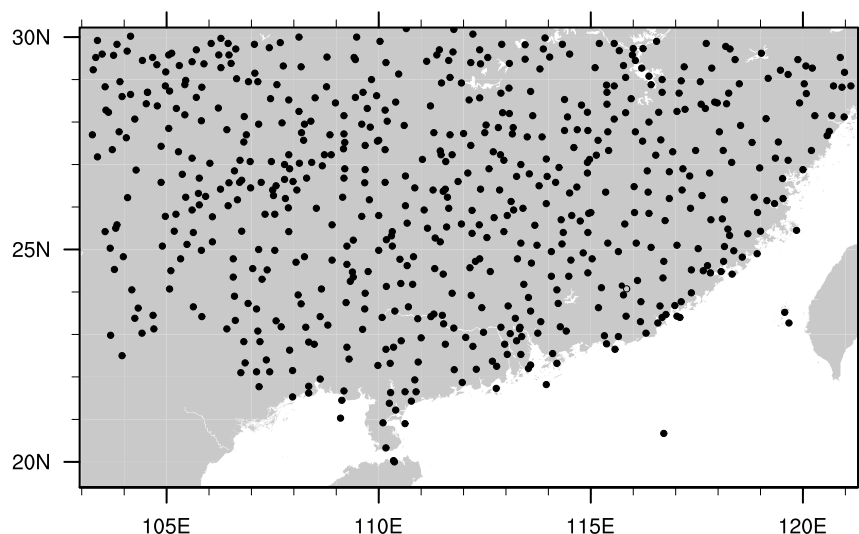


Figure 4. WRF domain (full domain) and a snapshot of radiosonde observation network (black dots).

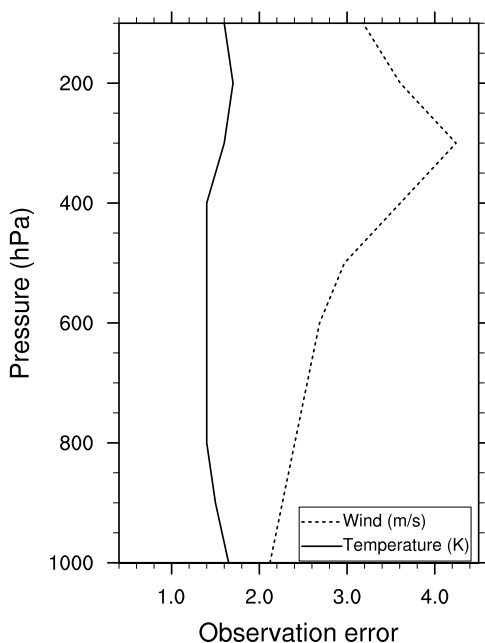


Figure 5. Vertical profile of the errors of observations for wind (dashed line) and temperature (solid line).

“truth”). We assumed the positions of the radiosonde stations to be the same as those of the surface stations. The positions of the fixed surface stations are obtained from the operational observation data sets from the China Meteorological Administration (CMA). Figure 4 shows the horizontal distribution of the radiosondes. The simulated observation time interval is set to 12 h. The vertical profile of the observation errors is shown in Figure 5. The random noise used to generate simulated observations is drawn from a Gaussian distribution with a mean of zero and standard deviations equal to the observation errors.

5.2. Sampling Strategy

[43] We choose 500 days between 1 January 1998 and 30 November 2007, with 7-day intervals. The model was then run for 60 h, starting at time 0000 UTC of these 500 days. The initial conditions and LBCs are generated using FNL analysis. One sample is chosen from one model simulation after 60 h of integration, to confirm some extent of independence of the samples. Therefore, the sample size is 500, and all experiments in this section use these 500 samples.

5.3. Experiments Setup

[44] The data assimilation time window is set to [0, 12] hours, which means that there are two observation time levels in each experiment. To evaluate the performance of 4DSVD for these realistically incomplete observations, the potential temperature T and u and v wind components are analyzed. In addition, to assess the ability of 4DSVD to reconstruct the nonobserved variables, the water vapor mixing ratio q is analyzed. The simulation started at 0000 UTC 1 December 2007 and lasted for 4 weeks. To save computational resources, the observations are assimilated every 48 h for 4 weeks; thus there are a total of 14 independent experiments.

