



## RESEARCH ARTICLE

# High-dimensional ensemble Kalman filter with localization, inflation, and iterative updates

Hao-Xuan Sun<sup>1</sup>  | Shouxia Wang<sup>2</sup> | Xiaogu Zheng<sup>3,4</sup> | Song Xi Chen<sup>5</sup> 

<sup>1</sup>Center for Big Data Research, Peking University, Beijing, China

<sup>2</sup>School of Mathematical Science, Peking University, Beijing, China

<sup>3</sup>Shanghai Zhangjiang Mathematics Institute, Pudong, China

<sup>4</sup>International Global Change Institute, Hamilton, New Zealand

<sup>5</sup>Department of Statistics and Data Science, Tsinghua University, Beijing, China

## Correspondence

Song Xi Chen, Department of Statistics and Data Science, Tsinghua University, Beijing 100084, China.  
Email: [songxichen@pku.edu.cn](mailto:songxichen@pku.edu.cn)

## Funding information

National Natural Science Foundation of China, Grant/Award Numbers: 12292980, 12292983, 92358303; National Key Scientific and Technological Infrastructure project “Earth System Science Numerical Simulator Facility” (EarthLab)

## Abstract

Accurate estimation of forecast-error covariance matrices is an essential step in data assimilation, which becomes a challenging task for high-dimensional data assimilation. The standard ensemble Kalman filter (EnKF) may diverge due to both the limited ensemble size and the model bias. In this article, we propose to replace the sample covariance in the EnKF with a statistically consistent high-dimensional tapering covariance matrix estimator to counter the estimation problem under high dimensions. A high-dimensional EnKF scheme combining covariance localization with the inflation method and an iterative update structure is developed. The proposed assimilation scheme is tested on the Lorenz-96 model with spatially correlated observation systems. The results demonstrate that the proposed method could improve the assimilation performance under multiple settings.

## KEYWORDS

data assimilation, ensemble Kalman filter, high-dimensional covariance estimation, localization length-scale selection

## 1 | INTRODUCTION

Data assimilation is a procedure to produce high-quality estimates of the state variables of a dynamic system based on a numerical model and observations of the system (Talagrand, 1997). The ensemble Kalman filter (EnKF) is a popular data assimilation scheme in atmospheric and oceanic science (Evensen, 1994). The EnKF facilitates the Kalman filter (KF: Kalman, 1960) without the computation and storage burden and assumption of linear systems demanded by the KF. A pillar of EnKF technology is high-quality estimates of the forecast-error covariance matrices based on the forecast ensemble. In general, a good estimation can be attained if the dimension ( $p$ ) of the state variables is small relative to the size of the ensembles ( $n$ ). The standard EnKF scheme estimates the forecast-error

covariance matrices by the sample covariance matrix of the ensemble forecast states. However, such an estimation is no longer statistically consistent if  $p/n \rightarrow 0$  (Bai *et al.*, 1988; Bai & Yin, 1993) when the dimension of the state vector is much larger than the ensemble size. In fact, this would be a common case in numerical weather prediction. Indeed, the excessive computation cost for generating ensembles from high-dimensional nonlinear systems prevents the ensemble size  $n$  from being large, which exacerbates the high dimensionality issue.

A statistically inconsistent forecast-error covariance estimator would at least cause the quality of data assimilation by the EnKF to be uncertain and would lose the statistical guarantee. Indeed, past EnKF works have found that the forecast-error covariance matrix is generally underestimated by the sample covariance matrix.

This underestimation is one of the main reasons for filter divergence (e.g., Anderson & Anderson, 1999; Constantinescu *et al.*, 2007). To compensate for this underestimation, inflation methods have been introduced to increase the ensemble spread. Such methods include additive inflation (Constantinescu *et al.*, 2007; Hamill & Whitaker, 2005) and multiplicative inflation (Anderson, 2007, 2009; Anderson & Anderson, 1999; Hodyss *et al.*, 2016; Li *et al.*, 2009; Liang *et al.*, 2012; Wang & Bishop, 2003; Wu *et al.*, 2013; Zheng, 2009). The inflation factor used in these methods can be optimized using the likelihood function of the observation-minus-forecast residuals (Dee *et al.*, 1999; Dee & Da Silva, 1999; Liang *et al.*, 2012; Zheng, 2009). Wu *et al.* (2013) and Zheng *et al.* (2013) developed an iterative update structure, which improves the estimation of the Kalman gain matrix using the newly specified analysis states. The iterative update structure combines additive and multiplicative inflation and can be viewed as an extra step to correct the ensemble prediction error (i.e., ensemble mean-minus-true state). In general, inflation approaches can account for not only the covariance estimation problem caused by the limited ensemble size, to some extent, but also the bias from inappropriate initial perturbations and errors of the forecast models. However, all these methods are based on the sample covariance matrix, a poor estimation of the forecast-error covariance matrix under high dimensions. In other words, the covariance estimation problem via high-dimensional data is not resolved directly.

Another approach to counter the filter divergence is covariance localization. Due to the limited ensemble size of numerical models, covariances between statistically independent variables can be estimated as non-zero. The key idea of localization is to diagnose and ignore the unphysical correlation systematically (Houtekamer & Mitchell, 1998). As proposed by Houtekamer and Mitchell (2001) and Hamill *et al.* (2001), localization methods are implemented by applying the elementwise (Schur) product on the sample covariance matrix with a certain local-supported correlation function. An ensemble size expansion approach via modulation product was developed by Bishop and Hodyss (2009) and Bishop *et al.* (2017). Such an approach showed an easy way of incorporating localization within an ensemble-based assimilation scheme. The localization parameters were configured manually in early practice. Recently, a vast adaptive localization parameter selection method has been developed; see Anderson and Lei (2013), Bishop and Hodyss (2011), Flowerdew (2015), and Morzfeld and Hodyss (2023).

It is notable that the localization technique virtually resembles the estimation of covariance matrices via high-dimensional data under certain structure assumptions. Recent statistical research has provided a deeper

understanding of the influence of high dimensions. Furrer and Bengtsson (2007) quantified that the ensemble size needs to grow proportionally to the square of the system dimension for bounded error growth. A covariance-shrinking (tapering) technique was developed to reduce the necessary ensemble size requirements. Bickel and Levina (2008) and Cai *et al.* (2010) proposed a statistically consistent banding and linearly tapering estimator for the high-dimensional covariance matrix, respectively, while Qiu and Chen (2015) developed corresponding bandwidth selection methods by minimizing the expected standardized squared Frobenius norm of the estimation error matrix. Such a selection scheme can be used to optimize the localization length-scale from the sample statistically. To some extent, the tapering estimator can be viewed as the localization of the sample covariance matrix under distance decay and sparsity conditions. Those studies have built a complete framework to deal with the covariance estimation problem via high-dimensional data and have extensive application prospects for data assimilation in the era of big data.

In this article, we apply the high-dimensional covariance matrix estimation method recently developed in statistics to the EnKF assimilation scheme. The sample covariance of the forecast ensemble in the EnKF is replaced by a high-dimensional tapering covariance estimator. Specifically, the localization length-scale is selected by minimizing the standardized squared Frobenius norm loss using the estimation of the forecast-error covariance. In general, the tapering estimator succeeds in maintaining statistical consistency under mild assumptions regarding the covariance matrix structure. Meanwhile, the inflation method and iterative update structure are incorporated to account for model bias. The proposed paradigm is finally tested on a Lorenz-96 (L96: Lorenz, 1996) model under multiple settings.

The main contributions of this article are the following. First, the high-dimensional covariance tapering estimator is introduced to take the effect of limited ensemble size fully into account when estimating the forecast-error covariance matrices. Compared with the aforementioned localization technique, the proposed method utilizes the structure of the state variables and can ensure statistical consistency of the estimation for any given tapering function. Second, the proposed EnKF scheme is incorporated with inflation and iterative updates. In general, the two techniques are capable of resisting forecast model biases and thus guarantee a more robust assimilation performance. Third, the proposed localization length-scale estimator does not rely on observations, which offers flexibility in practical use.

The article is organized as follows. In Section 2, we introduce high-dimensional localization using the

tapering estimator and propose a high-dimensional EnKF assimilation scheme. Section 3 presents the simulation results on the L96 model under different settings for the dimensions of model states, observations, and ensemble size, and compares the proposed high-dimensional EnKF scheme with existing EnKF schemes. Finally, a discussion and conclusions are given in Section 4. Details of derivations and additional experiments are provided in the Appendix and Supporting Information (SI).

## 2 | METHODOLOGY

### 2.1 | Ensemble Kalman filter (EnKF) and inflation methods

Using the notations of Ide *et al.* (1997), a possibly nonlinear discrete-time forecast and linear observational system is

$$\begin{cases} \mathbf{x}_i^t = M_{i-1}(\mathbf{x}_{i-1}^t) + \boldsymbol{\eta}_i, \\ \mathbf{y}_i^o = \mathbf{H}_i \mathbf{x}_i^t + \boldsymbol{\varepsilon}_i, \end{cases} \quad (1)$$

where  $i$  is the time index,  $\mathbf{x}_i^t$  is the  $p$ -dimensional true state vector at time step  $i$ ,  $M_{i-1}$  is a nonlinear forecast operator at time step  $i-1$ ,  $\mathbf{y}_i^o$  is a  $q_i$ -dimensional observation vector,  $\mathbf{H}_i$  is a linear observation operator, which is a  $q_i \times p$  matrix that maps the model state to the observation state, and  $\boldsymbol{\eta}_i$  and  $\boldsymbol{\varepsilon}_i$  are the model-error and observation-error vectors, respectively, which are assumed to be independent of each other and have mean zero and covariance matrices  $\mathbf{Q}_i$  and  $\mathbf{R}_i$ , respectively. The model error  $\boldsymbol{\eta}_i$  has three sources: (i) the errors in the parameters of  $M_{i-1}$ , (ii) the errors of the numerical schemes used to integrate  $M_{i-1}$ , and (iii) the effect of unresolved scales (Carrassi *et al.*, 2018).

