Can We Optimize the Assimilation Order in the Serial Ensemble Kalman Filter? A Study with the Lorenz-96 Model

SHUNJI KOTSUKI

Data Assimilation Research Team, RIKEN Advanced Institute for Computational Science, Kobe, Japan

STEVEN J. GREYBUSH

Department of Meteorology and Atmospheric Science, The Pennsylvania State University, University Park, Pennsylvania

TAKEMASA MIYOSHI

Data Assimilation Research Team, RIKEN Advanced Institute for Computational Science, Kobe, Japan, and Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland, and Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan

ABSTRACT

With the serial treatment of observations in the ensemble Kalman filter (EnKF), the assimilation order of observations is usually assumed to have no significant impact on analysis accuracy. However, Nerger derived that analyses with different assimilation orders are different if covariance localization is applied in the observation space. This study explores whether the assimilation order can be optimized to systematically improve the filter estimates. A mathematical demonstration of a simple two-dimensional case indicates that different assimilation orders can cause different analyses, although the differences are two orders of magnitude smaller than the analysis increments if two identical observation error variances are the same size as the two identical state error variances. Numerical experiments using the Lorenz-96 40-variable model show that the small difference due to different assimilation orders could eventually result in a significant difference in analysis accuracy. Several ordering rules are tested, and the results show that an ordering rule that gives a better forecast relative to future observations improves the analysis accuracy. In addition, the analysis is improved significantly by ordering observations from worse to better impacts using the ensemble forecast sensitivity to observations (EFSO), which estimates how much each observation reduces or increases the forecast error. With the EFSO ordering rule, the change in error during the serial assimilation process is similar to that obtained by the experimentally found best sampled assimilation order. The ordering has more impact when the ensemble size is smaller relative to the degrees of freedom of the dynamical system.

1. Introduction

Since first proposed by Evensen (1994), the ensemble Kalman filter (EnKF) has been explored extensively in the past two decades as a practical choice for advanced data assimilation methods in numerical weather prediction (NWP) and other applications. Advantages of the EnKF include flow-dependent background error covariances,

relative ease of implementation, an analysis ensemble that represents the analysis error, and straightforward application to nonlinear dynamics (e.g., Ott et al. 2004; Tippett et al. 2003; Li et al. 2009a,b). EnKF methods are categorized into two types: the perturbed-observation (PO) method (Burgers et al. 1998) and ensemble square root filters (EnSRFs; Whitaker and Hamill 2002). Within the category of EnSRF, several implementations have been proposed (Tippett et al. 2003). They have two different ways of treating observations: serial treatment and simultaneous treatment. Some EnSRF methods such as the ensemble adjustment Kalman filter (EAKF; Anderson 2001) and serial EnSRF (Whitaker and Hamill 2002) treat observations serially. On the other hand, the local EnKF

ⓐ Denotes content that is immediately available upon publication as open access.

Corresponding author: Shunji Kotsuki, shunji.kotsuki@riken.jp

(LEKF; Ott et al. 2004), the ensemble transform Kalman filter (ETKF; Bishop et al. 2001), and the local ETKF (LETKF; Hunt et al. 2007) are designed to treat observations simultaneously.

We usually assume that in the serial treatment of observations the assimilation order has no significant impact on data assimilation performance such as analysis accuracy and computational speed. In fact, with a linear observation operator and the Gaussian error assumption, assimilating two observations A and B in the order of (A, B) results in exactly the same analysis as assimilating the same observations in the reverse order (B, A) in any serial EnKF implementation *if* no covariance localization is applied (cf. a theoretical demonstration in section 2). However, Nerger (2015, hereafter N15) derived that the analyses with different assimilation orders are different *if* covariance localization is applied in the observation space. In practical EnKF applications, it is known that covariance localization is necessary to remove sampling errors in the ensemble-based error covariance between distant locations (Houtekamer and Mitchell 1998, 2001; Hamill et al. 2001). N15 pointed out that the analysis equation for the error covariance matrix is not fulfilled with the observation-space localization. Therefore, analyses with different assimilation orders may not be identical. With a simple two-dimensional example, N15 demonstrated that the analysis state depends on the assimilation order in the serial observation processing (cf. appendix of N15).

Previous studies explored better ordering rules for the serial observation treatment. Whitaker et al. (2008) proposed an ordering rule (hereafter, the W08 rule) based on the order of the error variance reduction. The W08 rule was designed for adaptive observation thinning by identifying observations with insignificant impacts in terms of the error variance reduction. More recently, N15 showed that different assimilation orders result in different analyses and error covariances with an idealized two-dimensional data assimilation example. N15 also tested several ordering rules, including random ordering, using the Lorenz-96 model (L96; Lorenz 1996; Lorenz and Emanuel 1998). N15 found that the W08 rule improved the analysis slightly but significantly. N15 pointed out that the W08 rule stabilized the filter for broad choices of inflation and localization parameters. N15 tested another ordering rule by sorting observations so that remaining observations are distributed uniformly, but found no significant improvement.

This study extends the previous studies and further explores the assimilation order in a serial EnKF. Here, we investigate if any new specific rules of selecting the assimilation order result in more accurate analyses, and how the serial EnKF with better and worse assimilation orders compares with the LETKF. Again, the LETKF assimilates all observations simultaneously and does not have the freedom to choose the assimilation order.

Section 2 demonstrates theoretically how the assimilation order may impact the analysis in the serial EnSRF with a simple two-dimensional example. Based on the simple two-dimensional example, we explore ordering rules to improve the analysis. Section 3 summarizes ordering methods, and numerical experiments using the L96 model are presented in section 4. Section 5 provides a discussion, followed by the conclusions in section 6.

2. Theoretical demonstration

Impact of localization

We first consider a simple two-dimensional problem to demonstrate how the assimilation order of observations could possibly impact the analysis. N15 showed a different two-dimensional example in which different assimilation orders result in different analyses and error covariances (cf. appendix of N15). Here, we consider a more general two-dimensional problem. See also appendix A for more details.

The state vector $\mathbf{x} = (x_1, x_2)^{\mathrm{T}}$ is estimated by taking two observations y_1 and y_2 with the corresponding observation operators $\mathbf{h}_1 = (1, 0)$ and $\mathbf{h}_2 = (0, 1)$, respectively, so that $y_i \, (i = 1, 2)$ observes the state variable x_i. If we assimilate observations simultaneously, the observation operator is the following:

$$\mathbf{H} = \begin{pmatrix} \mathbf{h}_1 \\ \mathbf{h}_2 \end{pmatrix}.$$

We assume that the two observations are independent, so that the error covariance matrix $\mathbf{R}$ is diagonal. The observation error covariance matrix $\mathbf{R}$ and the background (prior) error covariance matrix $\mathbf{P}^f$ are generally written in terms of variances and covariances as

$$\mathbf{R} = \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} \quad \text{and}$$

$$\mathbf{P}^f = \begin{pmatrix} v_1 & c \\ c & v_2 \end{pmatrix}.$$

In this two-dimensional experiment, we apply the Kalman filter equation to obtain the analysis $\mathbf{x}^a$:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^f), \tag{1}$$

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^{\mathrm{T}} (\mathbf{H}\mathbf{P}^f \mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1}, \tag{2}$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f(\mathbf{I} - \mathbf{K}\mathbf{H})^{\mathrm{T}} + \mathbf{K}\mathbf{R}\mathbf{K}^{\mathrm{T}}, \tag{3}$$

where $\mathbf{x}$ and $\mathbf{P}$ are the state vector and its error covariance matrix, respectively; superscript a and f denote analysis and forecast, respectively; $\mathbf{y}^o$ and $\mathbf{H}$ denote the observation vector and observational operator, respectively; and $\mathbf{K}$ is the Kalman gain. If $\mathbf{P}^f$ and $\mathbf{R}$ in Eq. (3) are the same as those in Eq. (2), Eq. (3) can be simplified to be

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f. \qquad (4)$$

The EnSRF solves these equations using an ensemble to estimate $\mathbf{P}^f$, so we can extend the following discussions to EnSRF without loss of generality.

Here, we define three localization methods in this two-dimensional example. The localization matrix $\mathbf{Z}$ is defined by

$$\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2) = \begin{pmatrix} 1 & L \\ L & 1 \end{pmatrix},$$

where L $(0 < L < 1)$ is the localization factor. The first method localizes the background error covariance $\mathbf{P}^f$ in the model space (hereafter, Ploc). The localization matrix $\mathbf{Z}$ is applied to $\mathbf{P}^f$ as

$$\mathbf{P}^f \circ \mathbf{Z} = \begin{pmatrix} v_1 & cL \\ cL & v_2 \end{pmatrix},$$

where the symbol $\circ$ denotes the elementwise product. With Ploc, the analysis error covariances $\mathbf{P}^a$ in Eqs. (3) and (4) are the same because the same background error covariance $\mathbf{P}^f \circ \mathbf{Z}$ is used in Eqs. (2) and (3). With Ploc, simultaneous and serial assimilations result in the same analysis $\mathbf{x}^a$ and error covariance $\mathbf{P}^a$ (cf. appendix A). However, we usually use observation space localization for high-dimensional systems because we cannot store $\mathbf{P}^f$ in memory explicitly (Houtekamer and Mitchell 2001; Campbell et al. 2010). In addition, increasing the rank of $\mathbf{P}^f$ with Ploc is also problematic. The localized $\mathbf{P}^f$ results in a high-rank $\mathbf{P}^a$ that cannot be represented by small ensembles. Therefore, one cannot apply Ploc with small ensembles even for low-dimensional systems such as the L96 model. The second method localizes the $\mathbf{P}^f\mathbf{H}^{\mathrm{T}}$ in Eq. (2) in the observation space (hereafter, Kloc). Kloc also localizes $\mathbf{HP}^f\mathbf{H}^{\mathrm{T}}$ in simultaneous assimilation methods. The serial EnKF usually uses Kloc, and the localization is applied as

$$\mathbf{P}^f \mathbf{h}_1^{\mathrm{T}} \circ \mathbf{z}_1 = \begin{pmatrix} v_1 \\ cL \end{pmatrix} \quad \text{or}$$

$$\mathbf{P}^f \mathbf{h}_2^{\mathrm{T}} \circ \mathbf{z}_2 = \begin{pmatrix} cL \\ v_2 \end{pmatrix},$$

respectively. In the simultaneous assimilation, Kloc is realized as

$$\mathbf{P}^f \mathbf{H}^{\mathrm{T}} \circ \mathbf{Z} = \begin{pmatrix} v_1 & cL \\ cL & v_2 \end{pmatrix}.$$

Note that the analysis error covariance $\mathbf{P}^a$ should be computed by Eq. (3) with Kloc because Eqs. (3) and (4)

are not equivalent with Kloc (cf. section 2c of N15). Finally, the third localization method inflates the observation error covariance $\mathbf{R}$ so that distant observations weigh less on the analysis (hereafter Rloc) as implemented in the LETKF (e.g., Hunt et al. 2007; Miyoshi and Yamane 2007). In the LETKF, analysis equations are solved independently at each component with the inflated observation error covariances. With Rloc, the localization matrices are defined independently for components x_1 and x_2 as

$$\mathbf{Z}_{x1}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1/L \end{pmatrix} \quad \text{and}$$

$$\mathbf{Z}_{x2}^{-1} = \begin{pmatrix} 1/L & 0 \\ 0 & 1 \end{pmatrix},$$

where subscripts x_1 and x_2 denote the localization matrix $\mathbf{Z}$ for components x_1 and x_2, respectively (Greybush et al. 2011). Using the localization matrices, the localized observation error covariances are realized as

$$\mathbf{R} \circ \mathbf{Z}_{x1}^{-1} = \begin{pmatrix} r_1 & 0 \\ 0 & r_2/L \end{pmatrix} \quad \text{and}$$

$$\mathbf{R} \circ \mathbf{Z}_{x2}^{-1} = \begin{pmatrix} r_1/L & 0 \\ 0 & r_2 \end{pmatrix}.$$

