Earth and Space Science

RESEARCH ARTICLE
10.1029/2022EA002254

A Mini-Batch Stochastic Optimization-Based Adaptive Localization Scheme and Its Implementation in NLS-i4DVar

Hongqin Zhang[1,2], Xiangjun Tian[1,3], and Xiaobing Feng[4]

[1]National Tibetan Plateau Data Center, State Key Laboratory of Tibetan Plateau Earth System and Resource Environment, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China, [2]ICCES, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China, [3]University of Chinese Academy Sciences, Beijing, China, [4]Department of Mathematics, The University of Tennessee, Knoxville, TN, USA

Key Points:

- Construct a cost function of the localization radius by estimating forecast and observation error statistics
- The adaptive localization scheme computes the "optimal" covariance localization radii adaptively and flow-dependently
- The mini-batch stochastic gradient descent method with a novel sampling strategy is used to minimize the cost function

Correspondence to:
X. Tian,
tianxj@itpcas.ac.cn

Citation:
Zhang, H., Tian, X., & Feng, X. (2022). A mini-batch stochastic optimization-based adaptive localization scheme and its implementation in NLS-i4DVar. *Earth and Space Science*, *9*, e2022EA002254. https://doi.org/10.1029/2022EA002254

Received 20 JAN 2022
Accepted 23 JUL 2022

Author Contributions:
Conceptualization: Xiangjun Tian
Formal analysis: Hongqin Zhang
Funding acquisition: Hongqin Zhang
Methodology: Hongqin Zhang, Xiangjun Tian
Project Administration: Hongqin Zhang
Software: Hongqin Zhang
Supervision: Xiangjun Tian, Xiaobing Feng
Validation: Hongqin Zhang
Writing – original draft: Hongqin Zhang
Writing – review & editing: Hongqin Zhang, Xiangjun Tian

Abstract This paper proposes a mini-batch stochastic optimization-based adaptive localization scheme for computing the "optimal" localization radius in data assimilation (DA) applications. After constructing a cost function of the localization radius by estimating forecast and observation error statistics, a mini-batch stochastic gradient descent method with a novel sampling strategy is proposed to minimize the cost function. The proposed stochastic optimization algorithm is further incorporated into the DA method NLS-i4DVar (the nonlinear least squares integral correcting four-dimensional variational DA method), which was developed by the authors in Tian et al. (2021, https://doi.org/10.1029/2021EA001767). It is utilized to compute the "optimal" covariance localization radii adaptively and flow-dependently inside NLS-i4DVar. The computational cost of NLS-i4DVar with the proposed adaptive localization scheme only increases slightly due to the use of the mini-batch stochastic optimization algorithm. Numerical experimental results using the shallow-water equations demonstrate that NLS-i4DVar with the proposed adaptive localization scheme shows substantial performance improvement over the standard NLS-i4DVar method.

1. Introduction

The objective of data assimilation (DA) aims to incorporate the observations into the Earth system models' state to improve their initial condition estimations and thus the models' predictive capabilities by taking advantage of the consistency constraints between the laws of time evolution and physical properties (Kalnay, 2003). It is well known that accurate specification of the background error covariance matrix, which will be denoted by **B** in this paper and usually has very high dimension in the order of $10^{7-9} \times 10^{7-9}$, plays a key role in improving the performance of the two major methods in DA (Tian et al., 2018), namely, the ensemble Kalman filter (EnKF: Evensen, 2007; Houtekamer & Mitchell, 1998; Whitaker et al., 2004) and four-dimensional variational data assimilation (4DVar: Courtier et al., 1994; Lewis & Derber, 1985; Navon et al., 2005).

The **B** matrix often can be approximated by the ensembles produced via the so-called method of the National Meteorological Center (NMC; now named the National Centers for Environmental Prediction) (Parrish & Derber, 1992) or the ensemble Kalman filter (Evensen, 2007). Under the limited consideration resource condition, the ensemble size is usually of the order 10^2, which is too small compared to the problem dimension (10^{7-9}). Such sampling errors may induce spurious long-distance error correlations resulting in poor conditioning of **B** matrix. Localization has proven to be an effective technique to ameliorate the spurious long-range correlations of the ensemble-estimated **B** matrix and to increase its rank, so that the system can "fit" the background innovations (Anderson, 2007; Anderson & Lei, 2013; Bishop & Hodyss, 2011; Blum et al., 2009; Hamill et al., 2001; Houtekamer et al., 2005; Kepert, 2008; Lei & Anderson, 2014; Zhen & Zhang, 2014).

Covariance localization is usually implemented as a Schür product between the ensemble-based **B** matrix and a decaying distance-dependent function, such as a Gaussian (Anderson, 2012) or the Gaspari-Cohn fifth-order piecewise polynomial (Gaspari & Cohn, 1999). The computation of the decaying distance-dependent function needs to predetermine an "optimal" localization radius (given empirically or through sensitivity experiments), which determines the influence distance of the observations. More importantly, the ensemble-estimated **B** matrix is usually flow-dependent, and its optimal localization radius is also not stagnant or unchangeable and depends on the ensemble size, observation properties, model dynamics and the model resolution, etc (Kirchgessner et al., 2014). A lot of effort has been devoted to adaptively estimating optimal localization radii from the data based on EnKF in order to optimize the assimilation performance according to some criteria (Bishop & Hodyss, 2011;

Ménétrier et al., 2015; Moosavi et al., 2018), for example, based on the ideas of statistics (Anderson, 2007; Lei & Anderson, 2014; Lei et al., 2015), according to the characteristics of correlation functions (Bishop & Hodyss, 2007) and sampling error (Anderson, 2012), as well as the method based on machine learning (Moosavi et al., 2018). Although some progress has been made in developing adaptive localization schemes, significant issues remain to be addressed and many improvements are still needed regarding the perspective of precision, applicability, and computational complexity.

On the other hand, covariance localization is an extremely expensive process when the optimization problem to be solved is of high dimension (namely, 10^{7-9}or higher). Its implementation in EnKF is generally conducted in the observation space by assimilating the observations one-by-one or batch-by-batch to alleviate its computational costs. In the ensemble-variational framework, covariance localization is conducted through the extended state/control variable strategy based on the local correlation matrix decomposition (Liu et al., 2009; Lorenc, 2003; Zhang & Tian, 2018). However, the matrix decomposition can be only roughly achieved by decomposing its low-resolution or low-rank approximations and interpolating it back to the expected high-resolution system (Liu et al., 2009). Zhang and Tian (2018) developed an efficient local correlation matrix decomposition method to approximate the full correlation matrix decomposition with sufficient accuracy, which was successfully used to improve the localization implementation of the nonlinear least squares four-dimensional variational data assimilation method (NLS-4DVar, Tian et al., 2018; Tian & Feng, 2015) to enhance its computational efficiency.

