

EFSR: Ensemble Forecast Sensitivity to Observation Error Covariance

DAISUKE HOTTA

University of Maryland, College Park, College Park, Maryland, and Meteorological Research Institute, Tsukuba, and Japan Meteorological Agency, Tokyo, Japan

EUGENIA KALNAY

University of Maryland, College Park, College Park, Maryland

YOICHIRO OTA

Japan Meteorological Agency, Tokyo, Japan

TAKEMASA MIYOSHI

University of Maryland, College Park, College Park, Maryland, and RIKEN Advanced Institute for Computational Science, Kobe, Japan

(Manuscript received 2 May 2017, in final form 20 September 2017)

ABSTRACT


Data assimilation (DA) methods require an estimate of observation error covariance \mathbf{R} as an external parameter that typically is tuned in a subjective manner. To facilitate objective and systematic tuning of \mathbf{R} within the context of ensemble Kalman filtering, this paper introduces a method for estimating how forecast errors would be changed by increasing or decreasing each element of \mathbf{R} , without a need for the adjoint of the model and the DA system, by combining the adjoint-based \mathbf{R} -sensitivity diagnostics presented by Daescu previously with the technique employed by Kalnay et al. to derive ensemble forecast sensitivity to observations (EFSO). The proposed method, termed EFSR, is shown to be able to detect and adaptively correct misspecified \mathbf{R} through a series of toy-model experiments using the Lorenz '96 model. It is then applied to a quasi-operational global DA system of the National Centers for Environmental Prediction to provide guidance on how to tune the \mathbf{R} . A sensitivity experiment in which the prescribed observation error variances for four selected observation types were scaled by 0.9 or 1.1 following the EFSR guidance, however, resulted in forecast improvement that is not statistically significant. This can be explained by the smallness of the perturbation given to the \mathbf{R} . An iterative online approach to improve on this limitation is proposed. Nevertheless, the sensitivity experiment did show that the EFSO impacts from each observation type were increased by the EFSR-guided tuning of \mathbf{R} .

1. Introduction

Data assimilation (DA) methods produce the best estimate of the current state of a dynamical system by combining the background and observations with an “optimal” weight. The optimal weight, denoted \mathbf{K} , is determined, implicitly (in variational methods) or explicitly [in ensemble Kalman filters (EnKFs)], based on

the background- and observation-error covariances, denoted, respectively, by \mathbf{B} and \mathbf{R} . An accurate specification of \mathbf{B} and \mathbf{R} is of vital importance, and several methodologies [e.g., EnKFs, ensemble variational methods (EnVar), or ensemble-variational hybrid methods] have been developed that allow us to use an adaptively estimated flow-dependent \mathbf{B} . However, in most DA methods that are in current use, \mathbf{R} remains an external parameter that needs to be prescribed a priori and thus is subject to empirical tuning. In this paper we focus on how to determine \mathbf{R} .

Reflecting the crucial importance of correctly specifying \mathbf{R} , a considerable amount of effort has been put forth over the decades toward accurately modeling it. As

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

Corresponding author: Daisuke Hotta, dhotta@mri-jma.go.jp

reviewed in detail by [Buehner \(2010\)](#), most previously proposed methods use the statistics of observation-minus-background departures ($O - B$), which contain contributions from both \mathbf{B} and \mathbf{R} , and separate \mathbf{R} from \mathbf{B} under some additional assumptions. For example, [Hollingsworth and Lönnberg's \(1986\)](#) method, the first of this type, assumes the diagonality of \mathbf{R} ; the so-called Desroziers method, one of the most popular methods of this kind ([Desroziers et al. 2005](#)), assumes optimality of the DA system (i.e., perfectly prescribed \mathbf{B} and \mathbf{R} and the perfectly computed Kalman gain \mathbf{K}) and uses observation-minus-analysis residuals ($O - A$) in addition to $O - B$ to check the optimality of the currently prescribed covariances. These approaches have been applied to many systems and data and have proven to be useful, but each has its limitations because no single assumption is applicable to every situation ([Buehner 2010](#)).

