

An Economical Approach to Four-dimensional Variational Data Assimilation

WANG Bin^{*1} (王斌), LIU Juanjuan^{1,2} (刘娟娟), WANG Shudong^{1,2} (王曙东), CHENG Wei^{1,2} (成巍),
LIU Juan^{1,2} (刘娟), LIU Chengsi^{1,2} (刘成思), Qingnong XIAO³, and Ying-Hwa KUO⁴

¹*State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics,
Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029*

²*Graduate School of the Chinese Academy of Sciences, Beijing 100049*

³*College of Marine Science, University of South Florida, St. Petersburg, FL 33701, USA*

⁴*Mesoscale and Microscale Meteorology Division, National Center for Atmospheric Research,
Boulder, CO 80307–3000, USA*

(Received 16 July 2009; revised 1 October 2009)

ABSTRACT

Four-dimensional variational data assimilation (4DVar) is one of the most promising methods to provide optimal analysis for numerical weather prediction (NWP). Five national NWP centers in the world have successfully applied 4DVar methods in their global NWP, thanks to the increment method and adjoint technique. However, the application of 4DVar is still limited by the computer resources available at many NWP centers and research institutes. It is essential, therefore, to further reduce the computational cost of 4DVar. Here, an economical approach to implement 4DVar is proposed, using the technique of dimension-reduced projection (DRP), which is called “DRP-4DVar.” The proposed approach is based on dimension reduction using an ensemble of historical samples to define a subspace. It directly obtains an optimal solution in the reduced space by fitting observations with historical time series generated by the model to form consistent forecast states, and therefore does not require implementation of the adjoint of tangent linear approximation.

To evaluate the performance of the DRP-4DVar on assimilating different types of mesoscale observations, some observing system simulation experiments are conducted using MM5 and a comparison is made between adjoint-based 4DVar and DRP-4DVar using a 6-hour assimilation window.

Key words: 4DVar, adjoint, dimension reduction, historical sample, observing system simulation experiment

Citation: Wang, B., J. J. Liu, S. D. Wang, W. Cheng, J. Liu, C. S. Liu, Q. N. Xiao, and Y.-H. Kuo, 2010: An economical approach to four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **27**(4), 715–727, doi: 10.1007/s00376-009-9122-3.

1. Introduction

With a growing number of observations available from remote sensing techniques, it is necessary for scientists to explore assimilation techniques that can take advantage of these asynchronous observations. One of the best choices for this purpose is four-dimensional variational data assimilation (4DVar) because it opti-

mally fits the observations in the assimilation window through the trajectory of the model solution while constraining the output with model dynamics and physics. The introduction of the adjoint method to meteorological data assimilation in the 1980s (e.g., Lewis and Derber, 1985; Le Dimet and Talagrand, 1986; Courtier and Talagrand, 1987) and the incremental approach by Courtier et al. (1994), made 4DVar feasible for

*Corresponding author: WANG Bin, wab@lasg.iap.ac.cn

scientific research and numerical weather predictions (NWP). It produces an optimal increment of initial condition (IC), \mathbf{x}'_a , at the initial time, t_0 , by minimizing the cost function, $J(\mathbf{x}')$, in the assimilation window, $[t_0, t_N]$:

$$\begin{cases} J(\mathbf{x}'_a) = \min_{\mathbf{x}'} J(\mathbf{x}') \\ J(\mathbf{x}') = \frac{1}{2}(\mathbf{x}')^T \mathbf{B}^{-1} \mathbf{x}' + \\ \quad \frac{1}{2}[\mathbf{y}'(\mathbf{x}') - \mathbf{y}'_{\text{obs}}]^T \mathbf{O}^{-1} [\mathbf{y}'(\mathbf{x}') - \mathbf{y}'_{\text{obs}}] \end{cases}, \quad (1)$$

where the superscript T denotes the transpose, and all the bold letters represent column vectors or matrices. Using this increment, one can obtain the optimal IC $\mathbf{x}_a = \mathbf{x}_b + \mathbf{x}'_a$. In Eq. (1), $\mathbf{x}' = \mathbf{x} - \mathbf{x}_b$ is regarded as a L_x -dimensional perturbation of the background field \mathbf{x}_b at t_0 , \mathbf{B} is the $L_x \times L_x$ background error covariance matrix (simply \mathbf{B} -matrix hereinafter), \mathbf{y}'_{obs} is the L_y -dimensional observation increment that is the difference between the observation, \mathbf{y}_{obs} , and the observation simulation, \mathbf{y}_b , of the basic state, and \mathbf{y}' is a simulation of observation increments through the observation operators and the prediction model. The expressions for $\mathbf{y}'_{\text{obs},i}$ and \mathbf{y}'_i are:

$$\begin{cases} \mathbf{y}'_{\text{obs}} = \begin{bmatrix} \mathbf{y}'_{\text{obs},1} \\ \mathbf{y}'_{\text{obs},2} \\ \vdots \\ \mathbf{y}'_{\text{obs},N} \end{bmatrix}, \\ \mathbf{y}' = \mathbf{y}'(\mathbf{x}') = \begin{bmatrix} \mathbf{y}'_1(\mathbf{x}') \\ \mathbf{y}'_2(\mathbf{x}') \\ \vdots \\ \mathbf{y}'_N(\mathbf{x}') \end{bmatrix} = \begin{bmatrix} \mathbf{y}'_1 \\ \mathbf{y}'_2 \\ \vdots \\ \mathbf{y}'_N \end{bmatrix}, \end{cases} \quad (2)$$

where

$$\begin{cases} \mathbf{y}'_i(\mathbf{x}') = \mathbf{y}_i(\mathbf{x}_b + \mathbf{x}') - \mathbf{y}_i(\mathbf{x}_b) \\ \mathbf{y}'_{\text{obs},i} = \mathbf{y}_{\text{obs},i} - \mathbf{y}_i(\mathbf{x}_b) \quad (i = 1, 2, \dots, N), \\ \mathbf{y}_i(\mathbf{x}) = \mathbf{H}[\mathbf{M}_{t_0 \rightarrow t_i}(\mathbf{x}, \tau)] \end{cases}, \quad (3)$$

$$\begin{cases} \mathbf{y}_{\text{obs}} = \begin{bmatrix} \mathbf{y}_{\text{obs},1} \\ \mathbf{y}_{\text{obs},2} \\ \vdots \\ \mathbf{y}_{\text{obs},N} \end{bmatrix}, \\ \mathbf{y}_b = \mathbf{y}(\mathbf{x}_b) = \begin{bmatrix} \mathbf{y}_1(\mathbf{x}_b) \\ \mathbf{y}_2(\mathbf{x}_b) \\ \vdots \\ \mathbf{y}_N(\mathbf{x}_b) \end{bmatrix} = \begin{bmatrix} \mathbf{y}_{b,1} \\ \mathbf{y}_{b,2} \\ \vdots \\ \mathbf{y}_{b,N} \end{bmatrix}. \end{cases} \quad (4)$$

Note that $\mathbf{y}'_{\text{obs},i}$, \mathbf{y}'_i , $\mathbf{y}_{\text{obs},i}$ and $\mathbf{y}_{b,i}$ are s_i -dimensional column vectors at time t_i ($i = 1, 2, \dots, N$; $L_y = \sum_{i=1}^N s_i$), N is the number of observation times in the assimilation window, $\mathbf{x}_i = \mathbf{M}_{t_0 \rightarrow t_i}(\mathbf{x}, \tau)$ is the nonlinear model prediction initiated from t_0 to t_i with \mathbf{x} as the IC and τ as the time step, \mathbf{H}_i is the observation operator at t_i , and \mathbf{O} is the covariance matrix of the observation error of \mathbf{y}_{obs} , composed of \mathbf{O}_i , which is an $s_i \times s_i$ observation error covariance matrix:

$$\mathbf{O} = \begin{bmatrix} \mathbf{O}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{O}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{O}_N \end{bmatrix}. \quad (5)$$

We assume that the observation errors are uncorrelated or white, and thus the \mathbf{O} matrix is diagonal. For the convenience of discussion, we also assume that the relationship between \mathbf{y}'_i and \mathbf{x}' is approximately linear according to the first expression of Eq. (3). So:

$$\mathbf{y}'_i = \mathbf{L}_i \mathbf{x}', \quad (6)$$

where $\mathbf{L}_i = \mathbf{H}'_i \mathbf{H}'_{t_0 \rightarrow t_i}$, \mathbf{H}'_i is the tangent linear operator of \mathbf{H}_i , and $\mathbf{H}'_{t_0 \rightarrow t_i}$ is the tangent linear model of $\mathbf{M}_{t_0 \rightarrow t_i}$. In this way, \mathbf{y}' also becomes an approximately linear variable with respect to \mathbf{x}' according to Eq. (6) and the second expression of Eq. (2). That is:

$$\mathbf{y}' = \begin{bmatrix} \mathbf{L}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{L}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{L}_N \end{bmatrix} \mathbf{x}'. \quad (7)$$

The optimal IC increment, \mathbf{x}'_a , from the above minimizations is consistent with the prediction model and fits the observations in the assimilation window through the model solution's trajectory in an optimal way, thereby playing an increasingly important role in NWP (Daley, 1991; Zou and Kuo, 1996; Wang et al., 2000; Xiao et al., 2000). This assimilation approach has to date been adopted by five operational NWP centers for global analysis (Rabier et al., 2000; Rawlins et al., 2007). However, the establishment of the adjoint model is still a tremendous effort, and the computational cost remains expensive.