NLS-4DVar (Tian et al., 2018; Tian & Feng, 2015) is an advanced four-dimensional ensemble variational (4DEnVar) algorithm that utilizes a Gauss-Newton iterative method (Dennis & Schnabel, 1996) to handle the nonlinearity of the 4DVar cost function; it is capable of enhancing the assimilation accuracy of the 4DEnVar while maintaining an overall structure that resembles traditional 4DEnVar algorithms (i.e., with no dependence on the adjoint model) and provides a flow-dependent background error covariance matrix **B**. Very recently, Tian et al. (2021) proposed an integral correcting 4DVar (i4DVar) approach by treating initial and model errors together as a whole and correcting them simultaneously and indiscriminately based on a less recognized property of the traditional 4DVar. The new i4DVar approach has the potential to be applicable to various DA problems from scientific and engineering fields because of its ease of implementation and superior performance compared with the traditional 4DVar (Tian et al., 2021). The implementation of the i4DVar optimization does not require much extra cost because it only slightly differs in the way how the forecast model is integrated (see Tian et al., 2021) during the optimization process. Thus, the ensemble nonlinear least squares-based approach can also be used to solve the i4DVar minimization problem (the resulting method is referred to as NLS-i4DVar).

The primary goal of this study is to propose a mini-batch stochastic optimization-based adaptive localization scheme for computing the "optimal" localization radius and to incorporate it into NLS-i4DVar. To the end, we first transform an objective function developed by Zheng (2009) based on a general approach of estimating forecast and observation error statistics by Dee (1995) and Dee and da Silva (1999) into a nonlinear optimization problem of seeking the optimal localization radius by introducing the localized background error covariance matrix **B**. Second, to reduce the computational cost and speed up the minimization algorithm, a mini-batch stochastic gradient decent optimization method (Bottou, 2012; LeCun et al., 2015) is proposed to solve the involved nonlinear optimization problem. Third, the Maximum Likelihood Estimation (MLE) of all the batches' optimal localization radii is then taken as the final optimal localization radius. A novel mini-batch sampling strategy is proposed to make best use of all the observational formations as much as possible. Finally, the adaptive and fast localization scheme is thus obtained by combining the adaptively optimizing process (for seeking the optimal localization radii) and the fast localization scheme based on an efficient local correlation matrix decomposition approach to avoid direct decomposition of the correlation matrix for improving localization implementation of the NLS-i4DVar.

The remainder of this paper is organized as follows. In Section 2, we introduce our mini-batch stochastic optimization-based adaptive localization scheme by detailing its four main steps. We then present an application of this new adaptive localization scheme to the localization implementation of the integral correcting 4DVar (i4DVar) method based on the ensemble nonlinear least squares-based approach (i.e., the NLS-i4DVar approach). In Section 3, we present a series of comparison experiments to evaluate the effectiveness and robustness of the proposed mini-batch stochastic optimization-based adaptive localization scheme and its potential for improving the localization schemes used in the NLS-i4DVar. Our numerical experiments are carried out on the 2D

shallow-water equation model. Finally, we finish the paper with a summary and some concluding remarks in Section 4.

2. Methodology

The methodology we propose consists of two parts: First, we formulate a mini-batch stochastic gradient descent (SGD) adaptive localization scheme; then, we incorporate this scheme into the NLS-i4DVar method.

2.1. Mini-Batch Stochastic Optimization-Based Adaptive Localization Scheme

As mentioned in the introduction, the proposed adaptive localization scheme consists of four steps as given below.

First, by incorporating the localized **B** matrix into the 2log-likelihood of $\mathbf{y}'_{obs}$ (Zheng, 2009, Equation 11) and following the pioneering work of Dee and da Silva (1999), we adopt the following loss function in the whole observational area:

$$L(r_0) = \ln\left\{\det\left[H\left(\mathbf{B}\circ\mathbf{C}(r_0)\right)H^{\mathrm{T}} + \mathbf{R}\right]\right\} + \left(\mathbf{y}'_{\mathrm{obs}}\right)^{\mathrm{T}}\left[H\left(\mathbf{B}\circ\mathbf{C}(r_0)\right)H^{\mathrm{T}} + \mathbf{R}\right]^{-1}\left(\mathbf{y}'_{\mathrm{obs},}\right) \tag{1}$$

where $\mathbf{y}'_{\mathrm{obs}} = \mathbf{y}_{\mathrm{obs}} - H\mathbf{x}$, $\mathbf{y}_{\mathrm{obs}}$ and $\mathbf{x}$ are the n_y/n_x-dimensional observation and state vectors, respectively, $H$is the observation operator, and $\mathbf{B} \in \Re^{n_x \times n_x}$ and $\mathbf{R} \in \Re^{n_y \times n_y}$ are the background and observation error covariance matrices, respectively. Here, $\mathbf{C}(r_0)$ is the spatial correlation matrix computed by Zhang and Tian (2018) and Tian et al. (2018).

$$\mathbf{C}(r_0) = \mathbf{C}_0\left(\frac{d_{i,j}}{r_0}\right). \tag{2}$$

where $d_{i,j}$ is the spatial spherical distance (computed by $d_{i,j} = r_e \times \arccos[\sin(y_1) \times \sin(y_2) + \cos(y_1) \times \cos(y_2) \times \cos(x_1 - x_2)]$, where r_e denotes the Earth radius, and (x_1,y_1) and (x_2,y_2) are the latitude and longitude coordinates in the radian of any two points) between the ith and jth grid points and r_0 is the covariance localization radius (to be predetermined). A common localization function used in production software is the fifth-order piecewise rational function (Gaspari & Cohn, 1999). Here, we use the Gaussian function $\mathbf{C}_0$ defined by

$$\mathbf{C}_0(u) = \exp\left(-u^2/2\right), \tag{3}$$

where $u = d_{i,j}/r_0$. Obviously, the minimizer (i.e., the optimal localization radius) of Equation 1 can be computed iteratively by performing evaluations of the cost function and its gradient using a suitable gradient descent algorithm (e.g., the limited memory Broyden-Fletcher-Goldfarb-Shanno [L-BFGS]; Liu & Nocedal, 1989). Unfortunately, it is not an easy task to compute the function values and its gradient due to their large scales in the whole observational area. Moreover, the analytical expression of the gradient is normally inaccessible. To overcome this difficulty, we use the idea of the mini-batch SGD algorithm (which is widely used in machine learning, LeCun et al., 2015) by dividing the whole (and large) observational vector into the mini-batch (small) ones.

Thus, in the **second** step, we propose the following modified mini-batch SGD algorithm and the mini-batch sampling strategy to solve the nonlinear optimization problem 1–3. The mini-batch sampling strategy is to divide the whole observation area into several subareas according to Figure 1 and randomly select observations in each subarea. More explicitly, we first divide the whole observational area into $i_{n_{\max}}$ (where $i_{n_{\max}}$ is the number of the total random batches) subareas and randomly select observations $n_{yL}(\leq n_y)$ (i.e., the local mini-batch observations) in each subarea, which is followed by solving the mini-batch version of the loss Equation 1 using the gradient descent method. The details are as follows:

for $i_n = 1, \cdots, i_{n_{\max}}$ **do**

1. Randomly select$n_{yL}(\leq n_y)$observations (the i_nth mini-batch observations in Figure 1) and construct the following mini-batch cost function

$$L(r_{0,L}) = \ln\left\{\det\left[\mathbf{B}_{y,L}\circ\mathbf{C}_y(r_{0,L}) + \mathbf{R}_L\right]\right\} + \left(\mathbf{y}'_{\mathrm{obs},L}\right)^{\mathrm{T}}\left[\mathbf{B}_{y,L}\circ\mathbf{C}_y(r_{0,L}) + \mathbf{R}_L\right]^{-1}\left(\mathbf{y}'_{\mathrm{obs},L,}\right), \tag{4}$$