The goal of the EnKF is to find the analysis state  $\bar{\mathbf{x}}_i^a$  that is sufficiently close to the true state  $\mathbf{x}_i^t$  by assimilating the observations into the forecasts. To overcome the divergence of the standard EnKF, the inflation scheme has been introduced to take into account the underestimation of forecast-error covariance matrices. The inflation factors are usually time-varying and are estimated using the information of the observation-minus-forecast residuals at each time step. Suppose the perturbed analysis states at time step  $i$  are  $\{\mathbf{x}_{i,j}^a\}_{j=1}^n$ , where  $n$  is the ensemble size and the analysis state  $\bar{\mathbf{x}}_i^a$  is defined as the ensemble mean  $n^{-1} \sum_{j=1}^n \mathbf{x}_{i,j}^a$ . Let  $\mathbf{P}_i = \mathbb{E}(\mathbf{x}_{i,1}^f - \mathbf{x}_i^t)(\mathbf{x}_{i,1}^f - \mathbf{x}_i^t)^T$  be the true forecast-error covariance matrix at time-step  $i$ . Denote by  $\mathbf{S}_{n,i}$  the sample covariance of forecast error, where we use the subscript  $n$  to emphasize the influence of ensemble size. Then, the EnKF incorporated with the inflation scheme is implemented through the following steps.

- (1). Run the full model forward in time to get the forecast states:

$$\mathbf{x}_{i,j}^f = M_{i-1}(\mathbf{x}_{i-1,j}^a); \quad \bar{\mathbf{x}}_i^f = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_{i,j}^f. \quad (2)$$

- (2.1). Compute the sample forecast-error covariance matrix,

$$\mathbf{S}_{n,i} = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{x}_{i,j}^f - \bar{\mathbf{x}}_i^f)(\mathbf{x}_{i,j}^f - \bar{\mathbf{x}}_i^f)^T, \quad (3)$$

and the perturbed observation-minus-forecast residuals,

$$\mathbf{d}_{i,j} = \mathbf{y}_i^o + \boldsymbol{\varepsilon}_{i,j}' - \mathbf{H}_i \mathbf{x}_{i,j}^f; \quad \mathbf{d}_i = \frac{1}{n} \sum_{j=1}^n \mathbf{d}_{i,j}, \quad (4)$$

where  $\boldsymbol{\varepsilon}_{i,j}'$  are sampled from  $N_{q_i}(\mathbf{0}_{q_i}, \mathbf{R}_i)$ .

- (2.2). Calculate the forecast-error covariance matrix as  $\hat{\mathbf{P}}_i = \mathbf{S}_{n,i}$  and inflate it by a factor  $\hat{\lambda}_i$  as described below around Equation (6).

- (2.3). Compute the perturbed analysis states

$$\mathbf{x}_{i,j}^a = \mathbf{x}_{i,j}^f + \hat{\lambda}_i \hat{\mathbf{P}}_i \mathbf{H}_i^T (\mathbf{H}_i \hat{\lambda}_i \hat{\mathbf{P}}_i \mathbf{H}_i^T + \mathbf{R}_i)^{-1} \mathbf{d}_{i,j}. \quad (5)$$

- (3). If  $i$  is not the ending time, set  $i = i + 1$  and repeat steps (1)–(2). Otherwise, the filtering ends.

The time-varying inflation factors  $\hat{\lambda}_i$  in Step (2.2) are obtained by the maximum-likelihood estimation (MLE) method (Liang *et al.*, 2012). Specifically, the MLE inflation scheme assumes  $\mathbf{d}_i \sim N_{q_i}(\mathbf{0}_{q_i}, \mathbf{H}_i \hat{\mathbf{P}}_i \mathbf{H}_i^T + \mathbf{R}_i)$  according to Equation (4) and employs  $-2 \log$  likelihood as the loss function, which is expressed as

$$L(\lambda_i) = \ln \det(\mathbf{H}_i \lambda_i \hat{\mathbf{P}}_i \mathbf{H}_i^T + \mathbf{R}_i) + \mathbf{d}_i^T (\mathbf{H}_i \lambda_i \hat{\mathbf{P}}_i \mathbf{H}_i^T + \mathbf{R}_i)^{-1} \mathbf{d}_i. \quad (6)$$

Minimizing Equation (6) with respect to  $\lambda_i$ , we obtain  $\hat{\lambda}_i$ , the MLE of the inflation factor. Indeed, as pointed out by Liang *et al.* (2012) and later by Zheng *et al.* (2013), MLE inflation is superior to the previous inflation methods, since it concerns the property of higher order moments. Other inflation factor estimation approaches are available in Wang and Bishop (2003), Anderson (2007, 2009), Li *et al.* (2009), Wu *et al.* (2013), and Hodyss *et al.* (2016). To reduce the computation cost, an efficient calculation method of the determinant and the inverse matrix in Equation (5) is given in Appendix A. Note that if we set the inflation factor in Step (2.2) to be 1, then the above assimilation scheme becomes the standard EnKF.

In the EnKF, the forecast-error covariance matrices are first estimated by the sample covariance matrix  $\mathbf{S}_{n,i}$ . However, the sample covariance is no longer statistically consistent when  $n$  is much smaller than the dimension of the state variables (Bai *et al.*, 1988; Bai & Yin, 1993). This may lead to filter divergence and is one of the reasons for proposing inflation methods. Recent

research in high-dimensional statistics has proposed several approaches to estimate the high-dimensional covariance matrix under mild assumptions about the covariance structure. Using the high-dimensional covariance matrix estimator to replace the sample covariance matrix would help to improve the performance of the conventional EnKF, which is the main purpose of this article.

## 2.2 | High-dimensional localization

To address fully the underestimation of forecast-error covariance matrices induced by the limited ensemble size and high-dimension state vector triggered by high-dimensional dynamic systems, we utilize recent results from high-dimensional statistical covariance estimation. As introduced in Equation (9) below, such a regularized estimation justifies the statistical consistency of the localization approach for forecast-error covariance estimation.

Specifically, it is assumed that the forecast-error covariance matrix  $\mathbf{P}_i$  belongs to the following bandable covariance matrix class:

$$\begin{aligned} \mathcal{U}(\epsilon, \alpha, C) = \left\{ \mathbf{P} = [\sigma_{\ell_1 \ell_2}]_{p \times p} : \right. \\ \text{(i) } \max_{\ell_2} \sum_{k_{\ell_1 \ell_2} > k} |\sigma_{\ell_1 \ell_2}| \leq Ck^{-\alpha} \text{ for all } k > 0; \\ \text{(ii) } 0 < \epsilon \leq \lambda_{\min}(\mathbf{P}) \leq \lambda_{\max}(\mathbf{P}) \leq \epsilon^{-1} \left. \right\}, \quad (7) \end{aligned}$$

for some positive constants  $\epsilon$ ,  $\alpha$ , and  $C$ , where  $\lambda_{\min}(\mathbf{P})$  and  $\lambda_{\max}(\mathbf{P})$  denote the smallest and largest eigenvalues of  $\mathbf{P}$ , respectively,  $k_{\ell_1 \ell_2}$  is the underlying distance between two sites or grids corresponding to the  $\ell_1$ th and  $\ell_2$ th elements of the model state vector, and  $\alpha$  is a parameter that quantifies the rate of decay to zero in the covariances between two locations. Then, a natural approach to estimate the forecast-error covariance  $\mathbf{P}_i$  is to shrink to zero the covariances at two locations with geo-distance that is sufficiently large. Such an idea is effectively the localization approach.

To achieve the localization, we define a tapering function  $g(z)$  on  $[0, \infty)$ , which is non-increasing, non-negative, and satisfies  $g(0) = 1$ ,  $g(z) > 0$  for  $z \in (0, 1)$  and  $g(z) = 0$  for  $z > 1$ . Then, for a matrix  $\mathbf{P} = [\sigma_{\ell_1 \ell_2}]_{p \times p}$ , a tapering operator induced by the tapering function  $g$  is

$$\mathbf{T}_g(\mathbf{P}, k_g) \equiv [\sigma_{\ell_1 \ell_2} g(k_{\ell_1 \ell_2}/k_g)]_{p \times p}, \quad (8)$$

where  $k_g$  is the parameter corresponding to the localization length-scale.

The tapering estimator can achieve a better estimation of the forecast-error covariance matrix  $\mathbf{P}_i$  for high-dimension models that allow  $p$  a lot larger than the ensemble size  $n$  (Bickel & Levina, 2008). Specifically,

suppose the ensemble members  $\{\mathbf{x}_{i,j}^f\}_{j=1}^n$  at time step  $i$  are drawn from a Gaussian distribution, as commonly assumed in EnKF. Let  $\mathbf{S}_{n,i}$  be the sample covariance matrix of the ensemble  $\{\mathbf{x}_{i,j}^f\}_{j=1}^n$  and denote by  $\|\cdot\|$  and  $\|\cdot\|_F$  the operator norm and the Frobenius norm of a matrix, respectively. Arrange the underlying distances  $\{k_{\ell_1 \ell_2}\}$  in ascending order and relabel them as  $\{k_1, \dots, k_L\}$ , then, according to Bickel and Levina (2008), if  $\sum_{\ell=1}^L g(k_{\ell}/k_{g,i}) \asymp (n^{-1} \log p)^{-1/(2(\alpha+1))}$ , we have

$$\|\mathbf{T}_g(\mathbf{S}_{n,i}, k_{g,i}) - \mathbf{P}_i\| = O_p \left\{ \left( \frac{\log p}{n} \right)^{\alpha/(2(\alpha+1))} \right\}, \quad (9)$$

where  $O_p(\cdot)$  refers to the stochastic boundedness, the specific meaning of which is given in Appendix S1 of the SI. Equation (9) means that the tapering estimator for  $\mathbf{P}_i$  is statistically consistent if  $\log p$  is a smaller order of the ensemble size  $n$ . Such a condition is more easily satisfied than requiring  $p$  to be a smaller order of  $n$ . In other words, a much smaller ensemble size  $n$  can be used for the same quality of estimation of  $\mathbf{P}_i$ .