Another approach, the ensemble Kalman filter (EnKF), proposed by Evensen (1994) and based on a representation ensemble to estimate background uncertainty, has also attracted much attention. It is comparatively easy to implement since it does not require the adjoint model. Many researchers have reported encouraging results with this approach (e.g. Houtekamer

and Mitchell, 1998; Anderson, 2001; Bishop et al., 2001; Whitaker et al., 2004; Houtekamer et al., 2005; Szunoygh et al., 2005; Pu and Hacker, 2009). As sequential algorithms, asynchronous observations were used at the analysis time (Houtekamer et al., 2005; Szunoygh et al., 2005). All observations can be analyzed together by adopting the time sequence of ensemble states between analysis times to account for model state correlations in time as well as in space (Lorenc, 2003), or exact observation times can be taken into account in a natural way (Hunt et al., 2004). Anderson (2003) mentioned that there are practical limitations on the number of observations to be included. More fundamentally, a finite ensemble size limits the ability to fit detailed observations. It may be better to consider hybrid methods. As Buehner (2005) suggested, the ensembles from the EnKF could be used to provide flow-dependent \mathbf{B} -matrices to 4DVar at no extra computational cost. Hunt et al. (2004) unified the extended Kalman filter and 4DVar approach to introduce a four-dimensional Ensemble Kalman filter (4DEnKF). Liu et al. (2008, 2009) proposed an ensemble-based 4DVar approach that used ensemble forecasts to produce the flow-dependent background error covariance and performed 4DVar minimization to produce the analysis. Zhang et al. (2009) examined the performance of coupling the deterministic 4DVar with an EnKF to produce a superior hybrid approach for data assimilation.

In this paper, a new economical approach to 4DVar is presented using a strategy similar to the implementation of EnKF. The basic principle of the approach is introduced in section 2. The \mathbf{B} -matrix is discussed in section 3, and the historical prediction samples are presented in section 4. These are followed by a direct localization strategy in section 5 and several observing system simulation experiments (OSSEs) in section 6. Finally, a summary and conclusions are provided in section 7.

2. Methodology

In order to efficiently obtain the optimal IC increment, \mathbf{x}'_a , using the minimization given by Eq. (1), we replaced the adjoint technique in MM5 with an easy and fast method. Equation (1) shows that the minimization of classical 4DVar is implemented in the space of model variables (simply as “x-space” hereinafter), which is actually a subspace of the L_x -dimension Euclidean space, \mathbf{E}_{L_x} , where L_x is the length of vector \mathbf{x}' . In general, L_x is a very big number whose value is about 10^6 – 10^8 . The huge size of the x-space is one of the key factors why 4DVar is so expensive. Some techniques have been adopted to speed

up the computations. For example, the ECMWF operational system performs its 4DVar at a lower resolution to reduce the number of degrees of freedom. The minimization for the new approach being reported in the present paper is realized in a much reduced projection space, and as a result the computational cost is reduced greatly. Therefore, the projection method is the focus in describing the new approach.

Before introducing the new approach, it is necessary to reformulate the cost function in Eq. (1) so that:

$$\begin{cases} J(\mathbf{x}') = \frac{1}{2}(\mathbf{x}')^T \mathbf{B}^{-1} \mathbf{x}' + \\ \quad \frac{1}{2}(\tilde{\mathbf{y}}'(\mathbf{x}') - \tilde{\mathbf{y}}'_{\text{obs}})^T (\tilde{\mathbf{y}}'(\mathbf{x}') - \tilde{\mathbf{y}}'_{\text{obs}}), \\ \tilde{\mathbf{y}}'(\mathbf{x}') = \mathbf{R}^{-1} \mathbf{y}'(\mathbf{x}'), \tilde{\mathbf{y}}'_{\text{obs}} = \mathbf{R}^{-1} \mathbf{y}'_{\text{obs}}, \end{cases} \quad (8)$$

where $\tilde{\mathbf{y}}'_{\text{obs}}$ and $\tilde{\mathbf{y}}'$ are called the weighted observation increment and simulation of the weighted observation increment, respectively, and \mathbf{R} is a diagonal matrix representing the observation error standard deviations under the usual assumption that \mathbf{O} is a diagonal matrix only including the observation error variance. Thus, \mathbf{R} and \mathbf{O} satisfy the following equality:

$$\mathbf{O} = \mathbf{R}\mathbf{R}^T. \quad (9)$$

We used $\tilde{\mathbf{y}}'_{\text{obs}}$ and $\tilde{\mathbf{y}}'$ in Eq. (8) because they are non-dimensional, which makes the cost function simpler.

Suppose m samples of $\tilde{\mathbf{y}}' : \tilde{\mathbf{y}}'_1, \tilde{\mathbf{y}}'_2, \dots, \tilde{\mathbf{y}}'_m$ are generated using the observation operator, \mathbf{H} , the matrix of the observation error standard deviations, \mathbf{R} , prediction model, \mathbf{M} , and the IC perturbation samples, $\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_m$, are related by:

$$\begin{aligned} \tilde{\mathbf{y}}'_k &= \tilde{\mathbf{y}}'(\mathbf{x}'_k) = \mathbf{R}^{-1}[\mathbf{H}(\mathbf{M}(\mathbf{x}_b + \mathbf{x}'_k, \tau)) - \\ &\quad \mathbf{H}(\mathbf{M}(\mathbf{x}_b, \tau))] \approx \mathbf{L}_{(\mathbf{x}_b, \tau, \mathbf{R})} \mathbf{x}'_k, \quad (10) \\ &\quad (k = 1, 2, \dots, m) \end{aligned}$$

where

$$\mathbf{L}_{(\mathbf{x}_b, \tau, \mathbf{R})} = \mathbf{R}^{-1} \begin{bmatrix} \mathbf{L}_1 & 0 & \dots & 0 \\ 0 & \mathbf{L}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{L}_N \end{bmatrix}, \quad (11)$$

according to Eq. (7), and \mathbf{H} and \mathbf{M} are defined by:

$$\left\{ \begin{array}{l} \mathbf{H} = \begin{bmatrix} \mathbf{H}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{H}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{H}_N \end{bmatrix}, \\ \mathbf{M} = \begin{bmatrix} \mathbf{M}_{t_0 \rightarrow t_1} & 0 & \cdots & 0 \\ 0 & \mathbf{M}_{t_0 \rightarrow t_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{M}_{t_0 \rightarrow t_N} \end{bmatrix}. \end{array} \right. \quad (12)$$