If we assimilate the two observations without localization ($L = 1$), simultaneous assimilation and serial assimilation result in the same solutions (cf. appendix A). However, analysis states and/or error covariances are different if we assimilate the two observations with localization ($0 < L < 1$) for Ploc, Kloc, and Rloc. It is important to evaluate how significant these differences are relative to the analysis increments and error variances. Here, we discuss the differences in the analysis and its error covariance obtained by Ploc, Kloc, and Rloc. We consider $v \sim O(\lambda)$, $r \sim O(\lambda)$, and $c \sim O(\lambda/2)$, respectively, where $O(\lambda)$ represents the order of magnitude of the system's error variance. We also approximate

$$(y_i - x_i) \sim O(\sqrt{2\lambda}), \quad i = 1, 2, \qquad (5)$$

using $v \sim O(\lambda)$, $r \sim O(\lambda)$ and Desroziers et al. (2005)'s diagnostic equation

$$\left\langle (\mathbf{y}^o - \mathbf{x})(\mathbf{y}^o - \mathbf{x})^{\mathrm{T}} \right\rangle = \mathbf{P} + \mathbf{R} \quad \text{if} \quad \mathbf{H} = \mathbf{I}, \qquad (6)$$

where $<\cdot>$ denotes the statistical expectation. We evaluate the differences of the analyses relative to the analysis increment, and the differences of the analysis error variances relative to the analysis error variance for the first component x_1 ($\mathbf{x}_{(1)}^a$ and $\mathbf{P}_{(1,1)}^a$) as follows:
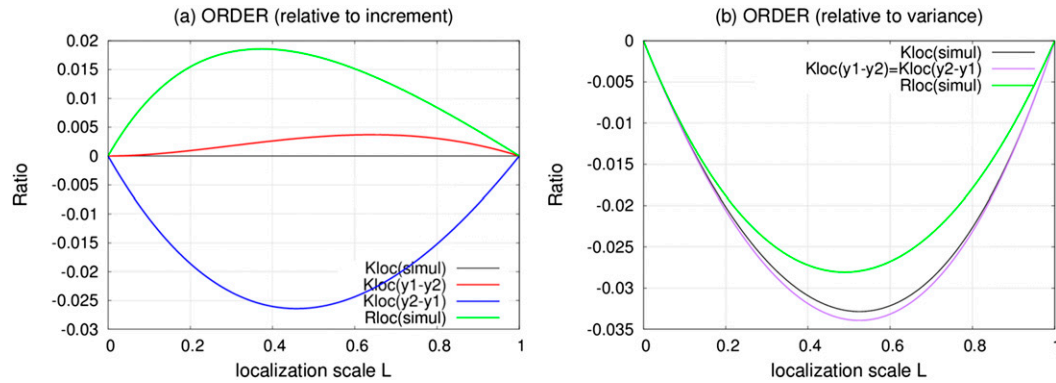
FIG. 1. Ratios of (a) the differences of the analyses relative to the analysis increment with Ploc [α in Eq. (7)], and (b) the differences of the analysis error variances relative to the analysis error variance with Ploc [β in Eq. (8)] where α and β represent the differences for the first component x_1. Black, red, blue, and green lines in (a) show the differences of the analyses between $\mathbf{x}^{a\mathrm{Ploc}}$ and $\mathbf{x}^{a\mathrm{Kloc}}_{\mathrm{simul}}$, $\mathbf{x}^{a\mathrm{Kloc}}_{y_1 \to y_2}$, $\mathbf{x}^{a\mathrm{Kloc}}_{y_2 \to y_1}$, and $\mathbf{x}^{a\mathrm{Rloc}}_{\mathrm{simul}}$, respectively. Black, green, and purple lines in (b) show the differences of the analysis error variances between $\mathbf{P}^{a\mathrm{Ploc}}$ and $\mathbf{P}^{a\mathrm{Kloc}}_{\mathrm{simul}}$, $\mathbf{P}^{a\mathrm{Rloc}}_{\mathrm{simul}}$, and $\mathbf{P}^{a\mathrm{Kloc}}_{y_1 \to y_2}$ ($=\mathbf{P}^{a\mathrm{Kloc}}_{y_2 \to y_1}$), respectively.

$$\alpha = \left(\mathbf{x}^a_{(1)} - \mathbf{x}^{a\mathrm{Ploc}}_{(1)}\right) / \left(\mathbf{x}^{a\mathrm{Ploc}}_{(1)} - \mathbf{x}^f_{(1)}\right), \tag{7}$$

$$\beta = \left(\mathbf{P}^a_{(1,1)} - \mathbf{P}^{a\mathrm{Ploc}}_{(1,1)}\right) / \mathbf{P}^{a\mathrm{Ploc}}_{(1,1)}, \tag{8}$$

where, $\mathbf{x}^a$ is $\mathbf{x}^{a\mathrm{Kloc}}_{\mathrm{simul}}$, $\mathbf{x}^{a\mathrm{Kloc}}_{y_1 \to y_2}$, $\mathbf{x}^{a\mathrm{Kloc}}_{y_2 \to y_1}$, or $\mathbf{x}^{a\mathrm{Rloc}}_{\mathrm{simul}}$, and $\mathbf{P}^a$ is $\mathbf{P}^{a\mathrm{Kloc}}_{\mathrm{simul}}$, $\mathbf{P}^{a\mathrm{Kloc}}_{y_1 \to y_2}$, $\mathbf{P}^{a\mathrm{Kloc}}_{y_2 \to y_1}$, or $\mathbf{P}^{a\mathrm{Rloc}}_{\mathrm{simul}}$, respectively. Subscripts (i) and (i, j) represent ith element of a vector, and ith-column–jth-row element of a matrix, respectively. Subscripts simul, $y_1 \to y_2$, and $y_2 \to y_1$ represent the simultaneous, serial order (y_1, y_2), and serial order (y_2, y_1), respectively. Superscripts Ploc, Kloc, and Rloc denote localization with Ploc, Kloc and Rloc, respectively.

Figure 1 compares α and β as a function of the localization factor L. Here we used $v = r = \lambda = 1$, $\mathbf{x} = (0, 0)^{\mathrm{T}}$, and $\mathbf{y}^o = (\sqrt{2\lambda}, \sqrt{2\lambda})^{\mathrm{T}} = (\sqrt{2}, \sqrt{2})^{\mathrm{T}}$, but we can obtain the same results with any choice of λ mathematically. For the analysis state $\mathbf{x}^a$, simultaneous assimilation with Kloc results in the same analysis with Ploc mathematically in this two-dimensional problem (i.e., $\mathbf{x}^{a\mathrm{Ploc}} = \mathbf{x}^{a\mathrm{Kloc}}_{\mathrm{simul}}$). However, they result in different analysis error covariances $\mathbf{P}^{a\mathrm{Ploc}}$ and $\mathbf{P}^{a\mathrm{Kloc}}_{\mathrm{simul}}$ (black line in Fig. 1b) because Ploc and Kloc use the same $\mathbf{K}$, but different $\mathbf{P}^f$ in Eq. (3). Serial assimilation with Kloc results in different analysis states ($\mathbf{x}^{a\mathrm{Kloc}}_{y_1 \to y_2}$ and $\mathbf{x}^{a\mathrm{Kloc}}_{y_2 \to y_1}$) from simultaneous assimilation with Kloc (red and blue lines in Fig. 1a). Figure 1a shows that the change in the blue line is larger than that in the red line, suggesting that the serial assimilation $y_2 \to y_1$ result in a larger difference in the analysis state for the component x_1 than the serial assimilation $y_1 \to y_2$. Namely, the difference in $\mathbf{x}^{a\mathrm{Kloc}}$ at the ith component between the simultaneous and serial assimilation becomes larger if the component is assimilated later. On the other hand, the analysis error covariances $\mathbf{P}^{a\mathrm{Kloc}}_{y_1 \to y_2}$ and $\mathbf{P}^{a\mathrm{Kloc}}_{y_2 \to y_1}$ are identical in this two-dimensional problem. Simultaneous and serial assimilation with Kloc

result in different analysis error variances (purple and black lines in Fig. 1b), but the difference is small compared to the difference between $\mathbf{P}^{a\mathrm{Ploc}}$ and $\mathbf{P}^{a\mathrm{Kloc}}_{\mathrm{simul}}$. If we focus on comparing the serial and simultaneous assimilation with Kloc, a larger difference is found in the analysis state than the error variance. Rloc results in the different analysis state and error covariance from Ploc.

The ratios of difference are less than 4.0%, suggesting that the differences be two orders of magnitude smaller than the analysis increment and error covariance when two identical observation error variances have the same order of magnitude as the two identical state error variances. The differences peak around $L = 0.5$ (Fig. 1). Figure 1 also indicates that the solutions are identical when $L = 1$ (i.e., no localization). Although the differences of $\mathbf{x}^a$ and $\mathbf{P}^a$ caused by the assimilation order are very small at a single assimilation step, the small differences may accumulate in time and potentially cause a significant impact on analysis accuracy as shown by N15. The differences may be larger if, for example, the observation error variance is significantly smaller or larger than the state error variance. Also, the differences may be much greater in larger dimensional systems such as the L96 or atmospheric models.

3. Ordering rules

With the serial treatment of observations, it is generally prohibitive to investigate all possible assimilation orders. We tested several possible assimilation ordering rules, and this paper discusses some assimilation orders that resulted in significant impacts on the analysis errors. Some other assimilated orders and their results are described in appendix B.
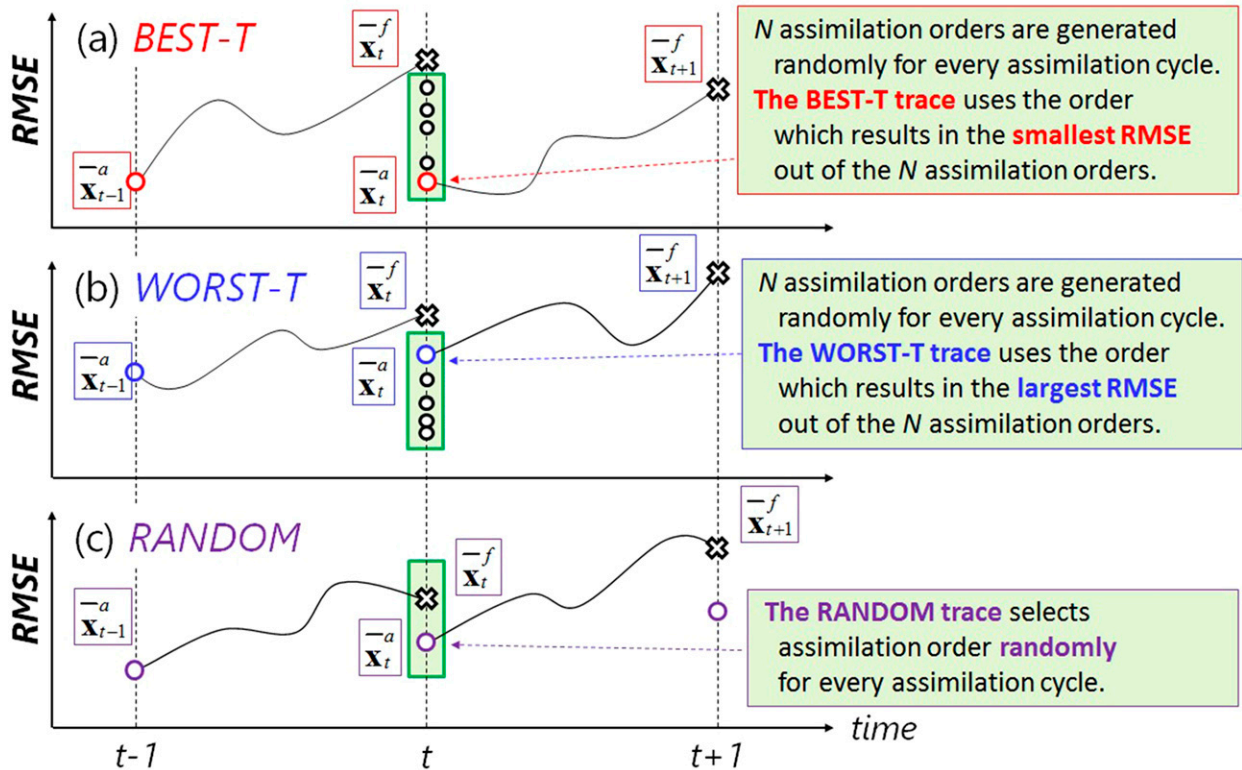
FIG. 2. Schematic of the (a) BEST-T, (b) WORST-T, and (c) RANDOM traces. In the BEST-T and WORST-T traces, N assimilation orders are generated randomly for every assimilation cycle. At each assimilation step, RMSEs relative to the true state $\mathbf{x}^t$ are computed for every assimilation order. The BEST-T and WORST-T traces use the order that results in the smallest and largest RMSEs out of the N assimilation orders. The RANDOM trace selects the assimilation order randomly for every assimilation cycle.