The question is how to choose  $k_g$  for a given tapering function  $g$ . We propose selecting it by minimizing the standardized squared Frobenius norm between the tapering estimator and  $\mathbf{P}_i$ . The choice of the standardized squared Frobenius norm is due to its easier traceability. Meanwhile, as shown in Qiu and Chen (2015), the localization length-scale selected by minimizing the standardized squared Frobenius norm is statistically consistent with the underlying localization length-scale. One may verify using the same approach of Qiu and Chen (2015) that

$$\begin{aligned} \|\mathbf{T}_g(\mathbf{S}_{n,i}, k_g) - \mathbf{P}_i\|_F^2 = \sum_{k_{\ell_1 \ell_2} \leq k_g} \left\{ g\left(\frac{k_{\ell_1 \ell_2}}{k_g}\right) \hat{\sigma}_{\ell_1 \ell_2} - \hat{\sigma}_{\ell_1 \ell_2} \right\}^2 \\ + \sum_{k_{\ell_1 \ell_2} > k_g} \hat{\sigma}_{\ell_1 \ell_2}^2, \quad (10) \end{aligned}$$

where  $\hat{\sigma}_{\ell_1 \ell_2}$  is the  $(\ell_1, \ell_2)$  element of  $\mathbf{S}_{n,i}$ . Then the objective function is

$$\begin{aligned} L_{g,i}(k_g) &\equiv p^{-1} \mathbb{E} \|\mathbf{T}_g(\mathbf{S}_{n,i}, k_g) - \mathbf{P}_i\|_F^2 \\ &= \frac{1}{p} \text{tr}(\mathbf{P}_i^2) + \frac{1}{p} (1 - n^{-1}) \tilde{L}_{g,i}(k_g), \quad (11) \end{aligned}$$

where the first term  $\frac{1}{p} \text{tr}(\mathbf{P}_i^2)$  is an ignorable constant and

$$\begin{aligned} \tilde{L}_{g,i}(k_g) = \sum_{k_{\ell_1 \ell_2} \leq k_g} \left[ \left\{ g\left(\frac{k_{\ell_1 \ell_2}}{k_g}\right) - 2g\left(\frac{k_{\ell_1 \ell_2}}{k_g}\right) \right\} \sigma_{\ell_1 \ell_2}^2 \right. \\ \left. + n^{-1} g\left(\frac{k_{\ell_1 \ell_2}}{k_g}\right) \sigma_{\ell_1 \ell_1} \sigma_{\ell_2 \ell_2} \right]. \quad (12) \end{aligned}$$



The detailed derivation is shown in Appendix S2 of the SI. By minimizing the objective function (12), the localization length-scale  $k_{g,i}$  can be selected by

$$\hat{k}_{g,i} = \arg \min_{k_g \in (\underline{k}, \bar{k})} \hat{L}_{g,i}(k_g), \quad (13)$$

where  $\hat{L}_{g,i}(k_g)$  is an estimator of  $\tilde{L}_{g,i}(k_g)$ . See Appendix S2 of the SI for its computation. Without extra information, the searching interval  $(\underline{k}, \bar{k})$  can be set to be  $\underline{k} = C^{-1}k_0(n^{-1} \log p)^{-1/2}$  and  $\bar{k} = Ck_0(n^{-1} \log p)^{-1/2}$  for some constant  $C$  and  $k_0$  that is relevant to the average distance between two adjacent grids. The larger  $C$ , the wider the searching interval would be. We set  $C = 10$  in the experiments for selecting the localization length-scale. In the meanwhile, the prior knowledge of the model state and the estimated localization length-scales in past steps may be helpful to narrow down the interval and reduce the computation cost.

Three tapering functions are considered in this study. They are the banding function  $g(z) = \mathbb{I}\{0 \leq z \leq 1\}$  from Bickel and Levina (2008), the linearly tapering function  $g(z) = \mathbb{I}\{0 \leq z \leq 1/2\} + (2 - 2z)\mathbb{I}\{1/2 < z \leq 1\}$  from Cai *et al.* (2010), and a fifth-order piecewise rational function  $g(z) = \phi(2z)$ , where

$$\phi(z) = \begin{cases} 1 - \frac{5}{3}z^2 + \frac{5}{8}z^3 + \frac{1}{2}z^4 - \frac{1}{4}z^5, & 0 \leq z \leq 1; \\ -\frac{2}{3}z^{-1} + 4 - 5z + \frac{5}{3}z^2 + \frac{5}{8}z^3 - \frac{1}{2}z^4 + \frac{1}{12}z^5, & 1 < z \leq 2; \\ 0, & z \geq 2. \end{cases} \quad (14)$$

The latter is a widely used localization function (Gaspari & Cohn, 1999). With an appropriate choice of the localization length-scale obtained via Equation (13), any of the three tapering functions can produce a statistically consistent estimator of the forecast-error covariance matrix as guaranteed by Equation (9).

As the tapering estimator  $T_g(\cdot, \cdot)$  may not result in a positive definite matrix, one can obtain a semi-positive definite version of  $T_g(\cdot, \cdot)$  via the eigendecomposition and truncate off those negative eigenvalues. In particular, computation saving can be made by applying the implicitly restarted Arnoldi method (IRAM: Sorensen, 1992) or the random singular value decomposition (SVD) algorithm (Halko *et al.*, 2011) to approximate the  $u \geq n$  largest eigenvalues and associated eigenvectors of  $T_g(\mathbf{S}_{n,i}, \hat{k}_{g,i})$ . Both methods can be used to recover a semi-positive tapering estimator while maintaining sufficient quality of data assimilation. The computational complexities, as well as results on the quality of the data assimilation with respect to using different  $u \geq n$ , have been provided in Appendix S3 of the SI via numerical experiments.

We use the tapering estimator  $T_g(\mathbf{S}_{n,i}, \hat{k}_{g,i})$  to replace the conventional sample covariance  $\mathbf{S}_{n,i}$  in Step (2.1) in Section 2.1 to tackle the estimation problem for high-dimensional systems with limited ensemble size. As noted earlier, another cause of filtering divergence is the error of the forecast model. This refers to the fact that  $\mathbf{M}_i$  in Equation (1) may be mis-specified. In addition to the high-dimensional tapering estimator, an extra step, which we call the iterative updates, is needed to take care of the model bias.

As the analysis state  $\bar{\mathbf{x}}_i^a$  is a better estimation of  $\mathbf{x}_i^t$  than the forecast state  $\bar{\mathbf{x}}_i^f$ , we will adopt the treatment of Wu *et al.* (2013) and Zheng *et al.* (2013) to replace the ensemble mean  $\bar{\mathbf{x}}_i^f$  in Equation (3) with the analysis state  $\bar{\mathbf{x}}_i^a$  and repeat Steps (2.1)–(2.3) using the tapering estimator. Such a replacement can enhance the quality of the forecast-error covariance estimation. It is worth noting that the analysis ensemble should be used to enhance the estimation of the forecast-error statistics (i.e., covariance) but not the forecast ensemble itself. A mathematical explanation is provided in Appendix S4 of the SI. For each repetition, the objective function (6) is recalculated. The process is repeated until the objective function (6) converges. Specifically, we denote by superscript  $r$  a variable to be updated in the  $r$ th round and let  $\delta$  be the threshold of the iteration stop condition. Then, the iterative update structure is implemented as follows.

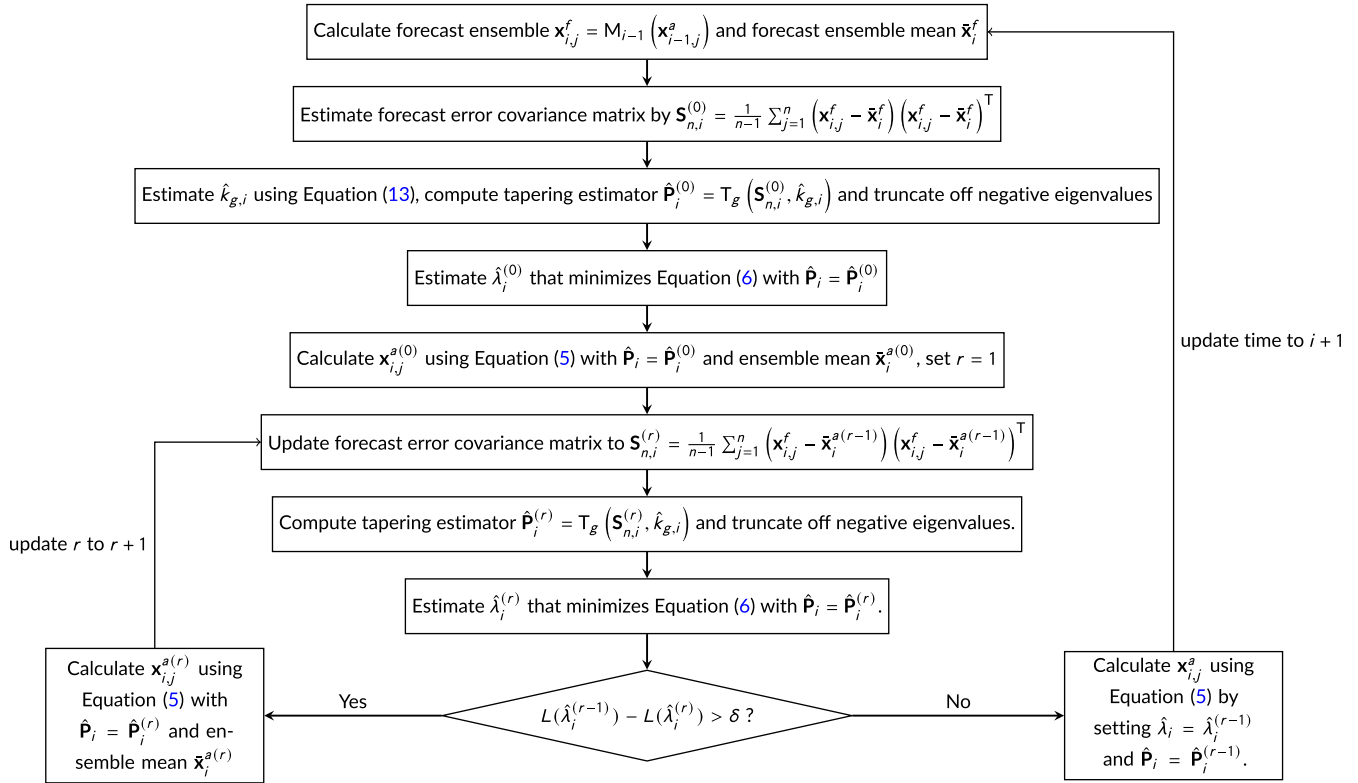
- (2.4). Initialize  $\mathbf{x}_{i,j}^{a(0)} = \mathbf{x}_{i,j}^a$ ,  $\bar{\mathbf{x}}_i^{a(0)} = \bar{\mathbf{x}}_i^a$ , and  $\hat{\mathbf{P}}_i^{(0)} = T_g(\mathbf{S}_{n,i}, \hat{k}_{g,i})$ . For the  $r$ th round, update

$$\mathbf{S}_{n,i}^{(r)} = \frac{1}{n-1} \sum_{j=1}^n \left( \mathbf{x}_{i,j}^f - \bar{\mathbf{x}}_i^{a(r-1)} \right) \left( \mathbf{x}_{i,j}^f - \bar{\mathbf{x}}_i^{a(r-1)} \right)^T$$

with the average analysis states  $\bar{\mathbf{x}}_i^{a(r-1)} = n^{-1} \sum_{j=1}^n \mathbf{x}_{i,j}^{a(r-1)}$  calculated in the  $(r-1)$ th round and estimate the inflation factor  $\lambda_i^{(r)}$  as the  $\hat{\lambda}_i^{(r)}$  that minimizes the objective function (6) with the substitute  $\hat{\mathbf{P}}_i^{(r)} = T_g(\mathbf{S}_{n,i}^{(r)}, \hat{k}_{g,i})$ . Generate  $\mathbf{x}_{i,j}^{a(r)}$  via Equation (5) and compute  $\bar{\mathbf{x}}_i^{a(r)}$ .