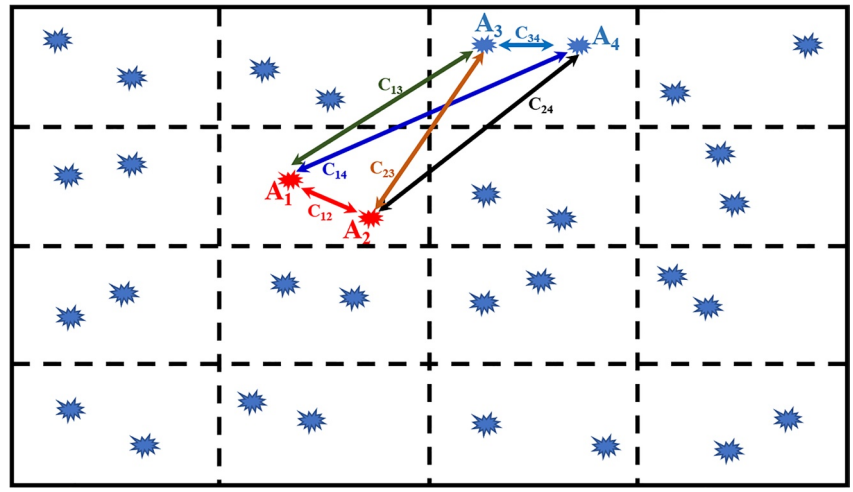

Figure 1. The diagram of a novel mini-batch sampling strategy.

where the subscript "L" denotes the i_nth local mini-batch and "y" means in the observation space. $\mathbf{B}_{y,L} = \left(\frac{\mathbf{P}_{y,L}^{\mathrm{T}}\mathbf{P}_{y,L}}{N-1}\right) \in \Re^{n_{y,L}\times n_{y,L}}$, $\mathbf{R}_L \in \Re^{n_{y,L}\times n_{y,L}}$, $\mathbf{y}'_{\mathrm{obs},L} \in \Re^{n_{y,L}}$, $\mathbf{C}_y(r_{0,L}) \in \Re^{n_{y,L}\times n_{y,L}}$, and $\mathrm{C}_{y,i,j}(r_{0,L}) = \mathrm{e}^{-\frac{(r_{i,j}/r_{0,L})^2}{2}}$ $(1 \le i \le n_{y,L}; 1 \le j \le n_{y,L})$, and $r_{i,j}$ is the spatial spherical distance between the ith and jth observations.

2. Minimize the mini-batch cost Equation 4 iteratively by performing its function evaluations and the second-order numerical gradient$L'(r_{0,L})$ defined by

$$L'(r_{0,L}) = \frac{L(r_{0,L} + \Delta r_{0,L}) - L(r_{0,L} - \Delta r_{0,L})}{2\Delta r_{0,L}} \tag{5}$$

using the L-BFGS algorithm (Liu & Nocedal, 1989) to obtain the optimal mini-batch localization radius $r_{0,L}^{i_n}$. Define the initial value of the next mini-batch localization radius as $r_{0,L} = r_{0,L}^{i_n}$.

end for

It is not hard to see that this mini-batch sampling strategy is capable of making best use of all the observational formations and achieving a satisfactory result at a very low computation cost.

The **third step** defines the MLE r_0^* of all the batches' optimal localization radii $r_{0,L}^{i_n}$ $(i_n = 1, \cdots, i_{n_{\max}})$ as the final optimal localization radius. This definition is based on the conviction that since each radius $r_{0,L}^{i_n}$ is optimal for its batch, then the best one out of all these optimal radii would be an accurate approximation to the MLE. In addition, this choice is perhaps the simplest and cheapest way to compute the MLE for our approach.

Finally, the adaptive and fast location scheme is obtained by combining the adaptively optimizing process (i.e., substituting the optimal localization radii r_0^* into the (Gaussian) spatial correlation matrix $\mathbf{C}(r_0)$) and the fast localization scheme proposed by Zhang and Tian (2018), which is further used to modify localization implementation of NLS-i4DVar. See Section 2.2 for further details.

2.2. The NLS-I4DVar Enhanced by the Adaptive and Fast Location Scheme

The integral correcting 4DVar (i4DVar, Tian et al., 2021) extends the constrained 4DVar's strategy of correcting the initial and model errors as a whole to other time points (with the same time intervalτ) in the assimilation window $[t_0, t_S]$ (the subindex 0 indicates the initial time point, $S + 1$ represents the total number of observation

time points in the assimilation window) by proposing the following i4DVar cost functional (for more details, see Tian et al., 2021):

$$J\left(\mathbf{x}'\right)=\frac{1}{2}\left(\mathbf{x}'\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\mathbf{x}'\right)+\frac{1}{2}\sum_{k=0}^{S}\left[H_k\left(\mathbf{x}_k^*\right)-\mathbf{y}_{\mathrm{obs},k}\right]^{\mathrm{T}}\mathbf{R}_k^{-1}\left[H_k\left(\mathbf{x}_k^*\right)-\mathbf{y}_{\mathrm{obs},k}\right] \tag{6}$$

under the constraint

$$\mathbf{x}_k^*\left(\mathbf{x}'\right)=\begin{cases}\mathbf{x}_k^*+\mathbf{x}', & k=t_0,t_0+\tau,\cdots,t_0+\left(\frac{t_S-t_0}{\tau}-1\right)\tau,\\ M_k\left(\mathbf{x}_{k-1}^*\right), & \text{otherwise.}\end{cases} \tag{7}$$

where $\mathbf{x}_{t_0}^*=\mathbf{x}_b$, $\mathbf{x}_b$ is the background field (b denotes the background value) and $M_k(\cdot)$ is the forecast model from t_{k-1} to t_k, the superscript T stands for matrix transpose, k denotes the kth observation time point, $H_k$denotes the observation operator, $\mathbf{y}_{\mathrm{obs},k}\in\Re^{n_{y,k}}$, $n_y=\sum_{k=0}^{S}n_{y,k}$, and matrices $\mathbf{B}$ and $\mathbf{R}_k\in\Re^{n_{y,k}\times n_{y,k}}$ are defined as the background and kth observation error covariances, respectively. Different from the traditional 4DVar, the assimilation window $[t_0,t_S]$ is further divided into t_S-t_0/τ sub-windows, τ is the length of each sub-window and the averaged "integral" correction term $\mathbf{x}'\in\Re^{n_x}$ is sequentially added at $t_0,t_0+\tau,\cdots,t_0+(t_S-t_0/\tau-1)\tau$ along the integration in the model. Obviously, when the correction term (or the analysis increment) $\mathbf{x}'$ at the initial time step is added, that is,

$$\mathbf{x}_k^*\left(\mathbf{x}'\right)=\begin{cases}\mathbf{x}_k^*+\mathbf{x}', & k=t_0,\\ M_k\left(\mathbf{x}_{k-1}^*\right), & \text{otherwise,}\end{cases} \tag{8}$$

we recover the usual 4DVar formulation.