Note that the assumption of tangent linear approximation here is influenced by nonlinearity of the observation operator and by the length of the assimilation window. For this reason, we generally used a small IC perturbation and a short assimilation window (e.g. 6 hours, or shorter). According to the practice of ensemble forecast, multiple simulations of $\tilde{\mathbf{y}}'$ with slightly different ICs of \mathbf{x}' can reduce the errors introduced by chaos or sensitivity-dependence on the ICs, and thus the ensemble mean of these simulations can offer a better forecast. Inversely, if the observation of $\tilde{\mathbf{y}}'$ is available, a weighted mean of the multiple simulations can be used to approach the observation and determine all the weight coefficients. In this way, the same weight coefficients can be used to obtain the weighted mean of multiple ICs according to the linear relation between $\tilde{\mathbf{y}}'$ and \mathbf{x}' in Eq. (7), which may provide a better IC for the forecast of $\tilde{\mathbf{y}}'$. Following this strategy, suitable IC perturbations, $\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_m$, are chosen to ensure the linear independence of the samples, $\tilde{\mathbf{y}}'_1, \tilde{\mathbf{y}}'_2, \dots, \tilde{\mathbf{y}}'_m$, and then a weighted mean of these samples is used to provide a forecast of $\tilde{\mathbf{y}}'$ as follows:

$$\tilde{\mathbf{y}}' = \alpha_1 \tilde{\mathbf{y}}'_1 + \alpha_2 \tilde{\mathbf{y}}'_2 + \cdots + \alpha_m \tilde{\mathbf{y}}'_m = \mathbf{P}_y \boldsymbol{\alpha}, \quad (13)$$

where \mathbf{P}_y is the $L_y \times m$ -dimension projection matrix composed of the samples $\tilde{\mathbf{y}}'_1, \tilde{\mathbf{y}}'_2, \dots, \tilde{\mathbf{y}}'_m$, and $\boldsymbol{\alpha}$ is a m -dimension column vector whose components are the weight coefficients, $\alpha_1, \alpha_2, \dots, \alpha_m$:

$$\left\{ \begin{array}{l} \mathbf{P}_y = (\tilde{\mathbf{y}}'_1, \tilde{\mathbf{y}}'_2, \dots, \tilde{\mathbf{y}}'_m), \\ \boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_m)^T. \end{array} \right. \quad (14)$$

It is easy to prove that the matrix $\mathbf{P}_y^T \mathbf{P}_y$ is of full rank when and only when the linear independence of the samples $\tilde{\mathbf{y}}'_1, \tilde{\mathbf{y}}'_2, \dots, \tilde{\mathbf{y}}'_m$ is true. According to the approximately linear relationship between \mathbf{x}' and \mathbf{y}' indicated in Eq. (7), the corresponding IC perturbation, \mathbf{x}' , can be expressed as the weighted mean of the

IC perturbations, $\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_m$. That is:

$$\mathbf{x}' = \alpha_1 \mathbf{x}'_1 + \alpha_2 \mathbf{x}'_2 + \cdots + \alpha_m \mathbf{x}'_m = \mathbf{P}_x \boldsymbol{\alpha}, \quad (15)$$

Where \mathbf{P}_x is the projection matrix consisting of the IC perturbations:

$$\mathbf{P}_x = (\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_m). \quad (16)$$

In this case, the original 4DVar defined in the x -space can be implemented in the m -dimension reduced space by minimizing a new cost function:

$$\left\{ \begin{array}{l} \mathbf{x}_a = \mathbf{x}_b + \mathbf{x}'_a = \mathbf{x}_b + \mathbf{P}_x \boldsymbol{\alpha}_a, \\ \tilde{J}(\boldsymbol{\alpha}_a) = \min_{\boldsymbol{\alpha} \in \mathbf{E}_m} \tilde{J}(\boldsymbol{\alpha}), \\ \tilde{J}(\boldsymbol{\alpha}) = \frac{1}{2} \boldsymbol{\alpha}^T \mathbf{B}_\alpha^{-1} \boldsymbol{\alpha} + \\ \quad \frac{1}{2} (\mathbf{P}_y \boldsymbol{\alpha} - \tilde{\mathbf{y}}'_{\text{obs}})^T (\mathbf{P}_y \boldsymbol{\alpha} - \tilde{\mathbf{y}}'_{\text{obs}}). \end{array} \right. \quad (17)$$

The solution to the above minimization problem is expressed as:

$$\boldsymbol{\alpha}_a = (\mathbf{B}_\alpha^{-1} + \mathbf{P}_y^T \mathbf{P}_y)^{-1} \mathbf{P}_y^T \tilde{\mathbf{y}}'_{\text{obs}}. \quad (18)$$

In order for the solution to Eq. (18) be found easily and of high quality, the IC perturbation samples, $\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_m$, are chosen to be only the most important growing modes of the atmosphere, of which the initial uncertainties may develop rapidly. In this way, m , the dimension of the matrices \mathbf{B}_α and $\mathbf{P}_y^T \mathbf{P}_y$, can be a very small number, and the inverse matrix $(\mathbf{B}_\alpha^{-1} + \mathbf{P}_y^T \mathbf{P}_y)^{-1}$ in Eq. (17) can be easily calculated.

3. Estimation of the \mathbf{B}_α matrix

The background error covariance matrix, \mathbf{B} , is a key part of variational data assimilation, and several approaches have been proposed and evaluated in the literature. EnKF uses an ensemble method to estimate \mathbf{B} . Assuming there are m ensemble members in the x -space, \mathbf{B} can then be estimated (Houtekamer et al., 1996; Fisher, 1999) as follows:

$$\left\{ \begin{array}{l} \mathbf{B} \approx \mathbf{b} \mathbf{b}^T, \\ \mathbf{b} = \frac{1}{\sqrt{m-1}} \times \\ \quad (\mathbf{x}'_1 - \bar{\mathbf{x}}', \mathbf{x}'_2 - \bar{\mathbf{x}}', \dots, \mathbf{x}'_m - \bar{\mathbf{x}}'), \\ \bar{\mathbf{x}}' = \frac{1}{m} (\mathbf{x}'_1 + \mathbf{x}'_2 + \cdots + \mathbf{x}'_m). \end{array} \right. \quad (19)$$

In this paper, the historical time series of consistent model forecast samples used to estimate \mathbf{B} with the ensemble method are projected to the reduced subspace. Therefore, the background error covariance matrix, \mathbf{B}_α , in the m -dimension reduced space can be

estimated in the same way:

$$\begin{cases} \mathbf{B}_\alpha = \mathbf{b}_\alpha \mathbf{b}_\alpha^T, \\ \mathbf{b}_\alpha = \frac{1}{\sqrt{m-1}} \times \\ \quad (\alpha_1 - \bar{\alpha}, \alpha_2 - \bar{\alpha}, \dots, \alpha_m - \bar{\alpha}), \\ \bar{\alpha} = \frac{1}{m}(\alpha_1 + \alpha_2 + \dots + \alpha_m), \end{cases} \quad (20)$$

where $\alpha_1, \alpha_2, \dots, \alpha_m$ are, respectively, the projection vectors of $\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_m$ in \mathbf{E}_m , or:

$$\left\{ \begin{array}{l} \alpha_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \\ \alpha_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \\ \dots, \\ \alpha_m = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}. \end{array} \right. \quad (21)$$

Actually, \mathbf{b}_α in Eq. (19) is just the projection of \mathbf{b} onto the sample space, because it is easy to prove the equality $\mathbf{b} = \mathbf{P}_x \mathbf{b}_\alpha$ using Eq. (14). Based on Eq. (20), \mathbf{b}_α is deduced to a very simple constant matrix only with respect to the sample number, m :

$$\mathbf{b}_\alpha = \frac{1}{\sqrt{m-1}} \times \begin{bmatrix} 1 - \frac{1}{m} & -\frac{1}{m} & \dots & -\frac{1}{m} \\ -\frac{1}{m} & 1 - \frac{1}{m} & \dots & -\frac{1}{m} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m} & -\frac{1}{m} & \dots & 1 - \frac{1}{m} \end{bmatrix}_{m \times m} \quad (22)$$

The problem is that \mathbf{b}_α is proven to be a singular matrix, of which the rank is only $(m - 1)$. That means

only $(m - 1)$ degrees of freedom of the m -dimension vector, α , are effective and the dimension of the projection space should be reduced to $(m - 1)$. The reduced rank of \mathbf{b}_α will lead to underestimation of \mathbf{B} . To avoid the dimension reduction of the projection space, a zero IC perturbation sample, $\mathbf{x}'_{m+1} = \mathbf{x}_b - \mathbf{x}_b = 0$, is introduced to build a $(m+1)$ -dimension Euclidean space, \mathbf{E}_{m+1} , in which a $(m+1)$ -dimension matrix, \mathbf{b}_α , with m as its rank could be estimated. After giving up the $(m+1)$ th row and $(m+1)$ th column of the matrix, \mathbf{b}_α , i.e. after abandoning the ‘‘borrowed’’ zero perturbation sample, \mathbf{x}'_{m+1} , \mathbf{b}_α is reduced to a full-rank m -dimension matrix:

$$\mathbf{b}_\alpha = \frac{1}{\sqrt{m}} \times \begin{bmatrix} 1 - \frac{1}{m+1} & -\frac{1}{m+1} & \dots & -\frac{1}{m+1} \\ -\frac{1}{m+1} & 1 - \frac{1}{m+1} & \dots & -\frac{1}{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m+1} & -\frac{1}{m+1} & \dots & 1 - \frac{1}{m+1} \end{bmatrix}_{m \times m} \quad (23)$$

The reconstructed \mathbf{b}_α can improve the aforementioned underestimation caused by the reduced rank. Actually, it is a kind of inflation technique.