- (2.5). If  $L(\hat{\lambda}_i^{(r-1)}) - L(\hat{\lambda}_i^{(r)}) > \delta$ , set  $r = r + 1$  and repeat Step (2.4). Otherwise, stop the iteration and update the perturbed analysis states as Equation (5) with  $\hat{\lambda}_i = \hat{\lambda}_i^{(r-1)}$  and  $\hat{\mathbf{P}}_i = \hat{\mathbf{P}}_i^{(r-1)}$ .

The iterative update structure can be viewed as an additive inflation method (Zheng *et al.*, 2013) to counter the model biases further. See the discussion in Appendix S4 of the SI. By combining the aforementioned inflation, localization, and iterative update technique, the EnKF can be more adaptive to high-dimensional models and state variables and meets the needs of most practical



**FIGURE 1** A flowchart of the proposed high-dimensional EnKF (HD-EnKF) assimilation scheme that conducts the localization, inflation, and iterative updates. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4846)]

applications. For this purpose, a high-dimensional EnKF (HD-EnKF) approach is finally developed. The proposed assimilation scheme has three new ingredients relative to the standard EnKF. The first one replaces the forecast-error sample covariance matrix (3) in Equation (5) by the tapering estimator (8) with the localization length-scale selected via Equation (13). The second one integrates the MLE inflation (multiplicative inflation), and the third iteratively updates the forecast-error covariance by replacing the ensemble means with the average analysis state (additive inflation). Both inflation techniques are necessary to counter the model bias. A flowchart of the HD-EnKF data assimilation scheme is shown in Figure 1.

### 3 | EXPERIMENT ON THE LORENZ-96 MODEL

#### 3.1 | Description of model and observation systems

The L96 Lorenz, 1996 model is a strongly nonlinear dynamical system governed by the equation

$$\frac{d\mathbf{x}(j)}{dt} = \{\mathbf{x}(j+1) - \mathbf{x}(j-2)\}\mathbf{x}(j-1) - \mathbf{x}(j) + F, \quad (15)$$

for  $j = 1, 2, \dots, p$ , where  $\mathbf{x}(j)$  denotes the  $j$ th element of the vector  $\mathbf{x}$  and we assume  $\mathbf{x}(-1) = \mathbf{x}(p-1)$ ,  $\mathbf{x}(0) = \mathbf{x}(p)$ ,  $\mathbf{x}(p+1) = \mathbf{x}(1)$ . The model is designed to mimic the time evolution of a meteorological quantity  $\mathbf{x}$  at  $p$  equally spaced grid points on a latitude circle.

We set  $p = 40, 100, 200$  and  $F = 8$ , respectively, under which the system behaves chaotically. The “true state” was generated by solving Equation (15) using the fourth-order Runge–Kutta time integration scheme (Butcher, 2016) with a time step of 0.05 non-dimensional unit. Such a setting is roughly equivalent to 6 h in real-world time if the characteristic time-scale of the dissipation in the atmosphere is supposed to be 5 days (Lorenz, 1996). The initial state was set to be  $\mathbf{x}_1^t(j) = F$  for  $j \neq \lfloor p/2 \rfloor$  and  $\mathbf{x}_1^t(\lfloor p/2 \rfloor) = F + 0.001$ .

In our experiments, simulated observations were available at every model grid point and were generated by adding random noise to the true states. For each synthetic observation, the random noise followed a multivariate normal distribution with mean zero and covariance matrix  $\mathbf{R}_i$ , the  $(\ell_1, \ell_2)$  element of which is  $0.5^{\min\{|\ell_1 - \ell_2|, p - |\ell_1 - \ell_2|\}}$ . Such a setting allows spatially correlated observation errors, which suits the case of potential correlation among observations like remote sensing data. We simulated observations every four time steps for 2000 steps in total and the results in the last 1000 steps were reported to

provide more robust conclusions. To demonstrate the effect of high dimensions, we implemented the EnKF assimilation schemes on different ensemble sizes  $n = 20, 30, 40$  for  $p = 40, 100, 200$ , respectively. The initial ensemble was generated by adding random noise following the normal distribution with mean zero and covariance matrix  $0.1\mathbf{I}_p$ .

### 3.2 | Assimilation results

For the L96 model with the multiple experimental settings outlined above, we evaluate the quality of data assimilation for the proposed HD-EnKF assimilation scheme, which combines the high-dimensional covariance matrix tapering estimators with the inflation method and the iterative update structure outlined in Section 2.2.

The proposed scheme is compared with the standard EnKF, the EnKF with both the MLE inflation factor and the iterative update structure as outlined in Sections 2.1 and 2.2, and the EnKF scheme using the tapering estimator for the forecast-error covariance matrix without incorporating the inflation. We also considered an assimilation scheme with a wide range of inflation factors and localization length-scale pairs. At each analysis step, the inflation factor and localization length-scale were selected by minimizing the difference between the analysis state and the true state. The inflation factor and localization length-scale thus selected would represent the best possible manually tuned versions of the inflation and localization EnKF methods. Thus, this is called the “Oracle” setting. Our strategy was to show that the proposed HD-EnKF, which automatically tunes the parameters at each analysis step, would have similar performance to those under the Oracle settings. The detailed algorithm can be seen in Appendix S5 of the SI. Moreover, the high-dimensional tapering estimators in the localization, HD-EnKF, and Oracle schemes were all based on the fifth-order piecewise rational function with the setting  $u = p$ .

Since the L96 model is a forced dissipative model with the forcing term  $F$  and behaves differently under different settings for the forcing term, to simulate the “mis-specified model” scenarios in forecasts we considered mis-specification on the forcing term  $F$ . Specifically, we used  $F' = 4, 5, \dots, 12$  for L96 to generate biased forecasts, while the true forcing term is  $F = 8$ .

For each method and each combination of the parameters  $(p, q, n, F')$ , the simulation experiments were replicated 50 times from two aspects. In one, the initial ensemble in each trial is generated by adding  $N_p(\mathbf{0}_p, 0.1\mathbf{I}_p)$  distributed random noise to  $\mathbf{x}_1$ ; the other was to add  $N_{q_i}(\mathbf{0}_{q_i}, \mathbf{R}_i)$  distributed random noise to the synthetic

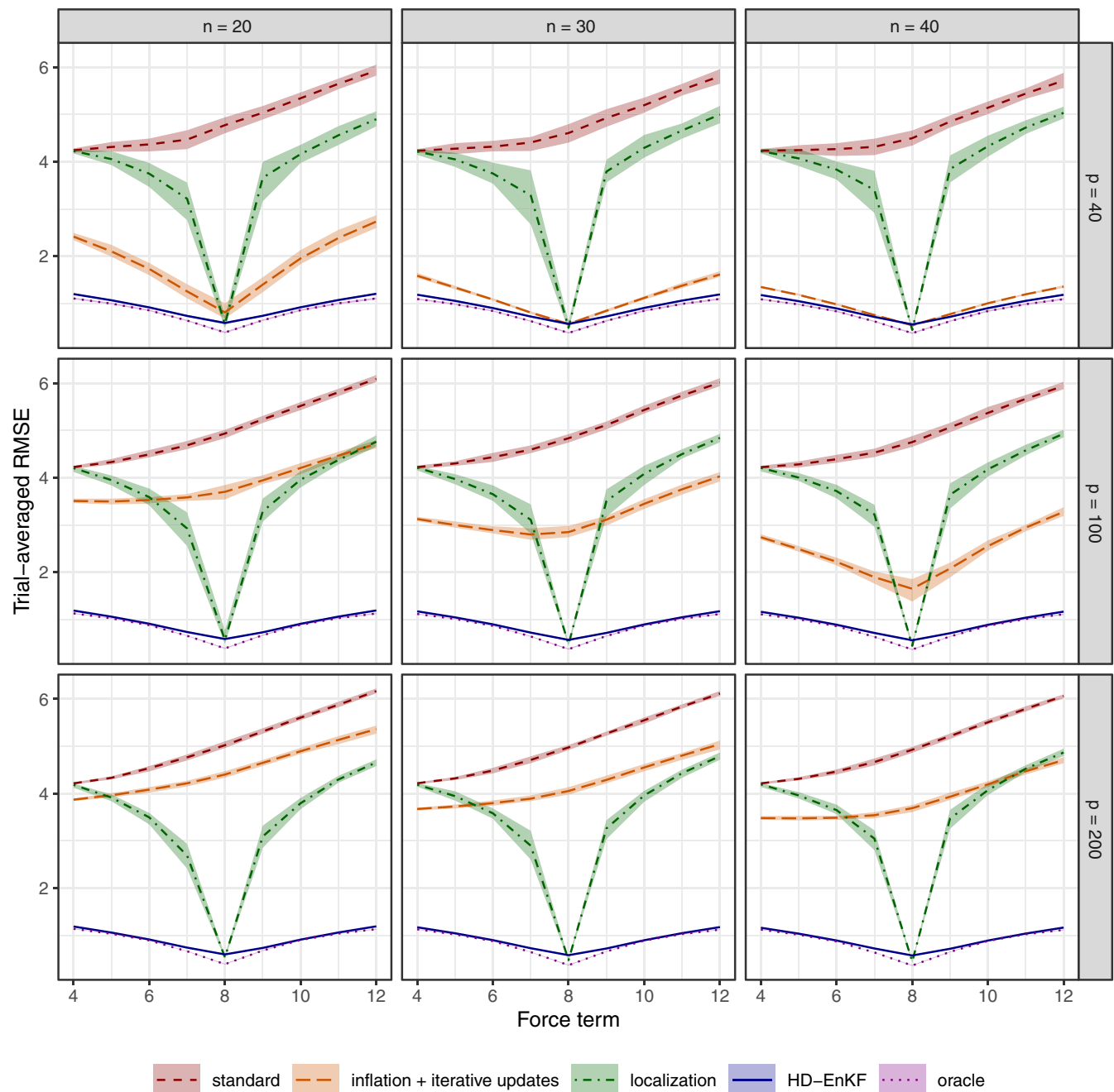
observations for each trial generated by  $\mathbf{H}_i\mathbf{x}_i$ . The assimilation performance was measured by the trial-averaged root-mean-square error (RMSE) over the last 1000 steps, that is,

$$\text{RMSE} = \frac{1}{50,000} \sqrt{\sum_{b=1}^{50} \sum_{i=1001}^{2000} \|\bar{\mathbf{x}}_i^{a,b} - \mathbf{x}_i^t\|_2^2}, \quad (16)$$

where  $\bar{\mathbf{x}}_i^{a,b}$  and  $\mathbf{x}_i^t$  denote the analysis state in the  $b$ th repetition and the true state at time step  $i$ , respectively. Meanwhile, the minimized trial-averaged objective function (6) over the last 1000 steps was also presented to reflect the closeness between the observed  $\mathbf{y}_i^o$  and the projected one from the forecast states.

