

Article

Evaluating the Role of the EOF Analysis in 4DVar Methods

Xingxia Kou ^{1,2,*}, Zhekun Huang ^{3,4} , Hongnian Liu ³, Meigen Zhang ^{2,*}, Si Shen ⁵ and Zhen Peng ³

¹ Institute of Urban Meteorology, China Meteorological Administration, Beijing 100089, China

² State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China

³ School of Atmospheric Sciences, Nanjing University, Nanjing 210093, China; zhekunhuang@163.com (Z.H.); liuhn@nju.edu.cn (H.L.); pengzhen@nju.edu.cn (Z.P.)

⁴ Huaxi Securities CO., LTD, Shenzhen 518048, China

⁵ National Center for Atmospheric Research, Boulder, MN 80302, USA; shensilasg@gmail.com

* Correspondence: xxkou@ium.cn (X.K.); mgzhang@mail.iap.ac.cn (M.Z.); Tel.: +86-010-6840-0635 (X.K.); +86-010-6237-9620 (M.Z.)

Received: 30 June 2017; Accepted: 9 August 2017; Published: 15 August 2017

Abstract: The four-dimensional variational data assimilation (4DVar) method is one of the most popular techniques used in numerical weather prediction. Nevertheless, the needs of the adjoint model and the linearization of the forecast model largely limit the wider applications of 4DVar. 4D ensemble-variational data assimilation methods (4DEnVars) exploit the strengths of the Ensemble Kalman Filter and 4DVar, and use the ensemble trajectories to directly estimate four-dimensional background error covariance. This study evaluates the role of the empirical orthogonal function (EOF) analysis in 4DEnVars. The widely-recognized 4DVar method DRP-4DVar (the Dimension-reduced projection 4DVar) is adopted as the representation of EOF analyses in this study. The performance of the Dimension-reduced projection 4DVar (DRP-4DVar), 4DEnVar (i.e., another traditional 4DEnVar scheme without EOF transformation), and the Ensemble Transform Kalman Filter (ETKF) was compared to demonstrate the effect of the EOF analysis in DRP-4DVar. Sensitivity experiments indicate that EOF analyses construct basis vectors in eigenvalue space and the dimension reduction in the DRP-4DVar approach helps improve computational efficiency and analysis accuracy. When compared with 4DEnVar and the ETKF, the DRP-4DVar demonstrates similar analysis root-mean-square error (RMSE) to 4DEnVar, whereas it surpasses the ETKF by 22.3%. In addition, sensitivity experiments of DRP-4DVar on the ensemble size, the assimilation window length, and the standard deviation of the initial perturbation imply that the DRP-4DVar with the optimized EOF truncation number is robust to a wide range of the parameters, but extremely low values should be avoided. The results presented here suggest the potential wide application of EOF analysis in the hybrid 4DEnVar methods.

Keywords: 4DEnVar; ensemble transform kalman filter; dimension-reduced projection 4DVar; space transformation

1. Introduction

The four-dimensional variational data assimilation (4DVar) method, the Ensemble Kalman Filter (EnKF), and “hybrid” methods are the most promising methods to provide optimal analysis for numerical weather prediction (NWP) [1,2]. The traditional 4DVAR technique could optimally fit observations at multiple times through the trajectory of the model solution while constraining the output with model dynamics and physics, but it requires the adjoint and tangent linear approximation of the forecast model due to the implicit expression of the control variables in the cost function. On the

other hand, the usual EnKF is easier to perform and could provide flow-dependent error estimates of the background errors by forecasting the statistical characteristics, but it lacks the dynamic constraint as in 4DVar. The 4DEnVar methods aim to retain the advantages of 4DVar and EnKF while avoiding the need of an adjoint or tangent linear model of the forecast model, and it has become increasingly popular. Lorenc and Kalnay et al. suggest the encouraging achievement of the 4DEnVar methods in advancing data assimilation [3,4]. Gustafsson also points out that the optimal approach is to combine the best ideas of 4DVar and EnKF [5]. Previous studies have evaluated the performance of 4DVar and EnKF with comparing experiments by Caya et al. in a cloud-resolving model, Yang et al. in a quasi-geostrophic model, Buehner et al. in global deterministic NWP, Poterjoy et al. in the Weather Research and Forecasting (WRF) model, Miyoshi et al. in the Japan Meteorological Agency's operational global analysis and prediction system, and Skachko et al. in a stratospheric chemical transport model [6–12]. With the development of hybrid methods, recent studies have shown favorable development of a hybrid method which combines the advantages of 4DVar and EnKF [8–10,13].

Many hybrid methods have been proposed using 4DVar or EnKF in recent years. Similar to the extension of the three dimensional variational technique (3DVar) to the 4DVar for assimilating asynchronous observations, Evensen and Van Leeuwen, Hunt et al., and Fertig et al. also extend EnKF to the Ensemble Kalman Smoother (EnKS) and a four-dimensional Local Ensemble Transform Kalman Filter (4D-LETKF), respectively [14–16]. With the idea of employing ensemble forecast to provide flow-dependent background error covariance matrices, Zhang et al. coupled 4DVar with EnKF to produce the Ensemble 4DVar method which runs both EnKF and 4DVar, and replaces the ensemble mean of EnKF with the 4DVar analysis [3,17,18].

Both four-dimensional ensemble-variational data assimilation methods (4DEnVars) and En4DVars use a hybrid combination of fixed background error covariance with localized covariance from an ensemble of current forecasts designed to provide flow-dependent error estimates. The fundamental difference between the methods is their modeling of the time evolution of errors: En4DVars also uses a linear model and its adjoint, while 4DEnVar uses a localized linear combination of nonlinear forecasts [19]. Furthermore, the 4DEnVar methods require no adjoint model, therefore they are easy to implement and computationally economical. For example, in Qiu's 4DEnVar, the analysis increment is expressed by a linear combination of basis vectors constructed by the singular value decomposition (SVD), and then the cost function is transformed to be an explicit expression of control variable without the need of an adjoint and tangent linear model [20,21]. Also, Liu et al. presented a 4DEnVar algorithm which directly uses the background ensemble perturbations as basis vectors [22,23]. In addition, Tian et al. developed a POD (proper orthogonal decomposition)-based ensemble four-dimensional variational data assimilation approach (referred to as POD-4DEnVar). The POD-4DVar method exploits the strengths of the EnKF and 4DVar, and the feasibility and effectiveness of POD-4DVar has been demonstrated in an idealized model, WRF, as well as global chemistry transport model Geos-Chem [24–27]. It should be noted that reduced order strategy is a promising way for efficient 4DVar assimilation [28–33], for example, the POD/DEIM developed by Stefanescu et al. [28]. Moreover, Wang et al. developed a dimension-reduced projection 4DVar (DRP-4DVar) method, which employed the empirical orthogonal function (EOF) to project perturbation ensemble to a reduced space spanned by the dominant EOF modes, and then used basis vectors to obtain an analysis increment which optimally combines the observations and backgrounds [34,35]. The above 4DEnVar methods share common features with previous hybrid methods (e.g., [18]), which combine the advantages of the flow-dependent background error covariance matrix (referred to as *B*-matrix) and simultaneous assimilated observations at multiple times.

As one of the representative 4DEnVar approaches, DRP-4DVar has been studied in a series of studies. Liu et al. demonstrated that the DRP-4DVar method reasonably generated a flow-dependent *B*-matrix capturing the weather trends during the assimilation window by using the fifth-generation Pennsylvania State University-National Center for Atmospheric Research Mesoscale Model (MM5) [36]. Zhao and Wang noted that the performance of the DRP-4DVar method has been improved with

the initial perturbation samples derived from the background error covariance of the WRF-3DVar assimilation system [37]. A series of numerical experiments were also conducted by using Lorenz-96 model to investigate the sensitivity of assimilation parameters (e.g., the assimilation window length and the flow-dependent background error covariance) of DRP-4DVar in comparison with the EnKF [38]. Zhao et al. developed a DRP-4DVar assimilation system based on MM5, and successfully assimilated the simulated sea level pressure observations to improve the typhoon-track forecasts in the observing system simulation experiments (OSSEs) [39]. However, previous studies paid little attention to the EOF-, SVD-, or POD-based techniques and their related parameters (e.g., truncation number) in the 4DVar method family [20,24,25,40]. Investigating and understanding the role of EOF is of great importance for the study of these 4DVars since the basis vectors expressing the analysis variables in the 4D space are constructed by the technique. This distinguishes the current paper from most others on hybrid data assimilation, which focus on comparing 4DVar, 4DVar, and 3DVar.

This study focuses on discussing the impact of EOF analysis in the DRP-4DVar method and the optimization of related parameters, by comparison with another traditional 4DVar scheme without the use of EOF transformation (referred to as 4DVar hereafter, [22]) and the Ensemble Transform Kalman Filter (ETKF, [41–43]) in theory and practical application. The ETKF and 4DVar methods share a common feature with the DRP-4DVar method, in that the analysis increments are expressed by expansion of the basis vectors, yet directly use the background ensemble perturbations in 3D space and 4D space, respectively, as the basis vectors. Thus, a slightly unusual presentation of the ETKF, the 4DVar, and the DRP-4DVar formulas is given, emphasizing the theoretical differences, particularly those related to the EOF analysis. The three methods are then applied to the Lorenz-96 model, illustrating the practical outcomes resulting from the theoretical differences. Further, the influence of the truncation number of EOF modes is investigated, since insufficient EOF modes might not represent the spatial structure and temporal evolution of the analysis variables, whereas surfeit EOF modes include noise and cost more computational resources. The analysis of DRP-4DVar, 4DVar, and the ETKF are compared to demonstrate the effect of the EOF analysis in DRP-4DVar. Finally, the sensitivity of the EOF analysis to parameters is examined, which are the ensemble size, the assimilation window length, and the ensemble spread of initial perturbations for each assimilation window. Clarifying the role of the EOF component and its relevant parameter optimization is helpful to improve the computational efficiency and result accuracy of the DRP-4DVar method. As the DRP-4DVar method is one of the representative approaches in the family, the discussion presented here can also facilitate the understanding and applying of other 4DVar hybrid methods. The nomenclature in this paper follows Recommended Nomenclature for EnVar Data Assimilation Methods by WMO's DAOS WG [38].