5.4. Results

[45] Figure 6 shows the RMS analysis errors of 4DSVD as a function of model levels. The RMS analysis errors of the potential temperature and wind are much smaller than the observation errors, implying that 4DSVD is able to assimilate the incomplete observations and generate good analyses for the model. In addition, the RMS analysis errors of water vapor mixing ratio are relative small. These results illustrate that 4DSVD is able to generate good analyses not just for observed variables, but also nonobserved variables. This performance exhibits the ability of 4DSVD to use the full model dynamics in constraining the assimilation. Figure 7 shows the details of the true fields and analysis fields of one case, at eta = 0.995 level and at time 0000 UTC 1 December 2007. The analyzed state is very similar to the true state.

[46] Figure 8 is the optimal truncation number of base vectors for all experiments. The optimal truncation numbers of the base vectors is found to be much smaller than the dimension of the analyzed variables. On the other hand, the optimal truncation numbers are different for different times. This implies that the optimal truncation number of base vectors is related to other factors in addition to the smallness of the Lyapunov dimension of the dynamic system’s chaotic attractor. Some discussions about how to determine the optimal truncation number of base vectors are presented in section 6.

6. Summary and Discussions

[47] In this paper we review a data assimilation method called CDA, and propose an implementation algorithm

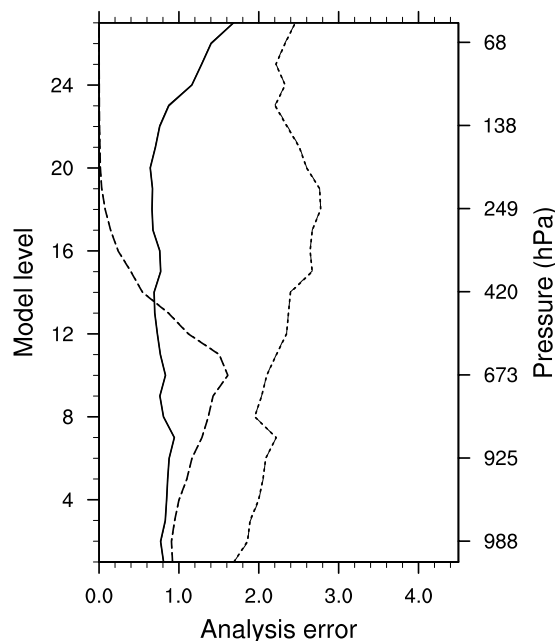


Figure 6. Vertical profiles as a function of model eta levels of the analysis errors for the potential temperature (solid line; unit: K), wind (short-dashed line; unit: m/s), and water vapor mixing ratio (long-dashed line; unit: g/kg). The pressure on the right axis was calculated from the eta levels, model top pressure of 50 hPa, and approximated surface pressure of 1000 hPa.

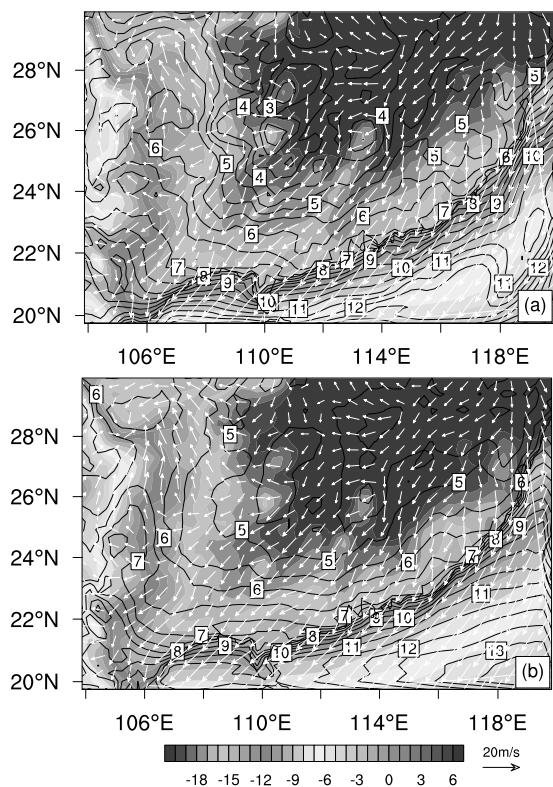


Figure 7. (a) True state and (b) analysis state at the $\eta = 0.995$ level of one case at 0000 UTC 2 December 2007. Contour indicates the water vapor mixing ratio (g/kg), color shading indicates perturbation potential temperature (K), and arrows indicate wind (m/s). The optimal truncation number of the base vectors for this case is 99. The RMS errors are 0.80 K, 1.68 m/s, and 0.92 g/kg for potential temperature, wind speed, and water vapor mixing ratio, respectively.

