Journal of the Meteorological Society of Japan

EARLY ONLINE RELEASE

The DOI for this manuscript is
DOI:10.2151/jmsj.2024-027
J-STAGE Advance published date: May 31st, 2024

A Hybrid Ensemble Kalman Filter to Mitigate Non-Gaussianity in Nonlinear Data Assimilation

Tadashi TSUYUKI[1]

Meteorological Research Institute
Japan Meteorological Agency, Tsukuba, Japan

Submitted: August 9, 2023
Revised: January 30, 2024
Re-revised: March 27, 2024

----

1) Corresponding author: Tadashi Tsuyuki, Observation and Data Assimilation Research Department, Meteorological Research Institute, 1-1 Nagamine, Tsukuba, 305-0052, JAPAN
Email: ttuyuki@mri-jma.go.jp
Tel: +81-29-853-8642

Abstract

Research on particle filters has been progressing with the aim of applying them to high-dimensional systems, but alleviation of problems with ensemble Kalman filters (EnKFs) in nonlinear or non-Gaussian data assimilation is also an important issue. It is known that the deterministic EnKF is less robust than the stochastic EnKF in strongly nonlinear regimes. We prove that if the observation operator is linear the analysis ensemble perturbations of the local ensemble transform Kalman filter (LETKF) are uniform contractions of the forecast ensemble perturbations in observation space in each direction of the eigenvectors of a forecast error covariance matrix. This property approximately holds for a weakly nonlinear observation operator. These results imply that if the forecast ensemble is strongly non-Gaussian the analysis ensemble of the LETKF is also strongly non-Gaussian, and that strong non-Gaussianity therefore tends to persist in high-frequency assimilation cycles, leading to the degradation of analysis accuracy in nonlinear data assimilation. A hybrid EnKF that combines the LETKF and the stochastic EnKF is proposed to mitigate non-Gaussianity in nonlinear data assimilation with small additional computational cost. The performance of the hybrid EnKF is investigated through data assimilation experiments using a 40-variable Lorenz-96 model. Results indicate that the hybrid EnKF significantly improves analysis accuracy in high-frequency data assimilation with a nonlinear observation operator. The positive impact of the hybrid

EnKF increases with the increase of the ensemble size.

Keywords hybrid ensemble Kalman filter; local ensemble transform Kalman filter; non-Gaussianity; nonlinear data assimilation; stochastic ensemble Kalman filter

1. Introduction

Data assimilation in high-dimensional nonlinear or non-Gaussian systems has been a challenge in meteorology and other geosciences (Bocquet et al. 2010). Although ensemble Kalman filters (EnKFs, Evensen 1994) have been widely used in data assimilation for numerical weather prediction and meteorological research, they are based on the Gaussian assumption, in which only the first- and second-order moments of a probability density function (PDF) are utilized, and may not work well in strongly non-Gaussian regimes. Research on particle filters (PFs, Gordon et al. 1993; Kitagawa 1996) that do not need the Gaussian assumption has been progressing with the aim of applying them to high-dimensional systems. Although it had been considered that the problem of weight degeneracy prevents the use of PFs for high-dimensional data assimilation (Snyder et al. 2008; van Leeuwen 2009), this limitation is currently disappearing owing to the recent efforts of a lot of investigators (van Leeuwen et al. 2019). Currently the localized PF (LPF) is attracting much attention (Penney and Miyoshi, 2016; Poterjoy, 2016; Poterjoy and Anderson 2016, Poterjoy et al. 2017; Farchi and Bocquet, 2018; Potthast et al., 2019; Kotsuki et al. 2022; Rojahn et al. 2023), and Kotsuki et al. (2022) presented a result that a Gaussian mixture extension of LPF (LPFGM) outperforms the local ensemble transform Kalman filter (LETKF, Hunt et al. 2007) in the accuracy of global analysis with an ensemble size of 40 and a realistic spatial distribution of radiosonde observations. Since a much larger ensemble would be needed to utilize some information on moments of a PDF higher than the second order (e. g., Nakano et al. 2007), the reason for the higher accuracy of LPFGM is possibly not because of the use of information on higher-order moments, but because of a problem with the LETKF.

Given the widespread use of EnKFs in meteorology, it is an important issue to alleviate problems with EnKFs in nonlinear or non-Gaussian data assimilation, especially because cumulus convection is strongly nonlinear. There are two methods for implementing ensemble Kalman filtering: the deterministic EnKF and the stochastic EnKF. The former EnKF generates an analysis ensemble by a linear transformation of a forecast ensemble (Anderson 2001; Bishop et al. 2001; Whitaker and Hamill 2002; Hunt et al. 2007), whereas the latter EnKF generates an analysis ensemble by assimilating perturbed observations (Burgers et al. 1998; Houtekamer and Mitchell 1998). In practice the deterministic EnKF is preferred over the stochastic EnKF, because the latter EnKF is less accurate due to sampling noise introduced by perturbed observations unless the ensemble size is sufficiently large (to name a few, Whitaker and Hamill 2002; Sakov and Oke 2008; Bowler et al. 2013). The LETKF belongs to the deterministic EnKF, and it is superior to the other EnKFs in computational efficiency because the analysis at each grid point can be independently computed in parallel. However, it is known that the deterministic EnKF is less robust to nonlinearity than the stochastic EnKF. Lawson and Hansen (2004) showed from geometric interpretation and ensemble diagnostics that the stochastic EnKF could better withstand regimes with nonlinear error growth. Lei et al. (2010) also derived a similar conclusion based on the stability analysis of the two EnKF methods against the small violation of Gaussian assumption. Anderson (2010) and Amezcua et al. (2012) showed the clustering of ensemble members but one member in nonlinear data assimilation with low-dimensional models. Although such a clustering is not observed in a more complex system, several studies showed that if the ensemble size is sufficiently large the stochastic EnKF is more accurate than the deterministic EnKF in nonlinear data assimilation (e. g., Lei and Bickel 2011; Tödler and Ahrens 2015; Tsuyuki and Tamura 2022).

The purpose of this study is twofold: to clarify the reason for less robustness of the LETKF to nonlinearity and to propose a hybrid EnKF that combines the LETKF and the stochastic EnKF for nonlinear data assimilation with small additional computational cost. We revisit the LETKF and the stochastic EnKF based on a decomposition of the ensemble transform matrix of the LETKF. We prove that if the observation operator is linear the analysis ensemble perturbations of the LETKF are uniform contractions of the forecast ensemble perturbations in observation space. This result implies that strong non-Gaussianity tends to persist in high-frequency assimilation cycles, leading to the degradation of analysis accuracy in nonlinear data assimilation. To mitigate non-Gaussianity, we introduce the hybrid EnKF in which a weighting average of the analysis ensembles of the two EnKFs is used as the analysis ensemble, if necessary, with adjustment of analysis spread. We could expect that the hybrid EnKF is more robust to non-Gaussianity than the LETKF with less sampling noise than the stochastic EnKF. To investigate the performance of the hybrid EnKF, we conduct data assimilation experiments using a 40-variable Lorenz-96 model (Lorenz, 1996). The results of experiments demonstrate a significantly better analysis accuracy of the hybrid EnKF in high-frequency assimilation cycles with a nonlinear observation operator.

The remainder of this paper is organized as follows. Section 2 is the revisit of the LETKF and the stochastic EnKF with a unifying framework, and Section 3 compares the performance of the two EnKFs with a nonlinear observation operator in a one-dimensional system. The hybrid EnKF is introduced in Section 4, and the design of data assimilation experiments is described in Section 5. The results of the experiments are presented in Section 6, and a summery and discussion are mentioned in Section 7.

2. Revisit of LETKF and stochastic EnKF

The derivation of ensemble Kalman filtering is usually based on the extended Kalman filter that adopts the tangent-linear approximation of an observation operator $H(\cdot)$. However, it is customarily to use a nonlinear observation operator as it is in EnKFs (e.g., Houtekamer and Mitchell, 2001; Hunt et al., 2007). Therefore, we begin the revisit of the LETKF and the stochastic EnKF with an extension of the analysis equation of Kalman filtering for a nonlinear observation operator:

$$\boldsymbol{x}^{\boldsymbol{a}} = \boldsymbol{x}^{\boldsymbol{f}} + \boldsymbol{K}\left(\boldsymbol{y}^{\boldsymbol{o}} - \boldsymbol{y}^{\boldsymbol{f}}\right), \tag{1}$$

where $\boldsymbol{x}^a$ and $\boldsymbol{x}^f$ are the analysis and forecast of the n-dimensional state variable $\boldsymbol{x}$, respectively, $\boldsymbol{y}^o$ is the observation of the m-dimensional variable $\boldsymbol{y}$, $\boldsymbol{y}^f \coloneqq H(\boldsymbol{x}^f)$ is the forecast of $\boldsymbol{y}$, and $\boldsymbol{K}$ is an $n \times m$ weight matrix. If there are no correlations between the forecast error of state variable $\Delta\boldsymbol{x}^f$ and the observation error $\Delta\boldsymbol{y}^o$ and between the forecast error of observed variable $\Delta\boldsymbol{y}^f$ and $\Delta\boldsymbol{y}^o$, the optimal value of $\boldsymbol{K}$ is given using the minimum mean square error criterion by

$$\boldsymbol{K} = \left\langle \Delta\boldsymbol{x}^f (\Delta\boldsymbol{y}^f)^{\mathrm{T}} \right\rangle \left(\boldsymbol{R} + \left\langle \Delta\boldsymbol{y}^f (\Delta\boldsymbol{y}^f)^{\mathrm{T}} \right\rangle\right)^{-1}, \tag{2}$$

where a pair of brackets denotes the expectation operator, the superscript T indicates the transpose of a vector or a matrix, and $\boldsymbol{R} \coloneqq \langle \Delta\boldsymbol{y}^o (\Delta\boldsymbol{y}^o)^{\mathrm{T}} \rangle$ is the observation error covariance matrix. As Eq. (1) with the minimum mean square error criterion is not based on the Bayes' theorem, it is suboptimal for nonlinear or non-Gaussian regimes.

2.1. LETKF

Let N be the ensemble size of ensemble Kalman filtering, and let us introduce an $n \times N$ matrix of forecast ensemble perturbations of the state variable with respect to the mean, $\boldsymbol{X}^f$, and an $m \times N$ matrix of forecast ensemble perturbations of the observed variable, $\boldsymbol{Y}^f$:

$$\boldsymbol{X}^f \coloneqq \left(\Delta\boldsymbol{x}^{f(1)}, \ \cdots, \Delta\boldsymbol{x}^{f(N)}\right), \qquad \boldsymbol{Y}^f \coloneqq \left(\Delta\boldsymbol{y}^{f(1)}, \ \cdots, \Delta\boldsymbol{y}^{f(N)}\right), \tag{3}$$

where $\left\{\Delta\boldsymbol{x}^{f(i)}\right\}_{i=1}^{i=N}$ and $\left\{\Delta\boldsymbol{y}^{f(i)}\right\}_{i=1}^{i=N}$ are the ensemble members of forecast perturbations. We can approximate Eq. (2) by

$$\boldsymbol{K} = \frac{\boldsymbol{X}^f (\boldsymbol{Y}^f)^{\mathrm{T}}}{N-1} \left[\boldsymbol{R} + \frac{\boldsymbol{Y}^f (\boldsymbol{Y}^f)^{\mathrm{T}}}{N-1}\right]^{-1}. \tag{4}$$

This matrix can be put in the following form using a variant of the Sherman–Morrison–

Woodbury formula (Golub and Van Loan, 2013):

$$\boldsymbol{K} = \frac{\boldsymbol{X}^f}{N-1}\left[\boldsymbol{I}_N + \frac{(\boldsymbol{Y}^f)^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{Y}^f}{N-1}\right]^{-1}(\boldsymbol{Y}^f)^{\mathrm{T}}\boldsymbol{R}^{-1}, \tag{5}$$

where $\boldsymbol{I}_N$ is the N-dimensional identity matrix. This representation of Kalman gain for a nonlinear observation operator was derived by Hunt et al. (2007), who adopted the minimization of a cost function with a linear approximation to obtain Eq. (5). Equation (1) with the minimum mean square error criterion does not need such an approximation, and it is straightforward as compared to the joint state-observation space method (Anderson, 2001). The mean of the analysis ensemble is given by Eq. (1) with $\boldsymbol{x}^f$ and $\boldsymbol{y}^f$ replaced with the corresponding ensemble means.

In the deterministic EnKF, an $n \times N$ matrix of analysis ensemble perturbations of the state variables, $\boldsymbol{X}^a$, is computed through ensemble transformation of $\boldsymbol{X}^f$. The LETKF adopts the following transformation using right multiplication:

$$\boldsymbol{X}^a = \boldsymbol{X}^f\boldsymbol{T}, \tag{6}$$

where $\boldsymbol{T}$ is called the ensemble transform matrix and given by

$$\boldsymbol{T} \coloneqq \left[\boldsymbol{I}_N + \frac{(\boldsymbol{Y}^f)^{\mathrm{T}}\boldsymbol{R}^{-1}\boldsymbol{Y}^f}{N-1}\right]^{-1/2}, \tag{7}$$

where $[\cdot]^{-1/2}$ denotes the inverse of the symmetric positive-definite square root of a positive-definite matrix (Golub and Van Loan, 2013). The matrix $\boldsymbol{T}$ has $\mathbf{1}_N$ as an eigenvector, where $\mathbf{1}_N$ is the N-dimensional vector of which components are all 1s, such that the sum of analysis ensemble perturbations of each state variable vanishes. Sakov and Oke (2008) mentioned that a general form of the ensemble transform matrix is given by multiplying $\boldsymbol{T}$ by a mean-preserving rotation matrix, which also has $\mathbf{1}_N$ as an eigenvector, from right. More generally, space inversion can also be applied to $\boldsymbol{T}$. We will return to this issue in Subsection 2.3.