2. Mathematical Formulations

2.1. The DRP-4DVar and the 4DVar Methods

By minimizing the incremental format of the standard 4DVar cost function $J(x')$, an optimum increment $(x^\alpha)'$ at the beginning time t_0 can be written as Equation (1) over an assimilation window of length L_w ($L_w = t_N - t_0$). Given the background state $x_{0,b}$ at t_0 (the beginning of the assimilation window) and the observations $y_{obs,i}$ at t_i ($i = 0, 1, \dots, N$), the minimization of the cost function provides the analysis increment and then the analysis initial condition.

$$J(x'_0) = \frac{1}{2} (x'_0)^T \mathbf{B}^{-1} (x'_0) + \frac{1}{2} [y'(x'_0) - y'_{obs}]^T \mathbf{O}^{-1} [y'(x'_0) - y'_{obs}], \quad (1)$$

$$x'_0 = x_0 - x_{0,b}, \quad (2)$$

$$\mathbf{y}'_{obs} = \begin{bmatrix} \mathbf{y}'_{obs,0} \\ \mathbf{y}'_{obs,1} \\ \vdots \\ \mathbf{y}'_{obs,N} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_{obs,0} - H_0(\mathbf{x}_{0,b}) \\ \mathbf{y}_{obs,1} - H_1(M_{t_0 \rightarrow t_1}(\mathbf{x}_{0,b})) \\ \vdots \\ \mathbf{y}_{obs,N} - H_N(M_{t_0 \rightarrow t_N}(\mathbf{x}_{0,b})) \end{bmatrix}, \quad (3)$$

$$\mathbf{y}'(\mathbf{x}_0') = \mathbf{y}(\mathbf{x}_0) - \mathbf{y}(\mathbf{x}_{0,b}) = \begin{bmatrix} y_0(\mathbf{x}_0) - y_0(\mathbf{x}_{0,b}) \\ y_1(\mathbf{x}_0) - y_1(\mathbf{x}_{0,b}) \\ \vdots \\ y_N(\mathbf{x}_0) - y_N(\mathbf{x}_{0,b}) \end{bmatrix} = \begin{bmatrix} H_0(\mathbf{x}_0) - H_0(\mathbf{x}_{0,b}) \\ H_1(M_{t_0 \rightarrow t_1}(\mathbf{x}_0)) - H_1(M_{t_0 \rightarrow t_1}(\mathbf{x}_{0,b})) \\ \vdots \\ H_N(M_{t_0 \rightarrow t_N}(\mathbf{x}_0)) - H_N(M_{t_0 \rightarrow t_N}(\mathbf{x}_{0,b})) \end{bmatrix}, \quad (4)$$

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

where $\mathbf{x}_{0,b}$ with a dimension of $L_x \times 1$, refers to the background or first guess at the initial time t_0 and \mathbf{B} denote the corresponding \mathbf{B} -matrix with a dimension of $L_x \times L_x$. \mathbf{y}'_{obs} represents the observation increment which is the difference between the observation \mathbf{y}_{obs} and the simulated observation $\mathbf{y}(\mathbf{x}_0)$ of the background $\mathbf{x}_{0,b}$ through the forecast model and the observation operators. $\mathbf{y}'(\mathbf{x}_0')$ represents the simulated observation increment of $\mathbf{x}'_{0,b}$. In the assimilation window $[t_0, t_N]$, $\mathbf{y}'(\mathbf{x}_0')$, $\mathbf{y}'_{obs,i}$ and $\mathbf{y}_{obs,i}$ are $S_i \times 1$ dimensional column vectors at t_i and the number of observation times is $N + 1$. Therefore, \mathbf{y}'_{obs} and $\mathbf{y}'(\mathbf{x}_0')$ have a dimension of $L_y \times 1$, where $L_y = \sum_{i=0}^N S_i$. $M_{t_0 \rightarrow t_i}$ denotes the model forecast integrated from t_0 to t_i . H_i denotes the observation operator at t_i . \mathbf{O} refers to the observation error covariance matrix, which include diagonal matrices \mathbf{O}_i on its diagonal line since observation errors are assumed to be uncorrelated.

The control variable \mathbf{x}'_0 in the incremental format of the 4DVar cost function is connected with \mathbf{x}'_i through forward model, so the gradient as well as the minimization of the cost function with respect to \mathbf{x}'_0 is difficult to compute. 4DVar utilizes optimal control theory [44,45] to minimize the cost function defined over the time interval by using an adjoint model to determine its gradient, but minimizing the cost function is still computationally expensive. Additionally, coding the adjoint model for 4DVar and maintaining the update with the model upgrading require tremendous effort.

The 4DEnVar methods (e.g., DRP-4DVar and 4DEnVar in this paper) substantially reduce computational cost and need no adjoint model [23,34]. The 4DEnVars use \mathbf{X}_{en} and \mathbf{Y}_{en} , ensemble perturbations generated by model integration to construct \mathbf{P}_y and \mathbf{P}_x , the basis vectors, and then express \mathbf{x}'_0 and \mathbf{y}'_b by a linear combination of the basis vectors with weighting coefficients given by α :

$$\mathbf{x}'_0 = \alpha_1 \mathbf{X}_1 + \alpha_2 \mathbf{X}_2 + \dots + \alpha_m \mathbf{X}_m = \mathbf{P}_x \alpha, \quad (6)$$

$$\mathbf{y}'_b = \mathbf{y}'(\mathbf{x}_0') = \alpha_1 \mathbf{Y}_{b,1} + \alpha_2 \mathbf{Y}_{b,2} + \dots + \alpha_m \mathbf{Y}_{b,m} = \mathbf{P}_y \alpha, \quad (7)$$

where m ($1 \leq m \leq K$) is the vector number of the basis, and K is the ensemble size of the perturbation ensemble. Meanwhile, similar to the ensemble methods, 4DEnVars estimate the \mathbf{B} -matrix in the m -dimension space spanned by the basis vectors:

$$\mathbf{B} = \frac{1}{m-1} (\mathbf{P}_x - \bar{\mathbf{P}}_x)(\mathbf{P}_x - \bar{\mathbf{P}}_x)^T = \mathbf{P}_x (\mathbf{b}_\alpha \mathbf{b}_\alpha^T) \mathbf{P}_x^T, \quad (8)$$

$$\bar{\mathbf{P}}_x = \frac{1}{m} \sum_{i=1}^m \mathbf{P}_x^{(i)}, \quad (9)$$

where the matrix \mathbf{b}_α is defined as $\mathbf{P}_x \mathbf{b}_\alpha = 1/\sqrt{m-1} (\mathbf{P}_x - \bar{\mathbf{P}}_x)$.

The 4DVar methods transform Equation (1) into Equation (10), which means transforming the implicit cost function of \mathbf{x}'_0 into one of the weighting coefficient matrix $\boldsymbol{\alpha}$, and then derives the explicit solutions of the analysis increment $\mathbf{x}'_{0,a}$ and its weighting coefficients matrix $\boldsymbol{\alpha}_{0,a}$. Thus, the iterative minimization and the adjoint model are no longer needed, and the computational cost is substantially reduced. With the information of observation increment \mathbf{y}'_{obs} , $\boldsymbol{\alpha}_{0,a}$ is derived in the space spanned by the basis vectors (Equation (11)).

$$J(\boldsymbol{\alpha}) = \frac{1}{2} \boldsymbol{\alpha}^T (\mathbf{b}_\alpha \mathbf{b}_\alpha^T)^{-1} \boldsymbol{\alpha} + \frac{1}{2} (\mathbf{y}'_{obs} - \mathbf{P}_y \boldsymbol{\alpha})^T \mathbf{O}^{-1} (\mathbf{y}'_{obs} - \mathbf{P}_y \boldsymbol{\alpha}), \quad (10)$$

$$\boldsymbol{\alpha}_{0,a} = [(\mathbf{b}_\alpha \mathbf{b}_\alpha^T)^{-1} + \mathbf{P}_y^T \mathbf{O}^{-1} \mathbf{P}_y]^{-1} \mathbf{P}_y^T \mathbf{O}^{-1} \mathbf{y}'_{obs}, \quad (11)$$

$$\mathbf{x}'_{0,a} = \mathbf{P}_x \boldsymbol{\alpha}_{0,a}. \quad (12)$$

Finally, the analysis $\mathbf{x}_{0,a}$ at the time t_0 is obtained by Equation (13), and then is employed as the initial condition for the forecast of the next time. Equation (13) indicates that the analysis initial condition $\mathbf{x}_{0,a}$ is derived through a linear combination of the basis vectors with weighting coefficients given by $\boldsymbol{\alpha}_{0,a}$.

$$\mathbf{x}_{0,a} = \mathbf{x}_{0,b} + \mathbf{P}_x \boldsymbol{\alpha}_{0,a}. \quad (13)$$

The 4DVar method directly uses \mathbf{X}_{en} and \mathbf{Y}_{en} as \mathbf{P}_y and \mathbf{P}_x , respectively. All information about the ensemble perturbations is kept and the analysis state is performed in the K -dimension space spanned by the ensemble perturbations.