Figure 2 displays the trial-averaged RMSEs of the five aforementioned EnKF schemes with respect to a range of mis-specified forcing levels. It shows that, as expected, the standard EnKF had the worst performance among the five data assimilation schemes, since it does not adjust for the high dimensionality in the forecast-error covariance and the model bias through localization and inflation. The EnKF with inflation and iterative updates improved upon the performance of the standard EnKF in all settings. However, the assimilation errors of the inflation method increased significantly as the dimension of the state vector  $p$  increased, which reflected the stress caused by the increase in dimension. The advantage of the EnKF with high-dimensional localization via the tapering estimation over the inflation method was apparent when the dimensionality was high and/or the ensemble size  $n$  was small while the model bias was small. The most striking results were the robust and promising performance of the HD-EnKF, which combined the high-dimensional tapering localization with inflation and iterative updates using the assimilated averages in the construction of the forecast-error covariance: the trial-averaged RMSEs were very close to those of the inflation and localization methods under the Oracle setting. These suggest that the HD-EnKF could counter the high dimensionality and the model bias quite well.

To gain insight into the performance when the model is severely biased, Figure 3 displays the analysis RMSEs at each time step for the L96 model with  $F'$  far away from 8, for  $n = 30$  and  $p = q = 40, 100, 200$ , respectively. It shows the same performance ordering of the five methods. A numerical experiment to differentiate the effects of inflation, localization, and iterative updates is provided in Appendix S6 of the SI. It is found that neither the inflation nor the localization alone could obtain assimilation performance comparable to the proposed HD-EnKF scheme, especially for large  $p$ . In other words, the combination of the high-dimensional statistical consistency of



**FIGURE 2** The trial-averaged RMSEs and their 5%–95% quantile bands over the last 1000 steps as a function of forcing term  $F'$  under  $p = q = 40, 100, 200$  and  $n = 20, 30, 40$  for the five EnKF schemes: the standard EnKF (dashed line); the EnKF with inflation and iterative updates (long-dashed line); the EnKF with localization (dot-dashed line); the HD-EnKF (solid line) and the Oracle (dotted line). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4846)]

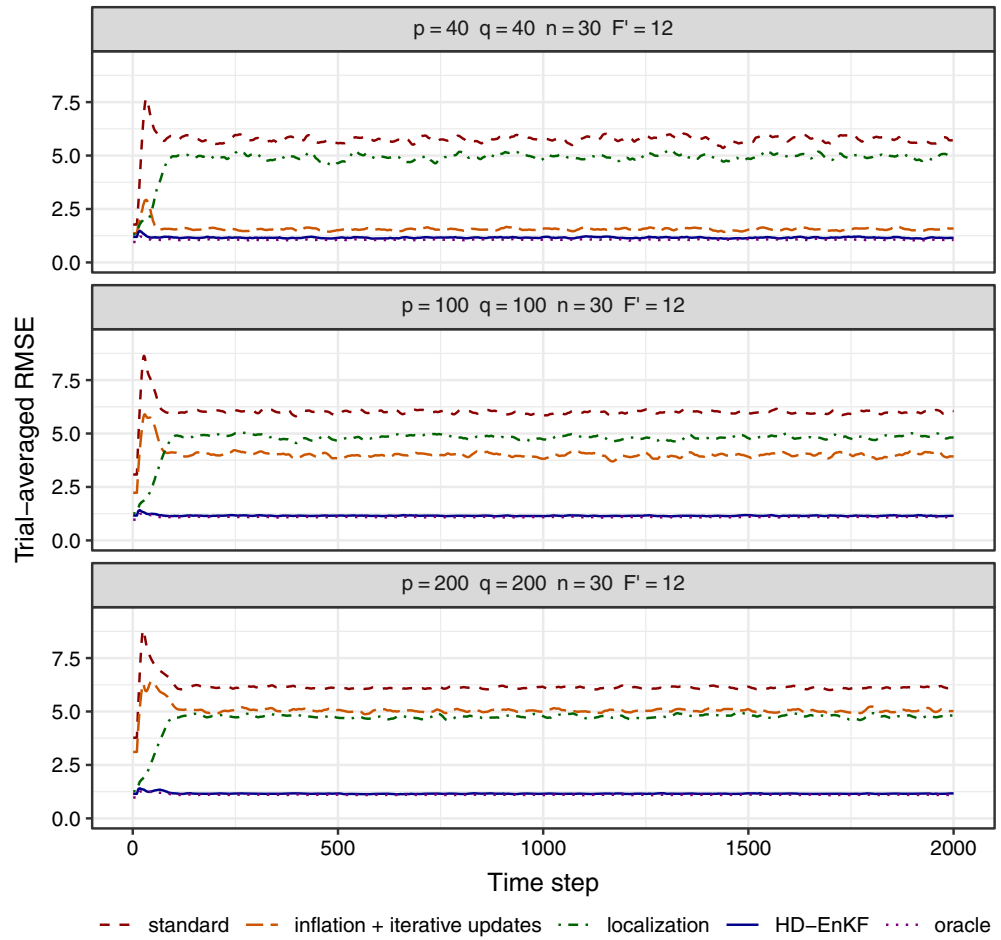
localization and the error correction capability of the inflation technique on model mis-specification is essential to the assimilation process.

Table 1 summarizes the trial-averaged RMSEs and corresponding standard deviations (SD), the value of the minimized objective function, and the selected localization length-scales over the last 1000 steps and 50 repetitions for all the  $(p, n)$  settings when  $F' = 12$ . The proposed

HD-EnKF scheme provided significantly smaller analysis RMSEs at 99% confidence. We note that the smaller the average value of the objective function, the larger the likelihood of observation-minus-forecast errors and the smaller the analysis RMSE. The result verified the efficiency of the iterative update structure to some extent. Besides this, the HD-EnKF generally achieved smaller RMSEs regardless of which of the three tapering functions is used.



**FIGURE 3** The trial-averaged RMSEs under  $p = q = 40, 100, 200$ ,  $n = 30$  and  $F' = 12$  at each time step for the five EnKF schemes: the standard EnKF (dashed line); the EnKF with inflation and iterative updates (long-dashed line); the EnKF with localization (dot-dashed line); the HD-EnKF (solid line) and the Oracle (dotted line). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]



## 4 | DISCUSSION AND CONCLUSIONS

The ultimate goal of data assimilation is to obtain high-quality analysis datasets for scientific research and for better forecasts by fusion of the forecast model and observations. To achieve such an objective, the EnKF inherits the approach of the KF, which re-weights the forecasts and observations via conditional imputation and develops an “online” estimation of the forecast-error covariance matrices through generated ensembles. In this way, the EnKF is not only applicable to nonlinear systems but also avoids the storage and iterative update of  $p \times p$  dimensional matrices.

For EnKF-based data assimilation schemes, the estimation of the forecast-error covariance matrices is key. Note that the forecast-error covariance has the decomposition

$$\begin{aligned} \mathbf{P}_i &= \mathbb{E}(\mathbf{x}_{i,1}^f - \mathbf{x}_i^t)(\mathbf{x}_{i,1}^f - \mathbf{x}_i^t)^T \\ &= \mathbb{E}(\mathbf{x}_{i,1}^f - \bar{\mathbf{x}}_i^f)(\mathbf{x}_{i,1}^f - \bar{\mathbf{x}}_i^f)^T + \mathbb{E}(\bar{\mathbf{x}}_i^f - \mathbf{x}_i^t)(\bar{\mathbf{x}}_i^f - \mathbf{x}_i^t)^T, \end{aligned} \quad (17)$$

where the first term is the variance of the ensemble and the second term reflects the model bias. This represents two pivotal issues that influence the quality of a data assimilation scheme. One is to account properly for the variation of the forecast ensemble. The limited ensemble size  $n$  reduces the quality of the variation estimation in the standard EnKF. From another perspective, it is the increased dimension that hinders the statistical consistency and accuracy of the standard sample covariance estimator. The proposed high-dimensional tapering estimator alleviates this problem. Another issue is the underestimation of  $\mathbf{P}_i$  due to neglect of the bias term. Inflation and iterative updates are designed to compensate for the omission.

The simulation results in Section 3 have demonstrated the necessity to address both issues. The standard EnKF does not account for either issue. As displayed in Figure S1, the absence of localization and inflation led to filter degeneration, with the ensemble spread being much smaller. Meanwhile, the model biases caused different patterns in the variance of analysis states between different trials. Specifically, let  $\bar{\mathbf{x}}_i^a = \frac{1}{50} \sum_{b=1}^{50} \bar{\mathbf{x}}_i^{a,b}$  be the trial-averaged analysis state; then,

$$\frac{1}{50} \sum_{b=1}^{50} \|\bar{\mathbf{x}}_i^{a,b} - \mathbf{x}_i^t\|_2^2 = \frac{1}{50} \sum_{b=1}^{50} \|\bar{\mathbf{x}}_i^{a,b} - \bar{\mathbf{x}}_i^a\|_2^2 + \|\bar{\mathbf{x}}_i^a - \mathbf{x}_i^t\|_2^2, \quad (18)$$

where the first term is the  $L_2$  difference between the trial-averaged analysis state and the analysis state of each trial and the second term is the  $L_2$  difference between the trial-averaged analysis state and the true state. Figure S2 displays the above two terms at a time step. Although a forecast operator under  $F' = 4$  underestimates the model states, the analysis states of the standard EnKF were more concentrated among different trials and thus led to a relatively smaller analysis RMSE. On the other hand, the EnKF with localization only deals with the first issue. As shown in Figure 2, the analysis RMSEs for large  $p$  were much smaller when the model bias was small, with the forcing term near  $F' = 8$ . However, the assimilation

became poor for more severely mis-specified models. The EnKF with inflation and iterative updates accounts for the second issue. When the effect of high dimensions is slight (i.e.,  $p = 40$ ), both inflation methods can reduce the analysis errors effectively. Meanwhile, the larger the model bias, the larger the analysis errors remain. A mathematical explanation is provided in Appendix S4 of the SI. This is evident in the increasing analysis errors as the forcing term  $F'$  deviates from the truth  $F = 8$ . However, the improvement offered by inflation approaches was rather limited as the dimension of the model states  $p$  increased.