4. Preparation of IC perturbation samples

According to the requirements given in section 2, the IC perturbation samples, \mathbf{P}_x , should be properly prepared so that the samples in \mathbf{P}_y are representative. In practice, this is rather difficult to do. This is because the samples in \mathbf{P}_y are required to satisfy the following two constraints: (1) the column vectors of \mathbf{P}_y should be linearly independent or orthogonal with each other: $\det(\mathbf{P}_y^T \mathbf{P}_y) \neq 0$ or $r(\tilde{\mathbf{y}}'_i, \tilde{\mathbf{y}}'_j) = 0$ ($i \neq j, i, j = 1, 2, \dots, m$); and (2) all the column vectors of \mathbf{P}_y are significantly correlated with $\tilde{\mathbf{y}}'_{\text{obs}}$, i.e. $|r(\tilde{\mathbf{y}}'_i, \tilde{\mathbf{y}}'_{\text{obs}})| > r_o$ ($i = 1, 2, \dots, m$), where $r(\mathbf{y}, \mathbf{z})$ represents the correlation coefficient between the vectors \mathbf{y} and \mathbf{z} , and r_o is the threshold of correlation coefficient for some significance level, β (e.g. $\beta = 0.001$), which controls the quality of the matrix \mathbf{P}_y . The larger the threshold, the closer to $\tilde{\mathbf{y}}'_{\text{obs}}$ the weighted mean forecast. Therefore, to prepare samples that are significantly correlated with $\tilde{\mathbf{y}}'_{\text{obs}}$ is very important for the proposed 4DVar.

The proposed method to prepare the IC perturbation samples is somewhat different from that of Qiu and Chou (2006). The main idea of this method is to use the historical time series of consistent model forecasts. For example, in this paper we used two 78-hour

forecasts, respectively initiated from 24 and 48 hours prior to the analysis time to generate the IC perturbation samples (i.e. the \mathbf{x}' samples) and the corresponding $\tilde{\mathbf{y}}'$ samples. Saving the model output every hour, 73 pairs of samples are generated from each forecast when the assimilation window is 6 hours. In this way, 146 pairs of samples can be prepared. Let's mark them as:

$$\begin{cases} \mathbf{X}' = (\hat{\mathbf{x}}'_1, \hat{\mathbf{x}}'_2, \dots, \hat{\mathbf{x}}'_{146}), \\ \mathbf{Y}' = (\hat{\mathbf{y}}'_1, \hat{\mathbf{y}}'_2, \dots, \hat{\mathbf{y}}'_{146}). \end{cases} \quad (24)$$

There are two steps to deal with these samples. The first is quality control: some of the samples of which the absolute value of the correlation coefficient with $\tilde{\mathbf{y}}'_{\text{obs}}$ is larger than or equivalent to the threshold r_0 are chosen. Under the constraint of the threshold, only

$$\begin{cases} \mathbf{E}_y = (\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m), \\ \mathbf{D} = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_m), \end{cases} \quad \mathbf{e}_i^T \mathbf{e}_j = \begin{cases} 1 & \text{when } i = j \\ 0 & \text{when } i \neq j \end{cases} \quad (i, j = 1, 2, \dots, m), \quad (27)$$

and $\mathbf{e}_i (i = 1, 2, \dots, m)$ is an f -dimension column vector. Finally, m pairs of samples for the new 4DVar are produced:

$$\begin{cases} \mathbf{P}_x = (\mathbf{x}'_1, \mathbf{x}'_2, \dots, \mathbf{x}'_m) = \mathbf{X}'_c \mathbf{E}_y, \\ \mathbf{P}_y = (\tilde{\mathbf{y}}'_1, \tilde{\mathbf{y}}'_2, \dots, \tilde{\mathbf{y}}'_m) = \mathbf{Y}'_c \mathbf{E}_y. \end{cases} \quad (28)$$

If necessary, one can further eliminate the samples from \mathbf{P}_y whose correlations with $\tilde{\mathbf{y}}'_{\text{obs}}$ are not significant. According to Eqs. (25) and (27), $\mathbf{P}_y^T \mathbf{P}_y$ is easily proven to be a diagonal matrix:

$$\mathbf{P}_y^T \mathbf{P}_y = \mathbf{D}. \quad (29)$$

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

$$\mathbf{x}'_a = \mathbf{P}_x (\mathbf{B}_\alpha^{-1} + \mathbf{D})^{-1} \mathbf{P}_y^T \tilde{\mathbf{y}}'_{\text{obs}}, \quad (30)$$

using the first expressions of Eqs. (16) and (17).

5. Localization

The new approach formulated in Eq. (30) directly obtains an optimal solution by reducing the dimensions of space to the sample space. The ensemble is composed of far fewer members than both the number of observational data and the degrees of freedom of model variables, which would lead to many spurious correlations between observation locations and model grids. A more practical and easier way to deal

some of the 146 pairs are left:

$$\begin{cases} \mathbf{X}'_c = (\hat{\mathbf{x}}'_{k_1}, \hat{\mathbf{x}}'_{k_2}, \dots, \hat{\mathbf{x}}'_{k_f}), \\ \mathbf{Y}'_c = (\hat{\mathbf{y}}'_{k_1}, \hat{\mathbf{y}}'_{k_2}, \dots, \hat{\mathbf{y}}'_{k_f}), \end{cases} \quad (25)$$

where f is the number of chosen samples ($f \leq 146$). The second step is to find m independent samples so that the correlations between them are as small as possible and that the matrix, \mathbf{P}_y , consisting of them satisfies $\det(\mathbf{P}_y^T \mathbf{P}_y) \neq 0$, or to seek m orthogonal dominant modes from \mathbf{Y}'_c using EOF, where ($m < f$):

$$(\mathbf{Y}'_c)^T \mathbf{Y}'_c = \mathbf{E}_y \mathbf{D} \mathbf{E}_y^T. \quad (26)$$

In the above equation, \mathbf{E}_y is a $f \times m$ -dimension matrix including m orthogonal eigenvectors of $(\mathbf{Y}'_c)^T \mathbf{Y}'_c$, \mathbf{D} is a $m \times m$ -dimension diagonal matrix including m eigenvalues of $(\mathbf{Y}'_c)^T \mathbf{Y}'_c$ on the diagonal line:

with this problem is through the localization technique, which ameliorates the spurious long-range correlations (Houtekamer and Mitchell, 2001). It is easy to realize localization in the new approach, just as in EnKF, although the assimilation is completed in the sample space. This is because the matrix $(\mathbf{B}_\alpha^{-1} + \mathbf{D})^{-1}$ in Eq. (30) will not change the spatial locations of all row vectors of \mathbf{P}_x or \mathbf{P}_y after the matrix multiples by \mathbf{P}_x or \mathbf{P}_y . In general, the number of observational data is less than the degrees of freedom of model variables, and thus the computational cost of $\mathbf{P}_y (\mathbf{B}_\alpha^{-1} + \mathbf{D})^{-1}$ is smaller than that of $\mathbf{P}_x (\mathbf{B}_\alpha^{-1} + \mathbf{D})^{-1}$. For this reason, the authors mark:

$$\hat{\mathbf{P}}_y = \mathbf{P}_y (\mathbf{B}_\alpha^{-1} + \mathbf{D})^{-1}, \quad (31)$$

and rewrite Eq. (30) as follows:

$$\mathbf{x}'_a = \mathbf{P}_x \hat{\mathbf{P}}_y^T \tilde{\mathbf{y}}'_{\text{obs}}. \quad (32)$$