called 4DSVD. Some simple experiments are carried out to test the performance of 4DSVD, using the Lorenz 28-variable model under the perfect model assumption. The performance of 4DSVD and 4DVAR are compared under the same conditions. Evaluating the performance of the 4DSVD using realistic observations in a more realistic atmospheric model, some OSSEs are designed with the WRF model.

[48] The 4DSVD method is able generate good analysis states that may be consistent with model dynamics. The performance of the 4DSVD is comparable to or even better than the performance of 4DVAR under some special situations, such as when the number observations are insufficient within fixed assimilation time window. It is also important that 4DSVD could save significant computational time cost compared with 4DVAR, while delivering nearly the same accuracy. Moreover, 4DSVD avoids having to estimate the background error covariance matrix; this calculation is difficult to estimate accurately, and its precision determines the accuracy of the 4DVAR analyses. The 4DSVD method is effective and efficient for assimilating observations under the perfect model scenario.

[49] The 4DSVD method is robust even when the observations are incomplete in a more realistic model like the

WRF; it can deliver good analyses not only for observed variables but also nonobserved variables. Moreover, the method is able to use full model dynamics to constrain the assimilation. The optimal truncation number of base vectors is much smaller than the dimension of the model state vector, which means that 4DSVD could significantly reduce the dimension of the data assimilation problem. The optimal truncation number of base vectors for the supporting attractor is related to other factors in addition to the smallness of the Lyapunov dimension of both the model system and observation chaotic attractors.

[50] The sampling strategy in section 5 suggests that the 4DSVD samples can be selected by an operational weather forecast system from historical operational forecast states. All the numerical results imply that 4DSVD is a good potential data assimilation method.

[51] Although we performed some ideal experiments under perfect model situations, and further tested the 4DSVD method with some OSSEs using more realistic models, the performance of 4DSVD under imperfect model and realistic observational situations requires further examination with more experiments. Many other important issues associated with 4DSVD also need to be studied further. A sophisticated strategy is needed for the selection of samples to generate the base vectors that support the chaotic attractors. The minimum number of sufficient samples also needs further investigation. One positive aspect is that through real applications, the operational model has generated many historical model states that could offer enough samples for 4DSVD. Selecting samples from these historical model states may be a good sample strategy for use with the OSSEs. In the future, we will focus on proposing a robust sample strategy. Further investigation is also needed to determine the optimal truncation number of the base vectors, because the optimal truncation number of the base vectors is related to other factors beyond the dimension of the chaotic attractor of the model system. The impact of 4DSVD on operational forecasts should be paid more attention. The ability of the 4DSVD to simultaneously assimilate nonconventional observational data will have to be evaluated; assessing, for example, the incorporation of satellite radiance and Doppler shifts, which are usually

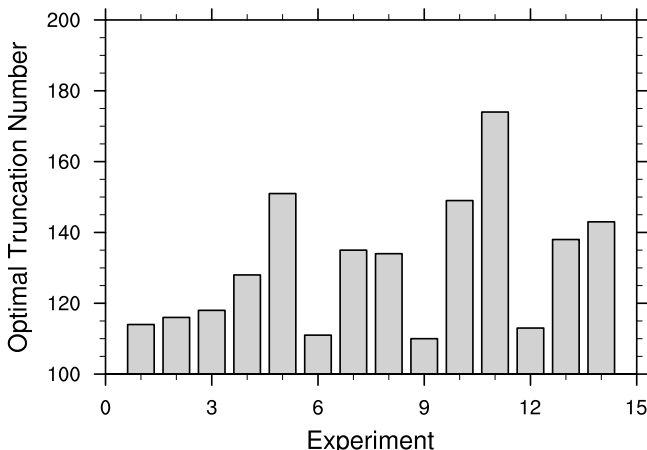


Figure 8. Optimal truncation numbers of the base vectors for all experiments.

assimilated into models using the 3DVAR [Gao and Droegemeier, 2004; Gao et al., 2001; Pu et al., 2008].