First, we consider three ordering rules to investigate the potential impact of the assimilation order experimentally (Figs. 2a–c). In the best and worst sampled traces relative to the truth (hereafter, BEST-T and WORST-T traces), N assimilation orders are generated randomly at every assimilation cycle. The BEST-T and WORST-T traces select the order that results in the smallest and largest root-mean-square errors (RMSEs) relative to the true state $\mathbf{x}^t$ out of the randomly selected N assimilation orders. This best-sampled and worst-sampled analysis then serves as the initial condition for the subsequent forecast, and the process is repeated. The RANDOM trace selects an assimilation order randomly at every assimilation step. Analysis accuracy of the RANDOM trace is considered to be an expected accuracy since the assimilation order is determined randomly without any special attention to the assimilation order. On the other hand, the BEST-T and WORST-T traces are designed to approximate the range of accuracy experimentally.

We consider additional ordering rules to find better and worse assimilation orders by minimizing and maximizing the root-mean-square differences (RMSDs) between the future observation and forecasts (hereafter BEST-FO and WORST-FO traces; Figs. 3a,b). Note that this manuscript defines the RMS error (RMSE) relative to the truth, and RMS difference (RMSD) relative to the observation. In the BEST-FO and WORST-FO traces, N' assimilation orders are generated randomly at every assimilation step. At each assimilation step, N' forecasts are launched from every initial condition of the N' assimilation orders until the end of a prescribed time window t_fcst + t_window. For every forecast, we compute the time-mean RMSDs relative to the observations $\mathbf{y}^o$. The BEST-FO and WORST-FO traces use the order that results in the smallest and largest RMSDs among the N' assimilation orders. Better analysis states, which have lower RMSE, statistically result in better forecasts relative to the future observations. To compute the time-mean RMSDs, we use observations over the time window (from t_fcst to t_fcst + t_window) to reduce the influence of observation noise. The observations before t_fcst are not used to compute the average RMSD because the observation errors are generally larger than the forecasted errors just after the initial time. The BEST-FO trace has the form of a smoother that utilizes future observations to improve the current analysis
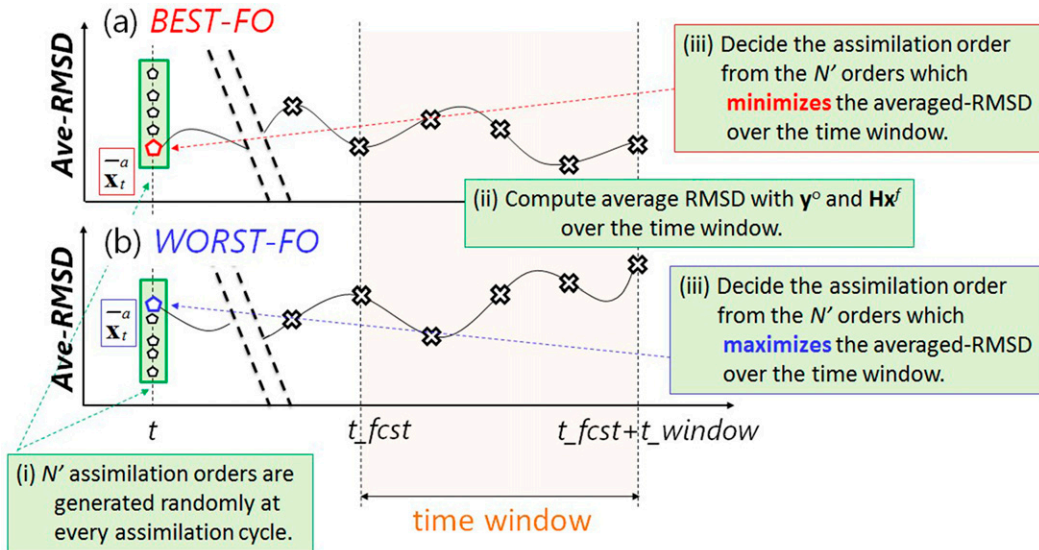
FIG. 3. Schematic of the (a) BEST-FO and (b) WORST-FO traces. In the BEST-FO and WORST-FO traces, N' assimilation orders are generated randomly for every assimilation cycle. At each assimilation step, average RMSDs over the time window relative to the observations $\mathbf{y}^o$ are computed from forecasts generated from every assimilation order. The BEST-FO and WORST-FO traces use the order that results in the smallest and largest RMSDs among the N' assimilation orders.

step. The BEST-FO and WORST-FO traces are applicable to reanalysis experiments in practice; they are, however, computationally expensive because of the extra forecasts. Note that random assimilation orders are generated at every assimilation step for the RANDOM, BEST-T, WORST-T, BEST-FO, and WORST-FO traces because the better and worse assimilation orders can be different at each analysis step depending on the forecasted state and error covariance. Also, it should be noted that we can investigate a finite number of observation orders among $p!$ possible orders for the RANDOM, BEST-T, WORST-T, BEST-FO, and WORST-FO traces.

To further address whether or not any specific rules of the assimilation order of observations result in more accurate analyses, we consider the following ordering rules. They are the ascending and descending orders of $|\mathbf{y}^o - \mathbf{H}\mathbf{x}^t|$ (observation minus truth; hereafter OMT), and $|\mathbf{H}\mathbf{x}^f - \mathbf{H}\mathbf{x}^t|$ (first guess minus truth; hereafter BMT). Here, the absolute value is used as the norm $|\ |$ (a.k.a. the l-1 norm). We apply the l-1 norm for each component of vectors $\mathbf{y}^o - \mathbf{H}\mathbf{x}^t$ and $\mathbf{H}\mathbf{x}^f - \mathbf{H}\mathbf{x}^t$. For example, ascending order of $|\mathbf{y}^o - \mathbf{H}\mathbf{x}^t|$ sorts observations from the smallest to largest $|\mathbf{y}^o - \mathbf{H}\mathbf{x}^t|$.

We also test the W08 rule, which determines the assimilation order as follows: 1) for each observation, assimilate the observation and compute an error variance reduction within the localization length scale, and 2) sort the observations according to the error variance reduction. The reduction of the error variance $s^{(i)}$ is defined as

$$s^{(i)} = \frac{\mathbf{H}\mathbf{P}^{a(i)}\mathbf{H}^{\mathrm{T}}}{\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}}}, \quad i = 1, \ldots, p, \tag{9}$$

where $\mathbf{P}^{a(i)}$ denotes the analysis error covariance by assimilating the ith observation. Following N15, we assimilate observations in the ascending order of $s^{(i)}$. Namely, the observation with the biggest impact on reducing $s^{(i)}$ is assimilated first. N15 reported that the ordering rule of W08 stabilized the analysis and reduced the analysis error slightly. The variance reduction becomes bigger as the signal-to-noise ratio (SN ratio) becomes larger. We can statistically expect a better impact from observations with a larger SN ratio. Therefore, the W08 rule would improve the analysis if better observations are assimilated earlier.

Finally, we test two ordering rules using the ensemble forecast sensitivity to observations (EFSO; Kalnay et al. 2012). The EFSO enables the estimation of how much each observation reduces or increases the forecast error verified against the analysis at a verification time t in the future. In the EFSO, the impact of the ith observation (J_i) is computed as follows [cf. Eq. (9) of Kalnay et al. (2012)]:

$$J_i \approx \frac{1}{m-1}\left(\mathbf{y}_0^o - \mathbf{H}\mathbf{x}_{0|-1}^f\right)_i^{\mathrm{T}}\left[\rho\mathbf{R}^{-1}\mathbf{H}\mathbf{X}_0^a\left(\mathbf{X}_{t|0}^{f\mathrm{T}}\right)\mathbf{C}\left(\mathbf{e}_{t|0}^f + \mathbf{e}_{t|-1}^f\right)\right]_i, \tag{10}$$

where m is the ensemble size, $\mathbf{X}$ is the ensemble perturbation matrix, ρ is the localization function, and $\mathbf{C}$ is a square matrix measuring the error. Subscript -1, 0, and t

denote time at the last data assimilation, current analysis and future verification time respectively. The variable $\mathbf{e}$ is the forecast error relative to the analysis at the verification time t, and given by $\mathbf{e}_{t|0}^f = \mathbf{x}_{t|0}^f - \mathbf{x}_t^a$ and $\mathbf{e}_{t|-1}^f = \mathbf{x}_{t|-1}^f - \mathbf{x}_t^a$, where $\mathbf{x}_{t|0}^f$ and $\mathbf{x}_{t|-1}^f$ denote forecasts started at time 0 and −1, respectively. Negative J means the forecast error reduction (i.e., better impact). The verification time t used in the experiments is described in section 4c. Kalnay et al. (2012) evaluated the impacts of the ith observation for the nth grid point $(\Delta e^2)_{n,i}$ [Eq. (9) of Kalnay et al. (2012)]. In that context, the observation impact J_i in Eq. (10) is given to be $J_i = \sum_{n=1}^{\mathrm{ngrid}} (\Delta e^2)_{n,i}$ where ngrid is the number of model grid points. Using the observation impacts estimated by the EFSO (J_i), observations are sorted from better to worse impact (EFSO-BtoW), or from worse to better impact (EFSO-WtoB). Note that the EFSO ordering rules use future observations to compute the analyses at the verification time for the forecast error verifications. However, the EFSO-BtoW and EFSO-WtoB orders can be determined uniquely, in contrast to the BEST-FO and WORST-FO traces, which also use future observations to determine the assimilation orders from randomly generated N' assimilation orders.

Appendix A (see Table A1) summarizes all ordering rules. Note that several ordering rules use the truth, and are inapplicable in practice. Although the true state is unknown in practice, we investigate if we could optimize the orders in the idealized situation such as an OSSE.