The proposed HD-EnKF scheme employs the high-dimensional tapering estimator to replace the sample covariance and retains the MLE inflation method and iterative update structure to counter the model biases. As indicated by the simulation results displayed in Figures 2 and 3, the proposed method outperformed

**TABLE 1** The trial-averaged RMSEs with standard deviations in parentheses, minimized trial-averaged objective function values, and average selected localization length-scales over the last 1000 steps under  $p = q = 40, 100, 200$ ,  $n = 20, 30, 40$ , and  $F' = 12$  for six EnKF schemes: standard EnKF (standard), EnKF with inflation and iterative updates (inflation + iterative updates), EnKF with localization (localization), and the proposed HD-EnKF embedding the high-dimensional tapering estimator induced by the banding function (BL), linearly tapering function (CZZ), and fifth-order piecewise rational function (GC).

EnKF scheme	RMSE	$L(\hat{\lambda}_i)$	$\hat{k}_{g,i}$	RMSE	$L(\hat{\lambda}_i)$	$\hat{k}_{g,i}$	RMSE	$L(\hat{\lambda}_i)$	$\hat{k}_{g,i}$
	$p = 40 \quad q = 40 \quad n = 20$			$p = 40 \quad q = 40 \quad n = 30$			$p = 40 \quad q = 40 \quad n = 40$		
standard	5.93 (0.069)	2173.91	–	5.81 (0.087)	2076.58	–	5.72 (0.09)	2005.57	–
inflation + iterative updates	2.74 (0.085)	287.22	–	1.62 (0.038)	78.15	–	1.36 (0.019)	48.80	–
localization	4.9 (0.102)	1436.41	11.6	5 (0.106)	1498.34	14.2	5.04 (0.086)	1519.89	16.4
HD-EnKF(BL)	1.36 (0.016)*	67.26	5.2	1.31 (0.014)*	61.63	5.9	1.3 (0.011)*	58.58	6.4
HD-EnKF(CZZ)	1.33 (0.013)*	63.73	6.7	1.29 (0.012)*	58.84	7.6	1.27 (0.012)*	55.89	8.3
HD-EnKF(GC)	1.21 (0.01)*	50.53	15	1.19 (0.009)*	47.36	17.1	1.19 (0.011)*	45.58	18.7
	$p = 100 \quad q = 100 \quad n = 20$			$p = 100 \quad q = 100 \quad n = 30$			$p = 100 \quad q = 100 \quad n = 40$		
standard	6.09 (0.043)	5823.40	–	6.01 (0.052)	5635.99	–	5.94 (0.05)	5480.51	–
inflation + iterative updates	4.72 (0.059)	2980.31	–	4.03 (0.062)	1929.16	–	3.28 (0.055)	1111.42	–
localization	4.76 (0.075)	3370.92	10.4	4.84 (0.056)	3490.36	12.8	4.93 (0.057)	3637.73	14.7
HD-EnKF(BL)	1.34 (0.009)*	165.74	4.6	1.3 (0.011)*	150.92	5.7	1.28 (0.009)*	145.04	6
HD-EnKF(CZZ)	1.3 (0.009)*	154.75	6.6	1.27 (0.009)*	142.98	7.3	1.25 (0.007)*	135.82	7.8
HD-EnKF(GC)	1.19 (0.007)*	120.17	14.7	1.17 (0.007)*	112.59	16.6	1.16 (0.007)*	107.97	17.9
	$p = 200 \quad q = 200 \quad n = 20$			$p = 200 \quad q = 200 \quad n = 30$			$p = 200 \quad q = 200 \quad n = 40$		
standard	6.17 (0.027)	11991.93	–	6.12 (0.033)	11750.98	–	6.06 (0.029)	11499.24	–
inflation + iterative updates	5.36 (0.05)	8465.45	–	5.05 (0.06)	7135.95	–	4.71 (0.043)	5872.27	–
localization	4.66 (0.047)	6437.26	9.7	4.79 (0.045)	6814.06	11.9	4.87 (0.045)	7058.22	13.7
HD-EnKF(BL)	1.34 (0.007)*	329.13	4.7	1.3 (0.006)*	304.08	5.6	1.29 (0.007)*	292.07	5.9
HD-EnKF(CZZ)	1.31 (0.006)*	309.09	6.4	1.27 (0.006)*	286.33	7.1	1.25 (0.006)*	271.84	7.7
HD-EnKF(GC)	1.18 (0.005)*	236.22	14.5	1.17 (0.005)*	221.35	16.2	1.16 (0.005)*	212.43	17.5

Note: The subscripts \* mark the trial-averaged RMSEs of the proposed HD-EnKF, which were significantly smaller at 99% confidence than those of the three existing EnKF methods (standard, inflation + iterative updates and localization) under the same  $(p, n)$  setting.

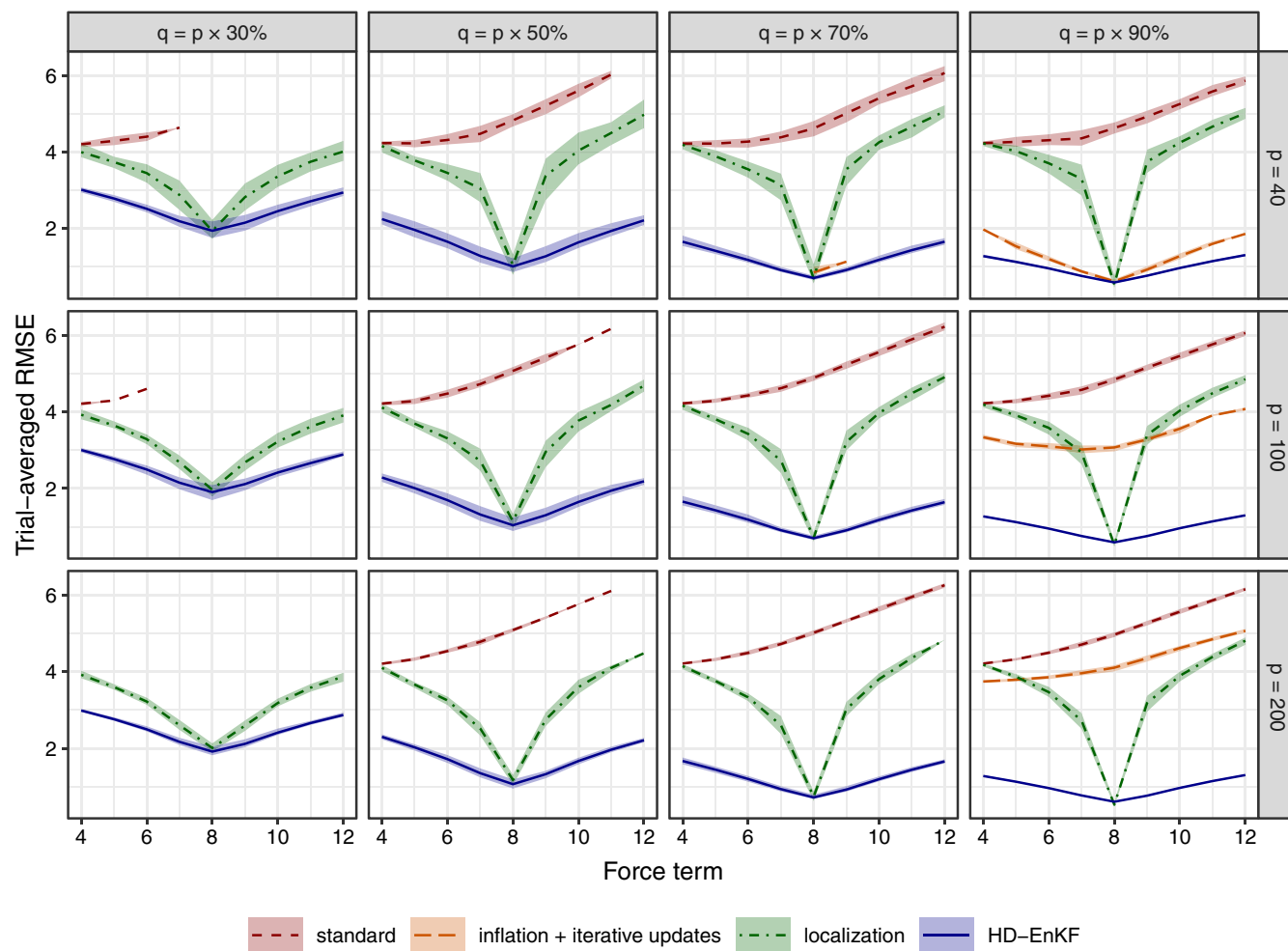
the other schemes and obtained results very close to the Oracle. This may be interpreted as, among all the assimilation schemes that involve the inflation factor and the localization length-scale, the HD-EnKF can achieve assimilation performance that is in very close proximity to the best-tuned versions. The proposed scheme combines the advantages of localization, inflation, and iterative update methods, and the superiority becomes much more apparent as the dimension  $p$  increases. Specifically, the HD-EnKF scheme provided relatively accurate analysis states with smaller analysis RMSEs under larger  $p/n$  ratio settings, much smaller than the aforementioned inflation scheme. On the other hand, for those “mild” settings like  $(p, n) = (40, 30)$  and  $(40, 40)$ , in which the effect of high dimensions is not so significant, the HD-EnKF scheme still offered better performance than the existing methods. For a high-dimensional chaotic system like L96, as shown by the analysis RMSEs of the L96 model under  $p = q = 200$  and  $n = 20, 30, 40$ , the simulation results reveal that employment of the high-dimensional covariance estimator would be necessary to estimate the forecast-error covariance matrices correctly and to guarantee better assimilation performance. It should also be emphasized that use of MLE-based inflation and iterative updates using the analysis mean is necessary to counter incorrectly specified models. As illustrated by Figures 2 and S3, the assimilation process would produce unsatisfactory results for the mis-specified models without the participation of the inflation method and iterative update structure. Since model biases are unavoidable in most applications, inflation steps are needed to address this issue.