Similar to Houtekamer and Mitchell (2001), the Schur product is applied to the matrix $\mathbf{P}_x \hat{\mathbf{P}}_y^T$ to filter out the remote correlation between observation locations and model grids more continuously, and the final increment of analysis is calculated using the formula:

$$\mathbf{x}'_a = \boldsymbol{\rho} \circ \left(\mathbf{P}_x \hat{\mathbf{P}}_y^T \right) \tilde{\mathbf{y}}'_{\text{obs}}, \quad (33)$$

where the Schur product of two matrices having the same dimension is denoted $\mathbf{A} = \mathbf{B} \circ \mathbf{C}$ and consists of the element-wise product such that $a_{i,j} = b_{i,j} \cdot c_{i,j}$. For providing the formula of the filtering

matrix, ρ , suppose the m -dimensional row vectors of \mathbf{P}_x and $\hat{\mathbf{P}}_y$ are, respectively, $\mathbf{p}_{x,1}, \mathbf{p}_{x,2}, \dots, \mathbf{p}_{x,L_x}$ and $\mathbf{p}_{y,1}, \mathbf{p}_{y,2}, \dots, \mathbf{p}_{y,L_y}$. Then, the matrix $\mathbf{P}_x \hat{\mathbf{P}}_y^T$ can be expressed as:

$$\mathbf{P}_x \hat{\mathbf{P}}_y^T = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,L_y} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,L_y} \\ \vdots & \vdots & \dots & \vdots \\ p_{L_x,1} & p_{L_x,2} & \cdots & p_{L_x,L_y} \end{bmatrix}, \quad (34)$$

and $d_{h,0}$ and $d_{v,0}$ are horizontal and vertical Schur radii, respectively. The matrix ρ filters out the small (and noisy) correlations associated with remote observations through the Schur product. This is the localization strategy. In addition, since ρ is smooth and monotonically decreasing, the Schur product tends to reduce and to smooth the effect of those observations at intermediate distances. The result is to smooth the analysis increments.

6. Observing system simulation experiment

For any given observations, the exact corresponding true state is not usually known, due to errors of the observations in space and time. When the observations are assimilated into the IC, there is no true basis for comparison. To effectively assess a new methodology for data assimilation, an Observation System Simulation Experiment (OSSE) is considered one of the best choices, for it can provide both the “true” state and the corresponding “observation”.

In this study, two OSSEs (OSSE_1 and OSSE_2) are designed to test the proposed method. Rainfall observations are important because the precipitation process depends upon many variables of the model through model dynamics and thermodynamics, and they may improve rainfall forecasts effectively through 4DVar (Zou and Kuo, 1996). Therefore, we use a six-hour assimilation window and focus on the evaluation of assimilating six-hour accumulated rainfall observations by the new 4DVar in OSSE_1. In contrast, conventional (temperature) observations will be assimilated in OSSE_2. The assimilation window is six hours, with observational data obtained at three and

where:

$$p_{i,j} = \mathbf{p}_{x,i} (\mathbf{p}_{y,j})^T \quad (i = 1, 2, \dots, L_x; j = 1, 2, \dots, L_y). \quad (35)$$

Marking the horizontal and vertical distances between the spatial locations of $\mathbf{p}_{x,i}$ and $\mathbf{p}_{y,j}$ as $d_{h,i,j}$ and $d_{v,i,j}$, respectively, then the elements of the matrix ρ can be calculated according to:

$$\rho_{i,j} = C_0(d_{h,i,j}/d_{h,0}) \cdot C_0(d_{v,i,j}/d_{v,0}) \quad (i = 1, 2, \dots, L_x; j = 1, 2, \dots, L_y), \quad (36)$$

where the filtering function C_0 is defined as:

$$C_0(r) = \begin{cases} -\frac{1}{4}r^5 + \frac{1}{2}r^4 + \frac{5}{8}r^3 - \frac{5}{3}r^2 + 1, & 0 \leq r \leq 1 \\ \frac{1}{12}r^5 - \frac{1}{2}r^4 + \frac{5}{8}r^3 + \frac{5}{3}r^2 - 5r + 4 - \frac{2}{3}r^{-1}, & 1 < r \leq 2 \\ 0, & 2 < r \end{cases} \quad (37)$$

six hours after the analysis time.

In order to establish a more realistic framework for these OSSEs, the experiments are designed using the fifth-generation Pennsylvania State University (PSU)–National Center for Atmospheric Research (NCAR) Mesoscale Model (MM5) as the forecast model for the assimilation. This model contains a full set of sub-grid-scale physical parameterizations, including the Dudhia’s simple ice scheme, the Anthes-Kuo cumulus parameterizations scheme, the MRF planetary boundary layer (Hong and Pan, 1996), and the cloud-radiation scheme [see Grell et al. (1994) for details]. For the experiments in the present study, the horizontal resolution of the model is 30 km, and the domain is 120 (grids in longitude) \times 100 (grids in latitude), which covers the whole of China. It had 24 vertical layers from $\sigma = 0$ to $\sigma = 1$.

First, a complete record of the assumed “true” atmosphere state over 30 hours is provided. In order to remove the influence of spin-up, the “truth” is initiated from 12 hours prior to the analysis time at 0000 UTC 13 June 2002, with $1^\circ \times 1^\circ$ National Centers for Environmental Prediction (NCEP) Final (FNL) Global Tropospheric Analyses. This simulation, marked as the “NATURE” run or “reference atmosphere,” is assumed to represent the “true” atmosphere and is used to simulate six-hour accumulated rainfall observations. Although the dataset used to initialize the “truth” is large-scale relative to the model resolution, some mesoscale structures can be generated in the “truth” at or after the analysis time by the model with a 30-km horizontal resolution through a 12-hour or longer spin-up process. For the OSSE to be meaningful, the assumed “observations” are located at 700 stations of

observation locations in the region (21° – 31° N, 105° – 125° E), which are interpolated from the “true” state. The observation errors are assumed to be white in space and time (during the assimilation window), and are given constant error standard deviation (0.2 mm for six-hour accumulated rainfall observation and 0.2 K for temperature observation) in the OSSEs.

The background fields used in OSSE_1 are produced from the $2.5^{\circ} \times 2.5^{\circ}$ ECMWF global analysis, and those in OSSE_2 are from a 24-hour forecast initialized by the $2.5^{\circ} \times 2.5^{\circ}$ ECMWF global analysis at 24 hours prior to the analysis time. The reason a 24-hour forecast is used as the background is to obtain a sizeable difference between the “true” state and the background field. A 30-hour integration through the assimilation window, which is initialized with the background field, is regarded as a control run (marked as “CTRL”), which is used to produce the basic states for assimilation and to evaluate the results from the assimilation experiment.

Second, qualified samples are chosen for assimilation. Two 78-hour forecasts with one output per hour are produced first, initialized with $2.5^{\circ} \times 2.5^{\circ}$ ECMWF global analysis, respectively at 24 and 48 hours prior to the analysis time. Each forecast includes 73 six-hour moving windows. The atmospheric state at the beginning of each window is used to prepare initial perturbation samples, while seven times outputs of atmospheric states contained in each window are applied to generate a sample of weighted observation increment simulation. In total, 146 pairs of samples can be produced from two forecasts and the sample matrices \mathbf{X}' and \mathbf{Y}' are prepared. According to the method described in section 4, the projection matrices \mathbf{P}_x and \mathbf{P}_y can then be obtained.