$$\mathbf{X}_{en} = \mathbf{x}_{en} - \mathbf{x}_{0,b}, \quad (14)$$

$$\mathbf{Y}_{en} = \mathbf{y}_{en} - \mathbf{y}_b, \quad (15)$$

where \mathbf{y}_b refers to $\mathbf{y}(\mathbf{x}_{0,b})$, \mathbf{x}_{en} and \mathbf{y}_{en} are ensemble of the prior members and the simulated observations, with a dimension of $L_x \times K$ and $L_y \times K$, respectively. The \mathbf{X}_{en} and \mathbf{Y}_{en} are generated by either historical sampling (e.g., [34]) or ensemble forecast for the assimilation window (e.g., [36,37,46]), and this study focuses on the latter scenario. Further, the background ensemble perturbations $\bar{\mathbf{X}}_{en}$ are assumed to be a Gaussian distribution with $\bar{\mathbf{X}}_{en} = 0$ and given standard deviations. The background covariance \mathbf{B} is approximately estimated by

$$\mathbf{B} = (K - 1)^{-1} \mathbf{X}_{en} (\mathbf{X}_{en})^T. \quad (16)$$

In the DRP-4DVar method, the basis vectors \mathbf{P}_y and \mathbf{P}_x are generated by projecting \mathbf{X}_{en} and \mathbf{Y}_{en} to the reduced dimension space spanned by \mathbf{U}_y (a $K \times m$ matrix), the leading m th EOFs of \mathbf{Y}_{en} .

$$\mathbf{P}_y = [\mathbf{Y}_{b,1}, \mathbf{Y}_{b,2}, \dots, \mathbf{Y}_{b,m}]_{L_y \times m} = \mathbf{Y}_{en} \mathbf{U}_y, \quad (17)$$

$$\mathbf{P}_x = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_m]_{L_x \times m} = \mathbf{X}_{en} \mathbf{U}_y. \quad (18)$$

The EOF analysis here reduces the dimension of \mathbf{Y}_{en} , through keeping the first m th EOF that make $\mathbf{Y}_{en} \mathbf{U}_y$ have maximum variability [41]. The validation of Equation (18) is under the tangent linear approximation. The contribution of the EOFs eigenvalues to the variance is measured by the corresponding eigenvalues in a $m \times m$ matrix $\boldsymbol{\Lambda}^2 = \text{Diag}(\lambda_1^2, \dots, \lambda_m^2)$.

$$\mathbf{Y}_{en}^T \mathbf{Y}_{en} = \mathbf{U}_y \boldsymbol{\Lambda}^2 \mathbf{U}_y^T, \quad (19)$$

$$\mathbf{U}_y = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m]_{K \times m}. \quad (20)$$

Therefore, the EOF analysis in the DRP-4DVar reduces the K -dimensional space to a smaller m -dimensional space, and filters the noise or redundant components that contributing little to the

variability. In addition, \mathbf{b}_α (Equation 8) is a singular matrix with a rank of $m - 1$, and leads to an underestimation of \mathbf{B} . To ameliorate such problem, an improved \mathbf{b}_α as Equation (21) is proposed by Wang et al. [34], and is proved necessary by Liu et al. [30].

$$\mathbf{b}_\alpha = \frac{1}{\sqrt{m}} \begin{bmatrix} 1 - \frac{1}{m+1} & -\frac{1}{m+1} & \cdots & -\frac{1}{m+1} \\ -\frac{1}{m+1} & 1 - \frac{1}{m+1} & \cdots & -\frac{1}{m+1} \\ \vdots & \vdots & \ddots & \vdots \\ -\frac{1}{m+1} & -\frac{1}{m+1} & -\frac{1}{m+1} & 1 - \frac{1}{m+1} \end{bmatrix}_{m \times m}. \quad (21)$$

2.2. The ETKF

The ETKF [41–43] shares the common feature with 4DEnVars, in that the analysis is expressed with weight coefficients that linearly combine the basis vectors, though the basis vectors are simply the background ensemble in three-dimensional space and the ensemble perturbations are updated as well. In particular, matrices and vectors related to L_y have $L_y = L_x$ in the following equations at analysis time.

At an analysis time, the background ensemble \mathbf{x}_b is represented by

$$\mathbf{x}_b = \bar{\mathbf{x}}_b + \mathbf{X}_{en,b}, \quad (22)$$

where $\bar{\mathbf{x}}_b$ is a column vector containing the mean of the background ensemble and $\mathbf{X}_{en,b}$ is a matrix whose columns are the background ensemble perturbations from the ensemble mean. For the mean analysis, the increments and \mathbf{y}'_b are represented by the linear combination of the background ensemble perturbations with the weight coefficients given by $\bar{\mathbf{w}}$

$$\mathbf{x}' = w_1 \mathbf{X}_1 + w_2 \mathbf{X}_2 + \dots + w_k \mathbf{X}_k = \mathbf{X}_{en,b} \bar{\mathbf{w}}, \quad (23)$$

$$\mathbf{y}'_b = w_1 \mathbf{Y}_{b,1} + w_2 \mathbf{Y}_{b,2} + \dots + w_k \mathbf{Y}_{b,k} = \mathbf{Y}_{en,b} \bar{\mathbf{w}}. \quad (24)$$

Since \mathbf{B} is estimated in the ETKF as Equation (25), the cost function for the analysis time then becomes an explicit function of weighting coefficient matrix $\bar{\mathbf{w}}$.

$$\mathbf{B} = (K - 1)^{-1} \mathbf{X}_{en,b} (\mathbf{X}_{en,b})^T, \quad (25)$$

$$J(\bar{\mathbf{w}}) = \frac{K - 1}{2} \bar{\mathbf{w}}^T \bar{\mathbf{w}} + \frac{1}{2} (\mathbf{y}'_{obs} - \mathbf{Y}_{en,b} \bar{\mathbf{w}})^T \mathbf{O}^{-1} (\mathbf{y}'_{obs} - \mathbf{Y}_{en,b} \bar{\mathbf{w}}). \quad (26)$$

Computing the gradient of the $J(\bar{\mathbf{w}})$ with respect to $\bar{\mathbf{w}}$ for the minimization, the weight coefficients for the analysis mean $\bar{\mathbf{w}}_a$ and the transform matrix $\tilde{\mathbf{B}}_a$ are obtained. It should be noted that the ETKF method transforms forecast perturbations into analysis perturbations by a transformation matrix $\tilde{\mathbf{B}}_a$ and then use $\tilde{\mathbf{B}}_a$ to solve the analysis state [41–43]. Similar to Equation (11), Equation (27) indicates the $K \times 1$ vector of weight $\bar{\mathbf{w}}_a$ is derived with information about observation increments, $\mathbf{y}_{obs} - H(\bar{\mathbf{x}}_b)$

$$\bar{\mathbf{w}}_a = (\tilde{\mathbf{B}}_a) (H(\mathbf{x}_b)^T \mathbf{O}^{-1} (\mathbf{y}_{obs} - H(\bar{\mathbf{x}}_b))), \quad (27)$$

$$\tilde{\mathbf{B}}_a = [(K - 1) \mathbf{I} + \mathbf{Y}_{en,b} \mathbf{O}^{-1} \mathbf{Y}_{en,b}]^{-1}. \quad (28)$$

The mean analysis is then computed from the background ensemble mean and a linear combination of the background ensemble perturbations.

$$\bar{\mathbf{x}}_a = \bar{\mathbf{x}}_b + \mathbf{x}'_a = \bar{\mathbf{x}}_b + \mathbf{X}_{en,b} \bar{\mathbf{w}}_a. \quad (29)$$

In addition, the analysis ensemble perturbations are converted from the background ensemble perturbation ensemble by the symmetric square root of $(K - 1) \tilde{\mathbf{B}}_a$.

$$\mathbf{W}_a = [(K - 1) \tilde{\mathbf{B}}_a]^{\frac{1}{2}}, \quad (30)$$

$$\mathbf{X}_{en,a} = \mathbf{X}_{en,b} \mathbf{W}_a. \quad (31)$$

Equations (29) and (31) indicate that the analysis ensemble is derived through a linear combination of the background ensemble, with weighting coefficients given by $\bar{\mathbf{w}}_a$ (a $K \times 1$ vector) for the mean analysis, and \mathbf{W}_a (a $K \times K$ matrix) for the analysis perturbations [42]. Therefore, the k -th analysis ensemble member is obtained by

$$\mathbf{x}_{a,k} = \bar{\mathbf{x}}_a + \mathbf{X}_{en,a,k} = \bar{\mathbf{x}}_b + \mathbf{X}_{en,b,k} [\bar{\mathbf{w}}_a + \mathbf{W}_{a,k}], \quad (32)$$

where $\mathbf{X}_{en,b,k}$, $\mathbf{X}_{en,a,k}$, and $\mathbf{W}_{a,k}$ refer to the k -th column corresponding matrices, respectively.

2.3. Comparison among the DRP-4DVar, the 4DEnVar, and the ETKF Methods

As shown in Figure 1, the DRP-4DVar, the 4DEnVar, and the ETKF methods theoretically differ in the construction of basis vectors and temporal evolution of the ensemble perturbations. The 4DEnVar and the ETKF methods directly employ the background ensemble perturbations as the basis vectors, and the major analysis of the two methods takes place in a K -dimensional space spanned by the background ensemble perturbations. $\mathbf{X}_{en,b}$, the ensemble perturbations, can also be regarded as a linear transformation from the original K -dimensional space to one spanned by $\mathbf{X}_{en,b}$. This helps avoid the mathematics conceptual difficulty in viewing them as basis vectors, since the columns of $\mathbf{X}_{en,b}$ are linearly dependent and the \mathbf{B} -matrix has rank at most $K - 1$ [40]. Yet constructing \mathbf{P}_y in 4DEnVar uses ensemble perturbations in four-dimensional space, whereas the ETKF uses ensemble perturbations in three dimensions.