2.2. Stochastic EnKF

The analysis ensemble of the stochastic EnKF is constructed in the following way:

$$\boldsymbol{x}^{a(i)} = \boldsymbol{x}^{f(i)} + \boldsymbol{K}\left(\boldsymbol{y}^o + \boldsymbol{\varepsilon}^{o(i)} - \boldsymbol{y}^{f(i)}\right), \qquad (i = 1, \cdots, N), \tag{8}$$

where $\left\{\boldsymbol{x}^{f(i)}\right\}_{i=1}^{i=N}$ and $\left\{\boldsymbol{y}^{f(i)}\right\}_{i=1}^{i=N}$ are the forecast ensembles of state and observed variables, respectively. The perturbations to observations $\left\{\boldsymbol{\varepsilon}^{o(i)}\right\}_{i=1}^{i=N}$ are given by

$$\boldsymbol{\varepsilon}^{o(i)} \coloneqq \boldsymbol{\varepsilon}^{o(i)*} - \frac{1}{N}\sum_{j=1}^{N} \boldsymbol{\varepsilon}^{o(j)*}, \qquad \boldsymbol{\varepsilon}^{o(i)*} \sim N(\mathbf{0}, \boldsymbol{R}), \tag{9}$$

where $N(\mathbf{0}, \boldsymbol{R})$ denotes the Gaussian distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{R}$. We can elaborate Eq. (9) by removing the correlation between $\left\{\boldsymbol{\varepsilon}^{o(i)}\right\}_{i=1}^{i=N}$ and $\left\{\boldsymbol{y}^{f(i)}\right\}_{i=1}^{i=N}$ for each observed variable and adjusting the resulting perturbations such that its variance is equal to the original value. This procedure is adopted in the data assimilation experiments in this study. Note that the numbering of ensemble members used in Eq. (8) is the same as in Eq. (3), and that the analysis ensemble mean of Eq. (8) is equal to that of the LETKF.

The analysis ensemble perturbations of the stochastic EnKF can be written as

$$\boldsymbol{X}^a = \boldsymbol{X}^f - \boldsymbol{K}\boldsymbol{Y}^f + \boldsymbol{K}\boldsymbol{E}^o, \tag{10}$$

where $\boldsymbol{E}^o$ represents the ensemble members of observation error perturbations and defined by

$$\boldsymbol{E}^o \coloneqq \left(\boldsymbol{\varepsilon}^{\boldsymbol{o}(1)}, \cdots, \boldsymbol{\varepsilon}^{\boldsymbol{o}(N)}\right). \tag{11}$$

Substitution of Eq. (5) into Eq. (10) yields

$$\boldsymbol{X}^a = \boldsymbol{X}^f \left[\boldsymbol{I}_N + \frac{(\boldsymbol{Y}^f)^{\mathrm{T}} \boldsymbol{R}^{-1} \boldsymbol{Y}^f}{N-1} \right]^{-1} + \boldsymbol{K}\boldsymbol{E}^o. \tag{12}$$

The first and second terms on the righthand side are hereafter referred to as the deterministic part and the stochastic part of stochastic EnKF, respectively. Comparison with Eqs. (6)–(7) reveals that the deterministic part is obtained by transforming $\boldsymbol{X}^f$ with the matrix $\boldsymbol{T}^2$. The addition of the stochastic part makes the expectation value of the analysis error covariance matrix equal to that of the LETKF. If the deterministic part is not Gaussian, this part makes the analysis ensemble more Gaussian. This property of the stochastic EnKF may be desirable for a better performance of EnKFs, which are based on the Gaussian assumption. However, the stochastic part introduces sampling noise to the stochastic EnKF.

2.3. Decomposition of matrix $\boldsymbol{T}$

The above discussion indicates that the ensemble transform matrix $\boldsymbol{T}$ plays a crucial role not only in the LETKF but also in the stochastic EnKF. In the following, we decompose this matrix using a complete orthonormal system in ensemble space to clarify a problem of the LETKF in nonlinear data assimilation. This decomposition is based on the property that for any real matrix A the set of positive eigenvalues of $A^{\mathrm{T}}A$ is the same as that of AA^{T}.

Let us apply the eigenvalue decomposition to a dimensionless forecast error covariance matrix in observation space:

$$\boldsymbol{P}_Y^f := \frac{\widehat{\boldsymbol{Y}}^f \left(\widehat{\boldsymbol{Y}}^f\right)^{\mathrm{T}}}{N-1} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{T}}, \tag{13}$$

where $\widehat{\boldsymbol{Y}}^f \coloneqq \boldsymbol{R}^{-1/2}\boldsymbol{Y}^f$, $\boldsymbol{U}$ is an orthogonal matrix consisting of eigenvectors, and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues given by

$$\boldsymbol{\Lambda} = \begin{cases} \mathrm{diag}\,[\lambda_1,\ \cdots,\lambda_{N-1},\ 0,\ \cdots,\ 0], & (N \le m) \\ \mathrm{diag}\,[\lambda_1,\ \cdots,\lambda_m], & (N \ge m+1) \end{cases}, \tag{14}$$

where $\{\lambda_i\}_{i=1}^{i=r}$ $(r \coloneqq \min\,(N-1, m\,))$ are assumed to be positive. These eigenvalues represent the ratio of the variance of forecast error to that of observation error. We transform $\widehat{\boldsymbol{Y}}^f$ to $\boldsymbol{Z}^f$ such that $\boldsymbol{Z}^f(\boldsymbol{Z}^f)^{\mathrm{T}}$ is a diagonal matrix:

$$\boldsymbol{Z}^f \coloneqq \boldsymbol{U}^{\mathrm{T}}\boldsymbol{R}^{-1/2}\boldsymbol{Y}^f = \begin{pmatrix} \left(\Delta\boldsymbol{z}_1^f\right)^{\mathrm{T}} \\ \vdots \\ \left(\Delta\boldsymbol{z}_m^f\right)^{\mathrm{T}} \end{pmatrix}, \tag{15}$$

where $\left\{\Delta\boldsymbol{z}_i^f\right\}_{i=1}^{i=m}$ are the column vectors of the forecast ensemble of each transformed variable. Equations (13)–(15) imply

$$\left(\Delta\boldsymbol{z}_i^f\right)^{\mathrm{T}}\Delta\boldsymbol{z}_j^f = (N-1)\lambda_i\delta_{ij}, \qquad (i,j = 1,\ \cdots,\ r), \tag{16}$$

$$\Delta\boldsymbol{z}_i^f = \boldsymbol{0}, \qquad (i = r+1,\ \cdots,\ m), \tag{17}$$

where δ_{ij} is the Kronecker delta, and we obtain the following orthonormal system in ensemble space:

$$\{\boldsymbol{v}_1,\ \cdots,\boldsymbol{v}_r\} \coloneqq \left\{ \frac{\Delta\boldsymbol{z}_1^f}{\sqrt{(N-1)\lambda_1}},\ \cdots,\frac{\Delta\boldsymbol{z}_r^f}{\sqrt{(N-1)\lambda_r}} \right\}. \tag{18}$$

A complete orthonormal system $\{\boldsymbol{v}_i\}_{i=1}^{i=N}$ can be constructed from this orthonormal system using the Gram–Schmidt orthogonalization. Then $\boldsymbol{T}$ can be decomposed by using $\{\boldsymbol{v}_i\}_{i=1}^{i=N}$ as

$$\boldsymbol{T} = \left[\boldsymbol{I}_N + \frac{(\boldsymbol{Z}^f)^{\mathrm{T}}\boldsymbol{Z}^f}{N-1}\right]^{-\frac{1}{2}} = \left[\sum_{i=1}^{r}(1+\lambda_i)\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}} + \sum_{i=r+1}^{N}\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\right]^{-\frac{1}{2}}$$

$$= \sum_{i=1}^{r} \frac{1}{\sqrt{1+\lambda_i}} \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}} + \sum_{i=r+1}^{N} \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}} = \boldsymbol{I}_N - \sum_{i=1}^{r} \left(1 - \frac{1}{\sqrt{1+\lambda_i}}\right) \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}}, \tag{19}$$

where the completeness condition $\sum_{i=1}^{N} \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}} = \boldsymbol{I}_N$ is used. It is obvious from this decomposition that $\boldsymbol{T}$ has $\mathbf{1}_N$ as an eigenvector, because $\left(\Delta \boldsymbol{z}_i^f\right)^{\mathrm{T}} \mathbf{1}_N = 0$ for $i = 1, \cdots, r$. Equation (19) implies that if the ensemble size N is larger than the number of observational data m, we can construct the ensemble transform matrix $\boldsymbol{T}$ by solving a smaller eigenvalue problem of $\boldsymbol{P}_Y^f$ defined by Eq. (13). As mentioned in Subsection 2.2, the ensemble transform matrix of the deterministic part of stochastic EnKF is given by $\boldsymbol{T}^{\mathbf{2}}$, the decomposition of which can be obtained by replacing $\sqrt{1+\lambda_i}$ with $1+\lambda_i$ in Eqs. (19).

To derive the analysis ensemble perturbations of each state variable, let us write the transposes of $\boldsymbol{X}^{\boldsymbol{a}}$ and $\boldsymbol{X}^f$ as

$$(\boldsymbol{X}^f)^{\mathrm{T}} = \left(\Delta \boldsymbol{x}_1^f, \ \cdots, \Delta \boldsymbol{x}_n^f\right), \qquad (\boldsymbol{X}^a)^{\mathrm{T}} = (\Delta \boldsymbol{x}_1^a, \ \cdots, \Delta \boldsymbol{x}_n^a), \tag{20}$$

where the subscript indicates the index of state variables. $\left\{\Delta \boldsymbol{x}_j^f\right\}_{j=1}^{j=n}$ and $\left\{\Delta \boldsymbol{x}_j^a\right\}_{j=1}^{j=n}$ are the forecast and analysis ensemble perturbations, respectively, of each state variable. Substitution of Eqs. (18) – (20) into the transpose of Eq. (6) yields

$$\Delta \boldsymbol{x}_j^a = \Delta \boldsymbol{x}_j^f - \sum_{i=1}^{r} \frac{1}{\lambda_i} \left(1 - \frac{1}{\sqrt{1+\lambda_i}}\right) \frac{\left(\Delta \boldsymbol{z}_i^f\right)^{\mathrm{T}} \Delta \boldsymbol{x}_j^f}{N-1} \Delta \boldsymbol{z}_i^f, \quad (j = 1, \ \cdots, \ n). \tag{21}$$

In this equation, $\{\lambda_i\}_{i=1}^{i=r}$ are the non-zero eigenvalues of the covariance matrix $\boldsymbol{P}_Y^f$, and $\left(\Delta \boldsymbol{z}_i^f\right)^{\mathrm{T}} \Delta \boldsymbol{x}_j^f / (N-1)$ is the covariance between $\Delta \boldsymbol{z}_i^f$ and $\Delta \boldsymbol{x}_j^f$. Therefore, if the ensemble size N goes to infinity, the factor of $\Delta \boldsymbol{z}_i^f$ in Eq. (21) becomes constant. It is well known that a linear combination of Gaussian random variables is Gaussian distributed. It follows that if the forecast ensemble is Gaussian and the observation operator is linear then the analysis ensemble generated by the ensemble transform matrix $\boldsymbol{T}$ is also Gaussian. Since this property holds under any mean-preserving rotation and space inversion, it is difficult to uniquely determine an appropriate ensemble transform matrix for data assimilation in linear Gaussian systems. Hunt et al. (2007) adopted the matrix $\boldsymbol{T}$ defined by Eq. (7) as the ensemble transform matrix of the LETKF on the basis that it makes the analysis ensemble perturbations as close as possible to the forecast ensemble perturbations (Wang et al. 2004; Ott et al. 2004; see Appendix). We can make the same choice by requiring that if there is no observation, in other words, if observation error variance goes to infinity, the analysis ensemble is the same as the forecast ensemble. Note that the stochastic EnKF satisfies this requirement. Equation (21) also implies that if the observation operator is nonlinear the analysis ensemble of state variables becomes non-Gaussian even if the forecast ensemble is Gaussian.

If the observation operator is linear, we can derive simple formulas for the relationship between the analysis ensemble perturbations and the forecast ensemble perturbations in observation space. Let us introduce the following transformed analysis ensemble perturbations:

$$\boldsymbol{Z}^a \coloneqq \boldsymbol{U}^T \boldsymbol{R}^{-1/2} \boldsymbol{Y}^a = \begin{pmatrix} (\Delta \boldsymbol{z}_1^a)^{\mathrm{T}} \\ \vdots \\ (\Delta \boldsymbol{z}_m^a)^{\mathrm{T}} \end{pmatrix}, \tag{22}$$

where $\boldsymbol{Y}^a$ is the analysis ensemble perturbations in observation space. Then we obtain

$$\boldsymbol{Z}^f = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{R}^{-1/2}\boldsymbol{H}\boldsymbol{X}^f, \qquad \boldsymbol{Z}^a = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{R}^{-1/2}\boldsymbol{H}\boldsymbol{X}^a, \tag{23}$$

where $\boldsymbol{H}$ is a linear observation operator.

For the LETKF, Eq. (6) can be put in the following form

$$\boldsymbol{Z}^a = \boldsymbol{Z}^f\boldsymbol{T}. \tag{24}$$

Substitution of Eqs. (15), (17), (19), and (22) into the transpose of Eq. (24) yields

$$(\Delta\boldsymbol{z}_1^a,\ \cdots,\Delta\boldsymbol{z}_m^a) = \left(\sum_{i=1}^{r}\frac{1}{\sqrt{1+\lambda_i}}\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}} + \sum_{i=r+1}^{N}\boldsymbol{v}_i\boldsymbol{v}_i^{\mathrm{T}}\right)\left(\Delta\boldsymbol{z}_1^f,\ \cdots,\Delta\boldsymbol{z}_r^f,\ \boldsymbol{0},\ \cdots,\ \boldsymbol{0}\right). \tag{25}$$

By using Eq. (18) and the orthogonality of $\{\boldsymbol{v}_i\}_{i=1}^{i=N}$, we finally obtain

$$\Delta\boldsymbol{z}_i^a = \begin{cases} \dfrac{1}{\sqrt{1+\lambda_i}}\Delta\boldsymbol{z}_i^f & (i = 1,\ \cdots,\ r) \\ \boldsymbol{0} & (i = r+1,\ \cdots,\ m) \end{cases}. \tag{26}$$

This result indicates that the analysis ensemble perturbations are uniform contractions of the forecast ensemble perturbations in observation space in each direction of the eigenvectors of $\boldsymbol{P}_Y^f$.