The ensemble nonlinear least squares-based i4DVar approach (NLS-i4DVar) assumes the optimal "integral" correction term $\mathbf{x}'\in\Re^{n_x}$ can be characterized as a linear combination of the model perturbations (MPs)$\mathbf{P}_x$; that is, $\mathbf{x}'=\mathbf{P}_x\boldsymbol{\beta}$, $\mathbf{P}_x=\left(\mathbf{x}_1',\cdots,\mathbf{x}_N'\right)$, $\boldsymbol{\beta}=(\beta_1,\cdots,\beta_N)^{\mathrm{T}}$, and $\mathbf{x}_j$ is jth ($j=1,\cdots,N$) ensemble and the background error covariance $\mathbf{B}$ is defined by $\mathbf{B}=(\mathbf{P}_x)(\mathbf{P}_x)^{\mathrm{T}}/N-1$. By substituting the above assumptions and minimizing the i4DVar cost function by the Gauss-Newton iteration method (Dennis & Schnabel, 1996), Tian et al. (2021) obtained

$$\boldsymbol{\beta}_\rho^i=\boldsymbol{\beta}_\rho^{i-1}+\sum_{k=0}^{S}\left(\rho_{y,k}<e>\boldsymbol{P}_{y,k}^*\right)^{\mathrm{T}}L_k'\left(\boldsymbol{x}'^{,i-1}\right)+\sum_{k=0}^{S}\left(\rho_{y,k}<e>\boldsymbol{P}_{y,k}^{\#}\right)^{\mathrm{T}}R_k^{-1}\left[y'_{\mathrm{obs},k}-L_k'\left(\boldsymbol{x}'^{,i-1}\right)\right] \tag{9}$$

$$x'^{,i}=\left(\rho_x<e>P_x\right)\boldsymbol{\beta}_\rho^i \tag{10}$$

for $i=1,\cdots,i_{max}$, where $i_{\max}$ is the maximum iteration number, $\mathbf{P}_{y,k}=\left(\mathbf{y}_1',\cdots,\mathbf{y}_N'\right)\in\Re^{n_{y,k}\times N}$, and $\mathbf{y}_{k,i}'=H_k\left[\mathbf{x}_{k,i}^*\left(\mathbf{x}_j'\right)\right]-H_k\left[\mathbf{x}_k^*(0)\right]$,

$$\mathbf{x}_{k,j}^*\left(\mathbf{x}_j'\right)=\begin{cases}\mathbf{x}_{k,j}^*+\mathbf{x}_j', & k=t_0,\\ M_k\left(\mathbf{x}_{k-1,j}^*\right), & \text{otherwise,}\end{cases} \tag{11}$$

$$L_k'\left(\mathbf{x}'\right)=H_k\left[\mathbf{x}_k^*\left(\mathbf{x}'\right)\right]-H_k\left[\mathbf{x}_k^*(0)\right], \tag{12}$$

and

$$\mathbf{y}'_{\mathrm{obs},k}=\mathbf{y}_{\mathrm{obs},k}-H_k\left[\mathbf{x}_k^*(0)\right]. \tag{13}$$

We also note that

$$\mathbf{P}^{*}_{y,k} = -(N-1)\left(\mathbf{P}_{y,k}\right)\left[\sum_{k=0}^{S}\left(\mathbf{P}_{y,k}\right)^{\mathrm{T}}\left(\mathbf{P}_{y,k}\right)\right]^{-1} \times \left[(N-1)\mathbf{I}_{N\times N} + \sum_{k=0}^{S}\left(\mathbf{P}_{y,k}\right)^{\mathrm{T}}\mathbf{R}_{k}^{-1}\left(\mathbf{P}_{y,k}\right)\right]^{-1}, \tag{14}$$

and

$$\mathbf{P}^{\#}_{y,k} = \left(\mathbf{P}_{y,k}\right)\left[(N-1)\mathbf{I}_{N\times N} + \sum_{k=0}^{S}\left(\mathbf{P}_{y,k}\right)^{\mathrm{T}}\mathbf{R}_{k}^{-1}\left(\mathbf{P}_{y,k}\right)\right]^{-1}, \tag{15}$$

where $\boldsymbol{\rho}_x \in \Re^{n_x\times r}$, $\boldsymbol{\rho}_x\boldsymbol{\rho}_x^{\mathrm{T}} = \mathbf{C}\left(r_0\right) \in \Re^{n_x\times n_x}$, and $\mathbf{C}_{i,j}\left(r_0\right) = \exp\left[-\left(\frac{d_{i,j}}{r_0^{*}}\right)^2/2\right]$, $\boldsymbol{\rho}_{y,k} \in \Re^{n_{y,k}\times r}$ is computed together with $\rho_x \in \Re^{n_x\times r}$ and r is the selected truncation mode number (Zhang & Tian, 2018). The operator "<e>" is defined by $\mathbf{D} = \mathbf{A} <e> \mathbf{B} = (\mathbf{A}\circ\mathbf{B}_1, \mathbf{A}\circ\mathbf{B}_2, \cdots, \mathbf{A}\circ\mathbf{B}_r)$, $\mathbf{A} \in \mathbb{R}^{m\times N}$, $\mathbf{B} \in \mathbb{R}^{m\times r}$, and $\mathbf{D} \in \mathbb{R}^{m\times(N\times r)}$. $\mathbf{B}_j$ is the jth column of $\mathbf{B}$. The operator "<o>" is the Schür product. It should be noted that the matrix decomposition of the correlation $\mathbf{C}\left(r_0\right) = \boldsymbol{\rho}_x\boldsymbol{\rho}_x^{\mathrm{T}}$ can be achieved by an efficient local correlation matrix decomposition approach proposed by Zhang and Tian (2018) in Appendix A. This efficient decomposition was further used to modify localization implementation in the NLS-i4DVar to enhance its computational efficiency (Tian et al., 2021; Zhang & Tian, 2018). Here, we accomplish the mini-batch SGD adaptive localization scheme and its implementation within the NLS-i4DVar. Additionally, the NLS-i4DVar also uses the following modified square root analysis scheme

$$\mathbf{P}_x = \mathbf{P}_x\mathbf{V}_2\Phi^{\mathrm{T}} \tag{16}$$

for the online ensemble update in the assimilation cycles (see Tian et al., 2020 for the definitions of $\mathbf{V}_2$ and Φ).

3. OSSEs Using the Shallow-Water Equation Model

In this section, we present several groups of observing system simulation experiments (OSSEs) based on the shallow-water equation model to demonstrate the potential merits and advantages of the NLS-i4DVar with the proposed adaptive localization scheme, especially to compare with the traditional (nonadaptive) localization implementation.