4. Numerical experiments

a. Model

We examine the impacts of different assimilation orders by performing idealized experiments with the 40-variable L96 model (Lorenz 1996; Lorenz and Emanuel 1998). L96 has been widely used for theoretical studies in data assimilation (e.g., Anderson 2001; Whitaker and Hamill 2002; Ott et al. 2004; Miyoshi 2011; N15) and is described by the following equation:

$$dX_j/dt = (X_{j+1} - X_{j-2})X_{j-1} - X_j + F, \quad i = 1, \ldots, 40, \tag{11}$$

where the boundary is cyclic: $X_{-1} = X_{39}$, $X_0 = X_{40}$, and $X_1 = X_{41}$. Following Lorenz and Emanuel (1998), the model is integrated with the forth-order Runge–Kutta scheme and a time step of 0.01 nondimensional units. The forcing F is fixed at 8.0, which makes the model behave chaotically. With $F = 8$, the unit time corresponds to 5 days in terms of the system's error growth rate (Lorenz 1996).

b. Data assimilation schemes

To investigate the potential impact of the assimilation order on analysis accuracy in the serial treatment of observations, we use the serial EnSRF proposed by Whitaker and Hamill (2002). The analysis ensemble mean $\bar{\mathbf{x}}^a$ and perturbations $\mathbf{E}^a$ by assimilating a single observation are obtained by

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{k}(y^o - \mathbf{h}\bar{\mathbf{x}}^f), \tag{12}$$

$$\mathbf{E}^a = (\mathbf{I} - \tilde{\mathbf{k}}\mathbf{h})\mathbf{E}^f, \tag{13}$$

$$\tilde{\mathbf{k}} = \left[1 + \sqrt{\frac{r}{r + \mathbf{h}\mathbf{E}^f(\mathbf{h}\mathbf{E}^f)^{\mathrm{T}}}}\right]^{-1} \mathbf{k}, \tag{14}$$

$$\mathbf{k} = \mathbf{E}^f(\mathbf{h}\mathbf{E}^f)^{\mathrm{T}}\left[\mathbf{h}\mathbf{E}^f(\mathbf{h}\mathbf{E}^f)^{\mathrm{T}} + r\right]^{-1}, \tag{15}$$

where y^o and r denote the scalar observation and its error variance, respectively. Here $\mathbf{E}^f$, $\mathbf{k}$, $\mathbf{k}(\sim)$, and $\mathbf{h}$ are the prior ensemble perturbations, Kalman gain for mean update, gain for perturbation update, and observational operator for the single observation, respectively (cf. Whitaker and Hamill 2002). Each column of $\mathbf{E}$ corresponds to the ensemble perturbations divided by the square root of $(m - 1)$. In the serial EnSRF, the Kalman gain $\mathbf{k}$ is localized as a function of distance between model points and observation points (Houtekamer and Mitchell 1998). The Kloc function f of distance d between grid points i and j is described as follows (Gaspari and Cohn 1999; Hamill et al. 2001; Greybush et al. 2011):

$$f^{\mathrm{Kloc}} = \exp\left\{-0.5\left[d(i,j)/\sigma_{\mathrm{loc}}\right]^2\right\}, \tag{16}$$

where σ_{loc} is the localization parameter to scale the width of localization. The Kloc function f becomes less than 0.002 beyond the finite distance at about 3.65 times σ_{loc}. The localization-scale parameter is manually optimized for each experiment. After assimilating one observation, $\mathbf{x}^a$ and $\mathbf{E}^a$ are considered as $\mathbf{x}^f$ and $\mathbf{E}^f$, and the next observation is assimilated. The process is repeated until all observations are assimilated at a single assimilation step.

We also perform numerical experiments with the LETKF that treats observations simultaneously. The LETKF solves the analysis equations at each grid point independently. The ensemble update of the LETKF is given by

$$\mathbf{E}^a = \mathbf{E}^f \mathbf{U}\Lambda^{-1/2}\mathbf{U}^{\mathrm{T}}, \tag{17}$$

where we define the eigenvalue decomposition:

$$(m - 1)\mathbf{I} + (\mathbf{H}\mathbf{E}^f)^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{E}^f = \mathbf{U}\Lambda\mathbf{U}^{\mathrm{T}} \tag{18}$$

TABLE 1. Experimental setting of observation error variances, manually optimized localization length scale σ, and manually optimized time window parameters (t_fcst and t_window) for the BEST-FO and WORST-FO traces.

	Expt 1	Expt 2
Observation error variance	1.0	Random order 0.05–2.0
Localization scale for the serial EnSRF (grid points)	5.0	5.0
Localization scale for the LETKF (grid points)	5.0	5.0
Time window parameters for the BEST-FO and WORST-FO traces (DAS)	t_fcst = 8 t_window = 16	

(cf. Hunt et al. 2007). Here $\mathbf{U}$ is the square matrix whose columns are the eigenvectors, and $\mathbf{\Lambda}$ is the diagonal matrix whose diagonal elements are the eigenvalues. The analysis ensemble mean $\bar{\mathbf{x}}^a$ is obtained by solving

$$\bar{\mathbf{x}}^a = \bar{\mathbf{x}}^f + \mathbf{E}^f\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}^{\mathrm{T}}(\mathbf{H}\mathbf{E}^f)^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\bar{\mathbf{x}}^f). \quad (19)$$

Error covariance localization within the LETKF is realized through observation localization (i.e., Rloc; Hunt et al. 2007; Miyoshi and Yamane 2007). Here, the inverse of the Kloc function is multiplied to the observational error variance, so that distant observations are treated as if they have larger errors and therefore weigh less on the analysis.

c. Experimental design

Perfect-model identical twin experiments are performed. The nature run is generated by running the L96 model for 8030 time units, which corresponds to 110 years in terms of the system's error growth rate. The observational data are simulated by adding uncorrelated white random noise with unit variance to the nature state at every grid point, so that all 40 grid points are observed. The observation and data assimilation frequency is chosen to be every 0.05 time units. Hereafter, this 0.05 time units is referred to as one data assimilation step (DAS). The ensemble size m is fixed at 8. To mitigate the impact of data assimilation parameters on discussion, we use the adaptive covariance inflation method of Miyoshi (2011). The prior inflation variance is fixed at 0.0016 (= 0.04^2). Following Miyoshi (2011), the inflation factors are estimated at each grid point adaptively. The localization scales are manually optimized for each experiment.

To investigate the sensitivity to the observation errors, two experiments are performed with different observation error settings (Table 1). Experiment 1 uses a spatially uniform observation error with its variance being 1.0. For Experiment 2, each observation has

different observation error variances from 0.05 to 2.0 with a 0.05 increment. The different observation error variances are assigned to random locations and are fixed through the experimental period. Data assimilation cycles for 160 600 DAS including a 14 600 DAS for the spinup are performed to manually optimize the localization length scale σ, so that the analysis RMSE are minimized. The optimized localization scales σ are also included in Table 1. For this optimization, assimilation orders are determined randomly at every assimilation step for the serial EnSRF. Special treatments on localization, such as a regulated localization technique (Nerger et al. 2012) and adaptive localization methods (e.g., Bishop and Hodyss 2009a,b), are beyond the scope of this paper.

With the serial EnSRF, we test the ordering rules considered in section 3. Randomly generated assimilation orders N for the BEST-T and WORST-T traces, and N' for the BEST-FO WORST-FO traces are set to be $N = 1000$ and $N' = 25$, respectively. Again, N and N' denote the number of assimilation orders for the BEST-T and WORST-T traces, and the BEST-FO and WORST-FO traces, respectively. Here, N' is chosen to be much smaller than N because the BEST-FO and WORST-FO traces require expensive additional forecast computations. The time window parameters (t_fcst and t_window) for the BEST-FO and WORST-FO traces are also optimized manually so as to minimize the RMSE and included in Table 1. For the RANDOM trace, we perform 100 parallel experiments because this experiment strongly depends on the random numbers. For the EFSO-BtoW and EFSO-WtoB, we determine the experimental setting following Kalnay et al. (2012; cf. sections 4a and 4c). The forecast time for the error verifications is fixed at 8 DAS. The l-2 norm is used to measure the forecast errors. The localization in the EFSO should account for the propagation of the information from observations at the analysis time (Hotta 2014). The localization center shifts +0.15 grids per DAS to consider the propagations (Kalnay et al. 2012).

d. Results

Figure 4 indicates the time series of RMSEs for experiment 1 of the BEST-T trace, WORST-T trace, and LETKF configurations for 120 consecutive DAS sampled from a total of 146 000 DAS. Again, we define the RMSE relative to the nature run (i.e., truth). This period is selected arbitrarily, not intentionally, to represent the general tendency of the time series for the entire 146 000 DAS. At the initial time of the spinup period, all experiments have the same prior state $\mathbf{x}^f$ and error covariance $\mathbf{P}^f$. After the first assimilation step, each experiment has its own state $\mathbf{x}$ and error covariance $\mathbf{P}$
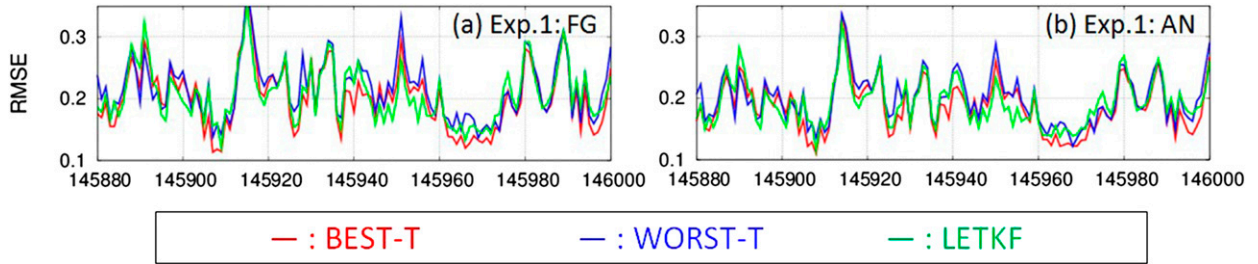
FIG. 4. Time series of RMSE for (a) forecast and (b) analysis steps of experiment 1. Red, blue, and green lines represent the RMSE of the BEST-T trace and WORST-T trace of serial EnSRF and the LETKF, respectively. RMSEs during the last 120 DAS are shown. This period is selected arbitrarily. The abscissa shows the number of assimilation steps.

cycled through data assimilation. Most of the time, RMSEs of the BEST-T trace appear to be lower than those of the WORST-T trace in both forecast and analysis steps. The WORST-T trace shows lower RMSEs than the BEST-T trace occasionally, because each experiment has its own data assimilation cycle. The results imply that the assimilation order of the serial EnSRF could potentially cause significant differences in analysis accuracy. The RMSEs of the LETKF are lower than those of the WORST-T trace in most of the time, while the RMSEs of the LETKF are slightly higher than those of the BEST-T trace in some periods.

Figure 5 shows the average RMSEs over 146 000 DAS for the two experiments. To test the robustness of the results, all experiments are repeated 20 times. Boxes and error bars represent the average and standard deviation of the 20 parallel experiments. The BEST-T trace always shows the lowest RMSE. By contrast, the WORST-T trace always shows the highest RMSE. The LETKF and average of the RANDOM trace show almost the same RMSEs. Since the RMSEs of the BEST-T and WORST-T traces are beyond the range spanned by the minimum and maximum RMSEs of the RANDOM trace, we infer the significance of the lowest and highest RMSEs of the BEST-T and WORST-T traces. The BEST-FO trace consistently shows lower RMSE than the RANDOM trace, while the WORST-FO trace shows higher RMSE than the RANDOM trace. The average RMSEs of the BEST-FO and WORST-FO traces are close to the minimum and maximum RMSEs of the RANDOM trace.

The assimilation orders with the specific five rules, from OMT-Asc to W08 in appendix A (see Table A1), have little impact on analysis accuracy; their RMSEs are within the range of the RANDOM traces. However, the lower RMSEs in the ascending order of $|\mathbf{y}^o - \mathbf{Hx}^t|$ (OMT-Asc) and the descending order of $|\mathbf{Hx}^f - \mathbf{Hx}^t|$ (BMT-Dsc) are observed consistently in the two experiments. Their RMSEs are almost equal to the minimum RMSE of the RANDOM trace. It implies that it may be better to assimilate the better observations

earlier to improve the analysis. The W08 rule also shows a small difference in RMSE compared to the average of the RANDOM trace, but slightly outperforms the RANDOM trace in the two experiments.