If the observation operator is weakly nonlinear, the following approximate equations hold:

$$\boldsymbol{Z}^f \approx \boldsymbol{U}^{\mathrm{T}}\boldsymbol{R}^{-1/2}\left.\frac{\partial H}{\partial \boldsymbol{x}}\right|_{x^f}\boldsymbol{X}^f, \qquad \boldsymbol{Z}^a \approx \boldsymbol{U}^{\mathrm{T}}\boldsymbol{R}^{-1/2}\left.\frac{\partial H}{\partial \boldsymbol{x}}\right|_{x^a}\boldsymbol{X}^a \tag{27}$$

Using Eq. (6), we obtain

$$\boldsymbol{Z}^a \approx \boldsymbol{Z}^f\boldsymbol{T} + \boldsymbol{U}^{\mathrm{T}}\boldsymbol{R}^{-1/2}\left(\left.\frac{\partial H}{\partial \boldsymbol{x}}\right|_{x^a} - \left.\frac{\partial H}{\partial \boldsymbol{x}}\right|_{x^f}\right)\boldsymbol{X}^f\boldsymbol{T}, \tag{28}$$

By the assumption of weak nonlinearity, the second term on the righthand side of this equation is sufficiently small compared to the first term. Then Eq. (24) approximately holds, and therefore Eq. (26) also approximately holds.

For the stochastic EnKF, we introduce a transformed matrix of observation error perturbations:

$$\boldsymbol{F}^o \coloneqq \boldsymbol{U}^{\mathrm{T}}\boldsymbol{R}^{-1/2}\boldsymbol{E}^o = \begin{pmatrix} (\boldsymbol{f}_1^o)^{\mathrm{T}} \\ \vdots \\ (\boldsymbol{f}_m^o)^{\mathrm{T}} \end{pmatrix}, \tag{29}$$

where

$$\langle (\boldsymbol{f}_j^o)^{\mathrm{T}} \boldsymbol{f}_i^o \rangle \coloneqq (N-1)\,\delta_{ij}. \tag{30}$$

If the observation operator is linear, we can derive the following equation using Eqs. (5), (7), (12), and (29) in addition to Eqs. (15), (17) – (19), (22), and (23):

$$\Delta \boldsymbol{z}_i^a = \begin{cases} \dfrac{1}{1+\lambda_i}\Delta \boldsymbol{z}_i^f + \dfrac{\lambda_i}{1+\lambda_i}\boldsymbol{f}_i^o & (i = 1, \cdots, r) \\ \boldsymbol{0} & (i = r+1, \cdots, m) \end{cases}. \tag{31}$$

This result indicates that the deterministic part of stochastic EnKF is twice contracted as compared to the LETKF. This twice contraction of forecast ensemble perturbations allows the addition of Gaussian perturbations to make the analysis ensemble perturbations more Gaussian. It is also found from Eqs. (16), (30), and (31) that as λ_i increases the stochastic part becomes more dominant. If the observation operator is weakly nonlinear, Eq. (31) approximately holds like Eq. (26) for the LETKF.

2.4. Example of LETKF analysis

Figure 1 presents an example of Eq. (26) for a system of two state variables, $\boldsymbol{x} = (x_1, x_2)^{\mathrm{T}}$, with an ensemble size of 10. The prior PDF is bimodal and given by Fig. 1

$$p(x_1, x_2) = \frac{1}{4\pi}\left\{\exp\left[-\frac{(x_1+2)^2}{2}\right] + \exp\left[-\frac{(x_1-2)^2}{2}\right]\right\}\exp\left[-\frac{{x_2}^2}{2}\right]. \tag{32}$$

The two modes are located at $(\pm 2, 0)^{\mathrm{T}}$, and the mean is $(0, 0)^{\mathrm{T}}$. This PDF is plotted with green contours in Fig. 1a. The forecast ensemble members are generated by independent random draws from the above PDF. Those members are plotted with green dots in the same panel and numbered from 0 to 9. These numbers are referred to in Fig. 1c. The forecast ensemble mean is $(-0.144, 0.121)^{\mathrm{T}}$. The state variables are assumed to be directly observed with a standard deviation of observation error of 0.5. The observations are $\boldsymbol{y}^o = (1,1)^{\mathrm{T}}$ and the likelihood function (red contours) is given by

$$p(1,1|x_1,x_2) = \frac{1}{2\pi(0.5)^2}\exp\left[-\frac{(x_1-1)^2}{2(0.5)^2}-\frac{(x_2-1)^2}{2(0.5)^2}\right]. \quad (33)$$

The posterior PDF calculated from Bayes' theorem is plotted with blue contours in Fig. 1b. This PDF is unimodal with mean $(1.193, 0.800)^{\mathrm{T}}$. The analysis ensemble members obtained by using Eqs. (1) and (5)–(7) are plotted with blue dots.

It is found that the distribution pattern of analysis ensemble members around the ensemble mean is very similar to that of forecast ensemble members with a significant reduction in spread. The analysis ensemble mean is $(0.928, 0.754)^{\mathrm{T}}$. It is shifted roughly by 0.3 in the direction of the forecast ensemble mean from the mean of the posterior PDF. The tilted coordinate axes plotted in Figs. 1a and 1b represent the directions of eigenvectors of $\boldsymbol{P}_Y^f$ with the origins set at each ensemble mean. Figure 1c plots the ratios between the analysis and forecast perturbations in each direction of the eigenvectors for each ensemble member. Those ratios are found to be constant in each direction, being consistent with Eq. (26). An example in which the observation operator is strongly nonlinear is presented in Section 3 for the LETKF and the stochastic EnKF.

2.5. Problem with LETKF

Lawson and Hansen (2004) presented histograms of analysis ensembles generated by the deterministic and stochastic EnKFs for a one-dimensional system with an ensemble size of 5 000. The state variable is assumed to be directly observed. Although they use the deterministic EnKF based on left multiplication, their analysis ensembles are the same as those generated by Eqs. (6)–(7). According to their Figs. 2 and 3, when the prior PDF is Gaussian, both EnKFs yield correct analysis ensembles. However, when the prior PDF is bimodal, this is not the case; the ensemble mean is inaccurate and the analysis spread tends to be overestimated. The reason for the latter result is probably because the minimum mean square error estimate is obtained under the specific assumption on analysis given by Eq. (1).

Their histograms for the deterministic EnKF are consistent with Eq. (26); the analysis ensemble perturbations are a uniform contraction of the forecast ensemble perturbations irrespective whether the prior PDF is Gaussian or bimodal. If the forecast ensemble at a certain analysis time is strongly non-Gaussian, the analysis ensemble at the same analysis time is also strongly non-Gaussian. In high-frequency assimilation cycles with a nonlinear numerical model, the error growth between the adjacent analysis times may be close to linear. Therefore, the forecast ensemble at the next analysis time will also be strongly non-Gaussian. By repeating these processes, strong non-Gaussianity tends to persist in high-frequency assimilation cycles. On the other hand, in low-frequency assimilation cycles, strong non-Gaussianity of the forecast ensemble is not likely to persist due to nonlinear error growth. Such persistent strong non-Gaussianity is unlikely to occur in the stochastic EnKF because of the addition of Gaussian perturbations.

In linear Gaussian or weakly nonlinear systems, when the ensemble size is small, the forecast ensemble could become strongly non-Gaussian due to sampling errors. However, the LETKF can yield an analysis with high accuracy using only the first- and second-order moments of the forecast ensemble with covariance inflation and localization. Therefore, the persistence of strong non-Gaussianity in high-frequency assimilation cycles may not cause a serious problem.

This is not the case in strongly nonlinear systems, in which information of moments of the forecast ensemble higher than the second-order is necessary for accurate analysis. EnKFs yield inaccurate analysis ensembles and tend to overestimate analysis spread even if the ensemble size is sufficiently large. In high-frequency assimilation cycles of the LETKF, those problems become worse because of the persistent strong non-Gaussianity, and its analysis may be less accurate than the stochastic EnKF. This would not occur in low-frequency assimilation cycles. The nonlinearity in data assimilation arises not only from the nonlinearity of a numerical model but also from the nonlinearity of an observation operator. When a region of sparse observations is present, the former nonlinearity becomes strong because the constraints by observations are weak in such a region.

3. EnKF analyses with a nonlinear observation operator

In this section, we examine the performance of the LETKF and the stochastic EnKF with a nonlinear observation operator $H(\boldsymbol{x}) = \max(\boldsymbol{x}, \boldsymbol{0})$, where the maximum operator shall be applied to each pair of components of the two argument vectors. The observations are generated by $\boldsymbol{y}^o = \max(\boldsymbol{x}^t + \boldsymbol{\varepsilon}^o, \boldsymbol{0})$, where $\boldsymbol{x}^t$ is the true value and $\boldsymbol{\varepsilon^o}$ is observation error, which is independent random draws from a Gaussian distribution with mean 0 and variance 1. Note that the observation error is added inside the maximum operator so that the observations are always non-negative like precipitation data. Those observational data therefore cannot be properly handled by EnKFs, because in the theory of Kalman filtering an observed value is assumed to be the sum of the true value and Gaussian random error.

The above observation operator is strongly nonlinear around x =0. Its likelihood function in a one-dimensional system is calculated as

$$p(y^o|x) = \frac{\theta(y^o)}{\sqrt{2\pi}} \exp\left[-\frac{(y^o - x)^2}{2}\right] + \frac{\delta(y^o)}{2}\left[1 - \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \tag{34}$$

where $\theta(\cdot)$ is the unit step function, $\delta(\cdot)$ is the delta function, and $\operatorname{erf}(\cdot)$ is the error function. The first term is the likelihood function for $y^o > 0$. The coefficient of the delta function can be determined from the condition that the integration of $p(y^o|x)$ from $-\infty$ to ∞ with respect to y^o is unity. The likelihood function for $y^o = 0$ is shown with the orange line in Fig. 2a. Since what matters in a likelihood function is a relative value, the coefficient of the delta function is plotted. This figure indicates that when $y^o = 0$ the state variable is likely to be negative.

We apply the LETKF and the stochastic EnKF with an ensemble size of 10 000 to the above one-dimensional system for the case of $y^o = 0$. The prior PDF is assumed to be Gaussian with mean 0 and variance 1, and the forecast ensemble is generated by independent random draws from the prior PDF. This PDF and the histogram of the forecast ensemble are shown in Fig. 2a. The analysis ensembles of the LETKF and the stochastic Fig. 2 EnKF are presented in Fig. 2b and Fig. 2c, respectively, along with the posterior PDF calculated from Bayes' theorem. Surprisingly, the posterior PDF is very close to Gaussian; the skewness and kurtosis of the posterior PDF is much smaller than unity (Table 1). The Table 1 analysis ensemble means of the two EnKFs are closer to the forecast ensemble mean than the mean of the posterior PDF. Their analysis spreads are slightly overestimated as compared to the posterior PDF.

Table 1 also reveals that the analysis ensembles are more skewed than the posterior PDF, and that the analysis ensemble of the stochastic EnKF is slightly more non-Gaussian than the LETKF. The latter result may be an unexpected one, since the stochastic EnKF is expected to yield a more Gaussian analysis ensemble through the addition of Gaussian perturbations. This result may be explained by considering the deterministic part of stochastic EnKF. If a forecast member is negative in state space, the corresponding analysis member remains the same as the forecast member, because the forecast member in observation space is equal to the observation $y^o = 0$. On the other hand, if a forecast member is positive, the corresponding analysis member is shifted toward the origin as in Kalman filtering of linear Gaussian systems. As a result, a discontinuity arises at x = 0 in the histogram of the analysis ensemble calculated from the deterministic part only. Although the addition of Gaussian perturbations wipes out this discontinuity, the analysis ensemble could become more non-Gaussian than the LETKF. If a nonlinear observation operator other than $H(\boldsymbol{x})$ = max($\boldsymbol{x}$, $\mathbf{0}$) is used, the analysis ensemble generated by the deterministic part of the stochastic EnKF may not be so strongly non-Gaussian, and the stochastic EnKF could yield a more Gaussian analysis ensemble than the LETKF.

4. Method of hybrid EnKF

The hybrid EnKF is based on a weighting average of the analysis ensembles of the LETKF and the stochastic EnKF. Note that the analysis ensemble mean of the stochastic EnKF is equal to that of the LETKF as mentioned in Subsection 2.2. Let $\boldsymbol{X}_L^a$ and $\boldsymbol{X}_S^a$ be the analysis ensemble perturbations of the LETKF and stochastic EnKF, respectively, in a local domain. The first step is the computation of a provisional value of the analysis ensemble perturbations as

$$\boldsymbol{X}_*^a = (1-w)\boldsymbol{X}_L^a + w\boldsymbol{X}_S^a \qquad (0 \leq w \leq 1), \tag{35}$$

where w is the weight of the stochastic EnKF. The analysis error covariance matrix of $\boldsymbol{X}^{a*}$ is calculated as

$$\boldsymbol{P}_*^a = (1-w)\frac{\boldsymbol{X}_L^a(\boldsymbol{X}_L^a)^{\mathrm{T}}}{N-1} + w\frac{\boldsymbol{X}_S^a(\boldsymbol{X}_S^a)^{\mathrm{T}}}{N-1} - \frac{w(1-w)}{N-1}(\boldsymbol{X}_L^a - \boldsymbol{X}_S^a)(\boldsymbol{X}_L^a - \boldsymbol{X}_S^a)^{\mathrm{T}} \tag{36}$$

Since $\langle \boldsymbol{X}_S^a (\boldsymbol{X}_S^a)^{\mathrm{T}} \rangle$ is equal to $\boldsymbol{X}_L^a (\boldsymbol{X}_L^a)^{\mathrm{T}}$, taking the expected value of Eq. (36) yields

$$\langle \boldsymbol{P}_*^a \rangle = \frac{\boldsymbol{X}_L^a (\boldsymbol{X}_L^a)^T}{N-1} - \frac{w(1-w)}{N-1} \langle (\boldsymbol{X}_L^a - \boldsymbol{X}_S^a)(\boldsymbol{X}_L^a - \boldsymbol{X}_S^a)^{\mathrm{T}} \rangle \tag{37}$$

This equation indicates that analysis spread is underestimated unless $w = 0$ or $w = 1$, because the matrix $\langle (\boldsymbol{X}_L^a - \boldsymbol{X}_S^a)(\boldsymbol{X}_L^a - \boldsymbol{X}_S^a)^{\mathrm{T}} \rangle$ is positive-semidefinite.