3.1. The 2D Shallow-Water Equation Model

The following 2D shallow-water equations, with the f-plane formulation, are utilized as the forecast model for the OSSEs:

$$\frac{\partial u}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial v} + fv - g\frac{\partial h}{\partial x} \tag{17}$$

$$\frac{\partial v}{\partial t} = -u\frac{\partial u}{\partial x} - v\frac{\partial u}{\partial v} - f\mathrm{u} - g\frac{\partial h}{\partial v} \tag{18}$$

$$\frac{\partial h}{\partial t} = -u\frac{\partial\left(h-h_s\right)}{\partial x} - v\frac{\partial\left(h-h_s\right)}{\partial v} - \left(H+h-h_s\right)\left(\frac{\partial u}{\partial x} + \frac{\partial u}{\partial v}\right) \tag{19}$$

where $f = 7.272\times10^{-5}\mathrm{s}^{-1}$ is the Coriolis parameter, $H = 3{,}000$ m is the basic state depth, $h_s = h_0\sin(4\pi x/L_x)[\sin(4\pi y/L_y)]^2$ is the terrain height, $h_0 = 250$m, and the lengths of the two sides of the computational domain are $D = L_x = L_y = 44d$m, respectively (where $d = \Delta_x = \Delta_y = 300$ km is the uniform grid size). The domain is partitioned into smaller square subdomains with 45 grid points in each coordinate direction, and periodic boundary conditions are imposed at $x = (0,L_x)$ and $y = (0,L_y)$. The second-order central finite difference and two-step backward difference schemes of (cf.) Matsuno (1966) are used to discretize the spatial and local time derivatives, respectively, in order to ensure computational stability (Qiu et al., 2007). The time step is taken as 360 s (6 min). The model state vector is represented by height h and the horizontal velocity components by u and v at the grid points.

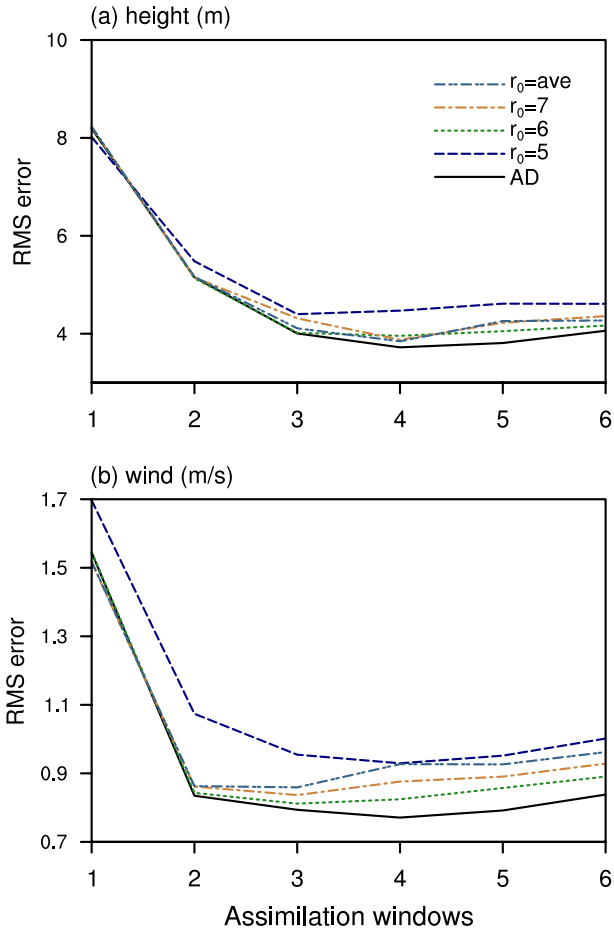

Figure 2. Time series of spatially and temporally averaged root mean square errors (RMSEs) of (a) height and (b) wind for the i4DVar + ADr_0 and i4DVar + r_0(=6×300, 5 × 300, 7 × 300 km, and the averaged radius = 6.9 × 300 km), respectively.

3.2. Experimental Setup

As in Tian et al. (2021), the "true" initial fields for all the OSSEs are first produced by integrating the perfect (i.e., $h_0 = 250$ m) shallow-water equation model from the following initial conditions:

$$h = 360\left[\sin\left(\frac{\pi y}{D}\right)\right]^2 + 120\sin\left(\frac{2\pi x}{D}\right)\sin\left(\frac{2\pi y}{D}\right) \quad (20)$$

$$u = -f^{-1}g\frac{\partial h}{\partial v} \text{ and } v = -f^{-1}g\frac{\partial h}{\partial x} \text{ at } t = -60\text{hr} \quad (21)$$

for 60 hr. Similarly, the background state $\mathbf{x}_b$ is also produced but using the imperfect model with $h_0 = 0$m. Obviously, $\mathbf{x}_b$ is significantly different from the true state (not shown) due to the two 60 hr model integrations under with $h_0 = 0$m and $h_0 = 250$m, respectively. Specifically, the spatially averaged root mean square errors (RMSEs) are 23.4 m, 1.53 m s^{-1}, and 2.58 m s^{-1} for h, u, and v, respectively.

There are six assimilation cycles (windows) in all the OSSEs. The length of each window is 12 hr, and "observations" are available every 3 hr (i.e., at 3, 6, 9, and 12 hr during each assimilation window). Each model grid has one randomly distributed observation site, resulting in a total of 44 × 44 observational sites at each observation time point. The "observations" are generated by adding random noise to the "true" values at the "observational" locations, which are obtained using a simple bilinear interpolation method. Therefore, the observation operator $H_k$is simply a bilinear interpolation operator. The ensemble number N is 70 and the default (optimal through sensitivity experiments) covariance localization radius is $r_0 = 6 \times 300$ km (Tian et al., 2021) and the 4D moving strategy (Tian & Feng, 2015) is adopted to produce the initial MPs for NLS-i4DVar. Finally, the following modified square root analysis scheme

$$\mathbf{P}_x = \mathbf{P}_x\mathbf{V}_2\Phi^{\mathrm{T}} \quad (22)$$

is utilized for the online ensemble update in the assimilation cycles (see Tian et al., 2020 for the definitions of $\mathbf{V}_2$and Φ).

3.3. Experimental Results

We first evaluate the performance of the NLS-i4DVar with (hereafter, i4DVar + ADr_0) and without (i4DVar + r_0) the proposed adaptive localization scheme. The spatially and temporally averaged RMSEs of *height* ($\overline{r^h_{\mathrm{mse}}}$) and *wind* ($\overline{r^{\mathrm{wind}}_{mse}}$) (see Tian & Zhang, 2019 for definitions of $\overline{r^h_{\mathrm{mse}}}$ and $\overline{r^h_{\mathrm{mse}}}$) for i4DVar + AD$r_0$ and i4DVar + r_0 [$r_0 = 6 \times 300$(the optimal radius through sensitivity experiments), 5 × 300 and 7 × 300 km, respectively] are compared for each assimilation window under the imperfect model scenario (i.e., $h_0 = 0$) in Figure 2. Both methods perform satisfactorily in terms of overall low RMS errors even under the imperfect model scenario, which obviously attribute to the capability of the integral correcting 4DVar by treating the initial and model errors together as a whole and by correcting them simultaneously and indiscriminately (Tian et al., 2021). In addition, the performance of the NLS-i4DVar could be effectively enhanced by the adaptive localization scheme, which contributes significantly to the substantially better performance of i4DVar + ADr_0 compared with i4DVar + r_0 when $r_0 = 6 \times 300$ km (Figure 2), although the latter radius ($r_0 = 6 \times 300$ km) is already the optimal value through several groups of sensitivity experiments. It is worth noting that i4DVar + r_0(=5 × 300) km, which is actually very close to the optimal value off only by one grid interval d(=300 km), performs significantly worse than i4DVar + ADr_0. Another case with $r_0 = 7 \times 300$ km is slightly better than i4DVar + r_0(=5 × 300) km with lower RMS errors for both the *wind* and *height* components, but still considerably larger than i4DVar + ADr_0(Figure 1). The average, minimum, and maximum values of all the six-window adaptively optimized radii are 6.9×, 5.5×,