Another relatively new approach, which has not been covered by [Buehner \(2010\)](#), originates from the adjoint sensitivity studies. The ground-breaking work by [Langland and Baker \(2004\)](#) introduced forecast sensitivity to observations (FSO), a technique that allows us to estimate how much each of the assimilated observations reduced or increased the forecast errors measured with some quadratic norm, without having to perform the expensive observing system experiments (OSEs). [Daescu \(2008\)](#) generalized the FSO technique and gave a formulation for forecast sensitivity to the background error covariance (\mathbf{B} sensitivity) and the observation error covariance (\mathbf{R} sensitivity). In parallel to FSO, [Daescu's \(2008\)](#) method allows us to estimate how much the forecast would be improved or degraded by adding small perturbations to each component of \mathbf{B} or \mathbf{R} without actually repeating DA experiments with different sets of \mathbf{B} and \mathbf{R} . [Daescu's R-sensitivity](#) method has been successfully applied to the global NWP systems of several operational centers including the National Aeronautics and Space Administration (NASA; [Daescu and Todling 2010](#)), the Japan Meteorological Agency (JMA; [Ishibashi 2010](#)), the Naval Research Laboratory (NRL; [Daescu and Langland 2013](#)), and the European Centre for Medium-Range Weather Forecasts (ECMWF; [Cardinali and Healy 2014](#)).

While being a powerful diagnostic tool, the applicability of the adjoint sensitivity methods such as FSO and [Daescu's](#) methods had been somewhat limited because these approaches require the adjoint of the forecast model, which is difficult to develop and/or maintain. For FSO, this limitation has been recently resolved by adapting it to EnKF ([Liu and Kalnay 2008](#); [Li et al. 2010](#); [Kalnay et al. 2012](#)). The most recent formulation of the ensemble-based FSO, or EFSO, proposed by [Kalnay](#)

[et al. \(2012\)](#), has been successfully implemented into a quasi-operational global EnKF system of the National Centers for Environmental Prediction (NCEP; [Ota et al. 2013](#)), a German convective-scale regional EnKF DA system ([Sommer and Weissmann 2014](#)), and JMA's global DA system.

The objective of this paper is to show that it is possible, by combining the derivations of EFSO in [Kalnay et al. \(2012\)](#) and the \mathbf{R} sensitivity in [Daescu \(2008\)](#), to formulate an ensemble version of forecast sensitivity to observation error covariance \mathbf{R} . We refer to these sensitivity diagnostics as the ensemble forecast sensitivity to \mathbf{R} (EFSR). This paper is structured as follows. [Section 2](#) derives the formulation of EFSR. [Section 3](#) presents the setup and the results of idealized experiments with a simple toy system that are designed to verify the validity of EFSR formulation. [Section 4](#) briefly describes the setup of realistic experiments using a lower-resolution version of the NCEP's global hybrid DA system. [Section 5](#) presents the results of EFSR diagnostics on this system and an \mathbf{R} -sensitivity experiment guided by the EFSR results. [Section 6](#) concludes the paper with an outlook on future directions.

2. EFSR formulation

In this section we introduce our EFSR formulation, building upon the derivations of [Daescu \(2008\)](#) and [Kalnay et al. \(2012\)](#).

a. Forecast sensitivity to each element of \mathbf{R}

1) GENERAL FORMULATION

Consider a DA problem at time $t = 0$. Our goal is to derive an expression for how the scalar error $e_{t|0}^f$ of the t -hour forecast would change by small variations in the observation error covariance matrix from \mathbf{R} to $\mathbf{R} + \mathbf{R}'$. We measure the forecast error with a quadratic norm:

$$e_{t|0}^f = \mathbf{e}_{t|0}^T \mathbf{C} \mathbf{e}_{t|0}, \quad (1)$$

with

$$\mathbf{e}_{t|0} = \mathbf{x}_{t|0}^f - \mathbf{x}_t^v, \quad (2)$$

where $\mathbf{x}_{t|0}^f$ is the forecast valid at time t initialized at time $t = 0$, \mathbf{x}_t^v is the verifying state at time t , and \mathbf{C} is a square positive-definite matrix that defines the error norm, which is discussed later. With FSR our interest is in quantifying the infinitesimal change to the scalar error aspect $e_{t|0}^f$ of the forecast $\mathbf{x}_{t|0}^f$ initialized with the analysis \mathbf{x}_0^a that would result from an infinitesimally small perturbation to \mathbf{R} , unlike FSO, which estimates the finite-amplitude

difference in the forecast error $e_{t|0}^f - e_{t|6}^f$ caused by the assimilation of observations at time 0. Accordingly, FSR formulation only involves analysis trajectory $\mathbf{x}_{t|0}^f$, in contrast to FSO formulation, which involves both analysis and background trajectories: $\mathbf{x}_{t|0}^f$ and $\mathbf{x}_{t|6}^f$.