Instead of directly using the background ensemble perturbations, DRP-4DVar employs the EOF analysis to extract the leading EOF vectors, explaining the major variance as the basis vectors. The EOF analysis allows the DRP-4DVar method to transform the basis from the K -dimensional space to a m -dimensional space ($m \leq K$). When $m = K$, the transformation maintains all the information about the ensemble perturbations, and the DRP-4DVar is equivalent to the 4DEnVar. When $m < K$, the DRP-4DVar only uses the leading EOF vectors, and filters the noise components in the ensemble perturbations. The noise components are the EOFs which contribute little to the variability, from the view of the variance analysis [35,47,48]. Further, since the sum of the columns of \mathbf{X}_{en} is zero and only the leading EOFs are retained, the sum of the columns of \mathbf{P}_x tends to be nonzero. The nonzero sum of the columns of \mathbf{P}_x also explains the difference between Equation (8) and Equation (16) (or Equation (25)) which, according to the same definition, estimate the \mathbf{B} -matrix in the DRP-4DVar and the 4DEnVar methods (or the ETKF), respectively. Therefore, when using $m < K$, the EOF analysis in the DRP-4DVar not only transforms the basis to the eigenvalue space, but also filters the noise in the ensemble perturbations.

As to the temporal evolution of the ensemble perturbations, the ETKF refreshes the background ensemble perturbations at each analysis step. The \mathbf{B} -matrix computed from the ensemble perturbations is flow-dependent. Whereas, at each assimilation window, the ensemble perturbations in the 4DEnVars are initialized independently by Gaussian random numbers, and then integrated by the model merely during the assimilation windows. The \mathbf{B} -matrix computed from the ensemble perturbations in the 4DEnVars is only partially flow-dependent. In order to obtain ensemble perturbations representative of the model dynamical structure, ensemble size, assimilation window length, and initial Gaussian random perturbations need to be optimized. Yet, the DRP-4DVar method is expected to be relatively robust because of the positive impact of the EOF analysis.

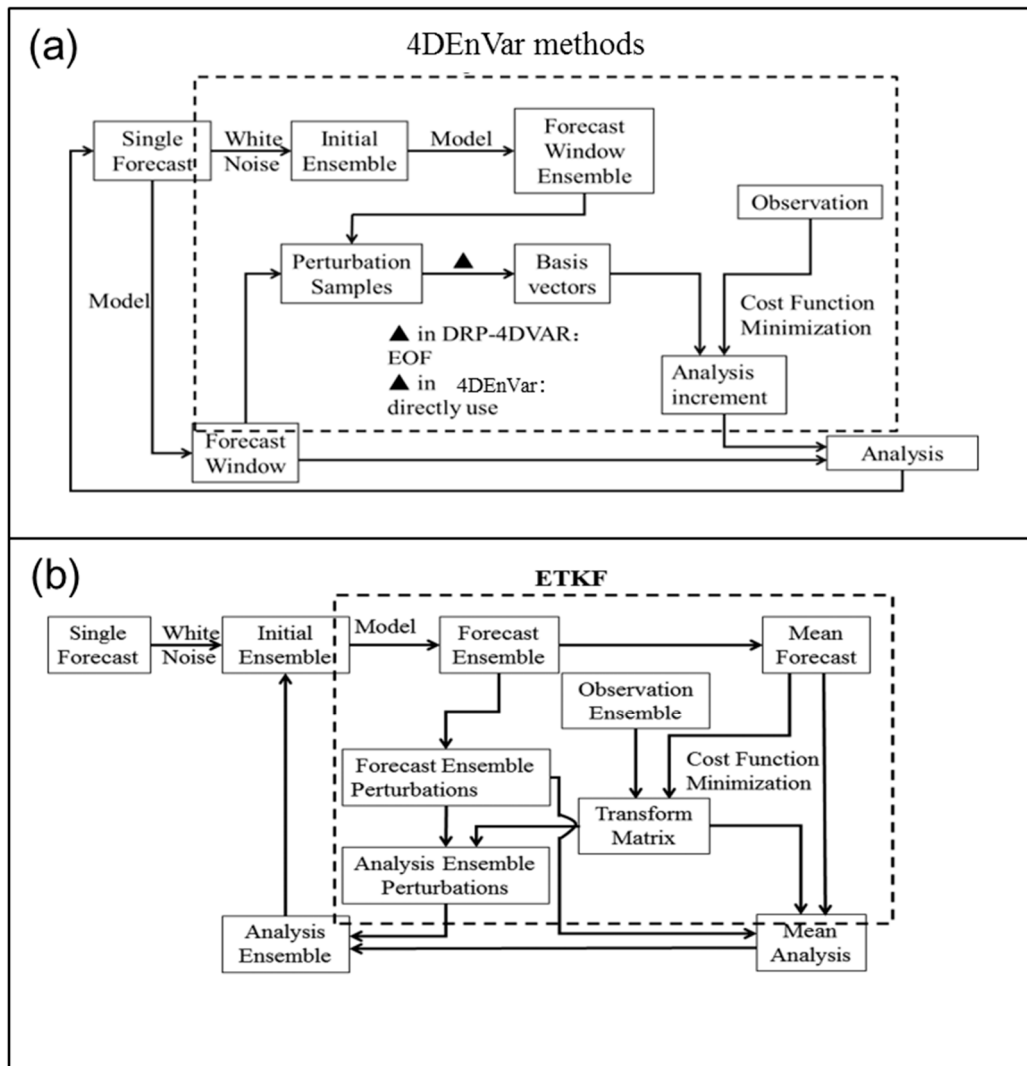


Figure 1. Schematic diagram of the 4DVar method (a) and the ETKF (b).

3. Experiment Design

A series of experiments were conducted using the Lorenz-96 model [49] in order to compare these three assimilation schemes and assess the impact of pertinent parameters. The ideal model simulates the simplified behavior of a meteorological variable in a periodic domain:

$$\frac{dx_j}{dt} = (x_{j+1} - x_{j-2})x_{j-1} - x_j + F, \quad (j = 1, \dots, 40), \quad (33)$$

$$x_{j+40} = x_j. \quad (34)$$

This section may be divided by subheadings. It should provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

Here x_j denotes the variable on the j -th grid point, 40 is the number of the grid points, and F denotes the forcing. The model solves Equation (33) through a fourth-order Runge–Kutta scheme with a time step of 0.05 (corresponding to 6 h). Although the Lorenz-96 model is simple, it is helpful in illustrating important features of assimilation methods and in providing instruction for practical optimization.

The true states include variables at the 40 grids of every model time step during a 1500 time-step integration with forcing $F = 8$. The observations are generated by adding white noise from a Gaussian distribution with zero mean and variance of 1.00 to the true states. The simulated observations consist of the variable at the 40 grid points at every model time step. The observation operators in such a situation are identity matrices.

Several sensitivity experiments are conducted in an imperfect model scenario with forcing parameter $F = 9$. Besides model errors, an initial bias of 2.00 is added to each grid point for examining both methods' capacity of handling error as close as possible to a realistic situation. Analysis is performed every model time step (or 6 h) for 1500 time steps. Exp-1 compares the performance of the DRP-4DVar, the 4DEnVar, and the ETKF methods. For the 4DEnVars, a single simulation is cycled in time through the forecast steps and analysis steps, and the sample ensemble at each analysis time is generated through ensemble forecasting at the beginning of each assimilation window with an ensemble size 80. The random perturbations, which are used to generate the initial ensemble of each assimilation window, are Gaussian distributed with mean of zero and standard deviation 0.10. The assimilation window has a 6 model time-step lengths (corresponding to 36 h). In the DRP-4DVar method, the leading 20 EOF vectors are extracted through the EOF analysis to form the reduced dimension space for major analysis. Meanwhile, the ETKF employs an ensemble with the ensemble size 100.

Both the localization and inflation techniques are essential to ameliorate the contaminations resulting from the spurious correlation over long spatial distances or between variables known to be uncorrelated [50]. For example, Amezcu and Goodliff found that with observations at every time step 4DVar outperforms 4DEnVar, but with localization, the localized 4DEnVar has similar results to the 4DVar [51,52]. To ameliorate problems related to sampling error in this work, the ETKF adopts the multiplicative covariance inflation technique [53], in which the background error covariance is multiplied by $(1 + \Delta)$, and Δ is the inflation factor. The manually optimized value $\Delta = 30\%$ is adopted, which yields the minimum mean root-mean-square errors (RMSEs) among 10%, 15%, 20%, 25%, 30%, 35%, and 40% inflation settings.

Following the Exp-1, sensitivity experiments Exp-2, Exp-3, Exp-4, and Exp-5 (see Table 1) are conducted to further illustrate the role of the EOF analysis and assess the impact of pertinent parameters (i.e., EOF truncation number, ensemble size, assimilation window length, and standard deviation of initial random perturbation for each assimilation window).

Table 1. Settings of sensitivity experiments for parameters related to the EOF analysis in the DRP-4DVar.

| Experiment | Sensitive Parameter | Parameter Setting |
|------------|--|---|
| Exp-2 | EOF truncation number, m | 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75 |
| Exp-3 | Ensemble size, K | 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100 |
| Exp-4 | Assimilation window length, L_w (Unit: hour) | 0, 6, 12, 18, 24, 30, 36, 42, 48, 54 |
| Exp-5 | Standard deviation of initial random perturbation for each assimilation window, SD_i | 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35 |

4. Experimental Results

As shown in Figure 2, the ETKF, 4DEnVar, and DRP-4DVar methods all successfully assimilated the observations and substantially improved result accuracy. The mean analysis RMSEs of the DRP-4DVar, the 4DEnVar, and the ETKF methods are 0.253, 0.310, and 0.386, respectively, which are much lower than observation error (approximately 1.00). Such results indicate that all the assimilation schemes are able to handle the model error and the initial bias. Meanwhile, the ensemble spread of the DRP-4DVar and the ETKF methods provide proper estimations of the uncertainty. The background

spread of the ETKF lies around the mean RMSE and within the \pm standard deviation line, and so does the ensemble spread of the DRP-4DVar method (represented by the mean ensemble spread during the assimilation window). However, the analysis space transforming and the noise filtering of the DRP-4DVar method render positive influence, as the mean analysis RMSE of the DRP-4DVar method is 0.133 (34%) and 0.057 (18%) lower than that of the ETKF and the 4DEnVar methods, respectively.