In weakly nonlinear regimes, it is desirable to adjust the spread of $\boldsymbol{X}_*^a$ to be equal to that of $\boldsymbol{X}_L^a$, because the LETKF may be more accurate than the stochastic EnKF. In strongly nonlinear regimes, however, the analysis spread of the LETKF in high-frequency assimilation cycles may be overestimated due to persistent strong non-Gaussianity as mentioned in Subsection 2.5. In this case, $\boldsymbol{X}_*^a$ can be used as the analysis ensemble perturbations of the hybrid EnKF, $\boldsymbol{X}^a$. Therefore, we introduce the following procedure to adjust the analysis spread. Consider the case where there is only one state variable at a grid point. The analysis ensemble perturbations at each grid point, $\Delta \boldsymbol{x}^a$, are computed by

$$\Delta \boldsymbol{x}^a = \left[(1-\alpha) + \alpha \frac{\sigma_L^a}{\sigma_*^a} \right] \Delta \boldsymbol{x}_*^a \qquad (0 \le \alpha \le 1), \tag{38}$$

where $\Delta \boldsymbol{x}_*^a$ are the corresponding ensemble perturbations of the provisional analysis, and σ_L^a and σ_*^a are the ensemble standard deviations of the LETKF and provisional analysis, respectively. If the parameter α is set to 0, analysis spread remains the same as that of the provisional ensemble. If α is set to 1, the spread becomes equal to that of the LETKF. When there are two or more state variables at a grid point, the ratio of standard deviations in Eq. (38) are to be replaced with that of a representative state variable, such as temperature or wind velocity in meteorology, so as not to destroy the dynamical balance represented in the provisional analysis ensemble. The above procedure is hereafter referred to as the analysis spread adjustment. The hybrid EnKF with $w = 0$ is the same as the LETKF, and the hybrid EnKF with $w = 1$ and $\alpha = 0$ is the same as the stochastic EnKF. Note that Eq. (38) has some resemblance to the relaxation-to-prior spread (RTPS) for covariance inflation (Whitaker and Hamill 2012). The RTPS relaxes the ensemble spread back to the forecast spread via

$$\Delta \boldsymbol{x}^a \leftarrow \left[(1-\alpha) + \alpha \frac{\sigma^f}{\sigma^a}\right] \Delta \boldsymbol{x}^a \qquad (0 < \alpha < 1), \tag{39}$$

at each grid point, where σ^f and σ^a are the forecast and analysis ensemble standard deviation at each grid point. On the other hand, Eq. (38) relaxes the analysis spread of the hybrid EnKF back to that of the LETKF to correct the underestimation of the analysis spread.

Figure 3 shows the workflow of the hybrid EnKF. Same as in the LETKF, the analysis can Fig. 3 be independently performed for each grid point, and observational data in a local domain centered on this grid point are assimilated using the R-localization (Greybush et al., 2011). The white boxes in the figure are common with the LETKF, and the colored three boxes are added to generate the stochastic analysis ensemble and to take a weighting average of the two analysis ensembles. Since the Kalman gain and the forecast ensemble in observation space are already computed in the LETKF, additional computational cost is small. Note that a discontinuity issue is crucial for the hybrid EnKF; if different perturbed observations are assimilated in neighboring local domains, the resulting analysis ensemble will become discontinuous between adjacent grid points. The same perturbations should be used for the same observations in neighboring local domains. One of the methods for this is to assign a different initialization parameter for random number generation to each observation.

5. Experimental design

5.1 Model

The governing equation of the Lorenz-96 model is

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F, \tag{40}$$

for $k = 1, \cdots, K$, satisfying periodic boundary conditions: $X_{-1} = X_{K-1}$, $X_0 = X_K$, and $X_1 = X_{K+1}$. The number of variables K and the forcing parameter F are set to 40 and 8, respectively. According to Lorenz and Emanuel (1998), the number of positive Lyapunov exponents of the model is 13, and the fractional dimension of the attractor, as estimated from the formula of Kaplan and Yorke (1979), is about 27.1. The leading Lyapunov exponent corresponds to a doubling time of 0.42.

Time integration of the model provides the truth that is used for the verification of analysis and the generation of observational data. The fourth-order Runge-Kutta scheme is used for time integration with a time step 0.01. The initial condition at each grid point is F plus an independent random number drawn from a Gaussian distribution with mean 0 and variance 4. The model is integrated from $t =$ 0 to $t =$ 5 050.

5.2 Observations

The data assimilation experiments are conducted using the above model in two cases: Case 1 where the state variables are directly observed and Case 2 where the nonlinear observation operator introduced in Section 3 is used. In Case 1, all the state variables $\boldsymbol{x} = (X_1, \cdots, X_{40})$ are directly observed, and the observation operator is given by $H(\boldsymbol{x}) = \boldsymbol{x}$. The observations $\boldsymbol{y}^o$ are generated by adding random errors $\boldsymbol{\varepsilon}^o$ to the truth $\boldsymbol{x}^t$: $\boldsymbol{y}^{\boldsymbol{o}} = \boldsymbol{x}^t + \boldsymbol{\varepsilon}^o$, where $\boldsymbol{\varepsilon}^o$ are independent random draws from a Gaussian distribution with mean 0 and variance 1. In Case 2, the observation operator is given by $H(\boldsymbol{x}) = \max(\boldsymbol{x}, \boldsymbol{0})$, and observations are generated by $\boldsymbol{y}^o = \max(\boldsymbol{x}^t + \boldsymbol{\varepsilon}^o, \boldsymbol{0})$. The observation error covariance matrix $\boldsymbol{R}$ used in the experiments is set to $\boldsymbol{I}_{40}$ in both cases.

All experiments are conducted for two values of the time interval between observations Δt: 0.05 and 0.50. Since the doubling time of the leading Lyapunov exponent of the Lorenz-96 model is 0.42, the case of $\Delta t =$ 0.05 is weakly nonlinear, whereas that of $\Delta t =$ 0.50 is strongly nonlinear. In addition, the former case corresponds to high-frequency assimilation cycles and the latter case corresponds to low-frequency assimilation cycles. Table 2 presents the characterization of each experiment.

Table 2

The observational data are prepared from $t =$ 0 to $t =$ 1 050 for $\Delta t =$ 0.05 and from $t =$ 0 to $t =$ 5 050 for $\Delta t =$ 0.50. This is because for the case of $\Delta t =$ 0.50 the analysis accuracy intermittently becomes very poor (Penny and Miyoshi, 2016) and, therefore, a much longer data assimilation period is required to obtain statistically stable results. All observations are prepared such that observations at the same analysis time are the same regardless of the value of Δt.

5.3 Data assimilation settings

We perform the data assimilation experiments with 10, 20, and 40 ensemble members. The ensemble size of 10 is not very small as compared with the number of positive Lyapunov exponents of the Lorenz-96 model, and an ensemble size of 40 is the same as the degrees of freedom of the model. The period of data assimilation is 1 050 for $\Delta t =$ 0.05 and 5 050 for $\Delta t =$ 0.50 for the reason mentioned above. The analyses from $t =$ 0 to $t =$ 50 are not used for verification to avoid adverse effects of spin up. The analyses at a time interval of 1 are used for verification to prepare almost independent samples. Therefore, the sample size is the same between the experiments with $\Delta t =$ 0.05 and those with $\Delta t =$ 0.50. Analysis accuracy is estimated by the root mean square error (RMSE) that is the square root of the squared error averaged over the grid points and the period of experiments. All experiments are conducted with the following five sets of tuning parameters of the hybrid EnKF: $(w, \alpha) =$ (0, 0), (0.5, 0), (0.5, 1), (1, 0), and (1, 1). Note that α = 0 and α = 1 indicate the experiment without and with the analysis spread adjustment, respectively. The hybrid EnKF with $(w, \alpha) =$ (1, 1) is different from the stochastic EnKF, because its analysis spread is adjusted to be equal to that of the LETKF. This adjustment suppresses sampling noise contained in the analysis error covariance matrix of the stochastic EnKF, and it is expected to result in better analysis accuracy in weakly nonlinear regimes. For some experiments, we change the value of w from 0 to 1 at a step of 0.1.

Unless the ensemble size is sufficiently large, ensemble Kalman filtering needs covariance localization and covariance inflation to optimize its performance. The correlation function defined by Eq. (4.10) of Gaspari and Cohn (1999) is taken for covariance localization. The parameter c in this equation is regarded as the localization radius r_L (unit: grid interval) in this study, at which radius the correlation coefficient decreases to 5/24. The radius of the local domain is set equal to r_L. The value of r_L is changed from 0 to 19 grid intervals in a step of 1 to obtain the most accurate analysis.

The adaptive inflation method proposed by Li et al. (2009) is applied to each local domain. This method is based on the innovation statistics by Desroziers et al. (2005). Li et al. (2009) imposed lower and upper limits in the “observed” inflation factor $\tilde{\Delta}^o$ before applying a smoothing procedure: 0.9 $\leq \tilde{\Delta}^o \leq$ 1.2. Since we conduct the data assimilation experiments over a much wider range of the time interval between observations, we optimize the upper limit of $\tilde{\Delta}^o$ leaving the lower limit at 0.9 to obtain the most accurate analysis. The candidates of the upper limit are 1.2, 1.5, 2.0, 3.0, 5.0, and infinity. In addition, although Li et al. (2009) set the error growth parameter κ to 1.03, we adopt a larger value κ = 1.1, because we found that the latter value led to a better analysis accuracy. In the adaptive inflation for Case 2, the observation operator of Case 1 is used instead of that of Case 2. This procedure is not correct when a predicted state variable is negative, but no serious difficulty arises because the range of $\tilde{\Delta}^o$ is limited. When we used the observation operator of Case 2, we found that analysis accuracy was considerably deteriorated. This was primarily because the constant observation error variance was used even if the observed value was zero.

6. Results

In the following, the localization radius r_L and the upper limit of $\tilde{\Delta}^o$ are optimized for each combination of ensemble size, Δt, w, and α, unless otherwise stated.

6.1 Case 1

We first compare the analysis ensembles of the LETKF and the stochastic EnKF before taking a weighting average in the hybrid EnKF. Figure 4 displays examples for the hybrid EnKF with $(w, \alpha) =$ (0.5, 1) at $t =$ 100 for $\Delta t =$ 0.05 and 0.50. The ensemble size is 10. Perturbations in x_1 and x_2 with respect to each ensemble mean are plotted. The two analysis members in the same color are generated from the same forecast member in this color. The correspondence between the LETKF analysis member and the forecast member is based on the uniform contraction property of the LETKF, and the correspondence for the stochastic EnKF is based on Eq. (8). Since the spreads of ensembles are very different between $\Delta t =$ 0.05 and $\Delta t =$ 0.50, different scales of axes are used in the two panels. It is found that the LETKF and stochastic EnKF analysis members that correspond to the same forecast member tend to be close to each other with some exceptions. This result indicates that the weighting average does not much change the analysis error covariance matrix.

Fig. 4

The analysis RMSEs of the LETKF, hybrid EnKFs, and stochastic EnKF for $\Delta t = 0.05$ and 0.50 are plotted in Fig. 5 against the ensemble size. For $\Delta t = 0.05$, the LETKF (red line) generates the most accurate analysis when the ensemble size is 10 and 20, and the stochastic EnKF (green line) is the least accurate for all ensemble sizes. As for the hybrid EnKFs, the RMSE increases with the increase of the weight. The analysis spread adjustment improves the accuracy of hybrid EnKFs for an ensemble size of 10, but no benefits are seen for an ensemble size of 40. Generally, hybrid EnKFs with $w = 0.5$ are more accurate than hybrid EnKFs with $w = 1$, and hybrid EnKFs with $\alpha = 1$ are more accurate than hybrid EnKFs with $\alpha = 0$. Those results can be explained by the suppression of sampling noise of the stochastic EnKF. Since the sampling noise decreases with the increase of ensemble size, the differences in analysis accuracy among the EnKFs decrease as the ensemble size increases. For $\Delta t = 0.50$, the hybrid EnKF with $(w, \alpha) = (0.5, 1)$ (cyan line) yields the most accurate analysis for all ensemble sizes. This hybrid EnKF has less sampling noise compared to the other hybrid EnKFs. In addition, as will be shown later, the forecast ensembles of hybrid and stochastic EnKFs are more Gaussian than that of the LETKF. Those two factors are considered to contribute to the above result. However, the positive impact on analysis accuracy is rather small; its RMSE is at most only 3% smaller than that of the LETKF. It is also found that the hybrid EnKF with $(w, \alpha) = (0.5, 0)$ (blue line) is less accurate than the other hybrid EnKFs. This result indicates that the analysis spread adjustment is necessary for the hybrid EnKF with $w = 0.5$.

Fig. 5

Since ensemble Kalman filtering is based on the Gaussian assumption, it may be of interest to compare the non-Gaussianity of forecast ensembles. The Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951) is used in this study to measure the difference of a forecast ensemble from a Gaussian distribution with the same mean and variance as the forecast ensemble. Since ensemble sizes are not very large in this study, a histogram of the forecast ensemble to compute the KL divergence is created using five equiprobable bins of the Gaussian distribution, and the KL divergence with respect to the Gaussian distribution is computed using the following equation:

$$D(p\|p_N) = \sum_{i=1}^{5} p_i \log \frac{p_i}{(p_N)_i} = \sum_{i=1}^{5} p_i \log \frac{p_i}{1/5}\,, \tag{41}$$

where p_i and $(p_N)_i$ are the probabilities of the ith bin of the forecast ensemble and the Gaussian distribution, respectively. If we adopt $D(p_N||p)$ instead of $D(p\|p_N)$, the value of KL divergence becomes infinite when $p_i = 0$ for a certain bin.