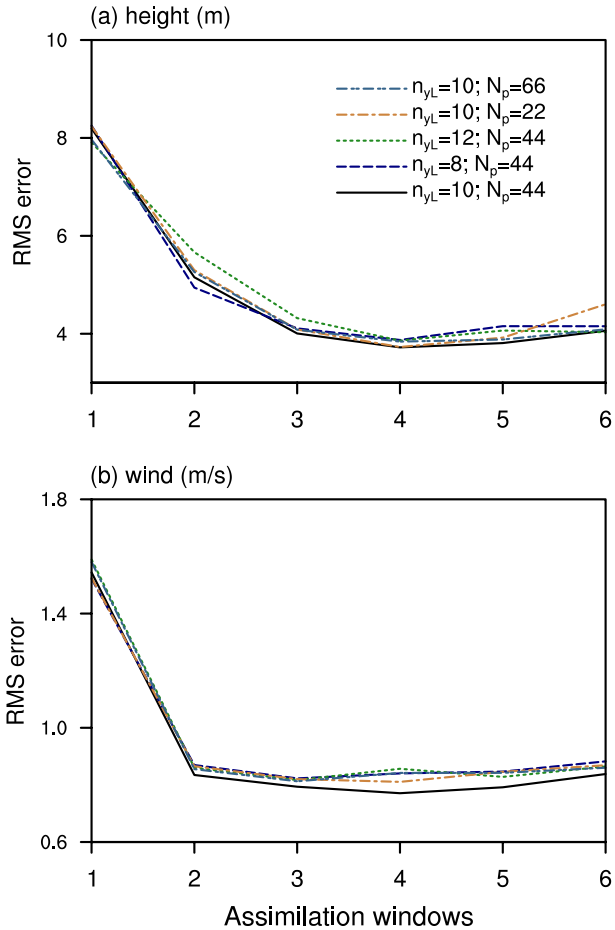

Figure 3. Time series of spatially and temporally averaged root mean square errors (RMSEs) of (a) height and (b) wind for the i4DVar + ADr_0(8 + 44), i4DVar + ADr_0(12 + 44), i4DVar + ADr_0(10 + 22), and i4DVar + ADr_0(10 + 66), respectively.

and 8.48 × 300 km, respectively. In particular, i4DVar + r_0(=6.9 × 300 km, the average value of all the six-window adaptively optimized radii) performs almost same as the case of r_0 = 7 × 300 km in terms of the *height* component but substantially worse than the latter (r_0 = 7 × 300 km) for the *wind* component. Again, this demonstrates that the optimal radii are indeed flow-dependent and change according to variations of atmospheric motion and incoming observation data.

To investigate the impact of the subdomain number N_P and the observational number $n_{yL}/2$ selected in each subdomain, four additional experiments are conducted to compare with the default i4DVar + ADr_0 (i.e., n_{yL} = 10andN_p = 44, referred to as i4DVar + AD (10 + 44)) in Figure 3: (1) n_{yL} = 8 and N_p = 44, referred to as i4DVar + ADr_0(8 + 44), (2)n_{yL} = 12 and N_p = 44, referred to as i4DVar + ADr_0(12 + 44), (3)n_{yL} = 10 and N_p = 22, referred to as i4DVar + ADr_0(10 + 22), and (4)n_{yL} = 10 and N_p = 66, referred to as i4DVar + ADr_0(10 + 66). In summary, the performance of the proposed adaptive localization scheme is not substantially improved by increasing the number of the subdomains and the observational number $n_{yL}/2$ selected in each subdomain. Noticeably, there is little difference between the RMS error curves of all the four additional N_P-n_{yL}-paired i4DVar + ADr_0 methods. This illustrates that, on the one hand, the proposed novel mini-batch sampling strategy shown in Figure 1 indeed works well to make full use of all the observations, and on the other hand, the appropriate combination of N_P and n_{yL} could reach the full potential of this adaptive localization scheme. Admittedly, the appropriate combination of N_P and n_{yL} can be only obtained through sensitivity experiments so far and clearly there is room for improvement.

We also compare the proposed adaptive localization scheme with the novel mini-batch sampling strategy (referred to as NMS) and with the commonly used random sampling strategy (referred to as CRS) by selecting 44 groups of 10 observational points within the observation area in Figure 4. For the mini-batch sampling strategy, we first divided the whole observation area into 44 subareas according to Figure 1 and randomly selected 10 observations in each subarea. For the commonly used sparse observation strategy, we randomly selected 10 observations for 44 times and minimized the mini-batch cost function (4) iteratively using the L-BFGS algorithm. As expected, the mini-batch sampling strategy substantially outperforms the CRS strategy (Figure 4). The results of these experiments show that the proposed mini-batch sampling strategy is indeed capable of achieving higher assimilation accuracy.

4. Summary and Concluding Remarks

A decaying distance-dependent function, such as a Gaussian or the Gaspari-Cohn fifth-order piecewise polynomial, is usually adopted to compute the Schür product appeared in the implementation of covariance localization. The localization radius plays a vital role in the computation of such functions. Furthermore, the background error covariance matrix **B** should be flow-dependent; consequently, covariance localization is also flow-dependent and its key parameter (i.e., the localization radius) should not be immutable. An optimization procedure must be used to determine the "optimal" localization radii (to achieve higher assimilation accuracy) adaptively according to model integrations and incoming observation data, which thus leads to adaptive localization schemes.

Any straightforward adaptive covariance localization process is expected to be computationally expensive due to the amount of computation involved. To circumvent this difficulty, we proposed to use a mini-batch stochastic optimization-based algorithm to minimize the objective function by introducing a localized **B** matrix and taking the localization radius as a variable to be optimized, which thus led to an adaptive covariance localization scheme of this study. Subsequently, a novel mini-batch sampling strategy was proposed to make best use of all the

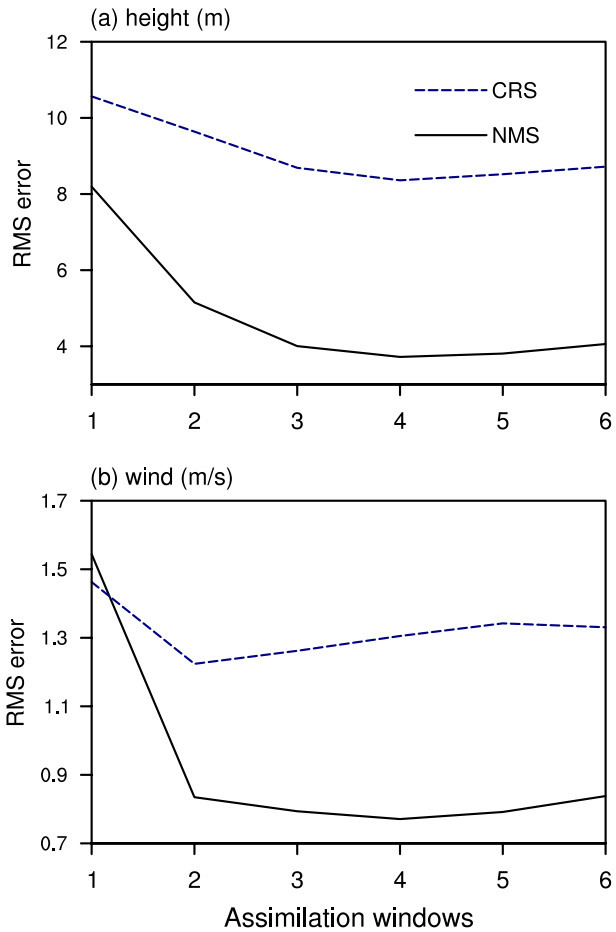

Figure 4. Time series of spatially and temporally averaged root mean square errors (RMSEs) of (a) height and (b) wind for the i4DVar + ADr_0 with the NMS and CRS sampling strategies, respectively.

observational formations as much as possible, which makes the proposed algorithm be able to obtain a satisfactory result at a low computational cost. We presented an application of the proposed adaptive covariance localization scheme to the localization implementation of NLS-i4DVar, which is an ensemble nonlinear least squares-based approach for solving the integral correcting i4DVar (developed recently by the authors).