[52] An interesting phenomenon is presented in section 4, showing that the analysis errors from 4DSVD and 4DVAR can have different dependencies on the observation time interval. There are several possible reasons for this. The 4DVAR method seeks the solution of the data assimilation problem in the error subspace. When the observation time interval is larger, the number of observations is not large enough for 4DVAR. In this case, the data assimilation problem in 4DVAR may be underdetermined, and its solution may be not unique, so the 4DVAR may not achieve the most optimal analysis state or extract enough information from the limited observations. However, when the observation time interval is smaller, there are enough observations for 4DVAR. In this case, the data assimilation problem in 4DVAR may be not underdetermined, and the method could extract more information from the observations to more accurately estimate the analysis state. The analysis errors decrease significantly with the decrease in observation time interval. The 4DSVD method significantly reduces the dimension data assimilation problem, which is not an underdetermined problem even when the observations are incomplete. On the other hand, 4DSVD can extract some information from the model-generated samples. A few observations can be enough for 4DSVD to estimate analysis states with small errors; more observations would not significantly reduce the analysis errors. These results imply that when the observations are limited, the performance of 4DSVD may be superior to 4DVAR. Many experiments of different types will be required to explain the reasons for these different dependencies, and to clarify the conditions under which 4DSVD can out-perform 4DVAR.

[53] **Acknowledgments.** The authors wish to thank Jenny Lin for her kind help in improving the quality of the paper with respect to English, as well as Z. Yu, R. Ding, D. Xiao, X. Zhao, and J. Feng for their valuable suggestions. The authors are thankful to the anonymous reviewers for their comments and suggestions, which greatly improved the presentation of the results. This work was jointly supported by the 973 Program (2006CB403600) and the National Natural Science Foundation of China (40325015 and 40221503).