Daescu (2008) showed that the sensitivity of $e_{t|0}^f$ with respect to the (i, j) element of \mathbf{R} can be expressed as

$$\frac{\partial e_{t|0}^f}{\partial R_{i,j}} = -(\mathbf{R}^{-1}\delta\mathbf{y}^{oa})_j \left(\frac{\partial e_{t|0}^f}{\partial \mathbf{y}^o} \right)_i, \quad (3)$$

with

$$\frac{\partial e_{t|0}^f}{\partial \mathbf{y}^o} = 2\mathbf{K}^T \mathbf{M}_{t|0}^T \mathbf{C} \mathbf{e}_{t|0}, \quad (4)$$

where $\delta\mathbf{y}^{oa} = \mathbf{y}^o - H(\mathbf{x}_0^o)$ is the $O - A$ residual, with $H(\cdot)$ denoting the observation operator, \mathbf{x}_0^o the analysis model state, and \mathbf{y}^o the observations, all valid at time 0; $\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}$ is the gain matrix with \mathbf{H} being the Jacobian of H linearized around the background model state $\mathbf{x}_0^b = \mathbf{x}_{0|6}^f$ valid at time 0; and $\mathbf{M}_{t|0}^T$ is the adjoint of the tangent linear forecast model from time 0 to t linearized around the analysis trajectory.

2) ADJOINT-BASED EVALUATION OF EQ. (4) WITHIN A 4D-VAR

In an operational system, the adjoint evaluation of \mathbf{K}^T in Eq. (4) is not straightforward since \mathbf{K} is extremely large (typically on the order of $\sim 10^9 \times 10^6$ elements as of 2017), so that it can never be explicitly stored on memory. Also, given the complexity of the DA code, writing its adjoint line by line, as was done by Zhu and Gelaro (2008), is a demanding task. Within the context of FSO calculations, a practical algorithm has been proposed that multiplies a vector by \mathbf{K}^T using the existing DA code without explicitly writing its adjoint (e.g., Trémolet 2008; Cardinali 2009), and we follow this approach in our AFSR calculations. The essence of this algorithm is to exploit the capacity of 4D-Var to implicitly evaluate the multiplication of a vector by the analysis error covariance matrix \mathbf{A} : in an optimal analysis, the Kalman gain matrix can be expressed as $\mathbf{K} = \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}$, so that the analysis equation becomes

$$\mathbf{x}_0^a - \mathbf{x}_0^b =: \delta\mathbf{x}^{ab} = \mathbf{K}\delta\mathbf{y}^{ob} = \mathbf{A}(\mathbf{H}^T\mathbf{R}^{-1}\delta\mathbf{y}^{ob}), \quad (5)$$

where $\delta\mathbf{y}^{ob} = \mathbf{y}^o - H(\mathbf{x}_0^b)$ is the $O - B$ innovation and $\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$ is the analysis error covariance matrix that is necessarily symmetric. Thus, the 4D-Var algorithm, which solves the analysis equation, Eq. (5), can be viewed as an algorithm that, given the input vector $\mathbf{v} \equiv \mathbf{H}^T\mathbf{R}^{-1}\delta\mathbf{y}^{ob}$, multiplies it with the matrix \mathbf{A}

and outputs $\mathbf{A}\mathbf{v}$. Then, by applying the same expression $\mathbf{K} = \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}$ to Eq. (4), and noting that \mathbf{A} and \mathbf{R}^{-1} are symmetric, we have

$$\begin{aligned} \frac{\partial e_{t|0}^f}{\partial \mathbf{y}^o} &= 2\mathbf{K}^T \mathbf{M}_{t|0}^T \mathbf{C} \mathbf{e}_{t|0} = 2(\mathbf{A}\mathbf{H}^T\mathbf{R}^{-1})^T \mathbf{M}_{t|0}^T \mathbf{C} \mathbf{e}_{t|0} \\ &= (\mathbf{R}^{-1}\mathbf{H})\mathbf{A}(\mathbf{M}_{t|0}^T 2\mathbf{C} \mathbf{e}_{t|0}). \end{aligned} \quad (6)$$