Third, OSSE_1 is used to compare DRP_4DVar (marked as “ASSM”) with classical 4DVar (marked as “4DVar_adj”). To ensure consistency between ASSM and 4DVar_adj, the experiments use the same “observations” from the “truth”, i.e. the station “observations” of six-hour accumulated rainfall. However, the background error covariance matrices used in these experiments are different because the available MM5 adjoint-based 4DVar system (Zou et al., 1998) only uses a simple and diagonal matrix in the grid-point space. In ASSM, the observational cost function shows a reasonable decrease from 8409 to 2735, noting that in 4DVar_adj the observational penalty decreases nearly one order of magnitude, which implies a much smaller weighting of background due to the use of a diagonal and overestimated \mathbf{B} matrix in the MM5 4DVar system. Note that one of the main purposes of an OSSE is to examine whether an approach can obtain an optimal IC that provides a better description of the

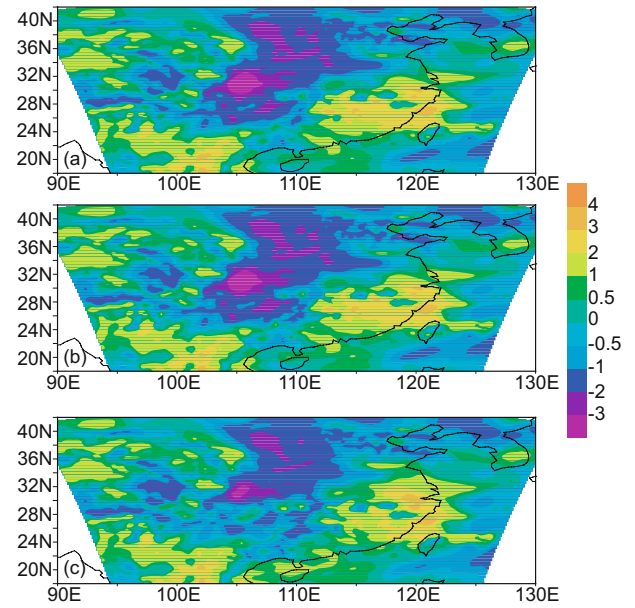


Fig. 1. Temperature difference on the $\sigma = 0.525$ level between the “true” state and (a) CTRL, (b) 4DVar_adj, and (c) ASSM at the analysis time in OSSE_1.

atmospheric state or is closer to the “true” state at the analysis time through incorporating the “observations” into the background. Figure 1 shows that the difference of temperatures on the $\sigma = 0.525$ level at the analysis time between ASSM and “truth” is smaller than those between 4DVar_adj and “truth” and between CTRL and “truth”. A significant decrease of the difference in the region (25° – 32° N, 102° – 112° E) can be found in ASSM compared with CTRL (see Figs. 1, c and a). However, there was no obvious decrease of the difference in 4DVar_adj (see Figs. 1, b and a). It means that the DRP-4DVar improves the accuracy of the IC more effectively than the MM5 adjoint-based 4DVar at the selected layer. Furthermore, to investigate the effectiveness and the overall performance of the DRP-4DVar, we compare the vertical profiles of RMSEs of some basic model variables from CTRL, 4DVar_adj and ASSM at the beginning and the end of the assimilation window, calculated at all horizontal model grid points of each layer. The same conclusion can be drawn that ASSM produces more accurate atmospheric states at the start and the end of the window than 4DVar_adj. As seen in Figs. 2 and 3, the vertical profiles of RMSEs of 4DVar_adj (dashed line with triangles) is much closer to those of CTRL (solid curve with circles), i.e. the accuracy of the simulated atmospheric states is not significantly improved. Meanwhile, it can be seen that ASSM (dashed line with circles) yields a smaller error than CTRL on most layers of the model. This implies that DRP-4DVar not only had a more reasonable fit to the observations, but

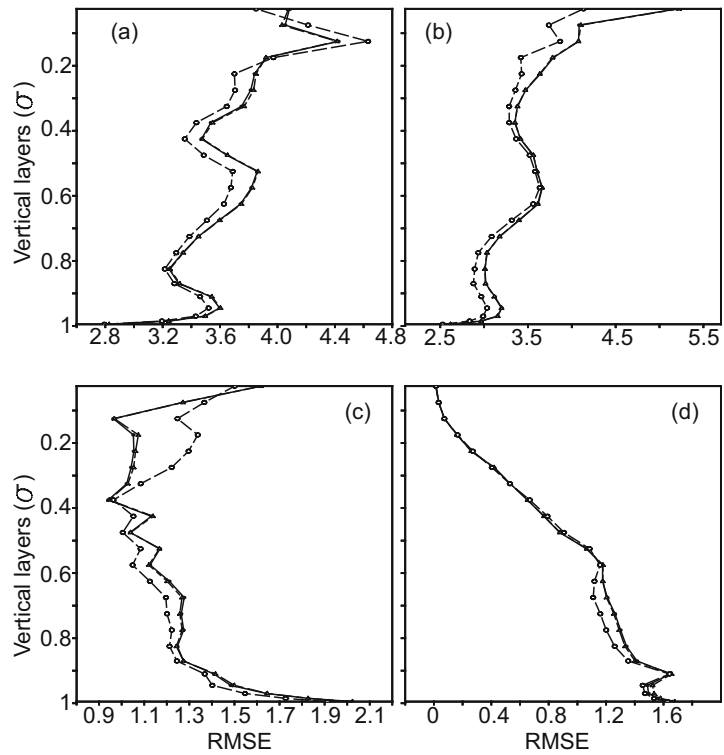


Fig. 2. Vertical profiles of RMSEs of (a) zonal wind (m s^{-1}), (b) meridional wind (m s^{-1}), (c) temperature ($^{\circ}\text{C}$), and (d) water vapor mixing ratio (g kg^{-1}) of CTRL (solid curve with dosed circles), 4DVar_adj (dashed line with triangles) and ASSM (dashed line with open) at the start of the assimilation window in OSSE_1.

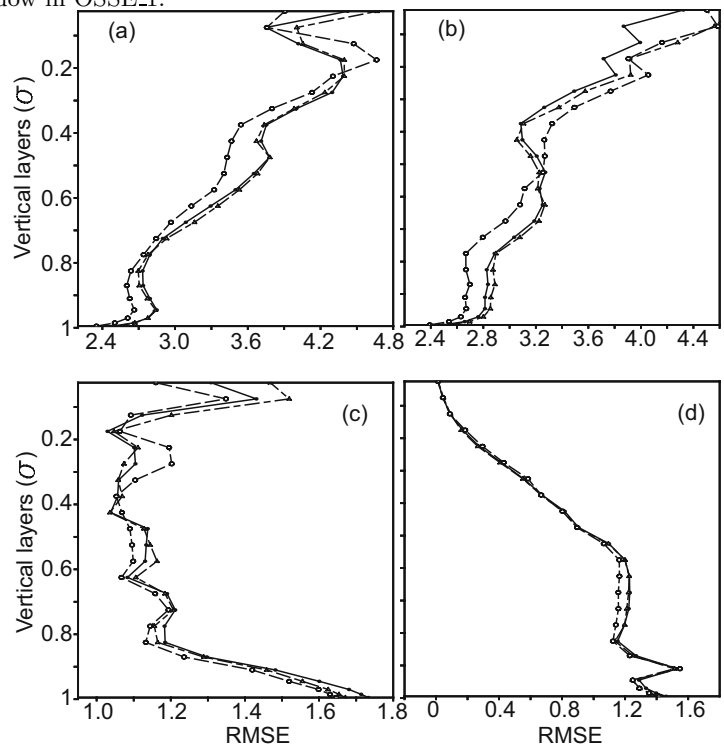


Fig. 3. Same as Fig. 2, except at the end of the assimilation window.

also obtains better atmospheric states in the assimilation window. In particular, meridional wind (V) and the water vapor mixing ratio (q) are improved on almost all layers. Moreover, the significant amelioration is mainly on the middle layers at the beginning and the end of the assimilation window. It seems that DRP-4DVar fits the observations mainly through changing the middle-layer atmospheric circulations, which may be closely associated with the descriptions of the cumulus convection process and the cloud microphysical process in the model. The results also suggest that DRP-4DVar is an effective assimilation approach, as it reduces the errors through benefiting both the immediate vicinity of the observations and the area outside the immediate vicinity [refer to Wee and Kuo (2004)], since the observational error is not negligible and the distribution of observations is sporadic and only at the land surface.