In practical applications, the observations of the model grids can be distributed unevenly. We simulated such cases by assimilating the synthetic observations that accounted for 30%, 50%, 70%, and 90% of the total number of model grids. The experiments were implemented under  $p = 40, 100, 200$  and  $n = 30$ , with the observed grids being randomly selected for each experimental setting. As indicated in Figure 4, the HD-EnKF succeeded in assimilating partial observations and had the smallest trial-averaged RMSEs. In addition, the filter divergence rate reported in Figure S4 revealed that the HD-EnKF achieved the least divergence, while the standard EnKF and the EnKF with inflation and iterative updates diverged in most of the trials when the number of observed grids was less than 50%. Another likely situation in practice is that the distributions of observation errors can be skewed. A simulation study is also included by introducing observation errors that follow a scaled Gamma distribution with shape parameter 4 and scale parameter 1. As displayed in Figure S5, although the trial-averaged RMSEs had small increases, the HD-EnKF outperformed the standard EnKF, the EnKF with inflation

and iterative updates, and the EnKF with localization, with the RMSEs largely the same as in the normally distributed case. The details of the experiments can be seen in S7 of the SI.

The advantages of the MLE-based inflation method have already been discussed in the existing literature. Through L96 experiments, Liang *et al.* (2012) suggested that MLE inflation factors outperform moment-estimation approaches (i.e., Wang & Bishop, 2003; Li *et al.*, 2009). Meanwhile, Wu *et al.* (2013) and Zheng *et al.* (2013) proved that the iterative update structure tended to be a better approach than one not updated, especially when the model is mis-specified. An example of a practical application is given by the global carbon flux estimation in Zhang *et al.* (2015). As shown in their fig. 3, the iterative update structure can produce more realistic forecast-error statistics. Therefore, the inflation method suggested in this article is applied to evaluate the localization parameter estimation method in the following study.

It is worth noting that the tapering function and the associated localization length-scale play an important role in the assimilation step. Flowerdew (2015) proposed minimizing the analysis error, while the posterior-optimal localization in Morzfeld and Hodyss (2023) estimated the localization parameters by working directly with the Kalman gain. However, both analysis error and the Kalman gain matrix are related to the observation error, which is physically independent of the forecast error. Hence, it is reasonable to not have the observation error involved in the localization parameter estimating procedures, as otherwise any observation deficiency may degrade the choice of localization parameters. The recently developed prior-optimal localization scheme (Morzfeld & Hodyss, 2023) constructs the localization matrix from the correlation coefficients of the ensemble. Such a scheme is decoupled with the observations and thus is compared with the HD-EnKF in a simulation study under  $p = q = 200$  and  $n = 20$ . As illustrated in Figure 5, the experimental results indicate that prior-optimal localization produced assimilation results with larger analysis errors. Such differences come partially from the different localization function settings, as prior-optimal localization underutilizes the structure of the model state, while Equation (11) in the HD-EnKF parameterizes the objective function according to the information regarding the underlying distance. Another reason may be that “prior-optimal localization dampens variances,” as revealed by eq. (13) in Morzfeld and Hodyss (2023), since it implied that prior optimality is achieved at the cost of biased estimation of the variance of the forecast errors. In contrast, statistical unbiasedness can be guaranteed in the proposed localization procedure, as shown in the derivation of Section 2.2 and Appendix S2 in the SI. From another viewpoint, the experimental results



**FIGURE 4** The trial-averaged RMSEs and their 5%–95% quantile bands over the last 1000 steps as a function of the forcing term  $F'$  under  $p = 40, 100, 200$ ,  $q = p \times 30\%, 50\%, 70\%, 90\%$  and  $n = 30$  for four EnKF schemes: the standard EnKF (dashed line); the EnKF with inflation and iterative updates (long-dashed line); the EnKF with localization (dot-dashed line) and the HD-EnKF (solid line). The disappearance for part of the lines corresponding to the standard EnKF and the EnKF with the inflation and iterative updates is due to the divergence of the algorithm. [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com/doi/10.1002/qj.4846)]

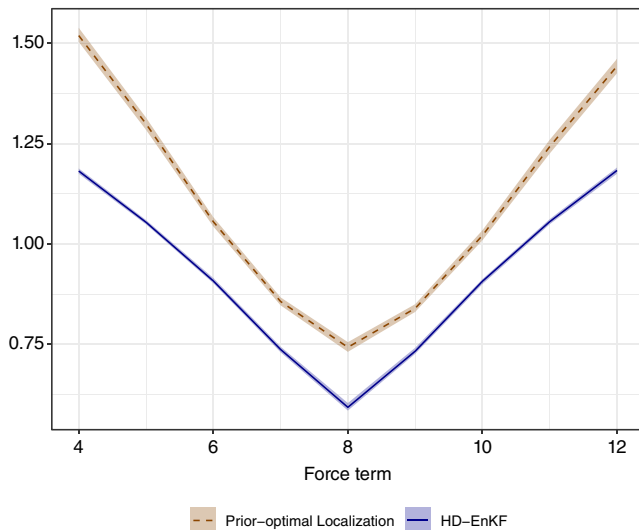
were also consistent with the conclusion of Morzfeld and Hodyss (2023), in that “a simple localization can be as effective as an optimal scheme.”

The computational cost of the proposed HD-EnKF can be reduced further in several ways. First, regarding the selection of the localization length-scale  $k_{g,i}$ , it is noted that the underlying localization length-scale is related to  $\alpha$ , the rate of decay of the covariance with respect to the distance. For a given system, the decay rate  $\alpha$  generally varies slowly over a period of time. This inspires us to estimate the localization length-scale as a constant for a certain period and update it with information about the prior choices. To demonstrate this approach, simulation experiments were implemented on the L96 models under  $p = q = 40, 100, 200$  and  $n = 30$ . For each trial, we first estimated the localization length-scale according to Equation (13) at each assimilation step for the first 1000

steps and then took the average of the estimated localization length-scales as the constant localization length-scale in the last 1000 steps. As displayed in Figure S6, the results showed small differences in the analysis RMSEs of the HD-EnKF schemes when using time-varying versus constant localization length-scales. For a fixed localization length-scale, a further computation saving can be achieved by judicious use of the modulation product proposed in Bishop and Hodyss (2009) and Bishop *et al.* (2017) and later used in Farchi and Bocquet (2019). Such a computation can avoid eigendecomposition at each analysis step.

Moreover, two implementations may be introduced to save the cost of obtaining the tapering covariance matrix  $T_g(\mathbf{S}_{n,i}, \hat{k}_{g,i})$  and calculating Equation (5). The first one is to consider a low-rank approximation of the tapering estimator. As introduced in Section 2.2, with further details in Appendix S3 of the SI, eigendecomposition is needed





**FIGURE 5** The trial-averaged RMSEs and their 5%–95% quantile bands over the last 1000 steps as a function of the forcing term  $F'$  for  $p = q = 200$  and  $n = 20$  for the HD-EnKF (solid line) and the EnKF with the inflation, the iterative updates and the prior-optimal localization (dashed line). [Colour figure can be viewed at [wileyonlinelibrary.com](https://onlinelibrary.wiley.com)]

to guarantee semi-positive definiteness. Indeed, a smaller number, say  $u$ , of the largest eigenvalues and associated eigenvectors can be solved quite efficiently and will help to reduce the cost of computing the analysis state. In practice, the number  $u$  can be decided empirically by the proportion of the sum of the largest eigenvalues in the total sum of all eigenvalues. An experiment is given in Appendix S3 of the SI. The result indicated that the number of largest eigenvalues that account for 90% of the total variance in general can lead to efficient assimilation. Another approach is to estimate the forecast-error covariance matrices and compute the analysis state in blockwise fashion. To be more specific, the model states are partitioned into blocks, with every two adjacent blocks overlapping at the boundaries. For each block, the analysis states are calculated separately and then combined. The blocking technique has been used in practice (Zhang *et al.*, 2015), and computation cost can be saved since the eigendecomposition and calculation in Equation (5) are concerned with much lower dimension. The details of the blockwise technique and the results of the experiments are given in Appendix S8 of the SI.

There is also an issue with respect to the assumption of Gaussian-distributed ensembles. Although this is assumed in most EnKF assimilation schemes, Gaussianity is usually not tested in practical implementations. Fortunately, the high-dimensional tapering estimator and localization length-scale selection methods are valid for general distributions to a certain extent, while the inflation factor estimation method is based on the likelihood function of Gaussian distributions. In future work, the robustness

of the assimilation performance under non-Gaussian ensembles may be studied.

## ACKNOWLEDGEMENTS

The research was partially supported by National Natural Science Foundation of China Grants 12292980, 12292983, and 92358303. This research was also partially supported by the National Key Scientific and Technological Infrastructure project “Earth System Science Numerical Simulator Facility” (EarthLab).

## CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

## DATA AVAILABILITY STATEMENT

The article has not used any empirical data. It only studies simulated data.

## ORCID

Hao-Xuan Sun  <https://orcid.org/0000-0001-5062-3968>

Song Xi Chen  <https://orcid.org/0000-0002-2338-0873>

## REFERENCES

- Anderson, J. (2009) Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A: Dynamic Meteorology and Oceanography*, 61, 72–83.
- Anderson, J. & Lei, L. (2013) Empirical localization of observation impact in ensemble Kalman filters. *Monthly Weather Review*, 141, 4140–4153.
- Anderson, J.L. (2007) An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A: Dynamic Meteorology and Oceanography*, 59, 210–224.
- Anderson, J.L. & Anderson, S.L. (1999) A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127, 2741–2758.
- Bai, Z. & Yin, Y. (1993) Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. *The Annals of Probability*, 21, 1275–1294.
- Bai, Z.D., Silverstein, J.W. & Yin, Y.Q. (1988) A note on the largest eigenvalue of a large dimensional sample covariance matrix. *Journal of Multivariate Analysis*, 26, 166–168.
- Bickel, P.J. & Levina, E. (2008) Regularized estimation of large covariance matrices. *The Annals of Statistics*, 36, 199–227.
- Bishop, C. & Hodyss, D. (2009) Ensemble covariances adaptively localized with ECO-RAP. Part 2: a strategy for the atmosphere. *Tellus A: Dynamic Meteorology and Oceanography*, 61, 97–111.
- Bishop, C.H. & Hodyss, D. (2011) Adaptive ensemble covariance localization in ensemble 4D-VAR state estimation. *Monthly Weather Review*, 139, 1241–1255.
- Bishop, C.H., Whitaker, J.S. & Lei, L. (2017) Gain form of the ensemble transform Kalman filter and its relevance to satellite data assimilation with model space ensemble covariance localization. *Monthly Weather Review*, 145, 4575–4592.
- Butcher, J.C. (2016) *Numerical methods for ordinary differential equations*. Chichester, UK: John Wiley & Sons.