We conducted a comprehensive performance evaluation of the NLS-i4DVar method with the proposed adaptive covariance localization scheme using several groups of OSSEs based on 2D shallow-water equations and compared the results with those obtained by the standard NLS-i4DVar method without adaptive localization. The experiments showed that the NLS-i4DVar with adaptive covariance localization performed substantially better than the standard NLS-i4DVar without adaptive localization. Moreover, while using the same number of samples, the proposed adaptive covariance localization scheme with the mini-batch sampling strategy performed much better than the commonly used random sampling strategy because NMS is capable of making best use of all the observational formations. On the other hand, the appropriate combination of the numbers of the subareas and the numbers of the random observations selected in each subarea can only be obtained via sensitivity experiments so far.

Incorporating the proposed adaptive covariance localization scheme into NLS-i4DVar has several advantages and great potential including its superior performance, which was achieved without noticeably increasing the computational cost, and its ease of implementation. Moreover, the proposed adaptive scheme can further enhance the NLS-i4DVar performance (which is an advanced ensemble nonlinear least squares-based approach for solving i4DVar and can correct the initial and model errors simultaneously and indiscriminately by treating them together as a whole). However, we like to note that the numerical evaluation experiments presented in this paper are only conducted based on the simple shallow-water equations. We will apply NLS-i4DVar with the adaptive localization scheme to more complex real-world models and continue exploring its potential applications in operational numerical weather prediction in future works. It is worth mentioning that the proposed adaptive localization scheme can be similarly adopted into other ensemble-based assimilation methods.

Appendix A: The Efficient Local Correlation Matrix Decomposition Approach

This approach is used to generated the $\boldsymbol{\rho}_m$ (used in model space) and $\boldsymbol{\rho}_o$ (used in observation space). Considering that the grid of some model state variables is not a regular grid, we first construct a regular grid to generate $\boldsymbol{\rho}$. Then, $\boldsymbol{\rho}_m$ and $\boldsymbol{\rho}_o$ are obtained according to interpolation.

The generation of $\boldsymbol{\rho}$ requires the following steps:

1. Construct the 1D correlation matrices $\mathbf{C}x,k(k = 1,2,\cdots,m_y)$, which vary with the latitude; $\mathbf{C}_y$; and $\mathbf{C}_z$ in three coordinate (i.e., x, y, and z) directions at a low resolution using the Gaussian function.
2. Construct empirical orthogonal function (EOF) of the one-dimensional matrices $\mathbf{C}_{x,k}$, $\mathbf{C}_y$, and $\mathbf{C}_z$ at a low resolution and obtain $\boldsymbol{\rho}_{x,k}$, $\boldsymbol{\rho}_y$, and $\boldsymbol{\rho}_z$ with r_x, r_y, and r_z columns, which named the truncation mode number and selected based on the cumulative relative root mean square error. Remarkably, r_x, r_y, and r_z are quite small at a low resolution.
3. $\boldsymbol{\rho}^*_{x,k}$, $\boldsymbol{\rho}^*_y$, and $\boldsymbol{\rho}^*_z$ of the high-resolution grid are obtained by spline interpolation.
4. Finally, the 3D $\boldsymbol{\rho}$ is generated using $\boldsymbol{\rho}^*_{x,k}$, $\boldsymbol{\rho}^*_y$, and $\boldsymbol{\rho}^*_z$ with the Kronecker product.

Then, $\boldsymbol{\rho}_m$ and $\boldsymbol{\rho}_o$ are generated by interpolation.

Data Availability Statement

The data used in this paper were obtained by the authors through numerical experiments using the shallow-water equation model and are all freely available at National Tibetan Plateau Data Center (http://data.tpdc.ac.cn), https://doi.org/10.11888/Atmos.tpdc.272233 (Zhang et al., 2022).

Acknowledgments
The work of H. Zhang was partially supported by the National Natural Science Foundation of China (Grant no. 42105150), the China Postdoctoral Science Foundation (Grant no. 2020M680645), and the Project supported by special research assistant of Chinese Academy of Sciences. The work of X. Tian was supported by the National Natural Science Foundation of China (Grant no. 41575100).

References

Anderson, J. (2007). Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Physical D*, *230*(1–2), 99–111. https://doi.org/10.1016/j.physd.2006.02.011

Anderson, J. (2012). Localization and sampling error correction in ensemble Kalman filter data assimilation. *Monthly Weather Review*, *140*(7), 2359–2371. https://doi.org/10.1175/mwr-d-11-00013.1

Anderson, J., & Lei, L. (2013). Empirical localization of observation impact in ensemble Kalman filters. *Monthly Weather Review*, *141*(11), 4140–4153. https://doi.org/10.1175/MWR-D-12-00330.1

Bishop, C., & Hodyss, D. (2011). Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Monthly Weather Review*, *139*(4), 1241–1255. https://doi.org/10.1175/2010MWR3403.1

Bishop, C. H., & Hodyss, D. (2007). Flow adaptive moderation of spurious ensemble correlations and its use in ensemble-based data assimilation. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences*, *133*(629), 2029–2044. https://doi.org/10.1002/qj.169

Blum, J., LeDimet, F., & Navon, I. (2009). Data assimilation for geophysical fluids. In R. Temam & J. Tribbia (Eds.), *Chapter in computational methods for the atmosphere and the oceans, volume 14: Special volume of handbook of numerical analysis*. Elsevier Science Ltd. publication date: NOV-2008 (Philippe G. Ciarlet, Editor).784.

Bottou, L. (2012). Stochastic gradient descent tricks. In G. Montavon, G. B. Orr, & KR. Müller (Eds.), *Neural networks: Tricks of the trade*, *Lecture notes in computer science* (Vol. 7700). Springer. https://doi.org/10.1007/978-3-642-35289-8_25

Courtier, P., Thepaut, J. N., & Hollingsworth, A. (1994). A strategy for operational implementation of 4DVar using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, *120*(519), 1367–1387. https://doi.org/10.1002/qj.49712051912

Dee, D. (1995). On-line estimation of error covariance parameters for atmospheric data assimilation. *Monthly Weather Review*, *123*(4), 1128–1145. https://doi.org/10.1175/1520-0493(1995)123<1128:oleoec>2.0.co;2

Dee, D., & da Silva, A. (1999). Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: Methodology. *Monthly Weather Review*, *127*(8), 1822–1834. https://doi.org/10.1175/1520-0493(1999)127<1822:mleofa>2.0.co;2

Dennis, J., & Schnabel, R. (1996). *Numerical methods for unconstrained optimization and nonlinear equations (classics in applied mathematics*. SIAM.378.