Also compared are the 30-hour forecasts (6 hours in the assimilation window and 24 hours outside of the window) of six-hour accumulated rainfall with the ICs from CTRL, ASSM and 4DVar_adj, respectively. Figure 4 shows that the RMSE of the 30-hour forecast with the IC from ASSM remains smaller than that from CTRL, which means DRP-4DVar can steadily improve the rainfall forecast over 30 hours. The RMSE of the 30-hour forecast with the IC from 4DVar_adj is smaller than that from CTRL most of the time, except at 24 hours. ASSM and 4DVar_adj show comparable performance, as indicated by their RMSEs over the forecast period. Better performance of 4DVar_adj during the assimilation window (i.e. the first 6 hours) can attribute to a larger reduction of its observational cost function using the diagonal and overestimated \mathbf{B} matrix. However, DRP-4DVar is much more timesaving than the MM5 adjoint-based 4DVar. The minimization process of 4DVar_adj was very expensive compu-

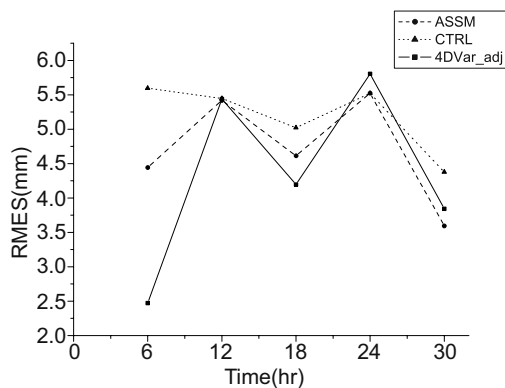


Fig. 4. RMSEs of 30-hour forecast of 6-hour accumulated rainfall from CTRL (dotted line with triangles), ASSM (dashed line with filled circles) and 4DVar_adj (solid line with squares).

tationally, which took 143 minutes for each iteration step. In this experiment, the assimilation was forced to stop after 20 iteration steps, which took a total of 2830 minutes. In contrast, all jobs of ASSM, including the preparation of samples and the assimilation, only took 90 minutes. Note that the above computational costs are all based on a single processor of IBM Personal Computer (PC) with Intel P4 (2.4 GHz, Linux environment).

Fourth, In OSSE_2, two times of temperature “observations” (at three and six hours after the analysis time) on the $\sigma = 0.525$ level are assimilated, and the observation penalty is approximately halved (from 4135 to 2319) in ASSM. It is encouraging that the assimilation also improves the description of the atmospheric state in the IC. Figures 5 and 6 show there are significant improvements in ASSM on the $\sigma = 0.525$ level at the analysis time and at the end of the assimilation window, especially in the region (21° – 31° N, 105° – 125° E). The overall performance can also be evalu-

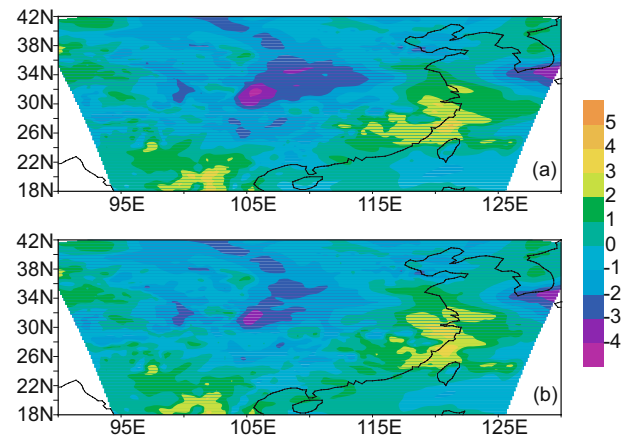


Fig. 5. Temperature difference on the $\sigma = 0.525$ level between the “true” state and (a) CTRL, (b) DRP-4DVar at the analysis time in OSSE_2.

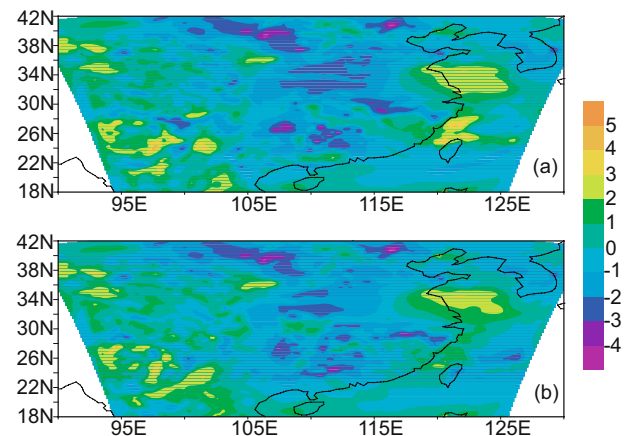


Fig. 6. Same as Fig. 5, except at the end of the assimilation window.

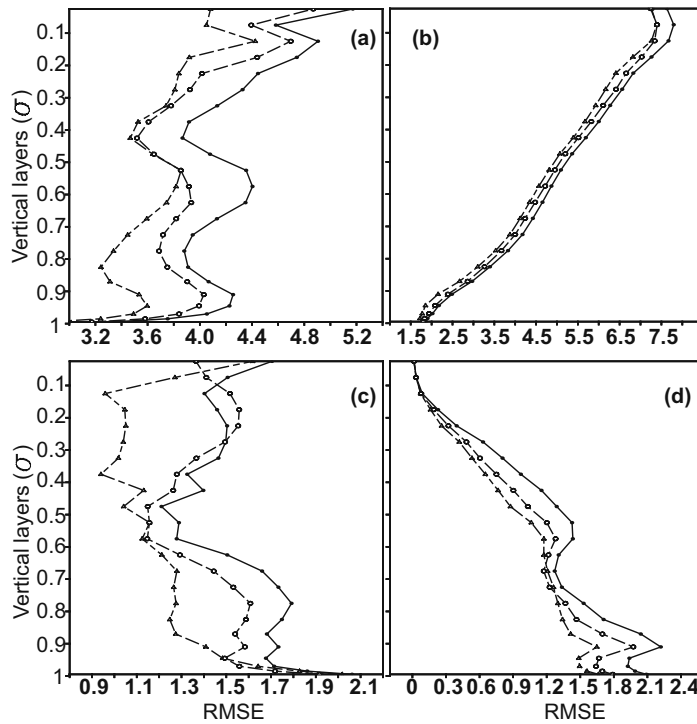


Fig. 7. Vertical profiles of RMSEs of (a) zonal wind (m s^{-1}), (b) vertical velocity (cm s^{-1}), (c) temperature ($^{\circ}\text{C}$), and (d) water vapor mixing ratio (g kg^{-1}) of CTRL (solid curve with filled circles), 4DVar_adj (dashed line with triangles) and ASSM (dashed line with circles) at the start of the assimilation window in OSSE2.

ated by referring to Fig. 7. This figure shows the vertical profiles of RMSEs of 4DVar_adj (dashed line with triangles) and ASSM (dashed line with circles) yield smaller errors than CTRL on most layers of the model. Better performance of 4DVar_adj during the assimilation window (i.e. the first six hours) can attribute to a larger reduction of its observational cost function using the diagonal and overestimated \mathbf{B} matrix. Figure 8 shows that the RMSE of the 24-hour

forecast with the IC from ASSM remains smaller than that from CTRL, which means DRP-4DVar can steadily improve the rainfall forecast over 24 hours. The RMSE of the 24-h forecast with the IC from 4DVar_adj is slightly smaller than that from ASSM at most times, except at six hours. But, basically, ASSM has a performance comparable to that of 4DVar_adj during the forecast period, except the final time. It is understandable that the insufficient sample representativeness in DRP-4DVar caused this error. However, DRP-4DVar is much more timesaving than the MM5 adjoint-based 4DVar, which is the elegant feature of the new approach.

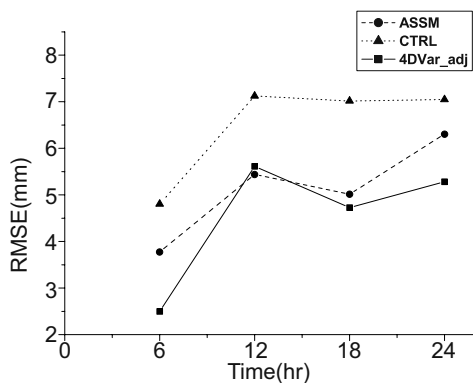


Fig. 8. RMSEs of 24-hour forecast of 6-hour accumulated rainfall from CTRL (dotted line with triangles), ASSM (dashed line with filled circles) and 4DVar_adj (solid line with squares).

7. Summary and conclusions

In this paper, a timesaving approach of 4DVar has been introduced based on the philosophy of dimension reduction projection, DRP-4DVar. The new approach minimizes the cost function of 4DVar in the low-dimension sample space, which is easy to implement and much more timesaving than the iterative-based 4DVar procedure using the adjoint technique. The DRP-4DVar method is summarized by the schematic diagram in Fig. 9.