In light of Eq. (6), and recalling that the J_o term in the cost function minimized in the incremental 4D-Var algorithm can be reorganized as

$$\begin{aligned} J_o(\delta\mathbf{x}) &= \frac{1}{2}(\delta\mathbf{y}^{ob} - \mathbf{H}\delta\mathbf{x})^T \mathbf{R}^{-1}(\delta\mathbf{y}^{ob} - \mathbf{H}\delta\mathbf{x}) \\ &= \frac{1}{2}(\mathbf{H}\delta\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}) - \delta\mathbf{x}^T (\mathbf{H}^T \mathbf{R}^{-1} \delta\mathbf{y}^{ob}) \\ &\quad + \frac{1}{2}(\delta\mathbf{y}^{obT} \mathbf{R}^{-1} \delta\mathbf{y}^{ob}) \\ &= \frac{1}{2}(\mathbf{H}\delta\mathbf{x})^T \mathbf{R}^{-1}(\mathbf{H}\delta\mathbf{x}) - \delta\mathbf{x}^T \mathbf{v} + \text{const.}, \end{aligned} \quad (7)$$

and that its gradient is $\nabla J_o = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \delta\mathbf{x} - \mathbf{v}$, the forecast error sensitivity to observations shown in Eq. (4) can be evaluated with the following procedure: first, compute the vector $\mathbf{u} := \mathbf{M}_{t|0}^T 2\mathbf{C} \mathbf{e}_{t|0}$ by integrating the adjoint model backward from time t to 0 from the ‘‘initial’’ conditions $\delta\mathbf{x} = 2\mathbf{C} \mathbf{e}_{t|0}$; then, ingest the vector \mathbf{u} into the 4D-Var algorithm Eq. (7) in place of $\mathbf{v} (= \mathbf{H}^T \mathbf{R}^{-1} \delta\mathbf{y}^{ob})$ in evaluating the J_o term and its gradient $\nabla J_o = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \delta\mathbf{x} - \mathbf{v}$. The resultant output is $\mathbf{A}\mathbf{u} = \mathbf{A}\mathbf{M}_{t|0}^T 2\mathbf{C} \mathbf{e}_{t|0}$. Finally, we compute the sensitivity vector $\partial e_{t|0}^f / \partial \mathbf{y}^o$ by applying \mathbf{H} and multiplying it with \mathbf{R}^{-1} . Note that in general the analysis is not optimal, so that \mathbf{A} is only an approximation of the analysis error covariance.

3) ENSEMBLE-BASED IMPLEMENTATION (EFSR)

Now, we proceed to derive an ensemble equivalent of Eqs. (3) and (4). The essential part of the derivation of EFSO by Kalnay et al. (2012) is to exploit the fact that, in EnKF, the Kalman gain \mathbf{K} is approximated by

$$\mathbf{K} = \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1} \approx \frac{1}{K-1}(\mathbf{X}^a \mathbf{X}^{aT})\mathbf{H}^T\mathbf{R}^{-1} = \frac{1}{K-1}\mathbf{X}^a \mathbf{Y}^{aT} \mathbf{R}^{-1}, \quad (8)$$

where K is the ensemble size, $\mathbf{X}^a = [\mathbf{x}_0^{a(1)} - \bar{\mathbf{x}}_0^a, \dots, \mathbf{x}_0^{a(K)} - \bar{\mathbf{x}}_0^a]$ is the matrix of the analysis perturbations with $\mathbf{x}_0^{a(i)}$ denoting the i th member analysis and $\bar{\mathbf{x}}_0^a = 1/K \sum_{i=1}^K \mathbf{x}_0^{a(i)}$ their ensemble mean, and $\mathbf{Y}^a = \mathbf{H}\mathbf{X}^a$ is the analysis perturbations mapped onto the observation space. In practice, when the observation operator H is nonlinear, \mathbf{Y}^a can be conveniently approximated by $\mathbf{H}\mathbf{X}^a \approx [H(\mathbf{x}_0^{a(1)}) - \overline{H(\mathbf{x}_0^a)}, \dots, H(\mathbf{x}_0^{a(K)}) - \overline{H(\mathbf{x}_0^a)}]$ with