Finally, the EFSO rules change the RMSE significantly. Again, the EFSO estimates how much each observation reduces or increases the forecast error. The mean RMSEs of the EFSO-WtoB and EFSO-BtoW are almost equal to that of the BEST-T and WORST-T traces. Namely, we had a significant improvement by ordering observations from worse to better impacts with the EFSO. Further discussion on this result is conducted in the next section. The above results and discussion are consistent across the two experiments.

The experiments demonstrate that different assimilation orders can change the analysis accuracy as shown by the BEST-T, WORST-T, BEST-FO, WORST-T, EFSO-BtoW, and EFSO-WtoB rules. These changes are supported by statistical inference (see appendix C for details). The W08 rule shows slightly lower RMSE than the average of RANDOM trace in the two experiments. The statistical inference, however, does not suggest significant improvement by the W08 rule.

5. Discussion

a. Changes in RMSE during the serial assimilation process

The numerical experiments demonstrate significant improvements by sorting observations from worse to better impacts with the EFSO. Here, we explore how the ordering rules with the EFSO improve and degrade the analysis experimentally. Using the same conditions (i.e., the same ensemble members, observations, and data assimilation parameters), we conduct the serial EnSRF for following assimilation orders: randomly selected, BEST-T, WORST-T, W08, EFSO-BtoW, and EFSO-WtoB orders. N15 investigated changes in RMSE and ensemble spread in the serial assimilation process.
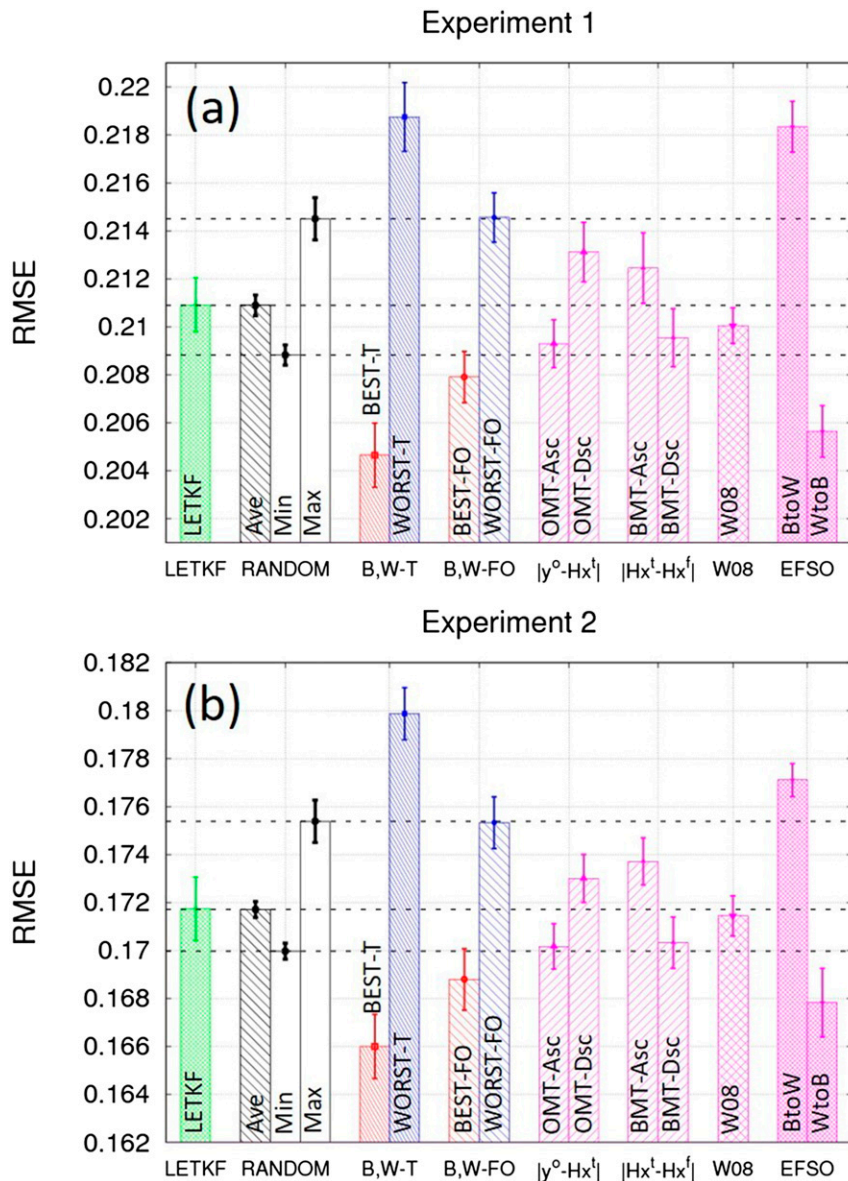
FIG. 5. Average RMSEs over 146 000 DAS of (a) experiment 1 and (b) experiment 2 for the LETKF, serial EnSRF with the BEST-T, WORST-T, RANDOM, BEST-FO, and WORST-FO traces. The average/minimum/maximum RMSEs of 100 runs of the RANDOM traces are shown in black bars (RANDOM-Ave/Min/Max). The magenta bars represent the RMSEs of serial EnSRF with assimilation orders using seven different rules: OMT-Asc, OMT-Dsc, BMT-Asc, BMT-Dsc, W08, EFSO-BtoW, and EFSO-WtoB. All 146 000 DAS experiments are performed 20 times using different realizations of random numbers and the nature run. Bars and error bars in the panels represent average and standard deviation of the 20 simulations. Dashed lines represent the minimum, average, and maximum RMSEs of the random traces.

Inspired by N15, we investigate the changes in RMSE and ensemble spread with the different assimilation orders. All computations in this subsection are conducted under the design of experiment 1. The initial conditions are prepared by the RANDOM trace with 1460 DAS for the spinup period.

Figures 6a and 6b show changes in RMSE and ensemble spread as a function of the number of assimilated observations. Because we use the same initial conditions, all RMSEs and ensemble spreads are identical at the start of the experiment. With the EFSO-BtoW, the RMSE decreases rapidly for the initial 10 observations,
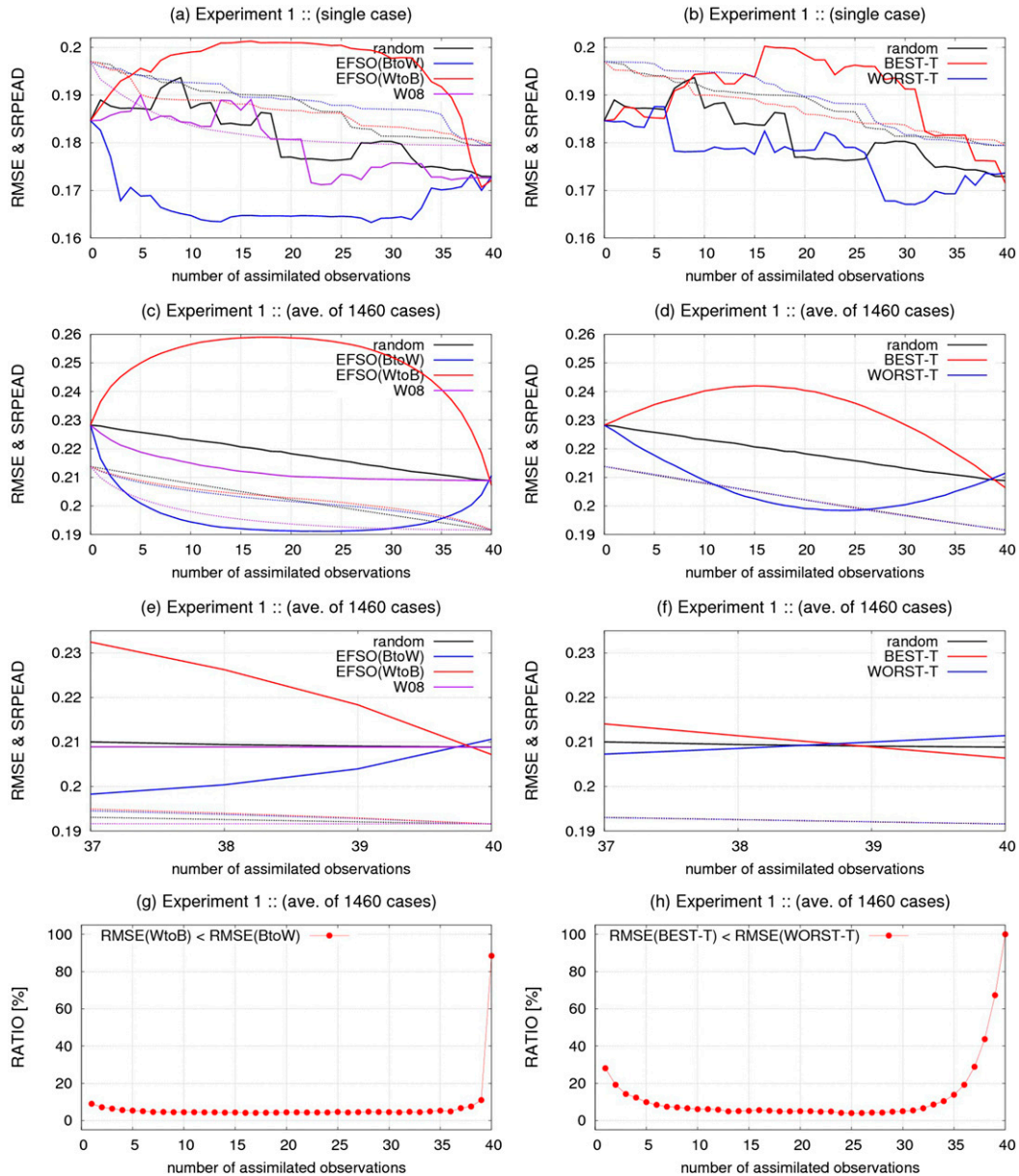
FIG. 6. Changes in RMSE and ensemble spread (solid and dashed lines) as a function of the number of assimilated observations for (a),(b) single case and (c),(d) average of 1460 cases. (e),(f) The last three observations of (c),(d). (a),(c),(e) RMSEs of the randomly selected (black), EFSO-BtoW (blue), EFSO-WtoB (red), and W08 (purple) orders. (b),(d),(f) RMSEs of the randomly selected (black), BEST-T (red), and WORST-T (blue) orders. (g),(h) The ratios at which the EFSO-WtoB is better than the ESFO-BtoW, and the BEST-T is better than the WORST-T among 1460 cases as a function of the number of assimilated observations.

and gradually increases after 28th observation (blue line in Fig. 6a). On the other hand, the RMSE increases with earlier observations, and decreases with later observations with the EFSO-WtoB. We repeat the serial data assimilations using 1460 initial conditions, and their averages are shown in Figs. 6c and 6d. The 1460 initial conditions are also prepared by the RANDOM trace

with 1460–2919 DAS for the spinup (i.e., 1460 different spinup periods). Figure 6c shows increase and decrease in RMSEs by earlier and later observations with the EFSO-WtoB. A similar curve is also observed by the BEST-T (Fig. 6d). On the other hand, the EFSO-BtoW and WORST-T orders show opposite behaviors that the RMSEs are decreased by earlier observations, and

TABLE 2. Manually optimized localization length scale σ and time window parameters (t_fcst and t_window) for BEST-FO and WORST-FO traces for ensemble sizes 4, 8, and 16.