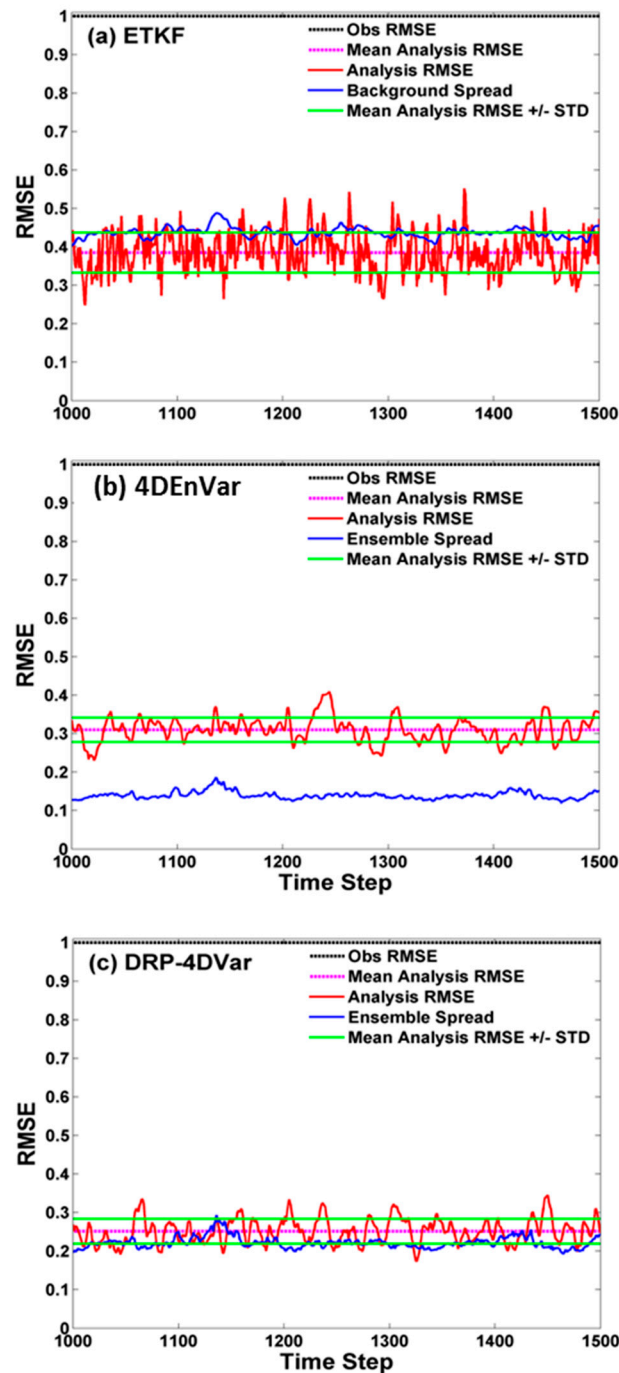


Figure 2. Time series of the analysis RMSE (red solid line) and background spread (blue solid line) of the ETKF experiment (a), the 4DEnVar experiment (b) and the DRP-4DVar (c) in Exp-1. Observation RMSEs (black dash line), mean analysis RMSEs (pink dash line), and Mean analysis RMSE \pm standard deviation of analysis RMSEs (green solid line) are computed over last 500 time steps.

To evaluate the analysis space transformation, a series test of the DRP-4DVar method with different m value is done in Exp-2. The RMSEs of the ETKF and the 4DEnVar methods in Exp-1 are used as benchmarks. Firstly, the difference among the ETKF, the 4DEnVar, and the DRP-4DVar methods with $m = 75$ in Figure 3 shows superiority of the 4DEnVar methods' simultaneous assimilation of asynchronous observations in this study. The DRP-4DVar method with $m = 75$ produces results similar to that of the 4DEnVar method in analysis RMSE (0.300 vs. 0.310), whereas it surpasses the ETKF by 0.086 (22.3%). Figure 4 illustrates that the cumulative sum of the variance explained of the first 75 EOFs is almost 100%. Further, Figure 5 shows that the 4DEnVar and the DRP-4DVar methods with $m = 75$ generate the almost same analysis increment during the assimilation window. Therefore, in accordance with the theoretical analysis in Section 2.3, the EOF analysis plays a role in transforming the basis from an ensemble space to an eigenvalue space. For a certain assimilation time window, Figure 5d–f shows in detail that the 4DEnVar methods tend to give more weight on the background state and smaller analysis increment than the ETKF. Considering the relatively large observation error, this inhibits improper correction of the background state and plausibly explains the better performance of the 4DEnVar method in analysis RMSE.

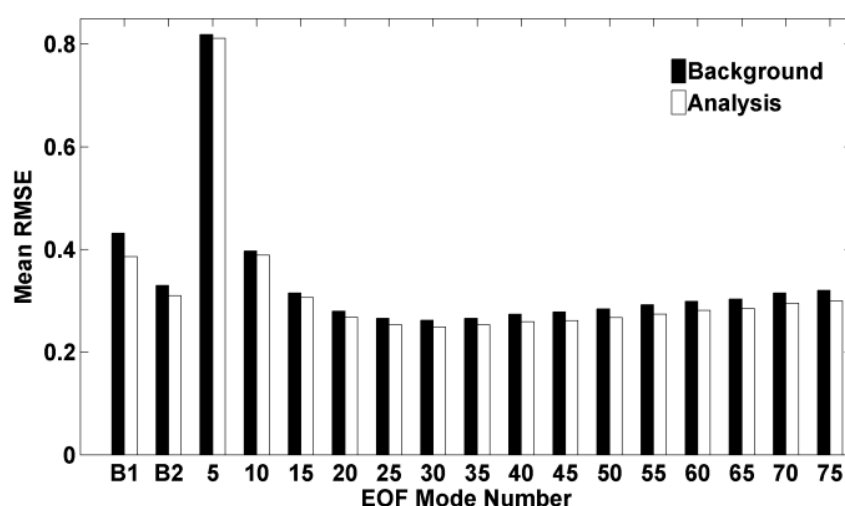


Figure 3. The mean RMSEs of the DRP-4DVar with m set as a series different value from 5 to 75. Other assimilation parameters of the DRP-4DVar experiments are the same as those of the DRP-4DVar experiment in Exp-1. The mean RMSEs of the ETKF and the 4DEnVar methods in Exp-1 are used as benchmarks, which are noted as B1 and B2, respectively. The black bars and white bars denote the mean background RMSE and the mean analysis RMSE during the last 500 time steps, respectively.

For impact of noise filtering, Figure 3 shows that both mean background RMSEs and mean analysis RMSEs in the DRP-4DVar method at first sharply decrease with the increasing of m until m reaches 30, and then increase slowly when m continues to increase. The DRP-4DVar method with a wide range of EOF truncation numbers surpasses the ETKF and the 4DEnVar methods in analysis RMSE. The first dramatic RMSE decrease suggests extremely low m values need to be avoided. Figure 4a shows that the first small part of EOFs corresponds to a large variance explained, whereas the following EOFs correspond to a long tail of decreasing values. This explains the first dramatic RMSE decrease, since every EOF mode of the first small part matters. When m is larger than 30, variances explained by the EOF modes are indistinguishable within their uncertainties. This means that their actual structures may not be particularly interesting, since any linear combination of these patterns is as significant as each of them [47]. When $m = 30$, filtering the noise contributes to a 0.057 analysis RMSE reduction. Figure 4b also shows that the cumulative variance explained increases rapidly above 90% at $m = 15$ and to almost 100% at $m = 35$. To further explain the results, Figure 5a–e illustrates that increasing the EOF truncation number improves the performance if m is small; while increasing the EOF truncation

number brings indistinct impact if m just corresponds to the sufficient variance explained, whereas it causes slight deterioration if m is overlarge. Therefore, choosing a relatively large EOF truncation number improves the accuracy and the optimal EOF truncation number in this study is $m = 30$.

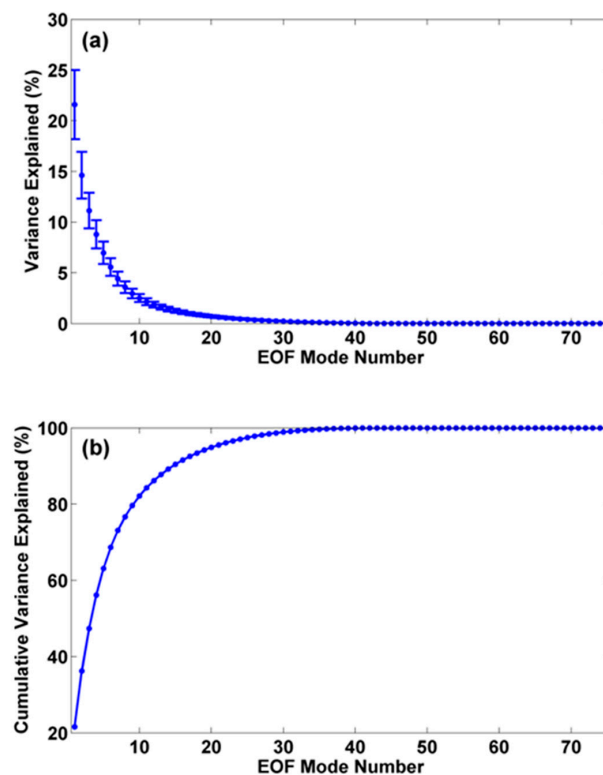


Figure 4. (a). The mean variance explained of the first 75 EOFs during the experimental period, i.e., the mean eigenvalue spectrum of the covariance matrix of the sample ensemble. Vertical bars show approximate 95% confidence limits given by the rule of thumb [54], which also indicate the uncertainty of the variance explained. (b). The cumulative variance explained by the first m th modes. The data are adopted from the DRP-4DVar experiment with $m = 75$ in Exp-2.