Evensen, G. (2007). *Data assimilation—the ensemble Kalman filter*. Springer.279.

Gaspari, G., & Cohn, S. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757. https://doi.org/10.1002/qj.49712555417

Hamill, T., Whitaker, J., & Snyder, C. (2001). Distance dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, *129*(11), 2776–2790. https://doi.org/10.1175/1520-0493(2001)129%3C2776:DDFOBE%3E2.0.CO;2

Houtekamer, P., & Mitchell, H. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, *126*(3), 796–811. https://doi.org/10.1175/1520-0493(1998)126%3C0796:DAUAEK%3E2.0.CO;2

Houtekamer, P., Pellerin, G., Buehner, M., Charron, M., Spacek, L., & Hansen, B. (2005). Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations. *Monthly Weather Review*, *133*(3), 604–620. https://doi.org/10.1175/MWR-2864.1

Kalnay, E. (2003). *Atmospheric modeling, data assimilation and predictability*. Cambridge university press.

Kepert, J. (2008). Covariance localisation and balance in an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, *135*(642), 1157–1176. https://doi.org/10.1002/qj.443

Kirchgessner, P., Nerger, L., & Bunse-Gerstner, A. (2014). On the choice of an optimal localization radius in ensemble Kalman filter methods. *Monthly Weather Review*, *142*(6), 2165–2175. https://doi.org/10.1175/mwr-d-13-00246.1

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lei, L., & Anderson, J. (2014). Empirical localization of observations for serial ensemble Kalman filter data assimilation in an atmospheric general circulation model. *Monthly Weather Review*, *142*(5), 1835–1851. https://doi.org/10.1175/MWR-D-13-00288.1

Lei, L., Anderson, J. L., & Romine, G. S. (2015). Empirical localization functions for ensemble Kalman filter data assimilation in regions with and without precipitation. *Monthly Weather Review*, *143*(9), 3664–3679. https://doi.org/10.1175/MWR-D-14-00415.1

Lewis, J., & Derber, J. (1985). The use of the adjoint equation to solve a variational adjustment problem with advective constraints. *Tellus*, *37*(4), 309–322. https://doi.org/10.3402/tellusa.v37i4.11675

Liu, C., Xiao, Q., & Wang, B. (2009). An ensemble-based four-dimensional variational data assimilation scheme. Part II: Observing system simulation experiments with Advanced Research WRF (ARW). *Monthly Weather Review*, *137*(5), 1687–1704. https://doi.org/10.1175/2008MWR2699.1

Liu, D. C., & Nocedal, J. (1989). On the limited memory BFGS method for large scale optimiza-tion. *Mathematical Programming*, *45*(1–3), 503–528. https://doi.org/10.1007/bf01589116

Lorenc, A. (2003). Modelling of error covariances by 4D-Var data assimilation. *Quarterly Journal of the Royal Meteorological Society*, *129*(595), 3167–3182. https://doi.org/10.1256/qj.02.131

Matsuno, T. (1966). Numerical integration of the primitive equations by a simulated backward difference method. *Journal of theMeteorological Society of Japan*, *44*, 76–84.

Ménétrier, B., Montmerle, T., Michel, Y., & Berre, L. (2015). Linear filtering of sample covariances for ensemble-based data assimilation. Part I: Optimality criteria and application to variance filtering and covariance localization. *Monthly Weather Review*, *143*(5), 1622–1643. https://doi.org/10.1175/mwr-d-14-00157.1

Moosavi, A., Attia, A., & Sandu, A. (2018). A machine learning approach to adaptive covariance localization. ArXiv e-prints.

Navon, M., Daescu, D. N., & Liu, Z. (2005). The impact of background error on incomplete observations for 4D-Var data assimilation with the FSU GSM. In *International conference on computational science 2005, lecture notes in computer science book series (LNCS)*. In *Journal of Geophysical Research: Atmospheres* (Vol. 3515, pp. 837–844). https://doi.org/10.1007/11428848_107

Parrish, D., & Derber, J. (1992). The National Meteorological Center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, *120*(8), 1747–1763. https://doi.org/10.1175/1520-0493(1992)120<1747:TNMCSS>2.0.CO;2

Qiu, C., Shao, A., Xu, Q., & Wei, L. (2007). Fitting model fields to observations by using singular value decomposition: An ensemble-based 4DVar approach. *Journal of Geophysical Research*, *112*, D11105. https://doi.org/10.1029/2006JD007994

Tian, X., & Feng, X. (2015). A non-linear least squares enhanced POD-4DVar algorithm for data assimilation. *Tellus*, *67*(1), 25340. https://doi.org/10.3402/tellusa.v67.25340

Tian, X., Han, R., & Zhang, H. (2020). An adjoint-free alternating direction method for four-dimensional variational data assimilation with multiple parameter Tikhonov regularization. *Earth and Space Science*, *7*(11), e2020EA001307. https://doi.org/10.1029/2020EA001307

Tian, X., & Zhang, H. (2019). A big data-driven nonlinear least squares four-dimensional variational data assimilation method: Theoretical formulation and conceptual evaluation. *Earth and Space Science*, *6*(8), 1430–1439. https://doi.org/10.1029/2019EA000735

Tian, X., Zhang, H., Feng, X., & Li, X. (2021). i4DVar: An integral correcting four-dimensional variational data assimilation method. *Earth and Space Science*, *8*(9), e2021EA001767. https://doi.org/10.1029/2021EA001767

Tian, X., Zhang, H., Feng, X., & Xie, Y. (2018). Nonlinear least squares En4DVar to 4DEnVar methods for data assimilation: Formulation, analysis, and preliminary evaluation. *Monthly Weather Review*, *146*(1), 77–93. https://doi.org/10.1175/MWR-D-17-0050.1

Whitaker, J., Compo, G., Wei, X., & Hamill, T. (2004). Reanalysis without radiosondes using ensemble data assimilation. *Monthly Weather Review*, *132*(5), 1190–1200. https://doi.org/10.1175/1520-0493(2004)132<1190:rwrued>2.0.co;2

Zhang, H., & Tian, X. (2018). An efficient local correlation matrix decomposition approach for the localization implementation of ensemble-based assimilation methods. *Journal of Geophysical Research: Atmospheres*, *123*, 3556–3573. https://doi.org/10.1002/2017JD027999

Zhang, H., Tian, X., & Feng, X. (2022). Evaluation data of adaptive localization algorithm based on 2-D shallow water equation model. National Tibetan Plateau Data Center, [Dataset]. https://doi.org/10.11888/Atmos.tpdc.272233

Zhen, Y., & Zhang, F. (2014). A probabilistic approach to adaptive covariance localization for serial ensemble square root filters. *Monthly Weather Review*, *142*(12), 4499–4518. https://doi.org/10.1175/MWR-D-13-00390.1

Zheng, X. (2009). An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. *Advances in Atmospheric Sciences*, *26*(1), 154–160. https://doi.org/10.1007/s00376-009-0154-5