$\overline{H(\mathbf{x}_0^a)} = 1/K \sum_{i=1}^K H(\mathbf{x}_0^{a(i)})$, in which case the last equality in Eq. (8) becomes an approximation. Substituting Eq. (8) into Eqs. (3) and (4) yields

$$\begin{aligned} \frac{\partial e_{t|0}^f}{\partial R_{ij}} &= -(\mathbf{R}^{-1} \delta \mathbf{y}^{oa})_j (2\mathbf{K}^T \mathbf{M}_{t|0}^T \mathbf{C} \mathbf{e}_{t|0})_i \\ &= -(\mathbf{R}^{-1} \delta \mathbf{y}^{oa})_j \frac{2}{K-1} [\mathbf{R}^{-1} \mathbf{Y}^a (\mathbf{M}_{t|0} \mathbf{X}^a)^T \mathbf{C} \mathbf{e}_{t|0}]_i \\ &\approx -(\mathbf{R}^{-1} \delta \mathbf{y}^{oa})_j \frac{2}{K-1} (\mathbf{R}^{-1} \mathbf{Y}^a \mathbf{X}_{t|0}^{fT} \mathbf{C} \mathbf{e}_{t|0})_i, \quad (9) \end{aligned}$$

where $\mathbf{X}_{t|0}^f = [\mathbf{x}_{t|0}^{f(1)} - \bar{\mathbf{x}}_{t|0}^f, \dots, \mathbf{x}_{t|0}^{f(K)} - \bar{\mathbf{x}}_{t|0}^f]$ is the matrix of forecast perturbations initialized at time 0 and valid at time t with $\mathbf{x}_{t|0}^{f(i)}$ denoting the i th member t -hour forecast from time 0 and $\bar{\mathbf{x}}_{t|0}^f = 1/K \sum_{i=1}^K \mathbf{x}_{t|0}^{f(i)}$ their ensemble mean, and the finite-difference approximation to the tangent linear time evolution of the perturbation, $\mathbf{X}_{t|0}^f \approx \mathbf{M}_{t|0} \mathbf{X}^a$, has been used in deriving the third from the second expression. Equation (9) is the formulation of our EFSR. Note that, unlike the adjoint-based formulation shown in Eq. (3), Eq. (9) is straightforward to evaluate: as with the EFSO of Kalnay et al. (2012), all the variables necessary to evaluate EFSR are readily available from the standard product of any ensemble DA system, except that the range of the ensemble forecast has to be extended to t hours to obtain $\mathbf{X}_{t|0}^f$.

In practical situations where the ensemble size is smaller than the system's number of degrees of freedom, covariance localization is necessary to suppress sampling errors, as with any ensemble-based methods. We localize the (cross-)covariance $1/(K-1) \mathbf{Y}^a \mathbf{X}_{t|0}^{fT}$ so that Eq. (9) becomes

$$\frac{\partial e_{t|0}^f}{\partial R_{ij}} = -(\mathbf{R}^{-1} \delta \mathbf{y}^{oa})_j \frac{2}{K-1} \{ \mathbf{R}^{-1} [\boldsymbol{\rho} \circ (\mathbf{Y}^a \mathbf{X}_{t|0}^{fT})] \mathbf{C} \mathbf{e}_{t|0} \}_i, \quad (10)$$

where the symbol \circ represents elementwise multiplication (Schur product) and $\boldsymbol{\rho}$ is a localization matrix. As discussed by Kalnay et al. (2012) and Ota et al. (2013), it is desirable to take into account in the representation of the localization factor $\boldsymbol{\rho}$ the effect of evolving error correlation structure. In the idealized experiments presented in section 3, we avoid localization by employing a large ensemble size. In the experiments with the NCEP's real system, we employ the simple localization advection scheme of Ota et al. (2013). For a long lead time, more sophisticated methods that account for time-evolving error correlation, such as the ensemble correlations raised to a power (ECO-RAP) scheme of Bishop and Hodyss (2009a,b) and a group-filter technique of Gasperoni and Wang (2015), would yield a better estimation.