Ensemble size m	4		8		16	
Expt	Expt 1	Expt 2	Expt 1	Expt 2	Expt 1	Expt 2
Localization scale for the serial EnSRF (grid points)	2.0	2.0	5.0	5.0	9.5	9.5
Localization scale for the LETKF (grid points)	2.0	2.0	5.0	5.0	9.0	8.5
Time window parameters for BEST-FO and WORST-FO traces (DAS)	t_fcst = 8		t_fcst = 8		t_fcst = 8	
	t_window = 8		t_window = 16		t_window = 20	

increased by later observations. The EFSO-WtoB outperforms the EFSO-BtoW at assimilation of the 40th observation, and the BEST-T trace outperforms WORST-T trace at assimilation of the 39th observation, respectively (Figs. 6e,f). It does not necessarily mean that only the 39th and 40th observations contribute toward the improvement and degradation. The differences in the final analysis are the result of accumulated impacts of the assimilated 40 observations.

Figures 6g and 6h show the ratios at which the EFSO-WtoB is better than the ESFO-BtoW, and the BEST-T is better than the WORST-T among the 1460 cases as a function of the number of assimilated observations. From the 1st to 39th observations, the EFSO-WtoB is worse than the EFSO-BtoW for more than 85% of the 1460 cases. On the other hand, the EFSO-WtoB outperforms the EFSO-BtoW after assimilating 40 observations for more than 85% of the 1460 cases, implying that the EFSO-WtoB statistically results in the better analysis than the EFSO-BtoW. This experiment assimilates the same observations from the same 1460 initial conditions. While their final analyses with the EFSO-WtoB and EFSO-BtoW show small differences in RMSE, the small differences accumulate in time and can cause significant impacts (Fig. 5). The similar curve is also observed in Fig. 6h comparing the BEST-T and WORST-T traces.

If we determine the assimilation order randomly, the RMSE reduces gradually during 40 observations (Figs. 6c,d). It implies that all observations in the serial processes are expected to reduce the error equivalently with the randomly selected order. The ensemble spread also decreases continuously with the randomly selected assimilation order. On the other hand, the W08 order shows the largest reduction in the ensemble spread by the earlier observations due to determination of the assimilation order according to the variance reduction. The W08 order also results in a larger reduction of RMSE by earlier observations compared to the randomly selected order. Interestingly, final analysis ensemble spreads after assimilating 40 observations are almost identical while their paths to the analysis are different (Figs. 6c–f). This example shows no significant difference in the analysis RMSE between the W08 and randomly selected orders. The EFSO-WtoB and BEST-T orders outperform the randomly selected order slightly.

This study obtains a superior performance for the EFSO-WtoB rule. We have not reached a mathematical explanation why the EFSO-WtoB outperforms the EFSO-BtoW. Further experiments with more realistic systems are needed to investigate whether or not the EFSO-BtoW is really a beneficial ordering rule to improve the analysis more generally. However, we experimentally demonstrate that the EFSO-WtoB order is similar to the BEST-T trace in terms of change in RMSE during the serial assimilation process. This may imply that the EFSO-WtoB order may be essential to improve the analysis for the L96 experiments.

One may wonder how the analysis accuracy would change if we assimilate only better observations using the EFSO. However, assimilating only better observations is not an easy issue. Figure 6 shows that the ensemble spread decreases nearly linearly with the number of assimilated observations while the error variances drop quickly with better observations. Therefore, if we assimilate only better observations, the ensemble spread becomes much larger than the actual error variance. We need to develop an algorithm to find a proper ensemble spread when assimilating only better observations, and these additional considerations are beyond the scope of this study.

b. Sensitivity to the ensemble size

Finally, we investigate the sensitivity of the assimilation order to the ensemble size. We perform experiments 1 and 2 with the LETKF, and serial EnSRF using RANDOM, BEST-T, WORST-T, BEST-FO, WORST-FO, W08, EFSO-BtoW, and EFSO-WtoB ordering rules by changing the ensemble sizes to $m = 4, 8$, and 16. The localization length scales and time window parameters (t_fcst and t_window) are manually optimized and are shown in Table 2. Note that the manually tuned localization scales for the serial EnSRF (Kloc) can be larger than those for the LETKF (Rloc) as pointed out by Greybush et al. (2011). Figure 7 shows the average RMSEs over 146 000 DAS in experiments 1 and 2. We repeated the 146 000
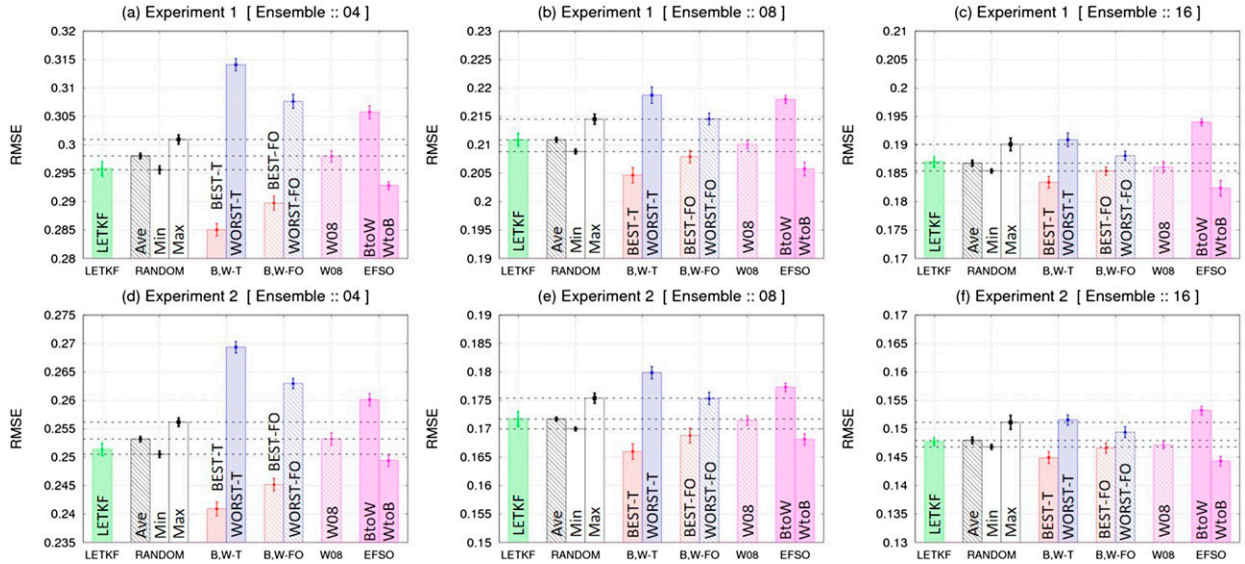
FIG. 7. As in Fig. 5, but showing the LETKF, RANDOM, BEST-T, WORST-T, BEST-FO, WORST-FO traces, W08, EFSO-BtoW, and EFSO-WtoB rules for (a)–(c) experiment 1 and (d)–(f) experiment 2. (a),(d); (b),(e); and (c),(f) show the results with ensemble sizes 4, 8, and 16, respectively.

DAS experiment 20 times with different realizations of random numbers. The standard deviations averaged over the 20 experiments are shown in Fig. 7.

With 4 and 8 ensemble members, the BEST-T and WORST-T trace shows the smallest and largest RMSEs, respectively. On the other hand, the EFSO-WtoB slightly outperforms the BEST-T trace with 16 ensemble members. Namely, EFSO found better and worse assimilation orders than the BEST-T and WORST-T traces without knowing the truth. Kalnay et al. (2012) pointed out that EFSO gives better estimations of observational impact as the ensemble size increases. With the better observational impact estimations by EFSO with 16 ensemble members, we obtain the better RMSE than the BEST-T. This implies that the EFSO-WtoB may be the best ordering rule to obtain better analysis statistically without knowing the truth for the L96 model experiments.

The BEST-FO and WORST-FO traces shows lower and higher RMSEs than the average of the RANDOM trace. The LETKF shows almost the same RMSE as the average of the RANDOM trace. The BEST-T, WORST-T, BEST-FO, and WORST-FO traces show large differences from the average of the RANDOM trace as the ensemble size decreases. This suggests that the assimilation order in the serial EnSRF have more impact on analysis accuracy when analyses are less accurate because of smaller ensemble size. On the other hand, we observe an insignificant difference between the RANDOM trace and W08 rule with four ensemble members. The improvement by the W08 may increase as the ensemble size increases.

6. Summary

In this study, we investigated the potential impact of the assimilation order of observations with the serial EnKF. We first demonstrated theoretically that the assimilation order causes different analyses in the serial treatment of observations *if* covariance localization is applied in observation space. This corroborates N15 who described that changing the assimilation order causes different analyses. With a two-dimensional problem, this study demonstrated that the impact is two orders of magnitude smaller than the analysis increments and error covariance if the two identical observation error variances are the same size as the two identical state error variances.

Based on the L96 model experiments with the serial EnSRF and LETKF, we reach the following conclusions. The experiments of the BEST-T and WORST-T traces of serial EnSRF suggest that modifying the assimilation orders of the serial EnSRF result in significant differences in analysis accuracy. Also, the LETKF shows no significant difference from the RANDOM trace of the EnSRF in our experimental design.

We improved the analysis significantly by assimilating observations from the worse to better order using the EFSO, which estimates how much each observation reduces or increases the forecast error. We experimentally demonstrated that the EFSO-WtoB order showed a similar change in RMSE to the BEST-T during the serial assimilation process. Another better ordering rule, the BEST-FO trace, is also designed using future observations to improve performance. It is based upon the premise that a better forecast is generally

produced by a better analysis, in this case a better observation order during the analysis. With an accurate EFSO, we obtained a better RMSE than the BEST-T, which finds better assimilation orders experimentally using the truth.

The ordering rules based on the BEST-FO trace can be applied practically in reanalysis experiments. On the other hand, we can apply the EFSO-WtoB rule for operational NWP forecasts. Hotta (2014) and Hotta et al. (2017) demonstrated with an operational global NWP system that we can evaluate impacts of observations by EFSO with only 6-h forecasts (cf. Fig. 1 of Hotta et al. 2017). They proposed a proactive quality control (PQC) method, which detects detrimental observations 6 h after the analysis using EFSO, and discussed that the PQC may be used in a real-time operational system. We can apply the EFSO-WtoB rule for operational NWP in the same manner. While additional experiments and verifications are needed with more realistic models, the EFSO-WtoB may be beneficial for operational data assimilation systems. Also, it would be worth investigating whether assimilating only better observations using EFSO can be used as an advanced observation thinning. On the other hand, we need additional computation to conduct the EFSO-WtoB and BEST-FO assimilation orders.

This study performed only idealized L96 experiments, and it is not trivial to determine if the findings can be directly applied to realistic geophysical data assimilation problems. However, we demonstrated a potential to improve the analysis significantly in the serial EnKF only by changing the assimilation order. In the practical geophysical reanalysis, the ensemble size is much smaller than the degrees of freedom of the dynamical system. Appli-

cation to high-dimensional, realistic systems remains a subject of future research.

APPENDIX A

Simple Two-Dimensional Problem

This appendix demonstrates detailed equations of the simple two-dimensional problem in section 2 (see Tables A1 and A2). Here we discuss three localization methods: for the background error covariance (Ploc), the Kalman gain (Kloc), and the observation error covariance (Rloc). The equations are derived manually, and also verified with a software "mathematica version 9.0."