The forecast KL divergences averaged over the grid points and the verification period are compared in Fig. 6 for an ensemble size of 40. This figure only shows general features, because the standard deviation of KL divergence is as large as the mean value. It is found that the KL divergences for $\Delta t =$ 0.50 are larger than those for $\Delta t =$ 0.05, and that the KL divergence of the hybrid and stochastic EnKFs are smaller than that of the LETKF. The latter result suggests that the forecast ensembles of the hybrid and stochastic EnKFs are more Gaussian than the LETKF. This result partly explains why the hybrid EnKF with $(w, \alpha) =$

Fig. 6

(0.5, 1) yields the most accurate analysis for $\Delta t = 0.50$, as mentioned previously.

In the above results, the weight of the hybrid EnKF is set to 0, 0.5 and 1.0. and the localization radius is optimized. Figure 7 plots the differences in analysis RMSE for $\Delta t =$ 0.50 between the LETKF with the optimal localization radius, which is shown by an open rectangle, and the hybrid EnKF with $\alpha = 1$ against the localization radius and the weight. Note that the hybrid EnKF with $w = 0$ is the LETKF, but the hybrid EnKF with $w = 1$ is different from the stochastic EnKF, because its analysis spread is adjusted to be equal to that of the LETKF. Warmer colors indicate that the hybrid EnKF is more accurate than the LETKF with the optimal localization radius. It is found that the optimal localization radius increases with larger ensemble sizes, and that if the localization radius and the weight are optimized the hybrid EnKF is more accurate than the LETKF as the ensemble size increases. The optimal weight of the hybrid EnKF tends to increase with the increase of ensemble size.

Fig. 7

In summary, when the observation operator is linear, the LETKF tends to yields the most accurate analysis in high-frequency assimilation cycles, whereas the hybrid EnKF with the analysis spread adjustment yields the most accurate analysis in low-assimilation cycles. However, its positive impact on analysis accuracy is rather small.

6.2 Case 2

The analysis RMSEs of the LETKF, hybrid EnKFs, and stochastic EnKF for $\Delta t = 0.05$ and $\Delta t = 0.50$ are plotted in Fig. 8 against the ensemble size. For $\Delta t = 0.05$, the RMSEs

Fig. 8

of the LETKF (red line) and the hybrid EnKF with $(w, \alpha) =$ (0.5, 1) (cyan line) increase with the increase of ensemble size. Similar problems for the deterministic EnKF in nonlinear systems were also documented in Mitchell and Houtekamer (2009), Anderson (2010), Lei and Bickel (2011), Tödler and Ahrens (2015), and Tsuyuki and Tamura (2022). If many observations of which observation operators are linear or weakly nonlinear are additionally assimilated, this problem would not occur. As the ensemble size increases, non-Gaussianity becomes more statistically significant, and it could exacerbate the adverse effect of persistent strong non-Gaussianity in high-frequency assimilation cycles. The stochastic EnKF (green line) and the hybrid EnKF with (w, α) =(1, 1) (light green line) do not exhibit such a tendency, but their RMSEs do not change much when the ensemble size is increased from 20 to 40. This might be a sign of saturation of accuracy like what is seen in the RMSE of the LETKF in Fig. 5a. The hybrid EnKF with (w, α) =(0.5, 0) (blue line) yields the most accurate analysis for all ensemble sizes. This result can be explained by the suppression of overestimation of analysis spread in the LETKF, the suppression of sampling noise in the stochastic EnKF, and its less non-Gaussianity than the LETKF. Those three factors are brought about by a weighting average without the analysis spread adjustment. For $\Delta t =$ 0.50, the RMSEs are much larger than those of Case 1. The differences in analysis accuracy among the EnKFs are not very different from Case 1, although the positive impact of the hybrid EnKF on analysis accuracy is slightly larger.

The forecast KL divergences averaged over the grid points and the verification period are compared in Fig. 9 for an ensemble size of 40. It is found that the forecast ensembles [Fig. 9] of the hybrid and stochastic EnKFs are more Gaussian than those of the LETKF. However, the KL divergence of the LETKF with $\Delta t = 0.05$ is smaller than that in Case 1. This is probably because the observation operator in Case 2 is strongly nonlinear around $x = 0$ and, therefore, the uniform contraction property of the LETKF does not hold very well. For $\Delta t = 0.50$, the KL divergence of the LETKF remains almost the same as that of Case 1, whereas the KL divergences of the hybrid and stochastic EnKFs are larger than those in Case 1.

Figure 10 plots the differences in analysis RMSE for $\Delta t = 0.05$ between the LETKF with [Fig. 10] the optimal localization radius and the hybrid EnKF with $\alpha = 0$ against the localization radius and the weight. The hybrid EnKF with $w = 1$ is the same as the stochastic EnKF. Although the optimal localization radius of the LETKF with an ensemble size of 40 is 7 grid intervals, the analysis RMSE of the LETKF is almost constant for localization radii from 6 to 19 gird intervals. When the localization radius and the weight are optimized, the hybrid EnKF is more accurate than the LETKF with the optimal localization radius as the ensemble size increases. The positive impact of the hybrid EnKF is much significant compared to that shown in Fig. 7. Figure 11 plots the differences in analysis RMSE for $\Delta t = 0.50$. The optimal [Fig. 11] localization radii are smaller than those of Case 1, and the positive impact of the hybrid EnKF is slightly larger than that of Case 1 when the ensemble size is 20 and 40. It is also found from Figs. 10 and 11 that the optimal weight of the hybrid EnKF tends to increase with the increase of ensemble size, similarly to Case 1.

In summary, when the observation operator is nonlinear, the hybrid EnKF without the analysis spread adjustment yields the most accurate analysis in high-frequency assimilation cycles with significant improvement over the LETKF. In low-assimilation cycles, the hybrid EnKF with the analysis spread adjustment yields the most accurate analysis with rather small improvement.

7. Summary and discussion

We first revisited the LETKF and the stochastic EnKF using a decomposition of the ensemble transform matrix. We proved that if the observation operator is linear the analysis ensemble perturbations of the LETKF are uniform contractions of the forecast ensemble perturbations in observation space in each direction of the eigenvectors of a forecast error covariance matrix. If the observation operator is weakly nonlinear, this property approximately holds. These results imply that if the forecast ensemble is strongly non-Gaussian the analysis ensemble is also strongly non-Gaussian, and that strong non-Gaussianity therefore tends to persist in high-frequency assimilation cycles, leading to the degradation of analysis accuracy in nonlinear data assimilation.

We next proposed the hybrid EnKF that combines the LETKF and the stochastic EnKF with small additional computational cost. The idea was that the hybrid EnKF could be more robust to nonlinearity than the LETKF and it has less sampling noise than the stochastic EnKF. We investigated the performance of the hybrid EnKF through data assimilation experiments using a 40-variable Lorenz-96 model with linear (Case 1) and nonlinear (Case 2) observation operators. In Case 1, the LETKF tends to yield the most accurate analysis in high-frequency assimilation cycles, whereas the hybrid EnKF with the analysis spread adjustment yields the most accurate analysis in low-assimilation cycles. However, its positive impact on analysis accuracy is rather small. In Case 2, the hybrid EnKF without the analysis spread adjustment yields the most accurate analysis in high-frequency assimilation cycles with significant improvement over the LETKF. In low-assimilation cycles, the hybrid EnKF with the analysis spread adjustment yields the most accurate analysis with rather small improvement. The positive impact of the hybrid EnKF increases with the increase of the ensemble size, and the optimal weight of the hybrid EnKF tends to increase with the increase of ensemble size.

Since ensemble Kalman filtering is based on the Gaussian assumption, the accuracy of EnKFs is expected to be improved by making forecast ensembles more Gaussian. However, we found from the experimental results for low-frequency assimilation cycles that its positive impact is rather small. Significant improvement is obtained from the hybrid EnKF without the analysis spread adjustment in Case 2 in high-frequency assimilation cycles. Strictly speaking, this hybrid EnKF cannot be called an EnKF, because the expected value of its analysis spread is different from the theoretical analysis spread of ensemble Kalman filtering. Relaxation of the Gaussian assumption in EnKFs may be one of the promising strategies for nonlinear or non-Gaussian data assimilation in high-dimensional systems.

The hybrid EnKF has two tuning parameters: the weight of the stochastic EnKF, w, and the degree of analysis spread adjustment, α. We could adaptively adjust those parameters according to the strength of nonlinearity. Since the hybrid EnKF is much beneficial in high-frequency assimilation cycles with strong nonlinearity, we may adopt it only in such a situation, where we can set α to zero and need to tune the value of w only.

It may be of some interest to compare the hybrid EnKF with the LETKF using the relaxation-to-prior perturbations (RTPP) for covariance inflation (Zhang et al. 2004). The RTPP relaxes the analysis perturbations back toward the forecast perturbations via

$$\boldsymbol{X}^a \leftarrow (1-\alpha)\boldsymbol{X}^a + \alpha\boldsymbol{X}^f \qquad (0<\alpha<1). \tag{42}$$

As the central limit theorem suggests, the PDF of a random variable generated by adding two non-Gaussian random variables of which PDFs are similar with different standard deviations tends to be less non-Gaussian. Therefore, the RTPP can partially mitigate non-Gaussianity in nonlinear data assimilation, but it does not work in a region of sparse observations, where $\boldsymbol{X}^a$ is close to $\boldsymbol{X}^f$. The hybrid EnKF adds Gaussian perturbations to a linear combination of the LETKF analysis and the analysis generated by the deterministic part of stochastic EnKF, and therefore can mitigate non-Gaussianity much more efficiently.

Data Availability Statement

The Python programs of the hybrid EnKF used in this study are available in J-STAGE Data. https://doi.org/10.34474/data.jmsj.xxxxxxx.

Acknowledgments

The insightful comments of the two anonymous reviewers were very helpful to revise the manuscript. This work was supported by the Ministry of Education, Culture, Sports, Science and Technology through the program for Promoting Research on the Supercomputer Fugaku (Large Ensemble Atmospheric and Environmental Prediction for Disaster Prevention and Mitigation; hp220167).

Appendix

We prove using Eq. (19) that the matrix $\boldsymbol{T}$ defined by Eq. (7) makes the analysis ensemble perturbations of the LETKF as close as possible to the forecast ensemble perturbations. Let Eq. (19) be put in the following form:

$$\boldsymbol{T} = \boldsymbol{I}_N - \sum_{i=1}^{r} \alpha_i \boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}}, \tag{A1}$$

where $0 < \alpha_i < 1$, and let $\boldsymbol{O}_N$ denote an arbitrary orthogonal matrix in N-dimensional space. We express the difference between $\boldsymbol{X}^a = \boldsymbol{X}^f \boldsymbol{T} \boldsymbol{O}_N$ and $\boldsymbol{X}^f$ by the Frobenius norm. The following inequality holds from a property of the norm.

$$\left\| \boldsymbol{X}^a - \boldsymbol{X}^f \right\| \le \| \boldsymbol{T}\boldsymbol{O}_N - \boldsymbol{I}_N \| \cdot \left\| \boldsymbol{X}^f \right\|, \tag{A2}$$

where

$$\| \boldsymbol{T}\boldsymbol{O}_N - \boldsymbol{I}_N \|^{\mathbf{2}} = \mathrm{tr}[(\boldsymbol{T}\boldsymbol{O}_N - \boldsymbol{I}_N)^{\mathrm{T}} (\boldsymbol{T}\boldsymbol{O}_N - \boldsymbol{I}_N)] = \mathrm{tr}[\boldsymbol{T}^{\mathbf{2}}] + N - 2\,\mathrm{tr}[\boldsymbol{T}\boldsymbol{O}_N]. \tag{A3}$$

Substitution of Eq. (A1) into the last term of Eq. (A3) yields

$$\mathrm{tr}[\boldsymbol{TO}_N] = \mathrm{tr}[\boldsymbol{O}_N] - \sum_{i=1}^{r} \alpha_i \,\mathrm{tr}\left[\boldsymbol{v}_i \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{O}_N\right] = \mathrm{tr}[\boldsymbol{O}_N] - \sum_{i=1}^{r} \alpha_i \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{O}_N \boldsymbol{v}_i \qquad \text{(A4)}$$

The trace of a matrix is invariant under orthogonal transformation. Then we can write $\mathrm{tr}[\boldsymbol{O}_N]$ as $\sum_{i=1}^{N} \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{O}_N \boldsymbol{v}_i$ and obtain

$$\mathrm{tr}[\boldsymbol{TO}_N] = \sum_{i=1}^{r} (1-\alpha_i) \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{O}_N \boldsymbol{v}_i + \sum_{i=r+1}^{N} \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{O}_N \boldsymbol{v}_i, \qquad \text{(A5)}$$

Since $1-\alpha_i > 0$ and $-1 \le \boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{O}_N \boldsymbol{v}_i \le 1$, $\mathrm{tr}[\boldsymbol{TO}_N]$ is maximum when $\boldsymbol{v}_i^{\mathrm{T}} \boldsymbol{O}_N \boldsymbol{v}_i = 1$ for $i = 1, \cdots, N$. This implies that $\|\boldsymbol{TO}_N - \boldsymbol{I}_N\|$ is minimum when $\boldsymbol{O}_N = \boldsymbol{I}_N$.

References

Amezcua, J., K. Ide, C. H. Bishop, and E. Kalnay, 2012: Ensemble clustering in deterministic ensemble Kalman filters. *Tellus A*, 64:1, 18039, DOI: 10.3402/tellusa.v64i0.18039.

Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903.

Anderson, J. L., 2010: A non-Gaussian ensemble filter update for data assimilation. *Mon. Wea. Rev.*, **138**, 4186–4198.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.*, **129**, 420–436.

Bocquet, M., C. A. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.*, **138**, 2997–3023.

Bowler, N. E., J. Flowerdew, and S. R. Pring, 2013: Tests of different flavours of EnKF on a simple model. *Quart. J. Roy. Meteor. Soc.*, **139**, 1505–1519.

Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 3385–3396.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10143–10162.