The computational cost required to evaluate Eq. (9) or (10) is not very expensive. Denoting the dimension of the system's state vector and the number of observations by N_{state} and N_{obs} , respectively, an explicit evaluation of Eq. (9) requires only $\sim N_{\text{state}} \times K$ operations (for multiplying the vector $\mathbf{C} \mathbf{e}_{t|0}$ by the matrix $\mathbf{X}_{t|0}^f$) and $\sim N_{\text{obs}} \times K$ operations [for multiplying the resultant $(K-1) \times 1$ vector by the matrix $\mathbf{R}^{-1} \mathbf{Y}^a$]. Assuming that \mathbf{R}^{-1} is diagonal, Eq. (10) can be evaluated for each observation (indexed by i), by first computing the contribution from the l th component of the state vector as $\rho_{il} (\mathbf{C} \mathbf{e}_{t|0})_l \sum_{k=1}^K (\mathbf{Y}^a)_{ik} (\mathbf{X}_{t|0}^f)_{lk}$ and then taking summation over l from 1 to N_{state} , which requires $\sim N_{\text{state}} \times K$ operations. This computation is repeated for $i = 1, \dots, N_{\text{obs}}$, amounting to a total of $\sim N_{\text{obs}} \times N_{\text{state}} \times K$ operations. This is more expensive than the case of Eq. (9) without localization by a factor of N_{obs} , but is still less expensive compared to the EnKF assimilation. In practice, the most expensive part of computing EFSR is generating extended-range (t hour) ensemble forecasts to obtain $\mathbf{X}_{t|0}^f$.

For convenience, we call the forecast sensitivity to observation error covariance matrix FSR (short for forecast sensitivity to \mathbf{R}), and refer to its original adjoint formulation by Daescu (2008) as AFSR and our ensemble formulation as EFSR.

We emphasize that, unlike other diagnostic methods for optimality of \mathbf{R} (e.g., Talagrand 1999; Desroziers et al. 2005), FSR diagnostics, neither the original adjoint version (Daescu 2008) nor our proposed ensemble version assume that \mathbf{B} and \mathbf{R} are correctly specified or that the observations/background are unbiased.

b. Sensitivity to scaling factors

In tuning the observation error covariance matrix \mathbf{R} , it is customary to classify observations into some subgroups among which observation error correlations can be neglected and then to scale the error covariances within each group by a single common factor. Daescu and Langland (2013) derived a formulation for forecast sensitivity to these scaling factors. Let the observations \mathbf{y}^o be partitioned into I subgroups $\{\mathbf{y}_i^o, i = 1, \dots, I\}$ and consider scaling each of them by the scaling factors $\{s_i^o, i = 1, \dots, I\}$:

$$\mathbf{R}_i \rightarrow s_i^o \mathbf{R}_i, \quad (11)$$

where \mathbf{R}_i is the subblock of \mathbf{R} corresponding to the subgroup \mathbf{y}_i^o of the observations. The scaling factors s_i^o are positive nondimensional scalars. Then, the forecast sensitivity to these scaling factors is given by

$$\frac{\partial e_{t|0}^f}{\partial s_i^o} = -[\mathbf{R}_i \boldsymbol{\Pi}_i (\mathbf{R}^{-1} \delta \mathbf{y}^{oa})]^T \boldsymbol{\Pi}_i \frac{\partial e_{t|0}^f}{\partial \mathbf{y}^o} = -(\boldsymbol{\Pi}_i \delta \mathbf{y}^{oa})^T \boldsymbol{\Pi}_i \frac{\partial e_{t|0}^f}{\partial \mathbf{y}^o}, \quad (12)$$

where we have introduced a *restriction* operator $\mathbf{\Pi}_i$ that projects the operand vector into its subvector corresponding to the subgroup \mathbf{y}_i^o of the observations. Equation (12) indicates that, for any subgroup of observations, the sensitivity of forecast errors to the scaling factor of observation error covariance for that subgroup can be computed as the $O - A$ residuals multiplied by the forecast sensitivity gradient to the corresponding observations summed up over all observations in that subgroup (with the sign flipped).

c. Sensitivity to the covariance inflation factor

It is interesting to note that from the EFSR formulation we can estimate the forecast sensitivity to the (globally constant) multiplicative background error covariance inflation factor. We consider scaling the background and the observation error covariance matrices \mathbf{B} and \mathbf{R} with the same scalar scaling factor s' :

$$\mathbf{R} \rightarrow s' \mathbf{R}, \quad \mathbf{B} \rightarrow s' \mathbf{B}. \tag{13}$$

From the expression for the Kalman gain matrix $\mathbf{K} = \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1}$, we know that \mathbf{K} does not change by this scaling. Therefore, the forecast error is also unchanged by this scaling.