If we assimilate the two observations without localization ($L = 1$), simultaneous assimilation and serial assimilation result in the same solutions as follows:

$$\mathbf{x}_a = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{(r_1 + v_1)(r_2 + v_2) - c^2} \left\{ \begin{array}{l} [v_1(r_2 + v_2) - c^2](y_1 - x_1) + cr_1(y_2 - x_2) \\ [v_2(r_1 + v_1) - c^2](y_2 - x_2) + cr_2(y_1 - x_1) \end{array} \right\}, \tag{A1}$$

$$\mathbf{P}^a = \frac{1}{(r_1 + v_1)(r_2 + v_2) - c^2}$$
$$\times \left\{ \begin{array}{cc} r_1[v_1(r_2 + v_2) - c^2] & cr_1 r_2 \\ cr_1 r_2 & r_2[v_2(r_1 + v_1) - c^2] \end{array} \right\}. \tag{A2}$$

If we localize $\mathbf{P}^f$ in the model space (Ploc), the simultaneous assimilation and serial assimilations in ei-

ther order result in the same Kalman gain $\mathbf{K}^a$, analysis state $\mathbf{x}^a$ and error covariance $\mathbf{P}^a$:

$$\mathbf{K}_{\text{simul}}^{a\text{Ploc}} = \frac{1}{(r_1 + v_1)(r_2 + v_2) - c^2 L^2}$$
$$\times \begin{bmatrix} v_1(r_2 + v_2) - c^2 L^2 & cLr_1 \\ cLr_2 & v_2(r_1 + v_1) - c^2 L^2 \end{bmatrix}, \tag{A3}$$

$$\mathbf{x}_{\text{simul}}^{a\text{Ploc}} = \mathbf{x}_{y_1 \to y_2}^{a\text{Ploc}} = \mathbf{x}_{y_2 \to y_1}^{a\text{Ploc}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{(r_1 + v_1)(r_2 + v_2) - c^2 L^2} \left\{ \begin{array}{l} [v_1(r_2 + v_2) - c^2 L^2](y_1 - x_1) + cLr_1(y_2 - x_2) \\ [v_2(r_1 + v_1) - c^2 L^2](y_2 - x_2) + cLr_2(y_1 - x_1) \end{array} \right\}, \tag{A4}$$

TABLE A1. The list of 13 ordering rules tested in this study.

Name	Using truth	Ordering rule
RANDOM trace		Randomly generated assimilation order
BEST-T trace	×	Best sampled trace relative to the truth
WORST-T trace	×	Worst sampled trace relative to the truth
BEST-FO trace		Best sampled trace relative to future observations
WORST-FO trace		Worst sampled trace relative to future observations
OMT-Asc	×	Ascending order of $|\mathbf{y}^o - \mathbf{Hx}^t|$
OMT-Dsc	×	Descending order of $|\mathbf{y}^o - \mathbf{Hx}^t|$
BMT-Asc	×	Ascending order of $|\mathbf{Hx}^f - \mathbf{Hx}^t|$
BMT-Dsc	×	Descending order of $|\mathbf{Hx}^f - \mathbf{Hx}^t|$
OMB-Asc		Ascending order of $|\mathbf{y}^o - \mathbf{Hx}^f|$
OMB-Dsc		Descending order of $|\mathbf{y}^o - \mathbf{Hx}^f|$
R-Asc		Ascending order of **R**
R-Dsc		Descending order of **R**
$\mathbf{HPH}^T$-Asc		Ascending order of $\mathbf{HP}^f\mathbf{H}^T$
$\mathbf{HPH}^T$-Dsc		Descending order of $\mathbf{HP}^f\mathbf{H}^T$
W08		According to error variance reduction (Whitaker et al. 2008)
EFSO-BtoW		Sorting observations from better to worse impact with EFSO
EFSO-WtoB		Sorting observations from worse to better impact with EFSO

TABLE A2. The *p* values (%) of the Wilcoxon signed rank test for each experiment relative to the RANDOM trace. Small *p* values <1% (italic numbers) represent rejection of the null hypothesis, and therefore imply there is a significant difference. Large *p* values correspond to no significant difference. Percentage values of 0.00 in the table indicate a *p* value < 0.005%.

	p values (%)	
Ordering rule	Expt 1	Expt 2
LETKF	57.87	10.60
BEST-T trace	*0.00*	*0.00*
WORST-T trace	*0.00*	*0.00*
BEST-FO trace	*0.00*	*0.00*
WORST-FO trace	*0.00*	*0.00*
OMT-Asc	*0.00*	*0.01*
OMT-Dsc	*0.00*	22.21
BMT-Asc	*0.04*	*0.30*
BMT-Dsc	2.07	*0.19*
OMB-Asc	52.51	44.47
OMB-Dsc	40.42	8.41
R-Asc	—	3.23
R-Dsc	—	3.32
$\mathbf{HPH}^T$-Asc	61.37	35.40
$\mathbf{HPH}^T$-Dsc	64.28	19.21
W08	80.35	31.30
EFSO-BtoW	*0.00*	*0.00*
EFSO-WtoB	*0.00*	*0.00*

$$\mathbf{P}_{\text{simul}}^{a\text{Ploc}} = \mathbf{P}_{y_1 \to y_2}^{a\text{Ploc}} = \mathbf{P}_{y_2 \to y_1}^{a\text{Ploc}} = \frac{1}{(r_1 + v_1)(r_2 + v_2) - c^2L^2}\left\{ \begin{matrix} r_1[v_1(r_2 + v_2) - c^2L^2] & cLr_1r_2 \\ cLr_1r_2 & r_2[v_2(r_1 + v_1) - c^2L^2] \end{matrix} \right\}. \tag{A5}$$

If we conduct serial data assimilation with Ploc, we assimilate the first observation y_1 and obtain $\mathbf{x}^a$ and $\mathbf{P}^a$:

$$\mathbf{k}_{y_1}^{\text{Ploc}} = \frac{1}{r_1 + v_1}\begin{bmatrix} v_1 \\ cL \end{bmatrix}, \tag{A6}$$

$$\mathbf{x}_{y_1}^{a\text{Ploc}} = \frac{1}{r_1 + v_1}\begin{bmatrix} r_1x_1 + v_1y_1 \\ (r_1 + v_1)x_2 + cL(y_1 - x_1) \end{bmatrix}, \tag{A7}$$

$$\mathbf{P}_{y_1}^{a\text{Ploc}} = \frac{1}{r_1 + v_1}\begin{bmatrix} v_1r_1 & cLr_1 \\ cLr_1 & v_2(r_1 + v_1) - c^2L^2 \end{bmatrix}. \tag{A8}$$

Next, these $\mathbf{x}^a$ and $\mathbf{P}^a$ are used as $\mathbf{x}^f$ and $\mathbf{P}^f$ for the second observation y_2, and we obtain the conclusive analysis $\mathbf{x}^a$ and $\mathbf{P}^a$ [Eqs. (A4) and (A5)].

The simultaneous assimilation and serial assimilations in each order result in different solutions with observation space localization. If we assimilate the two observations simultaneously with Kloc, the Kalman gain $\mathbf{K}^a$ and analysis state $\mathbf{x}^a$ are same as those with Ploc. However, the analysis error covariances $\mathbf{P}^a$ are different for Ploc and Kloc because of different background error covariances (i.e., with and without localization for $\mathbf{P}^f$; cf. section 2). The analysis error covariances $\mathbf{P}^a$ with Kloc for simultaneous assimilation is as follows:

$$\mathbf{P}_{\text{simul}}^{a\text{Kloc}} = \frac{1}{[(r_1 + v_1)(r_2 + v_2) - c^2L^2]^2}\begin{bmatrix} P_{(1,1)} & P_{(1,2)} \\ P_{(2,1)} & P_{(2,2)} \end{bmatrix}$$

$$P_{(1,1)} = r_1\left\{ c^4L^4 + v_1(r_1 + v_1)(r_2 + v_2)^2 + c^2L(r_2 + v_2)[r_1(L - 2) - 2Lv_1] \right\}$$

$$P_{(2,2)} = r_2\left\{ c^4L^4 + v_2(r_2 + v_2)(r_1 + v_1)^2 + c^2L(r_1 + v_1)[r_2(L-2) - 2Lv_2] \right\}$$

$$P_{(1,2)} = P_{(2,1)} = cr_1r_2[(r_1 + v_1)(r_2 + v_2) + c^2L^2(1 - 2L)]. \tag{A9}$$

In the serial data assimilation with Kloc, we assimilate the first observation y_1 and update $\mathbf{x}^a$ and $\mathbf{P}^a$:

$$\mathbf{k}_{y_1}^{\text{Kloc}} = \frac{1}{r_1 + v_1}\begin{bmatrix} v_1 \\ cL \end{bmatrix}, \tag{A10}$$

$$\mathbf{x}_{y_1}^{a\text{Kloc}} = \frac{1}{r_1 + v_1}\begin{bmatrix} r_1x_1 + v_1y_1 \\ (r_1 + v_1)x_2 + cL(y_1 - x_1) \end{bmatrix}, \tag{A11}$$

$$\mathbf{P}_{y_1}^{a\text{Kloc}} = \frac{1}{r_1 + v_1}\begin{bmatrix} v_1r_1 & cr_1 \\ cr_1 & Lc^2(L-2) + v_2(r_1 + v_1) \end{bmatrix}. \tag{A12}$$

Next, these $\mathbf{x}^a$ and $\mathbf{P}^a$ are used as $\mathbf{x}^f$ and $\mathbf{P}^f$ for the second observation y_2, and the conclusive analysis $\mathbf{x}^a$ and $\mathbf{P}^a$ are as follows:

$$\mathbf{k}_{y_1 \to y_2}^{a\text{Kloc}} = \frac{1}{Lc^2(L-2) + (r_1 + v_1)(r_2 + v_2)}$$
$$\times \begin{bmatrix} cLr_1 \\ Lc^2(L-2) + v_2(r_1 + v_1) \end{bmatrix}, \tag{A13}$$

$$\mathbf{x}_{y_1 \to y_2}^{a\text{Kloc}} = \frac{1}{Lc^2(L-2) + (r_1 + v_1)(r_2 + v_2)}$$
$$\times \begin{bmatrix} (r_2 + v_2)(r_1x_1 + v_1y_1) + cLr_1(y_2 - x_2) + \dfrac{Lc^2(L-2)(r_1x_1 + v_1y_1) - c^2L^2r_1(y_1 - x_1)}{(r_1 + v_1)} \\ (r_1 + v_1)(r_2x_2 + v_2y_2) + Lc^2(L-2)y_2 + cLr_2(y_1 - x_1) \end{bmatrix}, \tag{A14}$$

$$\mathbf{P}_{y_1 \to y_2}^{a\text{Kloc}} = \frac{1}{Lc^2(L-2) + (r_1 + v_1)(r_2 + v_2)}\left\{ \begin{matrix} r_1[Lc^2(L-2) + v_1(r_2 + v_2)] & cr_1r_2 \\ cr_1r_2 & r_2[Lc^2(L-2) + v_2(r_1 + v_1)] \end{matrix} \right\}. \tag{A15}$$

If we assimilate y_2 first, the conclusive analysis $\mathbf{x}^a$ is as follows:

$$\mathbf{x}_{y_2 \to y_1}^{a\text{Kloc}} = \frac{1}{Lc^2(L-2) + (r_1 + v_1)(r_2 + v_2)}$$
$$\times \begin{bmatrix} (r_2 + v_2)(r_1x_1 + v_1y_1) + Lc^2(L-2)y_1 + cLr_1(y_2 - x_2) \\ (r_1 + v_1)(r_2x_2 + v_2y_2) + cLr_2(y_1 - x_1) + \dfrac{Lc^2(L-2)(r_2x_2 + v_2y_2) - c^2L^2r_2(y_2 - x_2)}{(r_2 + v_2)} \end{bmatrix}. \tag{A16}$$

The analysis error covariance $\mathbf{P}_{y_2 \to y_1}^{a\text{Kloc}}$ is the same as $\mathbf{P}_{y_1 \to y_2}^{a\text{Kloc}}$.