Because of the space transforming and noise filtering role of the EOF analysis, the DPR-4DVar method is expected to be relatively robust to ensemble perturbations' representativeness of the model's dynamic structure. At first, results of Exp-3 (see Figure 6), which investigates the sensitivity to ensemble size, shows that proper EOF analysis setting is capable of ameliorating negative influence due to sampling error and generating analysis results with acceptable accuracy. Mean RMSEs of the DRP-4DVar experiments, as well as the RMSE decreasing speed, decreasing with the increasing ensemble size, and the mean RMSEs are indistinct when the ensemble size is larger than 75. The poorest performance of the DRP-4DVar in Exp-3 is still comparable with the 4DEnVar method and much better than the ETKF in Exp-1. This result indicates that the negative effect of the insufficient ensemble size is counteracted by the refinement of the EOF truncation number, since the 4DEnVar method has been proven to be equivalent to the DRP-4DVar method with $m = K$. Additionally, the stability of the performance of the DRP-4DVar method is achieved without additional covariance inflation or localization techniques used in previous studies (e.g., [22,25]). Subsequently, the mean RMSEs of the DRP-4DVar method in Exp-4 (see Figure 7) show a similar pattern to those in Exp-3. The DRP-4DVar method with an optimized EOF analysis setting is able to handle the impact of the assimilation window length, though extremely short window lengths need to be avoided. The RMSEs with assimilation windows longer than 24 h are lower than those of the 4DEnVar method in Exp-1, which is mainly because a longer assimilation window facilitates the development of ensemble perturbations representative of the model dynamics. Yet, the RMSEs' variation due to the increasing assimilation

window length is more dramatic. Extremely short assimilation window lengths lead to failure of the assimilation analysis when the RMSE is larger than 1.00. This result is also in accordance with previous studies on the assimilation window length in 4DVar and the DRP-4DVar methods [7,36].

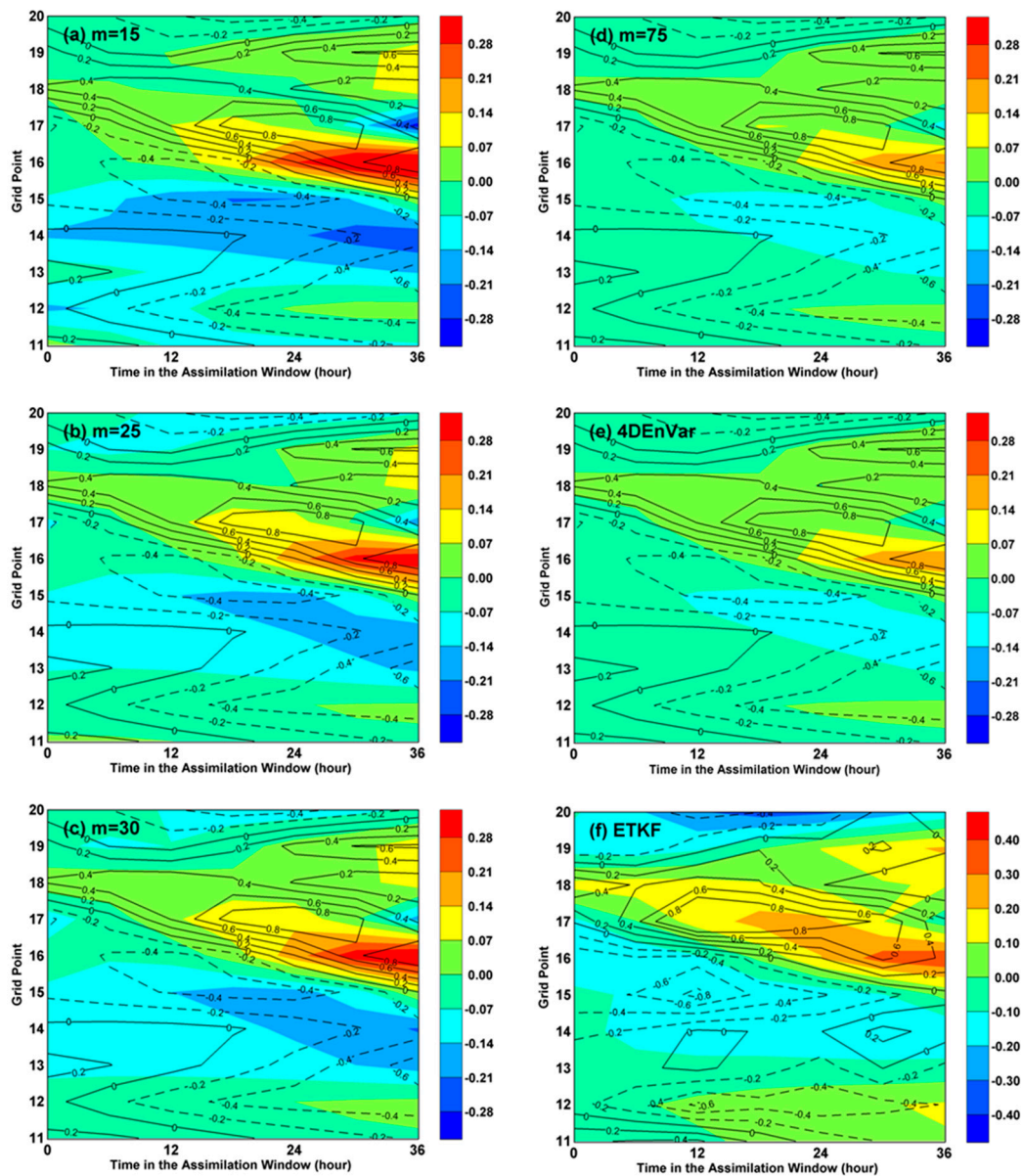


Figure 5. The analysis increments (correction, color shades) and forecast (background) errors (true state minus background mean, contours) of the ETKF, the 4DVar, and the DRP-4DVar methods with different m at the grid points 11~20 during the assimilation window on the 1200th model time step. (a–d) correspond to the DRP-4DVar method with m as 15, 25, 30, 75, (e) correspond to the 4DVar method, and (f) corresponds to the ETKF. The assimilation window length is 36 h (corresponding to 6 time-step length). The analysis y'_b of DRP-4DVar is generated by linear combination of the basis with different m as Equation (7). As to the DRP-4DVar method with the different m and the 4DVar method, the basis vectors and the background errors are computed from the ensemble perturbations from the same original background ensemble during the assimilation window. The original background ensemble is generated by the DRP-4DVar experiment with $m = 30$ in Exp-1. Meanwhile, since the ETKF sequentially assimilates observations, f employs the analysis y'_b and background error during the 1200–1206 time steps (corresponding to the 36 h assimilation window) of the ETKF experiment in EXP-1. Note that f has an order of magnitude different color scale.

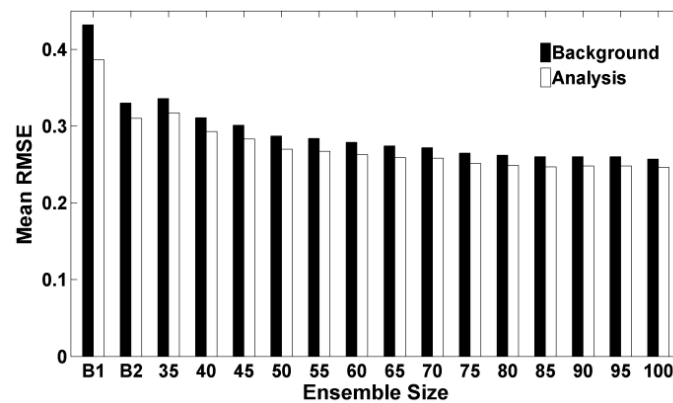


Figure 6. The mean RMSEs of DRP-4DVar with ensemble size of 35–100. Other assimilation parameters of the DRP-4DVar experiments are the same as those of the DRP-4DVar experiment in Exp-1. The mean RMSEs of the ETKF and the 4DEnVar methods in Exp-1 are used as benchmarks, which are noted as B1 and B2, respectively. The black bars and white bars denote the mean background RMSE and the mean analysis RMSE during the last 500 time steps, respectively.

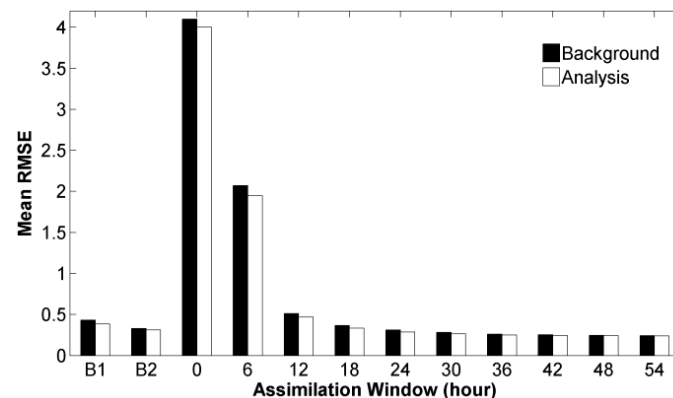


Figure 7. The mean RMSEs of the ETKF (noted as B1), the 4DEnVar (noted as B2), and the DRP-4DVar methods with assimilation window length set as a series of different value from 0 to 54 h. Other assimilation parameters of the DRP-4DVar experiments are the same as those of the DRP-4DVar experiment in Exp-1. The mean RMSEs of the ETKF and the 4DEnVar methods in Exp-1 are used as benchmarks, which are noted as B1 and B2, respectively. The black bars and white bars denote the mean background RMSE and the mean analysis RMSE during the last 500 time steps, respectively.