- Cai, T.T., Zhang, C.-H. & Zhou, H.H. (2010) Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38, 2118–2144.
- Carrassi, A., Bocquet, M., Bertino, L. & Evensen, G. (2018) Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *WIREs Climate Change*, 9, e535.
- Constantinescu, E.M., Sandu, A., Chai, T. & Carmichael, G.R. (2007) Ensemble-based chemical data assimilation. I: general approach. *Quarterly Journal of the Royal Meteorological Society*, 133, 1229–1243.
- Dee, D.P. & Da Silva, A.M. (1999) Maximum-likelihood estimation of forecast and observation error covariance parameters. Part I: methodology. *Monthly Weather Review*, 127, 1822–1834.
- Dee, D.P., Gaspari, G., Redder, C., Rukhovets, L. & Da Silva, A.M. (1999) Maximum-likelihood estimation of forecast and observation error covariance parameters. Part II: applications. *Monthly Weather Review*, 127, 1835–1849.
- Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99, 10143–10162.
- Farchi, A. & Bocquet, M. (2019) On the efficiency of covariance localisation of the ensemble Kalman filter using augmented ensembles. *Frontiers in Applied Mathematics and Statistics*, 5, 3.
- Flowerdew, J. (2015) Towards a theory of optimal localisation. *Tellus A: Dynamic Meteorology and Oceanography*, 67, 25257.
- Furrer, R. & Bengtsson, T. (2007) Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *Journal of Multivariate Analysis*, 98, 227–255.
- Gaspari, G. & Cohn, S.E. (1999) Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757.
- Halko, N., Martinsson, P.-G. & Tropp, J.A. (2011) Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53, 217–288.
- Hamill, T.M. & Whitaker, J.S. (2005) Accounting for the error due to unresolved scales in ensemble data assimilation: a comparison of different approaches. *Monthly Weather Review*, 133, 3132–3147.
- Hamill, T.M., Whitaker, J.S. & Snyder, C. (2001) Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Monthly Weather Review*, 129, 2776–2790.
- Hodyss, D., Campbell, W.F. & Whitaker, J.S. (2016) Observation-dependent posterior inflation for the ensemble Kalman filter. *Monthly Weather Review*, 144, 2667–2684.
- Houtekamer, P.L. & Mitchell, H.L. (1998) Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126, 796–811.
- Houtekamer, P.L. & Mitchell, H.L. (2001) A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129, 123–137.
- Ide, K., Courtier, P., Ghil, M. & Lorenc, A.C. (1997) Unified notation for data assimilation: operational, sequential and variational. *Journal of the Meteorological Society of Japan*, 75, 181–189.
- Kalman, R.E. (1960) A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82, 35–45.
- Li, H., Kalnay, E. & Miyoshi, T. (2009) Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quarterly Journal of the Royal Meteorological Society*, 135, 523–533.
- Liang, X., Zheng, X., Zhang, S., Wu, G., Dai, Y. & Li, Y. (2012) Maximum likelihood estimation of inflation factors on error covariance matrices for ensemble Kalman filter assimilation. *Quarterly Journal of the Royal Meteorological Society*, 138, 263–273.
- Lorenz, E.N. (1996) Predictability: a problem partly solved. In: *Proceedings on Seminar on Predictability*, Vol. 1. Reading, UK: ECMWF, pp. 1–18.
- Morzfeld, M. & Hodyss, D. (2023) A theory for why even simple covariance localization is so useful in ensemble data assimilation. *Monthly Weather Review*, 151, 717–736.
- Qiu, Y. & Chen, S.X. (2015) Bandwidth selection for high-dimensional covariance matrix estimation. *Journal of the American Statistical Association*, 110, 1160–1174.
- Sorensen, D.C. (1992) Implicit application of polynomial filters in a  $k$ -step Arnoldi method. *SIAM Journal on Matrix Analysis and Applications*, 13, 357–385.
- Talagrand, O. (1997) Assimilation of observations, an introduction. *Journal of the Meteorological Society of Japan*, 75, 191–209.
- Wang, X. & Bishop, C.H. (2003) A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *Journal of the Atmospheric Sciences*, 60, 1140–1158.
- Wu, G., Zheng, X., Wang, L., Zhang, S., Liang, X. & Li, Y. (2013) A new structure for error covariance matrices and their adaptive estimation in EnKF assimilation. *Quarterly Journal of the Royal Meteorological Society*, 139, 795–804.
- Zhang, S., Zheng, X., Chen, J.M., Chen, Z., Dan, B., Yi, X. et al. (2015) A global carbon assimilation system using a modified ensemble Kalman filter. *Geoscientific Model Development*, 8, 805–816.
- Zheng, X. (2009) An adaptive estimation of forecast error covariance parameters for Kalman filtering data assimilation. *Advances in Atmospheric Sciences*, 26, 154–160.
- Zheng, X., Wu, G., Zhang, S., Liang, X., Dai, Y. & Li, Y. (2013) Using analysis state to construct a forecast error covariance matrix in ensemble Kalman filter assimilation. *Advances in Atmospheric Sciences*, 30, 1303–1312.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Sun, H.-X., Wang, S., Zheng, X. & Chen, S.X. (2024) High-dimensional ensemble Kalman filter with localization, inflation, and iterative updates. *Quarterly Journal of the Royal Meteorological Society*, 1–15. Available from: <https://doi.org/10.1002/qj.4846>

## APPENDIX A. EFFICIENT CALCULATION

To achieve efficient calculation of  $\det(\mathbf{H}_i \hat{\lambda}_i \hat{\mathbf{P}}_i \mathbf{H}_i^T)$ , suppose  $\hat{\mathbf{P}}_i$  has a low-rank decomposition  $\hat{\mathbf{Z}}_i \hat{\mathbf{Z}}_i^T$ , where  $\hat{\mathbf{Z}}_i \in \mathbb{R}^{p \times u}$  and  $u \ll q_i$ . Let  $\hat{\mathbf{U}}_i \hat{\mathbf{\Sigma}}_i \hat{\mathbf{V}}_i^T$  be the SVD decomposition of

$\mathbf{R}_i^{-\frac{1}{2}} \mathbf{H}_i \hat{\mathbf{Z}}_i$ , where  $\hat{\mathbf{\Sigma}}_i = \text{diag}(\hat{\sigma}_{i1}, \dots, \hat{\sigma}_{iu})$ . Then

$$\det(\mathbf{H}_i \hat{\lambda}_i \hat{\mathbf{P}}_i \mathbf{H}_i^T) \quad (\text{A1})$$

$$\begin{aligned} &= \det(\mathbf{R}) \det \left\{ \mathbf{I}_q + \hat{\lambda}_i \left( \mathbf{R}_i^{-\frac{1}{2}} \mathbf{H}_i \hat{\mathbf{Z}}_i \right) \left( \mathbf{R}_i^{-\frac{1}{2}} \mathbf{H}_i \hat{\mathbf{Z}}_i \right)^T \right\} \\ &= \det(\mathbf{R}_i) \det \left\{ \hat{\mathbf{U}}_i \left( \mathbf{I}_q + \hat{\lambda}_i \hat{\mathbf{\Sigma}}_i \hat{\mathbf{\Sigma}}_i^T \right) \hat{\mathbf{U}}_i^T \right\} \\ &= \det(\mathbf{R}_i) \times \prod_{j=1}^u (1 + \hat{\lambda}_i \hat{\sigma}_{ij}^2), \end{aligned} \quad (\text{A2})$$

and

$$\log \det(\mathbf{H}_i \hat{\lambda}_i \hat{\mathbf{P}}_i \mathbf{H}_i^T) = \sum_{j=1}^u \log(1 + \hat{\lambda}_i \hat{\sigma}_{ij}^2) + \log(\det(\mathbf{R}_i)). \quad (\text{A3})$$

This implies that the computation of the determination is reduced to the SVD decomposition of a  $q_i \times u$  matrix.

For the calculation of  $(\mathbf{H}_i \hat{\lambda}_i \hat{\mathbf{P}}_i \mathbf{H}_i^T + \mathbf{R}_i)^{-1}$ , it follows from the Sherman–Morrison–Woodbury equation that

$$\begin{aligned} &(\mathbf{H}_i \hat{\lambda}_i \hat{\mathbf{P}}_i \mathbf{H}_i^T + \mathbf{R}_i)^{-1} \\ &= \mathbf{R}_i^{-1} - \hat{\lambda}_i \mathbf{R}_i^{-1} \mathbf{H}_i \hat{\mathbf{Z}}_i \left\{ \mathbf{I}_u + \hat{\lambda}_i (\mathbf{H}_i \hat{\mathbf{Z}}_i)^T \mathbf{R}_i^{-1} (\mathbf{H}_i \hat{\mathbf{Z}}_i) \right\}^{-1} \\ &\quad (\mathbf{H}_i \hat{\mathbf{Z}}_i)^T \mathbf{R}_i^{-1}. \end{aligned} \quad (\text{A4})$$

Therefore, only an  $u \times u$  inverse matrix needs to be computed rather than a  $q_i \times q_i$  inverse matrix. For the standard EnKF with inflation, one can see that  $\hat{\mathbf{Z}}_i = (\mathbf{x}_{i,1}^f - \bar{\mathbf{x}}_i^f, \dots, \mathbf{x}_{i,n}^f - \bar{\mathbf{x}}_i^f) / \sqrt{n-1}$  and  $u = n$ . For the proposed HD-EnKF with tapering estimators, low-rank decomposition can be achieved through truncation of the eigenvalues introduced in Appendix S3 of the SI. When the dimensions of the state variable ( $p$ ) and the observation ( $q_i$ ) are extremely large, computationally economical inflation methods (e.g., Wang & Bishop, 2003) and advanced computational techniques like random SVD (Farchi & Bocquet, 2019; Halko *et al.*, 2011) would need to be used to make the computations practical.