References

- Anderson, J. L. (2007), An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus, Ser. A*, 59, 210–224.
- Atlas, R. (1997), Atmospheric observations and experiments to assess their usefulness in data assimilation, *J. Meteorol. Soc. Jpn.*, 75, 111–130.
- Bergthorsson, P., B. Doos, S. Frykland, O. Hang, and R. Linquist (1955), Routine forecasting with the barotropics model, *Tellus, Ser. A*, 7, 329–340.
- Birgin, E. G., J. M. Martinez, and M. Raydan (2000), Nonmonotone spectral projected gradient methods on convex sets, *SIAM J. Optim.*, 10, 1196–1211.
- Bishop, C.H., B.J. Etherton, and S.J. Majumdar (2001), Adaptive sampling with the Ensemble Kalman filter. Part I: Theoretical aspects, *Mon. Weather Rev.*, 129, 420–436, doi:10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.
- Cao, Y., J. Zhu, I. M. Navon, and Z. Luo (2006), A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition, *Int. J. Numer. Methods Fluids*, 53, 1571–1583, doi:10.1002/flid.1365.
- Caya, A., J. Sun, and C. Snyder (2005), A comparison between the 4DVAR and the ensemble Kalman filter techniques for radar data assimilation, *Mon. Weather Rev.*, 133, 3081–3094, doi:10.1175/MWR3021.1.
- Cherry, S. (1996), Singular value decomposition analysis and canonical correlation analysis, *J. Clim.*, 9, 2003–2009, doi:10.1175/1520-0442(1996)009<2003:SVDAAC>2.0.CO;2.
- Courtier, P. (1997), Dual formulation of four-dimensional variational assimilation, *Q. J. R. Meteorol. Soc.*, 123, 2249–2261, doi:10.1256/smsqj.54413.
- Cressman, G. P. (1959), An operational objective analysis system, *Mon. Weather Rev.*, 87, 367–374, doi:10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2.
- Daley, R. (1991), *Atmospheric Data Analysis*, 457 pp., Cambridge Univ. Press, Cambridge, U.K.
- Duan, W. S., M. Mu, and B. Wang (2004), Conditional nonlinear optimal perturbations as the optimal precursors for El Niño–Southern Oscillation events, *J. Geophys. Res.*, 109, D23105, doi:10.1029/2004JD004756.
- Etherton, B. J. (2007), Preemptive forecasts using an ensemble Kalman filter, *Mon. Weather Rev.*, 135, 3484–3495, doi:10.1175/MWR3480.1.
- Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10,143–10,162, doi:10.1029/94JC00572.
- Evensen, G. (1997), Advanced data assimilation for strongly nonlinear dynamics, *Mon. Weather Rev.*, 125, 1342–1354, doi:10.1175/1520-0493(1997)125<1342:ADAFSN>2.0.CO;2.
- Evensen, G. (2006), *Data Assimilation: The Ensemble Kalman Filter*, 279 pp., Springer, New York.
- Evensen, G., and P. J. van Leeuwen (2000), An ensemble Kalman smoother for nonlinear dynamics, *Mon. Weather Rev.*, 128, 1852–1867.
- Farmer, J. D., E. Ott, and J. A. Yorke (1983), The dimension of chaotic attractors, *Physica D*, 7, 153–180.
- Gandin, L. S. (1963), *Objective Analysis of Meteorological Fields* (in Russian), Gidrometrol. Izdat, St. Petersburg, Russia., (Engl. Translation, Isrz. Program for Sci. Transl., Jerusalem, 1965.)
- Gao, J., and K. K. Droegemeier (2004), A variational technique for dealiasing Doppler radial velocity data, *J. Appl. Meteorol.*, 43, 934–940, doi:10.1175/1520-0450(2004)043<0934:AVTFDD>2.0.CO;2.
- Gao, J., and M. Xue (2007), An efficient dual-resolution approach for ensemble data assimilation and tests with simulated Doppler radar data, *Mon. Weather Rev.*, 136, 945–963.
- Gao, J., A. Shapiro, Q. Xu, and K. K. Droegemeier (2001), Three-dimensional simple adjoint velocity retrievals from single-Doppler radar, *J. Atmos. Oceanic Technol.*, 18, 26–38, doi:10.1175/1520-0426(2001)018<0026:TDSAVR>2.0.CO;2.
- Houtekamer, P. L., and H. L. Mitchell (2001), A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.*, 129, 123–137, doi:10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.
- Houtekamer, P. L., and H. L. Mitchell (2005), Ensemble Kalman filtering, *Q. J. R. Meteorol. Soc.*, 131, 3269–3289, doi:10.1256/qj.05.135.
- Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen (2005), Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, *Mon. Weather Rev.*, 133, 604–620, doi:10.1175/MWR-2864.1.
- Johnson, C., B. J. Hoskins, and N. K. Nichols (2006), A singular vector perspective of 4DVAR: The spatial structure and evolution of baroclinic weather systems, *Mon. Weather Rev.*, 134, 3436–3455, doi:10.1175/MWR3243.1.
- Judd, K. (2003), Nonlinear state estimation, indistinguishable states, and the extended Kalman filter, *Physica D*, 183, 273–281, doi:10.1016/S0167-2789(03)00180-5.
- Judd, K., and L. Smith (2001), Indistinguishable states I. Perfect model scenario, *Physica D*, 151, 125–141, doi:10.1016/S0167-2789(01)00225-1.
- Judd, K., L. Smith, and A. Weisheimer (2004), Gradient free descent: Shadowing, and state estimation using limited derivative information, *Physica D*, 190, 153–166, doi:10.1016/j.physd.2003.10.011.
- Kalnay, E. (2005), *Atmospheric Modeling, Data Assimilation and Predictability* (in Chinese), translated from English, edited by Z. Pu et al., 300 pp., China Meteorol. Press, Beijing.
- Kamen, E. W., and J. K. Su (1999), *Introduction to Optimal Estimation*, 380 pp., Springer, London.
- Krishnamurthy, V. (1993), A predictability study of Lorenz's 28-variable model as a dynamical system, *J. Atmos. Sci.*, 50, 2215–2229, doi:10.1175/1520-0469(1993)050<2215:APSOLV>2.0.CO;2.
- Kuo, Y.-H., and Y.-R. Guo (1989), Dynamic initialization using observations from a hypothetical network of profiles, *Mon. Weather Rev.*, 117, 1975–1998, doi:10.1175/1520-0493(1989)117<1975:DIUOFA>2.0.CO;2.
- Lahoz, W. A., R. Brugge, D. R. Jackson, S. Migliorini, R. Swinbank, D. Lary, and A. Lee (2005), An observing system simulation experiment to evaluate the scientific merit of wind and ozone measurements from the future SWIFT instrument, *Q. J. R. Meteorol. Soc.*, 131, 503–523, doi:10.1256/qj.03.109.
- Leeuwenburgh, O., G. Evensen, and L. Bertino (2005), The impact of ensemble filter definition on the assimilation of temperature profiles in the tropical Pacific, *Q. J. R. Meteorol. Soc.*, 131, 3291–3300, doi:10.1256/qj.05.90.
- Li, J. (1997), Further study on the properties of operators of atmospheric equations and the existence of attractor, *Acta Meteorol. Sin.*, 11, 216–223.
- Li, J., and J. Chou (1997), Existence of atmosphere attractor, *Sci. China Ser. D*, 40, 215–224.