Finally, we consider Rloc of the LETKF. Rloc multiplies $1/L$ to elements of the observation error covariance matrix (Greybush et al. 2011). The LETKF computes the local Kalman gains ($\mathbf{K}_{x1}^{\text{RLoc}}$ and $\mathbf{K}_{x2}^{\text{RLoc}}$) for x_1 and x_2 independently:

$$\mathbf{K}_{x1}^{\mathrm{RLoc}} = \frac{1}{(r_1 + v_1)(r_2 + v_2 L) - c^2 L} \left\{ \begin{array}{cc} v_1(r_2 + v_2 L) - c^2 L & cLr_1 \\ cr_2 & L[v_2(r_1 + v_1) - c^2] \end{array} \right\}. \tag{A17}$$

$$\mathbf{K}_{x2}^{\mathrm{RLoc}} = \frac{1}{(r_1 + v_1 L)(r_2 + v_2) - c^2 L} \left\{ \begin{array}{cc} L[v_1(r_2 + v_2) - c^2] & cr_1 \\ cr_2 L & v_2(r_1 + v_1 L) - c^2 L \end{array} \right\}. \tag{A18}$$

Using the Kalman gains, we obtain the analysis as follows:

$$\mathbf{x}_{\mathrm{simul}}^{a\mathrm{RLoc}} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \left\{ \begin{array}{c} \dfrac{[v_1(r_2 + v_2 L) - c^2 L](y_1 - x_1) + cLr_1(y_2 - x_2)}{(r_1 + v_1)(r_2 + v_2 L) - c^2 L} \\ \dfrac{[v_2(r_1 + v_1 L) - c^2 L](y_2 - x_2) + cLr_2(y_1 - x_1)}{(r_1 + v_1 L)(r_2 + v_2) - c^2 L} \end{array} \right\}, \tag{A19}$$

$$\mathbf{P}_{\mathrm{simul}}^{a\mathrm{RLoc}} = \left\{ \begin{array}{cc} \dfrac{r_1[v_1(r_2 + v_2 L) - c^2 L]}{(r_1 + v_1)(r_2 + v_2 L) - c^2 L} & - \\ - & \dfrac{r_2[v_2(r_1 + v_1 L) - c^2 L]}{(r_1 + v_1 L)(r_2 + v_2) - c^2 L} \end{array} \right\}. \tag{A20}$$

The LETKF assimilates observations simultaneously while the analyses are computed independently for each component of $\mathbf{x}$. The LETKF updates $\mathbf{P}^a$ implicitly, therefore, off-diagonal terms of $\mathbf{P}^a$ are not shown in Eq. (A20).

APPENDIX B

Other Possible Ordering Rules

This appendix describes six ordering rules and their results that are not described in the main manuscript. We tested assimilation ordering rules with ascending/descending order of $|\mathbf{y}^o - \mathbf{Hx}^f|$ (observation minus first guess; hereafter OMB), $\mathbf{R}$ (observation error variance), and $\mathbf{HP}^f\mathbf{H}^T$ (forecast error variance in the observation space). We consider that the ascending order of $|\mathbf{y}^o - \mathbf{Hx}^f|$ may result in a better analysis as does the ascending order of $|\mathbf{y}^o - \mathbf{Hx}^t|$ (Fig. 5). We select the ordering rules with $\mathbf{R}$ to investigate whether observations having smaller error should be assimilated earlier or later. We test the ordering rule with $\mathbf{HP}^f\mathbf{H}^T$ to investigate whether observations having smaller background error in the observation space should be assimilated earlier or later.

All assimilation orders are summarized in Table A1. We conducted experiment 1 and 2 for the six ordering rules and found they are not significantly different from that of the RANDOM trace.

APPENDIX C

Statistical Hypothesis Tests

To test the difference in analysis accuracy in the section 4d, statistical inference is performed with the Wilcoxon Signed-Rank Test (see our Fig. C1; Wilcoxon 1945). The Wilcoxon signed rank test is a nonparametric test for comparing paired samples. Since the RMSE has temporal correlations, 3650 samples (1 sample per 800 DAS) of the total 2 920 000 RMSEs over 2 920 000 DAS are used for the statistical inference. This way, the temporal correlations between RMSEs near in time are not included in the statistical inference. The RANDOM trace is selected arbitrarily from the 100 simulations for each 146 000 DAS. The level of the hypothesis testing is set to be 1% (or a 99% significance level). The null hypothesis is that there is no significant difference between the two experiments. Namely, the rejection of the null hypothesis corresponds to there being a significant difference between the two experiments.

We compute the p values for each experiment relative to the RANDOM trace (Table A2). The p values between the random trace and LETKF are large. Therefore, the analysis accuracy of the LETKF cannot be distinguished statistically from that of the serial EnSRF. Similarly, the ordering rules from OMB-Asc to $\mathbf{HPH}^T$-Des are not significantly different from the RANDOM trace.
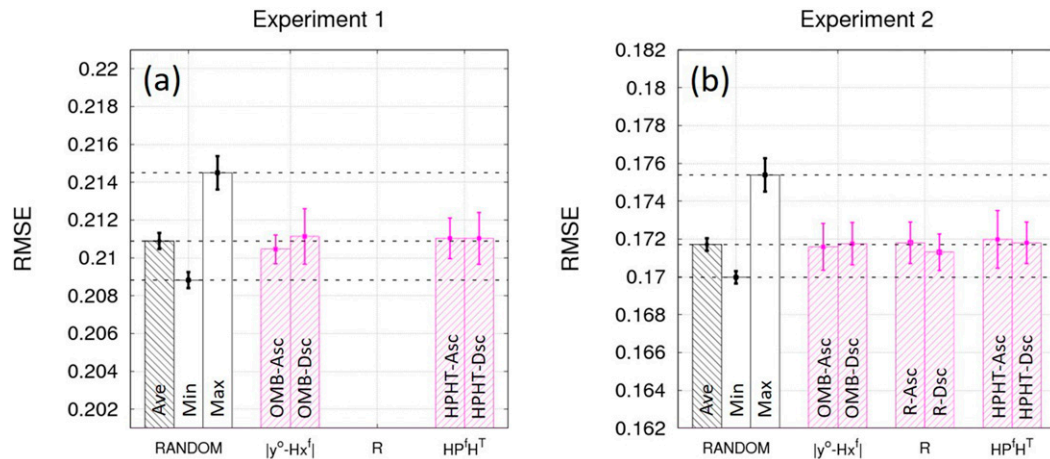
FIG. C1. As in Fig. 5, but showing the RANDOM trace, OMB-Asc, OMB-Dsc, **R**-Asc, **R**-Dsc, **HPH**[T]-Asc, and **HPH**[T]-Dsc rules for (a) experiment 1 and (b) experiment 2. No result is shown for **R**-Asc and **R**-Dsc in experiment 1 because of the spatially uniform observation error variance.

Also the analysis accuracy of the W08 cannot be distinguished statistically from the RANDOM trace. By contrast, the RMSEs of the BEST-T, WORST-T, BEST-FO, WORST-FO, EFSO-BtoW, and EFSO-WtoB are significantly different from that of the RANDOM trace at the 99.995% level. Relative to the OMT-Asc, OMT-Dsc, BMT-Asc, and BMT-Dsc, the RANDOM trace is statistically different in most of the experiments.

REFERENCES

Anderson, J., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, https://doi.org/10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.

Bishop, C. H., and D. Hodyss, 2009a: Ensemble covariances adaptively localized with ECO-RAP. Part 1: Tests on simple error models. *Tellus*, **61A**, 84–96, https://doi.org/10.1111/j.1600-0870.2008.00371.x.

——, and ——, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2:Astrategy for the atmosphere. *Tellus*, **61A**, 97–111, https://doi.org/10.1111/j.1600-0870.2008.00372.x.

——, 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, https://doi.org/10.1175/1520-0493(2001)129<0420:ASWTET>2.0.CO;2.

Burgers, G., P. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724, https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Campbell, W. F., C. H. Bishop, and D. Hodyss, 2010: Vertical covariance localization for satellite radiances in ensemble Kalman filters. *Mon. Wea. Rev.*, **138**, 282–290, https://doi.org/10.1175/2009MWR3017.1.

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, https://doi.org/10.1256/qj.05.108.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99**, 10 143–10 162, https://doi.org/10.1029/94JC00572.

Gaspari, G., and S. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, https://doi.org/10.1002/qj.49712555417.

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, https://doi.org/10.1175/2010MWR3328.1.

Hamill, T., J. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.*, **129**, 2776–2790, https://doi.org/10.1175/1520-0493(2001)129<2776:DDFOBE>2.0.CO;2.

Hotta, D., 2014: Proactive quality control based on ensemble forecast sensitivity to observations. Ph.D. dissertation, University of Maryland, 273 pp., https://doi.org/10.13016/M23W3P.

——, T.-C. Chen, E. Kalnay, Y. Ota, and T. Miyoshi, 2017: Proactive QC: A fully flow-dependent quality control scheme based on EFSO. *Mon. Wea. Rev.*, **145**, 3331–3354, https://doi.org/10.1175/MWR-D-16-0290.1.

Houtekamer, P., and H. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, **126**, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

——, and ——, 2001: A sequential ensemble Kalman filter for atmospheric data assimilation. *Mon. Wea. Rev.*, **129**, 123–137, https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Hunt, B. R., E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, **230**, 112–126, https://doi.org/10.1016/j.physd.2006.11.008.

Kalnay, E., Y. Ota, T. Miyoshi, and J. Liu, 2012: A simpler formulation of forecast sensitivity to observations: Application to ensemble Kalman filters. *Tellus*, **64A**, 18462, https://doi.org/10.3402/tellusa.v64i0.18462.

Li, H., E. Kalnay, and T. Miyoshi, 2009a: Simultaneous estimation of covariance inflation and observation errors within an

ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135**, 523–533, https://doi.org/10.1002/qj.371.

——, ——, ——, and C. M. Danforth, 2009b: Accounting for model errors in ensemble data assimilation. *Mon. Wea. Rev.*, **137**, 3407–3419, https://doi.org/10.1175/2009MWR2766.1.

Lorenz, E., 1996: Predictability—A problem partly solved. *Proc. Seminar on Predictability*, Reading, United Kingdom, ECMWF, 40–58.

——, and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, **55**, 399–414, https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2.

Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **139**, 1519–1535, https://doi.org/10.1175/2010MWR3570.1.

——, and S. Yamane, 2007: Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Mon. Wea. Rev.*, **135**, 3841–3861, https://doi.org/10.1175/2007MWR1873.1.

Nerger, L., 2015: On serial observation processing in localized ensemble Kalman filters. *Mon. Wea. Rev.*, **143**, 1554–1567, https://doi.org/10.1175/MWR-D-14-00182.1.

——, T. Janjić, J. Schröter, and W. Hiller, 2012: A regulated localization scheme for ensemble-based Kalman filters. *Quart. J. Roy. Meteor. Soc.*, **138**, 802–812, https://doi.org/10.1002/qj.945.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **56A**, 415–428, https://doi.org/10.3402/tellusa.v56i5.14462.

Tippett, M., J. Anderson, C. Bishop, T. Hamill, and J. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.*, **131**, 1485–1490, https://doi.org/10.1175/1520-0493(2003)131<1485:ESRF>2.0.CO;2.

Whitaker, J., and T. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, https://doi.org/10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.

——, ——, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP Global Forecast System. *Mon. Wea. Rev.*, **136**, 463–482, https://doi.org/10.1175/2007MWR2018.1.

Wilcoxon, F., 1945: Individual comparisons by ranking methods. *Biom. Bull.*, **1**, 80–83, https://doi.org/10.2307/3001968.