Exp-5 (see Figure 8) illustrates that the mean RMSEs of the DRP-4DVar method sharply decrease with the increasing of SD_I at first, and then increase slowly when SD_I continues to increase. Most mean RMSEs of the DRP-4DVar method are smaller than that of the 4DEnVar method in Exp-1. This result indicates the DRP-4DVar method is relatively stable to the variation of SD_I . However, similar to the situation of the EOF truncation number, the dramatic RMSE decrease suggests that choosing large SD_I values is relatively safe. Extremely low SD_I ($SD_I = 0.01$) values lead to the failure of assimilation (analysis RMSE = 5.046, not shown). Further, increasing mean ensemble spreads in Figure 8 might explain the pattern of RMSEs. Extremely low SD_I causes the ensemble spreads after model integration to underestimate the background RMSE, which means background states are over-weighted and little information from observations is taken to correct the uncertainty due to the imperfect model and erroneous initial condition. High SD_I leads to the overlarge ensemble spreads after model integration, so the observations are given too much weight even when the background state is much more accurate. Such impact of SD_I resembles that of the standard deviation of the white noise used in the additive covariance inflation [22,55].

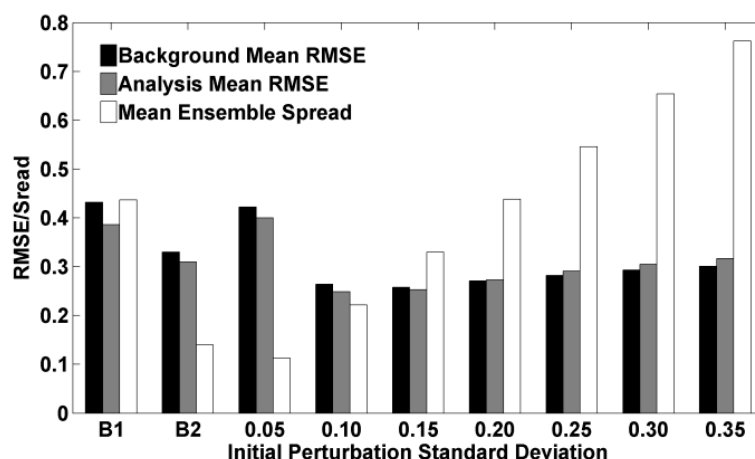


Figure 8. The mean RMSEs and ensemble spread of the DRP-4DVar method using initial random perturbation for each assimilation window with standard deviation of 0.05–0.35. Other assimilation parameters of the DRP-4DVar experiments are the same as those of the DRP-4DVar experiment in Exp-1. The mean RMSEs and ensemble spread of the ETKF and the 4DEnVar methods in Exp-1 are used as benchmarks, which are noted as B1 and B2, respectively. The ensemble spread of the DRP-4DVar and the 4DEnVar methods refers to the average of the ensemble spreads during an assimilation window. The black bars, gray bars, and white bars denote the mean background RMSE, the mean analysis RMSE and the mean ensemble during the last 500 time steps, respectively.

5. Summary and Conclusions

This study investigates the role of EOF analysis in the 4DEnVar scheme DRP-4DVar and the optimization of pertinent parameters, through comparisons among the DRP-4DVar, 4DEnVar, and the ETKF methods. The mathematical formula comparison and sensitivity experiments by using the Lorenze-96 model indicate that the EOF analysis plays an important role in space transforming and noise filtering in DRP-4DVar. The DRP-4DVar method with little dimension reduction produces results similar to that of the 4DEnVar method in analysis root-mean-square error (RMSE), whereas it surpasses the ETKF by 22.3%, and the further refinement of the EOF truncation number magnifies the superiority by 12.2% and slightly reduces the computational cost.

The space transforming and noise filtering role of the EOF analysis improves the computational efficiency and the analysis accuracy. The DRP-4DVar method with a wide range of the EOF truncation number surpasses the ETKF and the 4DEnVar methods in analysis RMSE. This explains the successful applications of the DRP-4DVar method as well as other 4DEnVar methods [20,24,34,38,56], without special optimization of the truncation number. Yet, optimizing the EOF truncation number makes sense. The DRP-4DVar method with the optimized EOF truncation number is robust to the variation of pertinent parameters, though extremely low parameter values should be avoided. Results of the sensitivity experiments on the parameters (i.e., the ensemble size, the assimilation window length, and the standard deviation of the initial perturbation for each assimilation window) also agree with previous studies on EnKF, 4DVar, 4DEnVar, and DRP-4DVar [7,22,36,55,57]. Thus, refinement of the EOF truncation number is attractive for cases with a high requirement of the computational efficiency and the analysis accuracy, particularly for applications on operational numerical system (e.g., [39]).

Clarifying the role of the EOF analysis and its relevant parameter optimization facilitates improvement of the computational efficiency and the analysis accuracy in the DRP-4DVar method, as well as other 4DEnVar methods. The results presented here suggest the potential wide application of EOF analysis in the 4DEnVar methods. Such applications are attractive since the 4DEnVar methods require no tangent linear model or adjoint model, and share the advantages of both EnKF and 4DVar. Our future work will focus on the development of regional chemical data assimilation systems with

the 4DEnVars scheme over East Asia, and comparing it with the current EnKF-based assimilation system [58].

Acknowledgments: This study was supported by the National Key Technologies Research and Development program of China (Grant No. 2016YFC0202103), and the National Natural Science Foundation of China (41575141).

Author Contributions: Meigen Zhang and Hongnian Liu conceived and designed the study; Zhekun Huang and Xingxia Kou performed the experiments; Si Shen and Zhen Peng analyzed the data, Xingxia Kou and Zhekun Huang wrote the paper. All authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Courtier, P.; Thépaut, J.N.; Hollingsworth, A. A strategy for operational implementation of 4DVar, using an incremental approach. *Q. J. R. Meteorol. Soc.* **1994**, *120*, 1367–1387. [[CrossRef](#)]
2. Evensen, G. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **1994**, *99*, 10143–10162. [[CrossRef](#)]
3. Lorenc, A.C. The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Q. J. R. Meteorol. Soc.* **2003**, *129*, 3183–3203. [[CrossRef](#)]
4. Kalnay, E.; Li, H.; Miyoshi, T.; Ballabrera, J. 4DVar or ensemble Kalman filter? *Tellus A* **2007**, *59*, 758–773. [[CrossRef](#)]
5. Gustafsson, N. Discussion on “4DVar or EnKF?”. *Tellus A* **2007**, *59*, 774–777. [[CrossRef](#)]
6. Caya, A.; Sun, J.; Snyder, C. A comparison between the 4DVar and the ensemble Kalman filter techniques for radar data assimilation. *Mon. Wea. Rev.* **2005**, *133*, 3081–3094. [[CrossRef](#)]
7. Yang, S.C.; Corazza, M.; Carrassi, A.; Kalnay, E.; Miyoshi, T. Comparison of local ensemble transform Kalman filter, 3DVAR, and 4DVar in a quasigeostrophic model. *Mon. Wea. Rev.* **2009**, *137*, 693–709. [[CrossRef](#)]
8. Buehner, M.; Houtekamer, P.L.; Charette, C.; Mitchell, H.L.; He, B. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.* **2010**, *138*, 1550–1566. [[CrossRef](#)]
9. Buehner, M.; Houtekamer, P.L.; Charette, C.; Mitchell, H.L.; He, B. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part II: One-month experiments with real observations. *Mon. Wea. Rev.* **2010**, *138*, 1567–1586. [[CrossRef](#)]
10. Poterjoy, J.; Zhang, F.Q. Inter-comparison and coupling of ensemble and variational data assimilation approaches for the analysis and forecasting of Hurricane Karl (2010). *Mon. Wea. Rev.* **2014**, *142*, 3347–3364. [[CrossRef](#)]
11. Miyoshi, T.; Sato, Y.; Kadowaki, T. Ensemble Kalman filter and 4D-Var intercomparison with the Japanese operational global analysis and prediction system. *Mon. Wea. Rev.* **2010**, *138*, 2846–2866. [[CrossRef](#)]
12. Skachko, S.; Errera, Q.; Chabrilat, S.; Ménard, R.; Christophe, Y. Comparison of EnKF and 4DVar methods in the context of stratospheric chemistry data assimilation. In Proceedings of the EGU General Assembly Conference Abstracts, Vienna, Austria, 7–12 April 2013.
13. Fairbairn, D.; Pring, S.R.; Lorenc, A.C.; Roulstone, I. A comparison of 4DVar with ensemble data assimilation methods. *Q. J. R. Meteorol. Soc.* **2013**, *140*, 281–294. [[CrossRef](#)]
14. Evensen, G.; van Leeuwen, P.J. An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.* **2000**, *128*, 1852–1867. [[CrossRef](#)]
15. Hunt, B.R.; Kalnay, E.; Kostelich, E.J.; Ott, E.; Patil, D.J.; Sauer, T.; Szunyogh, I.; Yorke, J.A.; Zimin, A.V. Four-dimensional ensemble Kalman filtering. *Tellus A* **2004**, *56*, 273–277. [[CrossRef](#)]
16. Fertig, E.J.; Harlim, J.; Hunt, B.R. A comparative study of 4D-VAR and a 4D ensemble Kalman filter: Perfect model simulations with Lorenz-96. *Tellus A* **2007**, *59*, 96–100. [[CrossRef](#)]
17. Buehner, M. Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 1013–1043. [[CrossRef](#)]
18. Zhang, F.; Zhang, M.; Hansen, J.A. Coupling ensemble Kalman filter with four-dimensional variational data assimilation. *Adv. Atmos. Sci.* **2009**, *26*, 1–8. [[CrossRef](#)]