- Lions, J. L., R. Teman, and S. Wang (1995), Mathematical theory for the coupled atmosphere-ocean models (CAO III), *J. Math. Pures Appl.*, *74*, 105–163.
- Lions, J. L., O. P. Manley, R. Teman, and S. Wang (1997), Physical interpretation of the attractor dimension for the primitive equations of atmospheric circulation, *J. Atmos. Sci.*, *54*, 1137–1143, doi:10.1175/1520-0469(1997)054<1137:PIOTAD>2.0.CO;2.
- Liu, Z.-Q., and F. Rabier (2003), The potential of high-density observations for numerical weather prediction: A study with simulated observations, *Q. J. R. Meteorol. Soc.*, *129*, 3013–3035.
- Lorenz, A. C. (1986), Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc.*, *112*, 1177–1194, doi:10.1002/qj.49711247414.
- Lorenz, E. N. (1965), A study of the predictability of a 28-variable atmospheric model, *Tellus, Ser. A*, *17*, 321–333.
- Meng, Z., and F. Zhang (2007), Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation. Part II: Imperfect model experiments, *Mon. Weather Rev.*, *135*, 1403–1423, doi:10.1175/MWR3352.1.
- Miller, R. N. (1990), Tropical data assimilation experiments with simulated data: The impact of the Tropical Ocean and Global Atmosphere thermal array for the ocean, *J. Geophys. Res.*, *95*, 11,461–11,482, doi:10.1029/JC095iC07p11461.
- Miller, R. N., M. Ghil, and F. Gauthiez (1994), Advanced data assimilation in strongly nonlinear dynamical systems, *J. Atmos. Sci.*, *51*, 1037–1056, doi:10.1175/1520-0469(1994)051<1037:ADAIN>2.0.CO;2.
- Mu, M., and Z. Zhang (2006), Conditional nonlinear optimal perturbations of a two-dimensional quasigeostrophic model, *J. Atmos. Sci.*, *63*, 1587–1604, doi:10.1175/JAS3703.1.
- Mu, M., H. Xu, and W. Duan (2007a), A kind of initial errors related to “spring predictability barrier” for El Niño events in Zebiak-Cane model, *Geophys. Res. Lett.*, *34*, L03709, doi:10.1029/2006GL027412.
- Mu, M., W. Duan, and B. Wang (2007b), Season-dependent dynamics of nonlinear optimal error growth and El Niño–Southern Oscillation predictability in a theoretical model, *J. Geophys. Res.*, *112*, D10113, doi:10.1029/2005JD006981.
- Nan, S., and J. Li (2003), The relationship between the summer precipitation in the Yangtze River valley and the boreal spring Southern Hemisphere annular mode, *Geophys. Res. Lett.*, *30*(24), 2266, doi:10.1029/2003GL018381.
- Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke (2004), A local ensemble Kalman filter for atmospheric data assimilation, *Tellus, Ser. A*, *56*, 415–428.
- Panofsky, H. (1949), Objective weather-map analysis, *J. Appl. Meteorol.*, *6*, 386–392.
- Pu, Z.-X., and S. A. Braun (2001), Evaluation of bogus vortex techniques with four-dimensional variational data assimilation, *Mon. Weather Rev.*, *129*, 2023–2039, doi:10.1175/1520-0493(2001)129<2023:EOBVTW>2.0.CO;2.
- Pu, Z.-X., and W.-K. Tao (2004), Mesoscale assimilation of TMI rainfall data with 4DVAR: Sensitivity studies, *J. Meteorol. Soc. Jpn.*, *82*(5), 1389–1397, doi:10.2151/jmsj.2004.1389.
- Pu, Z.-X., X. Li, C. S. Velden, S. D. Abersson, and W. T. Liu (2008), The impact of aircraft dropsonde and satellite wind data on numerical simulations of two landfalling tropical storms cloud systems and processes experiment, *Weather Forecasting*, *23*, 62–79.