Now, we consider scaling \mathbf{B} by $s^{b'}$ and the block submatrices of $\{\mathbf{R}_i, i = 1, \dots, I\}$ by $\{s_i^{o'}, i = 1, \dots, I\}$. The variation of the forecast error $e_{t|0}^f$ caused by this scaling can be written as

$$e_{t|0}^{f'} = \frac{\partial e_{t|0}^f}{\partial s^b} s^{b'} + \sum_i \frac{\partial e_{t|0}^f}{\partial s_i^{o'}} s_i^{o'}. \tag{14}$$

As indicated in the previous paragraph, if the scaling factors $s^{b'}$ and $\{s_i^{o'}, i = 1, \dots, I\}$ are all identical (denoted s'), then the resulting change in the forecast error must be zero. Thus,

$$0 = \frac{\partial e_{t|0}^f}{\partial s^b} s' + \sum_i \frac{\partial e_{t|0}^f}{\partial s_i^{o'}} s' \quad \text{and} \tag{15}$$

$$\Rightarrow \frac{\partial e_{t|0}^f}{\partial s^b} = - \sum_i \frac{\partial e_{t|0}^f}{\partial s_i^{o'}}. \tag{16}$$

Daescu and Todling (2010) presented the above equation and explained it as an intrinsic consequence of variational DA schemes. Daescu and Langland (2013) give an alternative proof of this equation by directly deriving the expression for $\partial e_{t|0}^f / \partial s^b$ in their appendix. Now, within the context of EnKF, noting that the scaling factor for the background error covariance matrix can be interpreted as a globally constant multiplicative inflation factor, we can interpret Eq. (16) as telling us that

the forecast sensitivity to the inflation factor can be estimated as the sum of the forecast sensitivity to all observation error covariance scaling factors (with the sign flipped).

3. Toy-model experiments using the Lorenz '96 model

a. Model and DA system

The model and DA system that we use are essentially identical to those used by Kalnay et al. (2012). As the forecast model, we use the Lorenz '96 model (Lorenz 1996). It is an N -dimensional ODE system defined by

$$\frac{dx_j}{dt} = x_j(x_{j+1} - x_{j-2}) - x_j + F, \quad j = 1, \dots, N \tag{17}$$

with a set of cyclic boundary conditions $x_{-1} = x_{N-1}$, $x_0 = x_N$, and $x_1 = x_{N+1}$. As in the original study by Lorenz and Emanuel (1998), we adopt $N = 40$ and $F = 8.0$. Kalnay et al. (2012) and Liu and Kalnay (2008) used different values of F for the nature run and DA cycles to simulate model errors, but here we use the same parameter $F = 8.0$ for both the nature run and the forecast (an ‘‘identical twin’’ setting). The forecast model Eq. (17) is integrated by the standard fourth-order Runge–Kutta scheme with time step $\Delta t = 0.01$.

As the DA system, we adopt the local ensemble transform Kalman filter (LETKF; Hunt et al. 2007) with member size $K = 40$. Since the member size is equal to the dimension of the state space, there is no need for covariance localization in our experiments. To avoid filter divergence, however, we applied multiplicative covariance inflation (Anderson 2001) with a constant inflation parameter $\alpha = 1.25$ (i.e., each background perturbation is inflated by a factor of 1.25) at each assimilation cycle. Although this parameter is not tuned to an optimal value, the system worked without problems. In fact, we could use Eq. (16) to tune the inflation if necessary; see section 3e for an example of such an attempt. The cycling interval (the assimilation window) is 0.05 in nondimensional time. Lorenz and Emanuel (1998) propose to dimensionalize the time by interpreting nondimensional time of 0.2 as 24 h. Hence, our cycling interval is equivalent to 6 h in dimensional time.

In our experiments, we assimilate observations available at every grid point. For the j th grid point, the observations are generated for every analysis time by adding independent Gaussian pseudorandom numbers with variance $\sigma_j^{o, \text{true}^2}$. The true observation error covariance can thus be assumed to be

TABLE 1. The true and prescribed observation error variances for the experiments performed using the Lorenz '96 model.