Farchi, A., and M. Bocquet, 2018: Review article: Comparison of local particle filters and new implementations. *Nonlinear Processes Geophys.*, **25**, 765–807.

Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757.

Golub, G. H., and C. F. Van Loan, 2013: *Matrix Computations*, *4th edition*, Johns and Hopkins University Press, Baltimore, 756 pp.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith, 1993: Novel approach to nonlinear/non-Gaussian Bayesian state estimation, *IEE Proceedings F*, **140**, 107–113.

Greybush, S. J., E. Kalnay, T. Miyoshi, K. Ide, and B. R. Hunt, 2011: Balance and ensemble Kalman filter localization techniques. *Mon. Wea. Rev.*, **139**, 511–522.

Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D Nonlinear Phenom.*, **230**, 112-126.

Kaplan, J. L., and J. A. Yorke, 1979: Chaotic behavior of multidimensional difference equations. *Lecture Notes in Mathematics*, H.-O. Peitgen and H.-O. Waters, Eds., Springer Verlag, 204 –227.

Kitagawa, G., 1996: Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *J. Comput. Graph. Statist.*, **5**, 1–25.

Kotsuki, S., T. Miyoshi, K. Kondo, and R. Potthast, 2022: A local particle filter and its Gaussian mixture extension implemented with minor modifications to the LETKF. *Geosci. Model Dev.*, https://doi.org/10.5194/gmd-2022-69.

Kullback, S., and R. A. Leibler, 1951: On information and sufficiency. *Ann. Math. Stat.*, **22**, 79–86.

Lawson, W. G., and J. A. Hansen, 2004: Implications of stochastic and deterministic filters as ensemble-based data assimilation methods in varying regimes of error growth. *Mon. Wea. Rev.*, **132**, 1966–1981.

Lei, J., P. Bickel, and C. Snyder, 2010: Comparison of ensemble Kalman filters under non-Gaussianity. *Mon. Wea. Rev.*, **138**, 1293–1306.

Lei, J., and P. Bickel, 2011: A moment matching ensemble filter for nonlinear non-Gaussian data assimilation. *Mon. Wea. Rev.*, **139**, 3964–3973.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 523–533.

Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proceedings of the ECMWF Seminar on Predictability*, Reading, UK, ECMWF, 18 pp. [Available at https://www.ecmwf.int/node/10829.]

Lorenz, E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414.

Mitchell, H. L., and P. L. Houtekamer, 2009: Ensemble Kalman filter configurations and their performance with the logistic map. *Mon. Wea. Rev.*, **137**, 4324–4343.

Nakano, S., G. Ueno, and T. Higuchi, 2007: Merging particle filter for sequential data assimilation. *Nonlinear Processes Geophys.*, **14**, 395–408.

Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus A*, **56**, 415–428.

Penny, S. G., and T. Miyoshi, 2016: A local particle filter for high-dimensional geophysical systems. *Nonlinear Processes Geophys.*, **23**, 391–405.

Poterjoy, L., 2016: A localized particle filter for high-dimensional nonlinear systems. *Mon. Wea. Rev.*, **144**, 59-76.

Poterjoy, J., and J. L. Anderson, 2016: Efficient assimilation of simulated observations in a high-dimensional geophysical system using a localized particle filter. *Mon. Wea. Rev.*, **144**, 2007–2020.

Poterjoy, J., R. A. Sobash, and J. L. Anderson, 2017: Convective-scale data assimilation for the weather research forecasting model using the local particle filter. *Mon. Wea. Rev.*, **145**,1897–1918.

Potthast, R., A. Walter, and A. Rhodin, 2019: A localized adaptive particle filter within an operational NWP framework. *Mon. Wea. Rev.*, **147,** 345–362.

Rojahn, A., N. Shenk, P. J. van Leeuwen, and R. Potthast, 2023: Particle filtering and Gaussian mixtures – On a localized mixture coefficients particle filter (LMCPF) for global NWP. *J. Meteor. Soc. Japan*, **101**, 233–253.

Sakov, P., and P. R. Oke, 2008: Implications of the form of the ensemble transformation in the ensemble square root filters. *Mon. Wea. Rev.*, **136**, 1042–1053.

Snyder, C., T. Bengtsson, P. Bickel, and J. Anderson, 2008: Obstacles to high-dimensional particle filtering. *Mon. Wea. Rev.*, **136**, 4629–4640.

Tödter, J., and B. Ahrens, 2015: A second-order exact ensemble square root filter for nonlinear data assimilation. *Mon. Wea. Rev.*, **143**, 1347–1367.

Tsuyuki, T., and R. Tamura, 2022: Nonlinear data assimilation by deep learning embedded in an ensemble Kalman filter. *J. Meteor. Soc. Japan*, **100**, 533–553.

van Leeuwen, P. J., 2009: Particle filtering in geophysical systems. *Mon. Wea. Rev.*, **137**, 4089–4114.

van Leeuwen, P. J., H. R. Kunsch, L. Nerger, R. Potthast, and S. Reich, 2019: Particle filters for high-dimensional geoscience applications: A review. *Quart. J. Roy. Meteor. Soc.*, **145**, 2335–2365.

Wang, X., C. H. Bishop, and S. J. Julier, 2004: Which is better, an ensemble of positive–negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.* **132**,1590–1605.

Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140**, 3078–3089.

Zhang, F., C. Snyder, and J. Sun, 2004: Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble Kalman filter. *Mon. Wea. Rev.*, **132**, 1238–1253.

List of Figures

Fig. 1. Example of LETKF analysis based on Eqs. (6)–(7) in a two-dimensional system with an ensemble size of 10. (a) Prior PDF (green contours), likelihood (red contours), forecast ensemble members (green dots), and their ensemble mean (green cross). The ensemble members are numbered from 0 to 9. (b) Posterior PDF (blue contours), analysis ensemble members (blue dots), and their ensemble mean (blue cross). (c) Ratios of each pair of analysis and forecast members. The horizontal axis is the member's number indicated in (a). The tilted axes plotted in (a) and (b) indicate the directions of eigenvectors of $\boldsymbol{P}_Y^f$ with the origins set at each ensemble mean. The contour intervals are set to relative to the maximum value of each PDF.

Fig. 2. Forecast and analysis ensembles with a nonlinear observation operator in a one-dimensional system with an ensemble size of 10 000 when the observation is 0. (a) Prior PDF (light green line), likelihood function (orange line), and forecast ensemble (green bars). (b) Posterior PDF (cyan line) and analysis ensemble of LETKF (blue bars). (c) Posterior PDF and analysis ensemble of stochastic EnKF.

Fig. 3. Workflow of the hybrid EnKF. See text for details.

Fig. 4. Ensemble members of forecasts (crosses), LETKF analysis (dots) and stochastic EnKF analysis (triangles) before taking a weighting average in Case 1 at $t = 100$ for (a) $\Delta t = 0.05$ and (b) $\Delta t = 0.50$. The ensemble size is 10 and the values of parameters are $w = 0.5$ and $\alpha = 1$. Perturbations in x_1 and x_2 with respect to each ensemble mean are plotted. The analysis members in the same color correspond to the forecast member in this color. The upper limit of $\tilde{\Delta}^o$ and localization radius are optimized for each experiment.

Fig. 5. Analysis RMSEs of LETKF (red), hybrid EnKF with $w = 0.5$ and $\alpha = 1$ (cyan), hybrid EnKF with $w = 0.5$ and $\alpha = 0$ (blue), hybrid EnKF with $w = 1$ and $\alpha = 1$ (light green), and stochastic EnKF (green) in Case 1 for (a) $\Delta t = 0.05$ and (b) $\Delta t = 0.50$. They are plotted against ensemble size. The upper limit of $\tilde{\Delta}^o$ and localization radius are optimized for each experiment.

Fig. 6. Forecast KL divergences with respect to the Gaussian distribution of LETKF (red), hybrid EnKF with $w = 0.5$ and $\alpha = 1$ (cyan), hybrid EnKF with $w = 0.5$ and $\alpha = 0$ (blue), hybrid EnKF with $w = 1$ and $\alpha = 1$ (light green), and stochastic EnKF (green) in Case 1 with ensemble size 40. The upper limit of $\tilde{\Delta}^o$ and localization radius are optimized for each experiment.

Fig. 7. Differences in analysis RMSE for $\Delta t = 0.50$ between LETKF with the optimal localization radius, which is shown by an open rectangle on the abscissa, and hybrid EnKF with $\alpha = 1$ in Case 1. The ensemble size is (a) 10, (b) 20, and (c) 40. They are plotted against localization radius and weight, and warmer colors indicate that hybrid EnKF is more accurate than the LETKF with the optimal localization radius. The upper limit of $\tilde{\Delta}^o$ is set to infinity as the optimal value.

Fig. 8. Same as Fig. 5 except for Case 2.

Fig. 9. Same as Fig. 6 except for Case 2.

Fig. 10. Same as Fig. 7 except for Case 2, $\Delta t = 0.05$, and hybrid EnKF with $\alpha = 0$. The upper limit of $\tilde{\Delta}^o$ is set to 1.2 as the optimal value. Analysis RMSE of LETKF is almost constant for localization radii from 6 to 19.

Fig. 11. Same as Fig. 7 except for Case 2. The upper limit of $\tilde{\Delta}^o$ is set to 5.0 in (a) and (c), and to infinity in (b) as the optimal value.

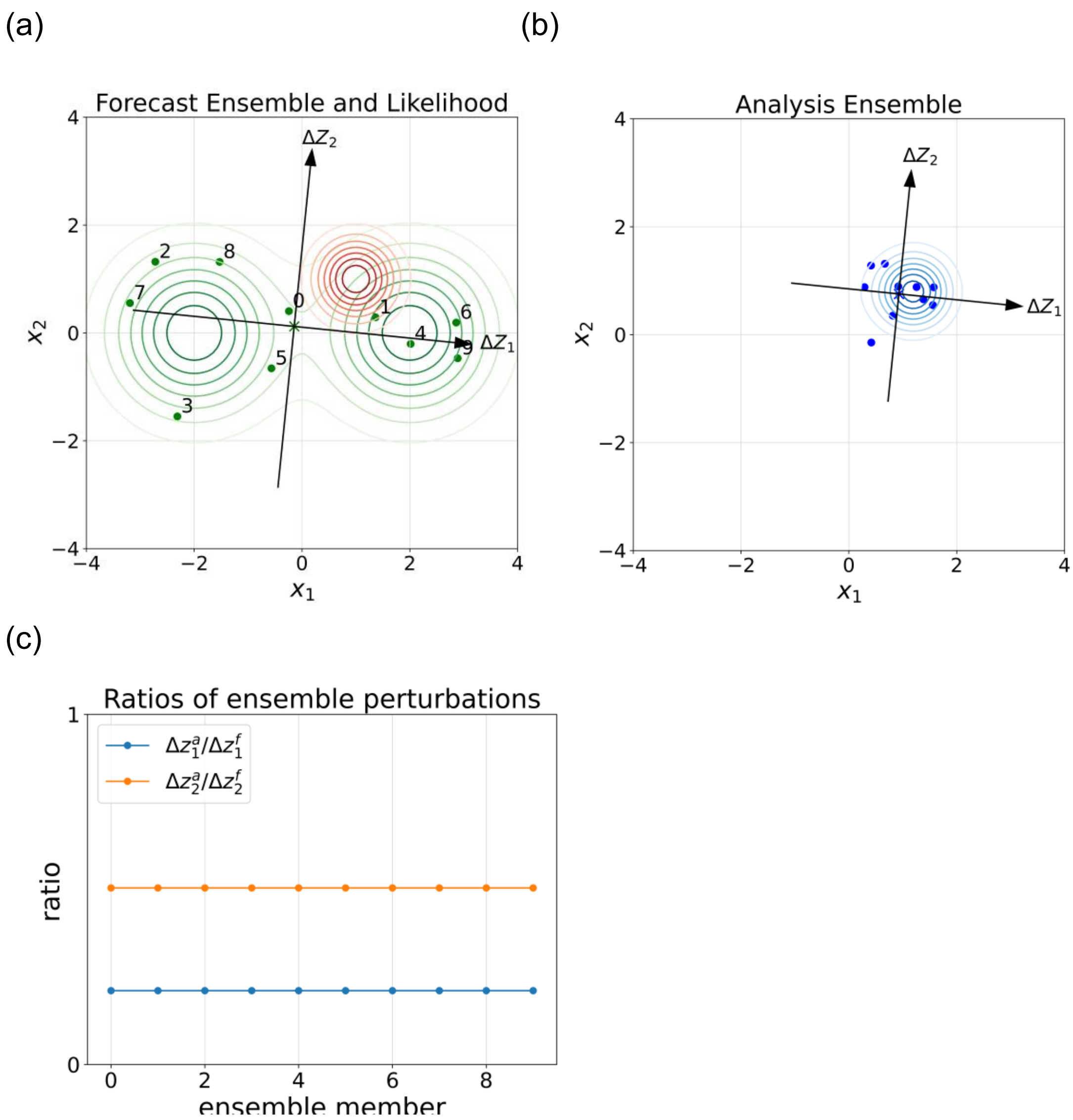


Fig. 1. Example of LETKF analysis based on Eqs. (6)–(7) in a two-dimensional system with an ensemble size of 10. (a) Prior PDF (green contours), likelihood (red contours), forecast ensemble members (green dots), and their ensemble mean (green cross). The ensemble members are numbered from 0 to 9. (b) Posterior PDF (blue contours), analysis ensemble members (blue dots), and their ensemble mean (blue cross). (c) Ratios of each pair of analysis and forecast members. The horizontal axis is the member's number indicated in (a). The tilted axes plotted in (a) and (b) indicate the directions of eigenvectors of $\boldsymbol{P}_Y^f$ with the origins set at each ensemble mean. The contour intervals are set to relative to the maximum value of each PDF.