- Qiu, C., and J. Chou (2006), Four-dimensional data assimilation method based on SVD: Theoretical aspect, *Theor. Appl. Climatol.*, *83*, 51–57, doi:10.1007/s00704-005-0162-z.
- Qiu, C., L. Zhang, and A. Shao (2007), An explicit four-dimensional variational data assimilation method, *Sci. China Ser. D*, *50*, 1232–1240.
- Reinhold, B. B., and R. T. Pierrehumbert (1982), Dynamics of weather regimes: Quasi-stationary waves and blocking, *Mon. Weather Rev.*, *110*, 1105–1145, doi:10.1175/1520-0493(1982)110<1105:DOWRQS>2.0.CO;2.
- Shu, Q., M. W. Kemplowski, and M. McKee (2005), An application of ensemble Kalman filter in integral-balance subsurface modeling, *Stochastic Environ. Res. Risk Assess.*, *19*, 361–374, doi:10.1007/s00477-005-0242-8.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers (2005), A description of the advanced research WRF version 2, *NCAR Tech. Note NCAR/TN-468+STR*, Nat. Cent. for Atmos. Res., Boulder, Colo.
- Teman, R. (1991), Approximation of attractors, large eddy simulations and multiscale methods, *Proc. R. Soc., Ser. A*, *434*, 23–29.
- Wallace, J. M., C. Smith, and C. S. Bretherton (1992), Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies, *J. Clim.*, *5*, 561–576, doi:10.1175/1520-0442(1992)005<0561:SVDOWS>2.0.CO;2.
- Wang, B., R. Wu, and T. Li (2003), Atmosphere-warm ocean interaction and its impacts on Asian-Australian monsoon variation, *J. Clim.*, *16*, 1195–1211, doi:10.1175/1520-0442(2003)16<1195:AOIAII>2.0.CO;2.
- Wang, H., and R. Fu (2002), Cross-equatorial flow and seasonal cycle of precipitation over South America, *J. Clim.*, *15*, 1591–1608, doi:10.1175/1520-0442(2002)015<1591:CEFASC>2.0.CO;2.
- Wang, J., J. Li, and J. Chou (2008), Comparison and error analysis of two 4-dimensional singular value decomposition data assimilation schemes (in Chinese), *Chin. J. Atmos. Sci.*, *32*(2), 277–288.
- Wang, X., D. M. Barker, C. Snyder, and T. M. Hamill (2008), A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: Observing System Simulation Experiment, *Mon. Weather Rev.*, in press. (Available at http://www.cdc.noaa.gov/people/xuguang.wang/wrfhybda_part1.pdf)
- Xue, M., M. Tong, and K. K. Droegemeier (2006), An OSSE framework based on the ensemble root Kalman filter for evaluating the impact of data from radar networks on thunderstorm analysis and forecasting, *J. Atmos. Oceanic Technol.*, *23*, 46–66, doi:10.1175/JTECH1835.1.
- Zhang, B., and J. Chou (1992), The applications of empirical orthogonal functions to numerical simulations of climate, *Sci. China Ser. B*, *35*, 92–101.
- Zheng, X. (2003), Data assimilation, an important and challenging research topic for environmental statisticians, in *Sustainable Environments: A Statistical Analysis*, pp. 198–207, Oxford Univ. Press, New York.
- Zhu, J., H. Wang, and M. Kamachi (2002), The improvement made by a modified TLM in 4DVAR with a geophysical boundary layer model, *Adv. Atmos. Sci.*, *19*, 563–582, doi:10.1007/s00376-002-0001-4.

J. Li and J. Wang, National Key Laboratory of Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, P.O. Box 9804, Beijing 100029, China. (ljp@lasg.iap.ac.cn)