19. Lorenc, A.C.; Bowler, N.E.; Clayton, A.M.; Pring, S.R. Comparison of Hybrid-4DVar and Hybrid-4DVar Data Assimilation Methods for Global NWP. *Mon. Wea. Rev.* **2015**, *143*, 212–229. [[CrossRef](#)]
20. Qiu, C.J.; Shao, A.M.; Xu, Q.; Wei, L. Fitting model fields to observations by using singular value decomposition: An ensemble-based 4DVar approach. *J. Geophys. Res.* **2007**, *112*, D11105. [[CrossRef](#)]
21. Golub, G.H.; van Loan, C.F. *Matrix Computations*, 4th ed.; The Johns Hopkins University Press: Baltimore, MD, USA, 2013; pp. 210–260. ISBN 978-1421407944.
22. Liu, C.S.; Xiao, Q.N.; Wang, B. An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test. *Mon. Wea. Rev.* **2008**, *136*, 3363–3373. [[CrossRef](#)]
23. Liu, C.S.; Xiao, Q.N.; Wang, B. An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with Advanced Research WRF (ARW). *Mon. Wea. Rev.* **2009**, *137*, 1687–1704. [[CrossRef](#)]
24. Tian, X.J.; Xie, Z.H.; Dai, A.G. An ensemble-based explicit four-dimensional variational assimilation method. *J. Geophys. Res.* **2008**, *113*, D21124. [[CrossRef](#)]
25. Tian, X.J.; Xie, Z.H.; Sun, Q. A POD based ensemble four-dimensional variational assimilation method. *Tellus A* **2011**, *63*, 805–816. [[CrossRef](#)]
26. Zhang, B.; Tian, X.J.; Sun, H.J.; Chen, F.; Zhang, Y.C.; Zhang, L.F.; Fu, S.M. PODEn4DVar-based radar data assimilation scheme: Formulation and preliminary results from real-data experiments with advanced research WRF (ARW). *Tellus A* **2015**, *67*, 26045. [[CrossRef](#)]
27. Tian, X.J.; Xie, Z.H.; Liu, Y.; Cai, Z.N.; Fu, Y.; Zhang, H.; Feng, L. A joint data assimilation system (Tan-Tracker) to simultaneously estimate surface CO₂ fluxes and 3-D atmospheric CO₂ concentrations from observations. *Atmos. Chem. Phys.* **2014**, *14*, 13281–13293. [[CrossRef](#)]
28. Stefanescu, R.; Sandu, A.; Navon, I.M. POD/DEIM Reduced-Order Strategies for efficient four dimensional variational data assimilation. *J. Comput. Phys.* **2015**, *295*, 569–595. [[CrossRef](#)]
29. Altaf, M.U.; Ambrozic, M.; McCabe, M.F.; Hoteit, I. A study of reduced-order 4DVAR with a finite element shallow water model. *Int. J. Numer. Method. Fluids* **2016**, *80*, 631–647. [[CrossRef](#)]
30. Ambrozic, M. A Study of Reduced Order 4D-VAR with a Finite Element Shallow Water Model. Master's Thesis, Delft University of Technology Faculty of Electrical Engineering, Delft, The Netherlands, 19 November 2013.
31. Tissot, G.; Cordier, L.; Benard, N.; Noack, B.R. 4D-Variational Data Assimilation for Pod Reduced-Order Models. Available online: <http://www.dl.begellhouse.com/references/3ce1b491115b5c16,15f13bde2c0586bc,68fd7b9c15cacc0a.html> (accessed on 30 August 2013).
32. Tissot, G.; Cordier, L.; Noack, B.R. 4D-Var Identification of DMD Reduced-Order Models. Available online: <http://documents.irevues.inist.fr/handle/2042/57594> (accessed on 28 August 2015).
33. Cao, Y.H.; Zhu, J.; Navon, I.M.; Luo, Z.D. A reduced-order approach to four-dimensional variational data assimilation using proper orthogonal decomposition. *Int. J. Numer. Method. Fluids* **2007**, *53*, 1571–1583. [[CrossRef](#)]
34. Wang, B.; Liu, J.J.; Wang, S.D.; Cheng, W.; Liu, J.; Liu, C.S.; Xiao, Q.N.; Kuo, Y.H. An economical approach to four-dimensional variational data assimilation. *Adv. Atmos. Sci.* **2010**, *27*, 715–727. [[CrossRef](#)]
35. Navarra, A.; Simoncini, V. *A Guide to Empirical Orthogonal Functions for Climate Data Analysis*; Springer: New York, NY, USA, 2013; pp. 39–64. ISBN 978-90-481-3702-2.
36. Liu, J.J.; Wang, B.; Wang, S.D. The structure of background-error covariance in a four-dimensional variational data assimilation system: Single-point experiment. *Adv. Atmos. Sci.* **2010**, *27*, 1303–1310. [[CrossRef](#)]
37. Zhao, J.; Wang, B. Sensitivity of the DRP-4DVar performance to perturbation samples obtained by two different methods. *J. Meteor. Res.* **2010**, *24*, 527–538.
38. Liu, J.J.; Wang, B.; Xiao, Q.N. An evaluation study of the DRP-4DVar approach with the Lorenz-96 model. *Tellus A* **2011**, *63*, 256–262. [[CrossRef](#)]
39. Zhao, Y.; Wang, B.; Liu, J.J. A DRP-4DVar Data Assimilation Scheme for Typhoon Initialization Using Sea Level Pressure Data. *Mon. Wea. Rev.* **2012**, *140*, 1191–1203. [[CrossRef](#)]
40. Qiu, C.J.; Zhang, L.; Shao, A.M. An explicit four-dimensional variational data assimilation method. *Sci. China Ser. D* **2007**, *50*, 1232–1240. [[CrossRef](#)]
41. Bishop, C.H.; Etherton, B.J.; Majumdar, S.J. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.* **2001**, *129*, 420–436. [[CrossRef](#)]

42. Hunt, B.R.; Kostelich, E.J.; Szunyogh, I. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D* **2007**, *230*, 112–126. [[CrossRef](#)]
43. Wang, X.G.; Bishop, C.H. A comparison of breeding and Ensemble Transform Kalman Filter ensemble forecast schemes. *J. Atmos. Sci.* **2002**, *60*, 1140–1158. [[CrossRef](#)]
44. Lorenc, A.C. Recommended nomenclature for EnVar data assimilation methods. In *Research Activities in Atmospheric and Oceanic Modelling*; Met Office: Exeter, UK, 2013.
45. Le Dimet, F.X.; Talagrand, O. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. *Tellus A* **1986**, *38*, 97–110. [[CrossRef](#)]
46. Zhao, J.; Wang, B.; Liu, J.J. Impact of Analysis-time Tuning on the Performance of the DRP-4DVar Approach. *Adv. Atmos. Sci.* **2011**, *28*, 207–216. [[CrossRef](#)]
47. Hannachi, A.; Jolliffe, I.T.; Stephenson, D.B. Empirical orthogonal functions and related techniques in atmospheric science: A review. *Int. J. Climat.* **2007**, *27*, 1119–1152. [[CrossRef](#)]
48. Yang, S.C.; Kalnay, E.; Hunt, B.R.; Bowler, E. Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter. *Q. J. R. Meteor. Soc.* **2009**, *135*, 251–262. [[CrossRef](#)]
49. Lorenz, E.N. Predictability: A problem partly solved. In *Proceedings of the Proc. Seminar on Predictability*, Reading, UK, 10–13 September 1996.
50. Evensen, G. *Data Assimilation: The Ensemble Kalman Filter*, 2nd ed.; Springer: New York, NY, USA, 2009; pp. 237–253. ISBN 978-3-642-03710-8.
51. Amezcua, J.; Goodliff, M.; Van Leeuwen, P.J. A weak-constraint 4DensembVar. Part I: Formulation and simple model experiments. *Tellus A* **2017**, *69*, 1271564. [[CrossRef](#)]
52. Goodliff, M.; Amezcua, J.; Van Leeuwen, P.J. A weak-constraint 4DensembVar. Part II: Experiments with larger models. *Tellus A* **2017**, *69*, 1271565. [[CrossRef](#)]
53. Anderson, J.L. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **2001**, *129*, 2884–2903. [[CrossRef](#)]
54. North, G.R.; Bell, T.L.; Cahalan, R.F. Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.* **1982**, *110*, 699–706. [[CrossRef](#)]
55. Yang, S.C.; Baker, D.; Li, H.; Cordes, K.; Huff, M.; Nagpal, G.; Okereke, E.; Villafane, J.; Kalnay, E.; Duane, G.S. Data Assimilation as Synchronization of Truth and Model: Experiments with the Three-Variable Lorenz System. *J. Atmos. Sci.* **2006**, *63*, 2340–2354. [[CrossRef](#)]
56. Cheng, H.; Jardak, M.; Alexe, M.; Sandu, A. A hybrid approach to estimating error covariances in variational data assimilation. *Tellus A* **2010**, *62*, 288–297. [[CrossRef](#)]
57. Zhang, M.; Zhang, F.Q.; Huang, X.Y.; Zhang, X. Intercomparison of an ensemble Kalman filter with three- and four-dimensional variational data assimilation methods in a limited-area model over the month of June 2003. *Mon. Wea. Rev.* **2011**, *139*, 566–572. [[CrossRef](#)]
58. Huang, Z.K.; Peng, Z.; Liu, H.N.; Zhang, M.G.; Ma, X.G.; Yang, S.C.; Lee, S.D.; Kim, S.Y. Development of MAQ for East Asia CO₂ data assimilation under an EnKF framework: A first result. *Chin. Sci. Bull.* **2014**, *59*, 3200–3208. [[CrossRef](#)]