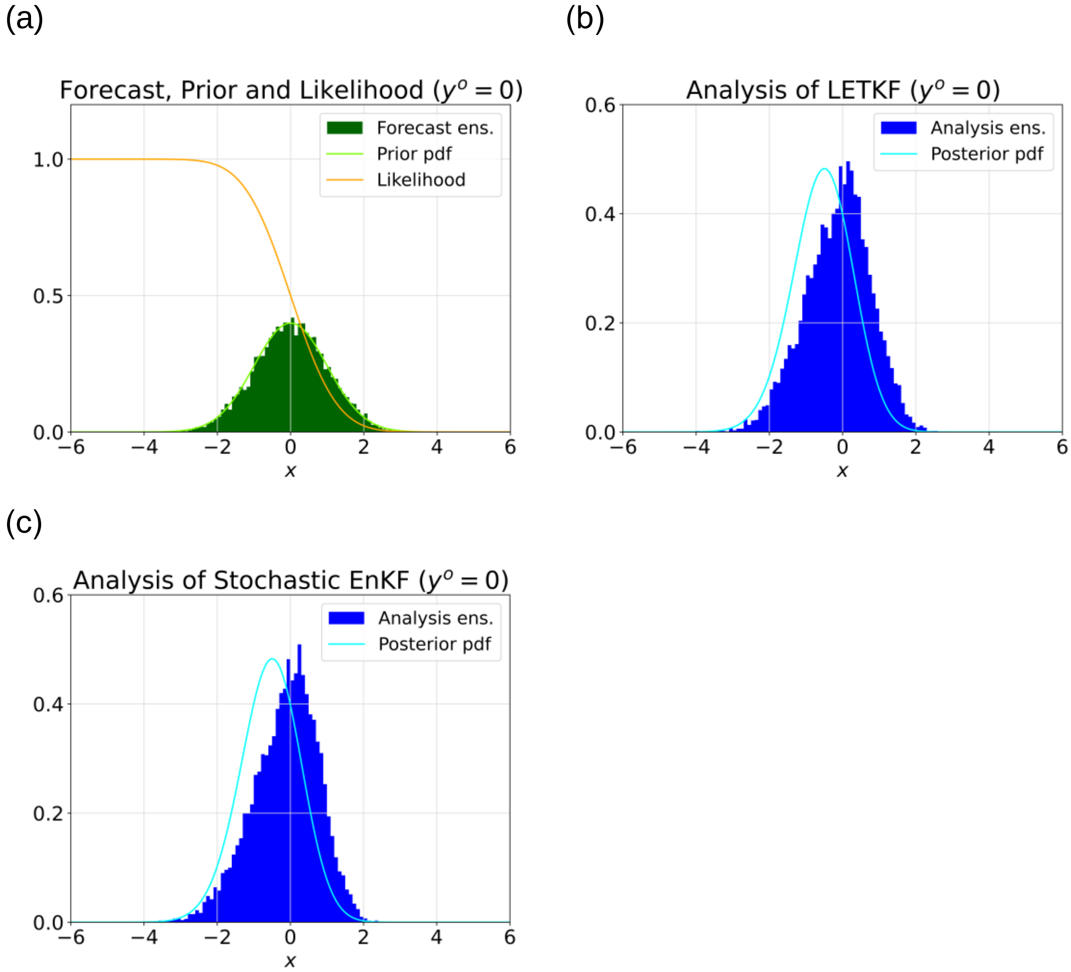


Fig. 2. Forecast and analysis ensembles with a nonlinear observation operator in a one-dimensional system with an ensemble size of 10 000 when the observation is 0. (a) Prior PDF (light green line), likelihood function (orange line), and forecast ensemble (green bars). (b) Posterior PDF (cyan line) and analysis ensemble of LETKF (blue bars). (c) Posterior PDF and analysis ensemble of stochastic EnKF.

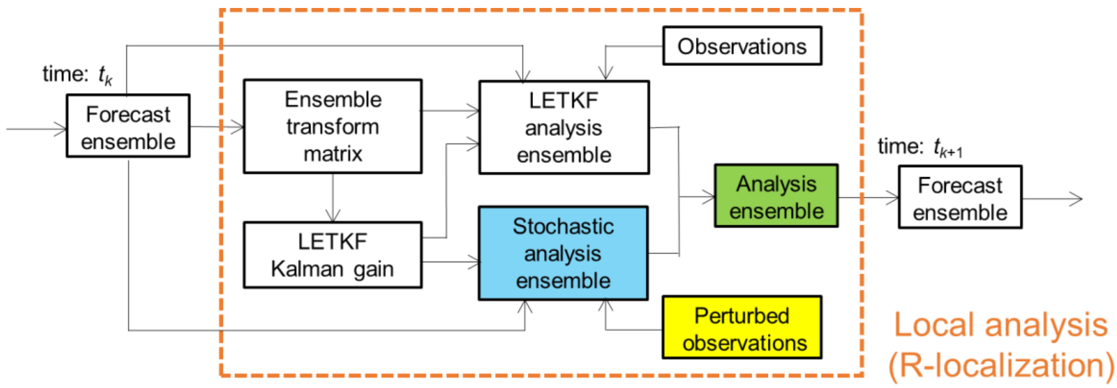
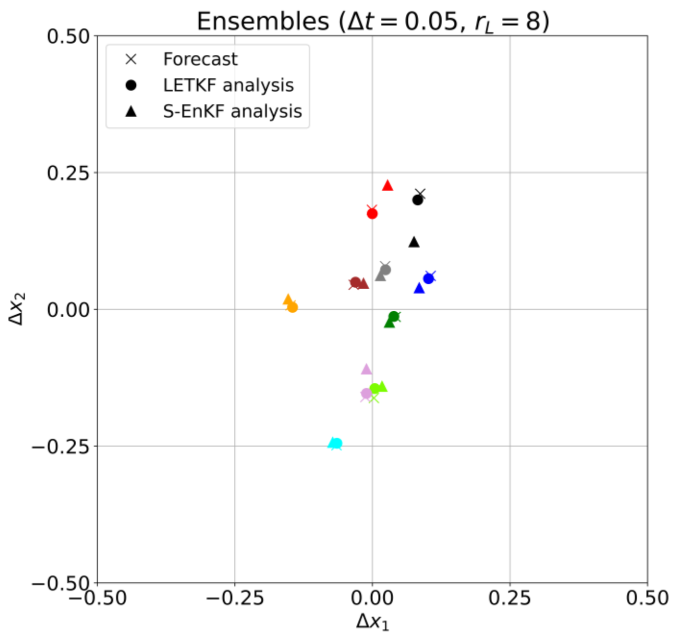


Fig. 3. Workflow of the hybrid EnKF. See text for details.

(a)



(b)

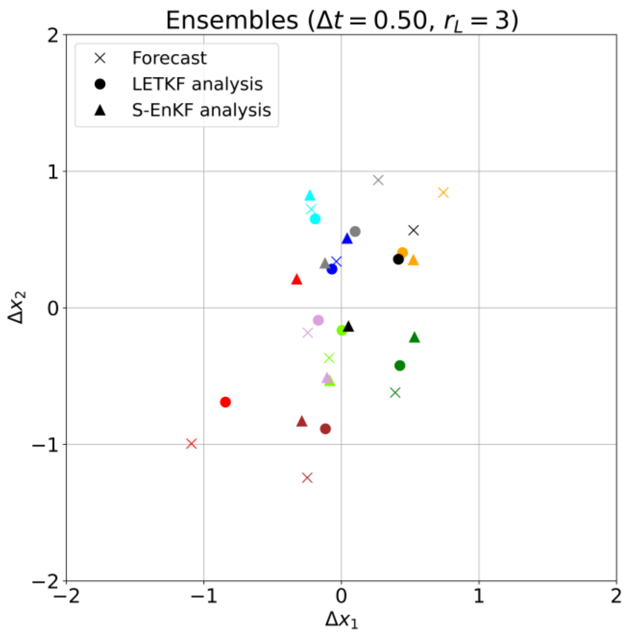


Fig. 4. Ensemble members of forecasts (crosses), LETKF analysis (dots) and stochastic EnKF analysis (triangles) before taking a weighting average in Case 1 at $t =$ 100 for (a) $\Delta t =$ 0.05 and (b) $\Delta t =$ 0.50. The ensemble size is 10 and the values of parameters are $w =$ 0.5 and $\alpha =$ 1. Perturbations in x_1 and x_2 with respect to each ensemble mean are plotted. The analysis members in the same color correspond to the forecast member in this color. The upper limit of $\tilde{\Delta}^o$ and localization radius are optimized for each experiment.

(a) (b)

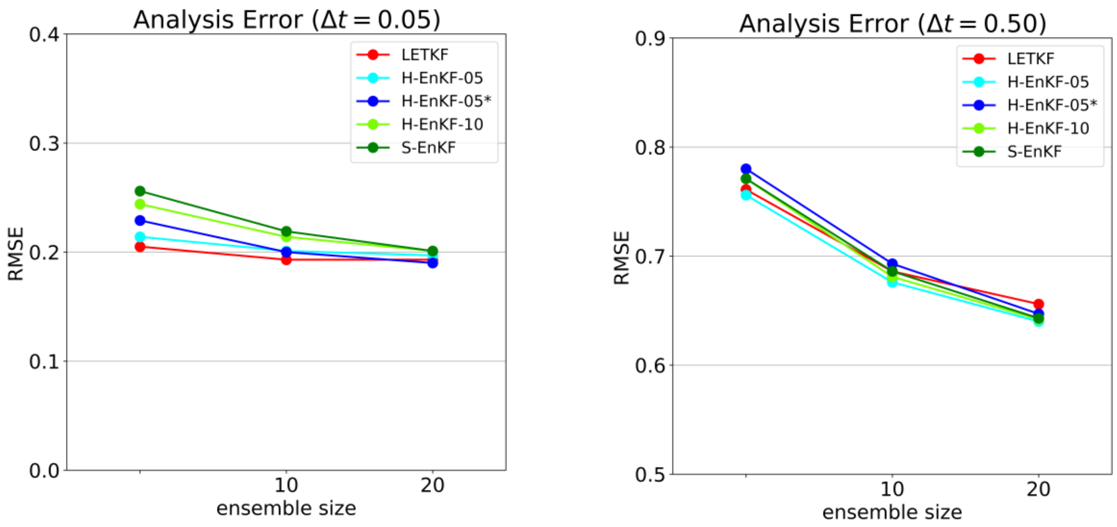


Fig. 5. Analysis RMSEs of LETKF (red), hybrid EnKF with $w = 0.5$ and $\alpha = 1$ (cyan), hybrid EnKF with $w = 0.5$ and $\alpha = 0$ (blue), hybrid EnKF with $w = 1$ and $\alpha = 1$ (light green), and stochastic EnKF (green) in Case 1 for (a) $\Delta t = 0.05$ and (b) $\Delta t = 0.50$. They are plotted against ensemble size. The upper limit of $\tilde{\Delta}^o$ and localization radius are optimized for each experiment.

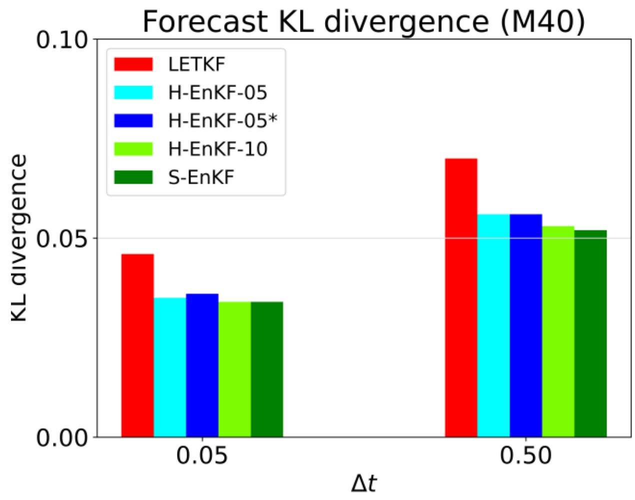


Fig. 6. Forecast KL divergences with respect to the Gaussian distribution of LETKF (red), hybrid EnKF with $w = 0.5$ and $\alpha = 1$ (cyan), hybrid EnKF with $w = 0.5$ and $\alpha = 0$ (blue), hybrid EnKF with $w = 1$ and $\alpha = 1$ (light green), and stochastic EnKF (green) in Case 1 with ensemble size 40. The upper limit of $\tilde{\Delta}^o$ and localization radius are optimized for each experiment.