This new approach shares some common features

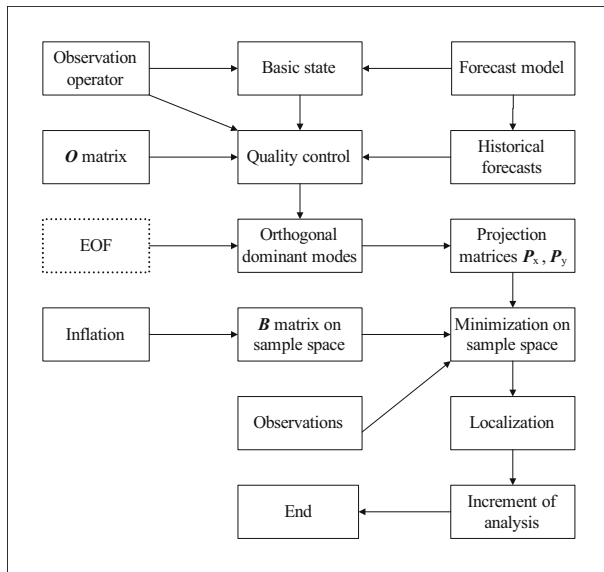


Fig. 9. Schematic diagram of DRP-4DVar.

with EnKF; for example, they both estimate error covariance by ensemble members, obtain solutions on a reduced space, and are model-independent and thus can easily be applied to any other models without the programming cost involved in the adjoint data assimilation system. However, there are also significant differences between these two approaches. First, EnKF is a sequential assimilation approach in which the data and the forecast model are used sequentially in time to produce the analysis fields, while DRP-4DVar is a non-sequential method, just like the classical 4DVar which simultaneously combines the background and all the observations during the assimilation window in an optimal way. Second, the reduced spaces are different. EnKF is performed on the observational space, while DRP-4DVar is implemented on the sample space and thus is smaller and timesaving. Third, DRP-4DVar uses a different technique to diminish the underestimation of background error covariance.

Comparisons between DRP-4DVar and the MM5 adjoint-based 4DVar show that DRP-4DVar not only has a more reasonable fit to the observations but also obtains better atmospheric states in the assimilation window. Both approaches have comparable performance on 30-hour forecasts of six-hour accumulated rainfall, but DRP-4DVar is much more timesaving than the MM5 adjoint-based 4DVar.

Two OSSEs are carried out with the assimilation of six-hour accumulated rainfall observations and temperature respectively, which provides a proper framework for assessing the performance of this new method for data assimilation. The decrease of observational cost function indicates an effective incorporation of six-hour accumulated rainfall observations and con-

ventional observations to the IC through DRP-4DVar and a good fit to these observations through the trajectory of the model solution. The RMSEs of some basic model variables with significant reduction on middle layers and slight decrease on lower layers after the assimilation demonstrate the effectiveness of the new approach, which can improve the model atmosphere at both ends of the assimilation window despite the sporadic distribution of observations in space and time. Not only can it improve simulations around the locations of observations, but it can also benefit data-void regions.

An important advance of the new approach is that the flow-dependent \mathbf{B} matrix is estimated based on a number of IC-reliant historical forecast samples, although its projection on the sample space is a constant and modeled matrix. Further improvement of \mathbf{B} can be achieved by including an analog prediction sample of which the corresponding simulated observation increment is highly correlated with the real observation increment.

Acknowledgements. The authors acknowledge the Ministry of Science and Technology of China for funding the 973 project (Grant No. 2004CB418304), and the Ministry of Finance of China and the China Meteorological Administration for the Special Project of Meteorological Sector [Grant No. GYHY(QX)2007-6-15].

REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.
- Anderson, J. L., 2003: A local least squares framework for ensemble filtering. *Mon. Wea. Rev.*, **131**, 634–642.
- Bishop, C. H., B. Etherton, and S. J., Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, **129**, 420–436.
- Buehner, M., 2005: Ensemble-derived stationary and flow-dependent background error covariances: Evaluation in a quasi-operational setting for NWP setting. *Quart. J. Roy. Meteor. Soc.*, **131**, 1013–1043.
- Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, **120**, 1367–1387.
- Courtier, P., and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Quart. J. Roy. Meteor. Soc.*, **113**, 1329–1347.
- Daley, R., Ed., 1991: *Atmospheric Data Analysis*. Cambridge University Press, Cambridge, UK, 457pp.
- Evensen, G., 1994: Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo

- methods to forecast error statistics. *J. Geophys. Res.*, **99**(C5), 10143–10162.
- Fisher, M., 1999: Background error statistics derived from an ensemble of analyses. ECMWF Research Department Technical Memorandum, No. 79, 1–12.
- Grell, G. A., J. Dudhia, and D. R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). Tech. Note NCAR/TN-3981STR, NCAR/MMM, Boulder, Colorado/USA, 122pp.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.
- Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.
- Houtekamer, P. L., L. Lefavre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system Simulation Approach to Ensemble Prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.
- Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with the ensemble Kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133**, 604–620.
- Hong, S.-Y., and H.-L. Pan, 1996: Non-local boundary layer vertical diffusion in a medium-range forecast model. *Mon. Wea. Rev.*, **124**, 2322–2339.
- Hunt, B. R., and Coauthors, 2004: Four-dimensional ensemble Kalman filtering. *Tellus*, **56A**, 273–277.
- Le Dimet, F.-X., and O. Talagrand, 1986: Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus*, **38A**, 97–110.
- Lewis, J. M., and J. C. Derber, 1985: The use of the adjoint equation to solve a variational adjustment problem with advective constraints. *Tellus*, **37A**, 309–322.
- Liu, C. S., Q. Xiao, and B. Wang, 2009: An ensemble-based four-dimensional variational data assimilation scheme: Part II: Observing system simulation experiments with the Advanced Research WRF (ARW). *Mon. Wea. Rev.*, **137**, 1687–1704.
- Liu, C. S., Q. Xiao, and B. Wang, 2008: An Ensemble-based four-dimensional variational data assimilation scheme: Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.*, **136**, 3363–3373.
- Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4DVAR. *Quart. J. Roy. Meteor. Soc.*, **129**, 3183–3203.
- Pu, Z. X., and J. Hacker, 2009: Ensemble-based Kalman filters in strongly nonlinear dynamics. *Adv. Atmos. Sci.*, **26**(3), 373–380.
- Qiu, C.-J., and J.-F. Chou, 2006: Four-dimensional data assimilation method based on SVD: Theoretical aspect. *Theoretical and Applied Climatology*, **83**, 51–57.
- Rabier, F., J. Jarvinen, E. Klinker, J. F. Mahfouf, and A. Simmons, 2000: The ECMWF implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.*, **126**, 1143–1170.
- Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. Li, G. W. Inverarity, A. C. Lorenc, and T. J. Payne, 2007: The Met Office global four-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.*, **133**, 347–362.
- Szunoygh, I., E. J. Kostelich, G. Gyarmati, D. J. Patil, B. R. Hunt, E. Kalnay, E. Ott, and J. A. Yorke, 2005: Assessing a local ensemble Kalman filter: Perfect model experiments with the National Centers for Environmental Prediction global model. *Tellus*, **57A**, 528–545.
- Wang, B., X. Zou, and J. Zhu, 2000: Data assimilation and its application. *Proceedings of the National Academy of Sciences, USA*, **97**, 11143–11144.
- Wee, T.-K., and Y.-H. Kuo, 2004: Impact of a digital filter as a weak constraint in MM5 4DVAR: An observing system simulation experiment. *Mon. Wea. Rev.*, **132**, 543–559.
- Whitaker, J. S., G. P. Compo, X. Wei, and T. M. Hamill, 2004: Reanalysis without radiosondes using ensemble data assimilation. *Mon. Wea. Rev.*, **132**, 1190–1200.
- Xiao, Q., X. Zou, and B. Wang, 2000: Initialization and simulation of a landfalling hurricane using a variational bogus data assimilation scheme. *Mon. Wea. Rev.*, **128**, 2252–2269.
- Zhang, F. Q., M. Zhang, and J. A. Hansen, 2009: Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.*, **26**(1), 1–8, doi: 10.1007/s00376-009-0001-8.
- Zou, X., and Y.-H. Kuo, 1996: Rainfall assimilation through an optimal control of initial and boundary conditions in a limited-area mesoscale model. *Mon. Wea. Rev.*, **124**, 2859–2882.
- Zou, X., W. Huang, and Q. Xiao, 1998: A user's guide to the MM5 adjoint modeling system. NCAR Tech. Note TN-437+IA, Boulder, Colorado, 1070pp.