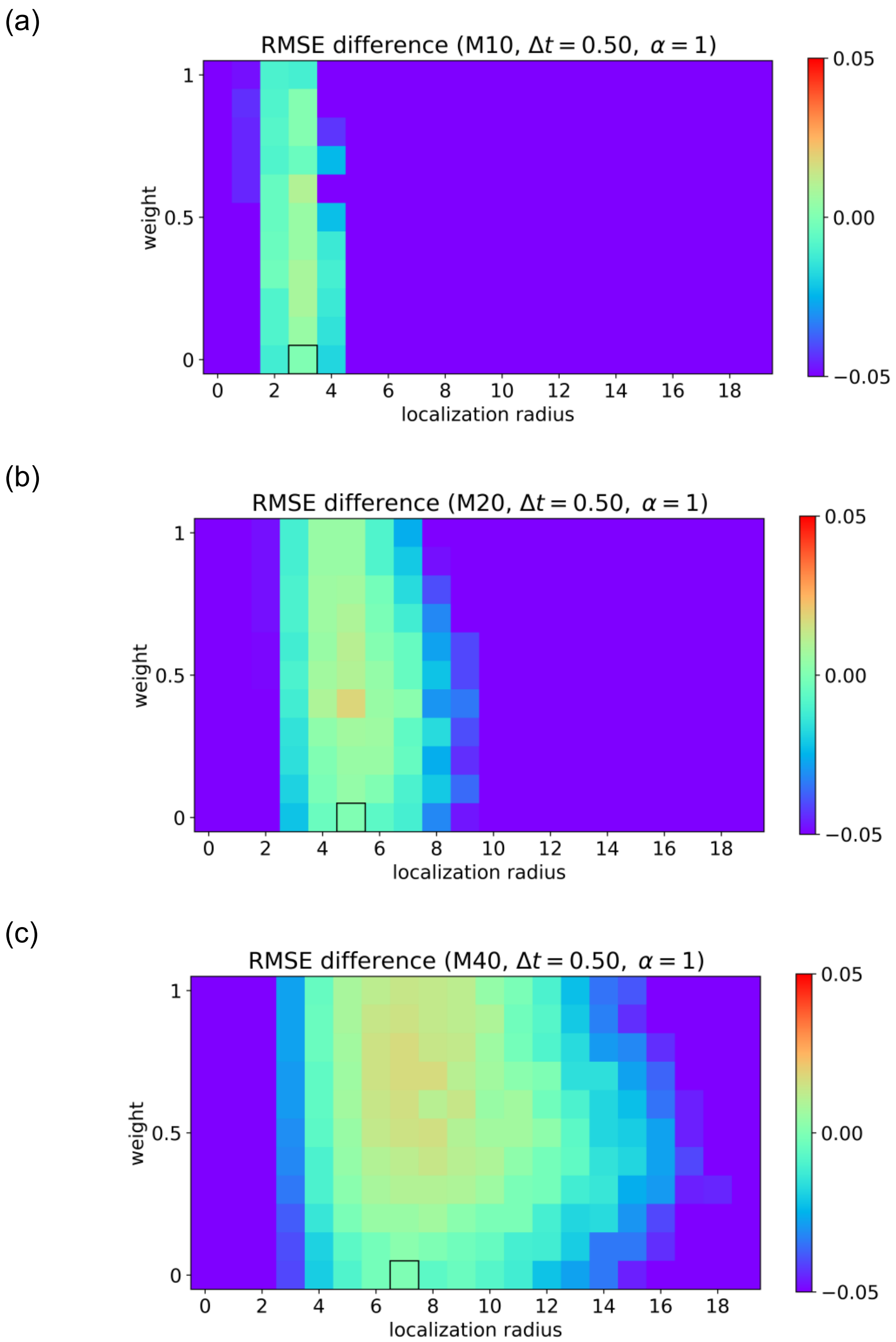


Fig. 7. Differences in analysis RMSE for $\Delta t = 0.50$ between LETKF with the optimal localization radius, which is shown by an open rectangle on the abscissa, and hybrid EnKF with $\alpha = 1$ in Case 1. The ensemble size is (a) 10, (b) 20, and (c) 40. They are plotted against localization radius and weight, and warmer colors indicate that hybrid EnKF is more accurate than the LETKF with the optimal localization radius, The upper limit of $\tilde{\Delta}^o$ is set to infinity as the optimal value.

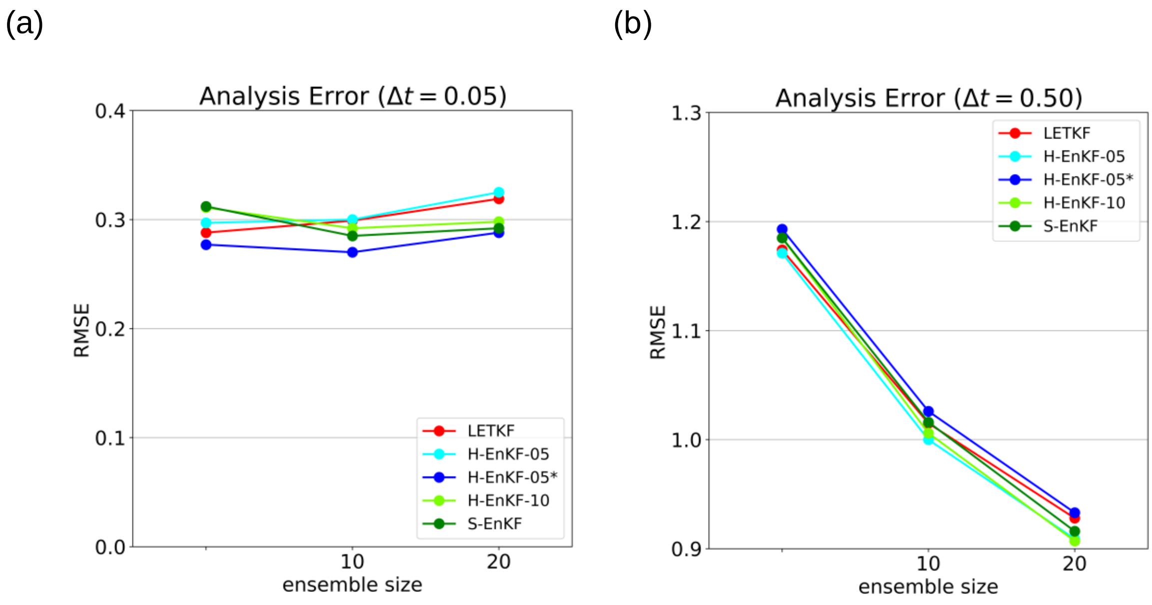


Fig. 8. Same as Fig. 5 except for Case 2.

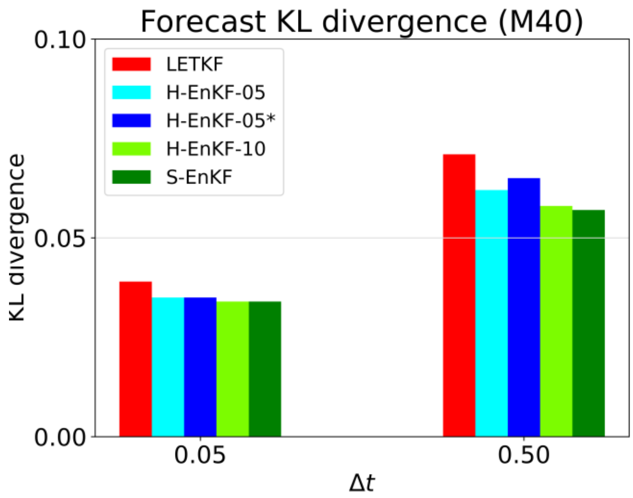
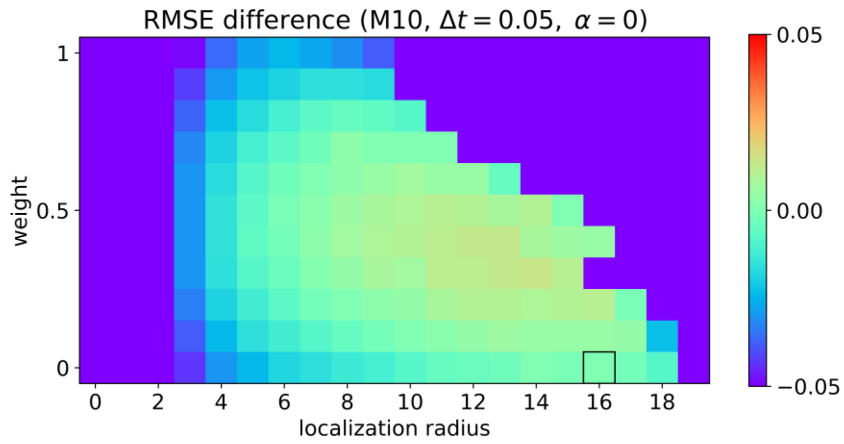
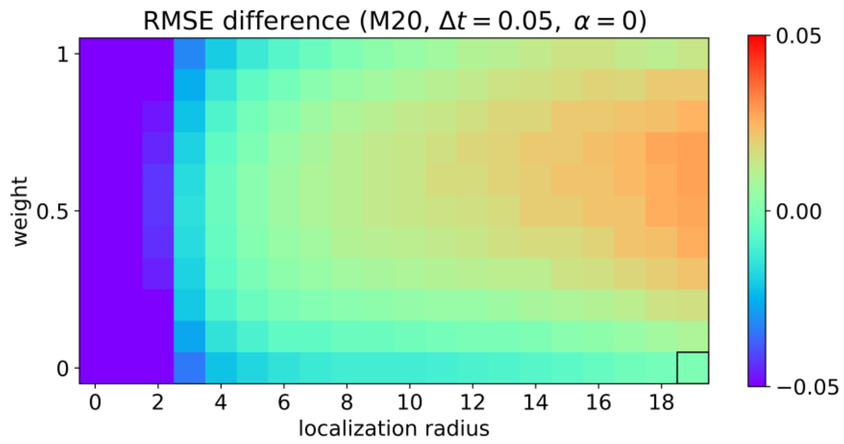


Fig. 9. Same as Fig. 6 except for Case 2.

(a)



(b)



(c)

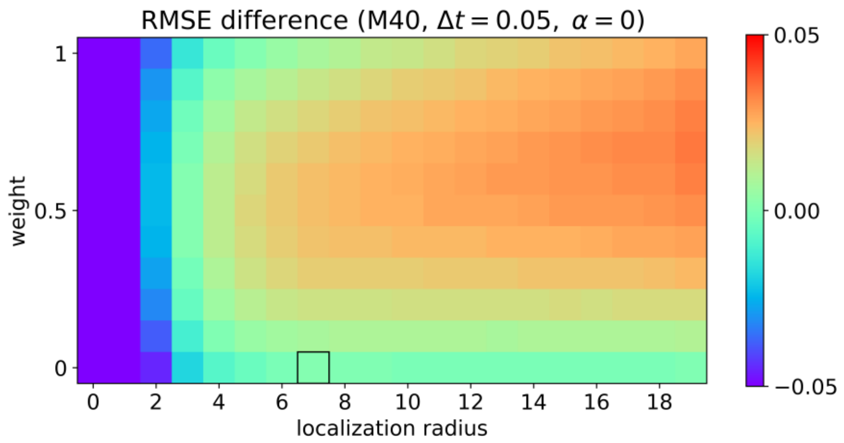
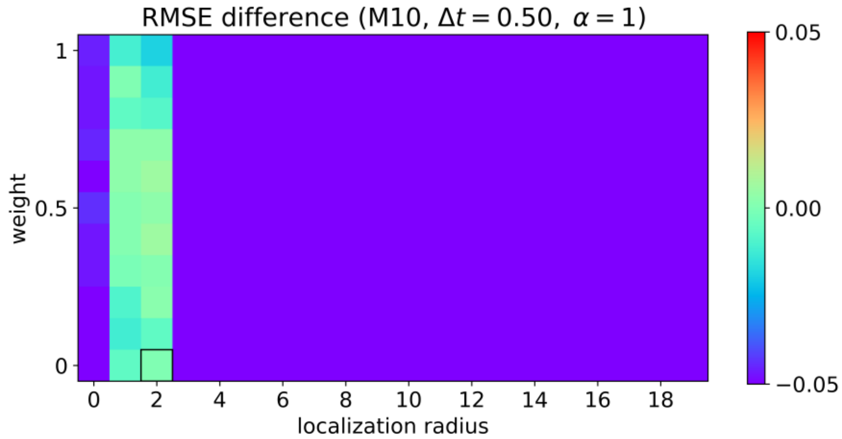
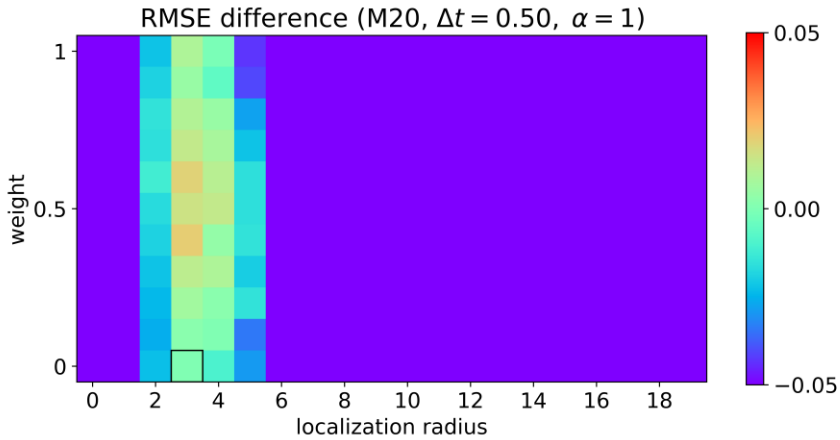


Fig. 10. Same as Fig. 7 except for Case 2, $\Delta t = 0.05$, and hybrid EnKF with $\alpha = 0$. The upper limit of $\tilde{\Delta}^o$ is set to 1.2 as the optimal value. Analysis RMSE of LETKF is almost constant for localization radii from 6 to 19.

(a)



(b)



(c)

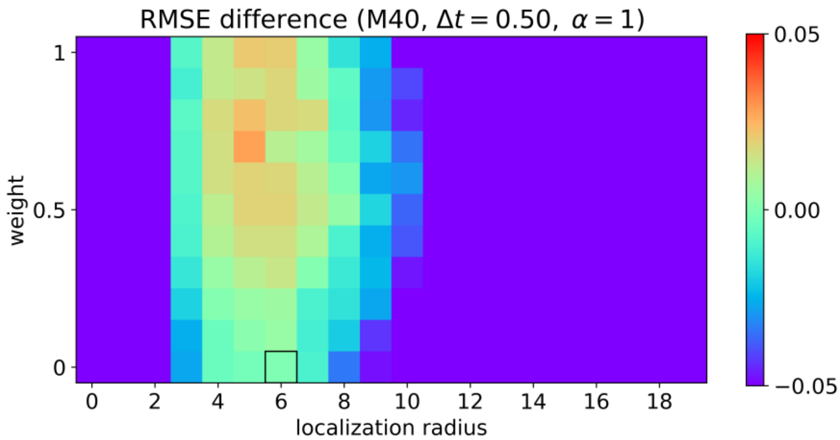


Fig. 11. Same as Fig. 7 except for Case 2. The upper limit of $\tilde{\Delta}^o$ is set to 5.0 in (a) and (c), and to infinity in (b) as the optimal value.

List of Tables

Table 1. Statistics of the posterior PDF and the analysis ensembles of the LETKF and the stochastic EnKF shown in Fig. 2.

	Posterior PDF	LETKF	Stochastic EnKF
Mean	–0.564	–0.134	–0.134
Std. Dev.	0.826	0.898	0.898
Skewness	–0.137	–0.305	–0.445
Kurtosis	0.062	0.037	0.214

Table 2. Characterization of each experiment.

	$\Delta t = 0.05$	$\Delta t = 0.50$
Case 1	High frequency Weakly nonlinear	Low frequency Strongly nonlinear
Case 2	High frequency Strongly nonlinear	Low frequency Strongly nonlinear