Expt	True observation error variance	Prescribed observation error variance
SPIKE	$\sigma_j^{o,\text{true}^2} = \begin{cases} 0.8^2, & j = 11 \\ 0.2^2, & j \neq 11 \end{cases}$	$\sigma_j^{o^2} = 0.2^2$ everywhere
STAGGERED	$\sigma_j^{o,\text{true}^2} = \begin{cases} 0.1^2, & j: \text{odd} \\ 0.3^2, & j: \text{even} \end{cases}$	$\sigma_j^{o^2} = 0.2^2$ everywhere

$$\mathbf{R}^{\text{true}} = \text{diag}(\sigma_1^{o,\text{true}^2}, \dots, \sigma_{40}^{o,\text{true}^2}). \quad (18)$$

Throughout the experiments, the observation error covariance prescribed to the DA system \mathbf{R} is also assumed to be diagonal:

$$\mathbf{R} = \text{diag}(\sigma_1^{o^2}, \dots, \sigma_{40}^{o^2}). \quad (19)$$

b. Experimental design

First, we produced the nature (or “truth”) by running the forecast model Eq. (17) from an initial condition randomly generated from the uniform distribution in $[0, 1]$. The nature run is integrated from time $t = 0$ to 730 (which corresponds to 10 yr in dimensional time), generating truth for 14 600 cycles.

The initial background ensemble at time $t = 0$ is generated by picking up 40 truth states at 40 randomly chosen distinctive dates. Each DA experiment is run for 14 600 cycles (10 yr) and the first 1460 cycles (1 yr), regarded as a spinup period, are excluded from verification.

To examine the ability of AFSR and EFSR to detect the misspecification of observation error variances $\sigma_j^{o^2}$, $j = 1, \dots, 40$, we conducted two pairs of “identical twin” experiments. Each pair consists of two DA cycle runs: one with correctly specified \mathbf{R} (i.e., identical to the truth; hereafter referred to as the correct- \mathbf{R} run) and the other with incorrectly specified \mathbf{R} (hereafter referred to as the incorrect- \mathbf{R} run). The true and prescribed observation error variances for each experiment are summarized in Table 1. For each of the experiments, we compute the sensitivity vector by both the adjoint [Eqs. (3) and (6)] and ensemble methods [Eq. (9)]. As in the DA system, no covariance localization is performed for EFSR estimations. As the forecast lead time, we adopt 24 h (0.2 in nondimensional time). For evaluating forecast errors with Eqs. (1) and (2), we use the analysis as the verifying state \mathbf{x}_t^a and the error is measured with the Euclidian norm. We do not show the AFSR results because they were nearly identical to the EFSR results in all cases. The root-mean square of their normalized differences was less than 0.2% for all experiments. We note however that it is unclear whether the high consistency between the ensemble and adjoint diagnostics,

as obtained in our idealized toy system, also holds in a more realistic system because these two approaches rely on different approximations. In particular, in the EFSR formulation, the validity of the finite-difference approximation $\mathbf{X}_{j|0}^f \approx \mathbf{M}_{j|0} \mathbf{X}^a$, and the propagation of the localization function that is only crudely accounted for, may become questionable especially when the evaluation lead time is long; similarly, in the AFSR formulation, the validity of the tangent linearity assumption for perturbation growth may become difficult to maintain as the lead time gets longer and the perturbation grows to attain a sizable finite amplitude.

c. The SPIKE experiment

The SPIKE experiment is inspired by Liu and Kalnay (2008) and Kalnay et al. (2012), who examined the capacity of EFSO to capture the negative impacts from the observations at the 11th grid point that have larger observation errors than the others. In this experiment, all observations but the one at the 11th grid point have the error variance 0.2^2 , while at the 11th grid point, it is 0.8^2 . In the incorrect- \mathbf{R} run, they are all prescribed as 0.2^2 . With this experiment, we intend to see whether the AFSR or EFSR can detect the misspecification of the error variance at the 11th grid point to provide useful guidance on how to correct it. We also examine whether the FSR diagnostics do not signal “false alarms” when the specification of \mathbf{R} is correct.

We first examine the analysis errors with respect to the truth to ensure that the system did not suffer from any malfunction (a “filter divergence” in particular). Figure 1a shows the root-mean-square errors (RMSEs) of analysis verified against the truth averaged over the 9 yr for the correct- \mathbf{R} and incorrect- \mathbf{R} runs along with the observation errors. Both runs are successful in that the analysis is more accurate than the observations, which ensured that a filter divergence did not occur. As expected, the analysis becomes substantially less accurate in the incorrect- \mathbf{R} run than in the correct- \mathbf{R} run, especially in the vicinity of the “bad” observation (the 11th grid point).

We now examine the FSR diagnosed by ensemble- and adjoint-based methods. Figure 1b shows the EFSR-based forecast sensitivity to scaling factors of \mathbf{R} [Eq. (12)] for each observation. In the incorrect- \mathbf{R} run (filled

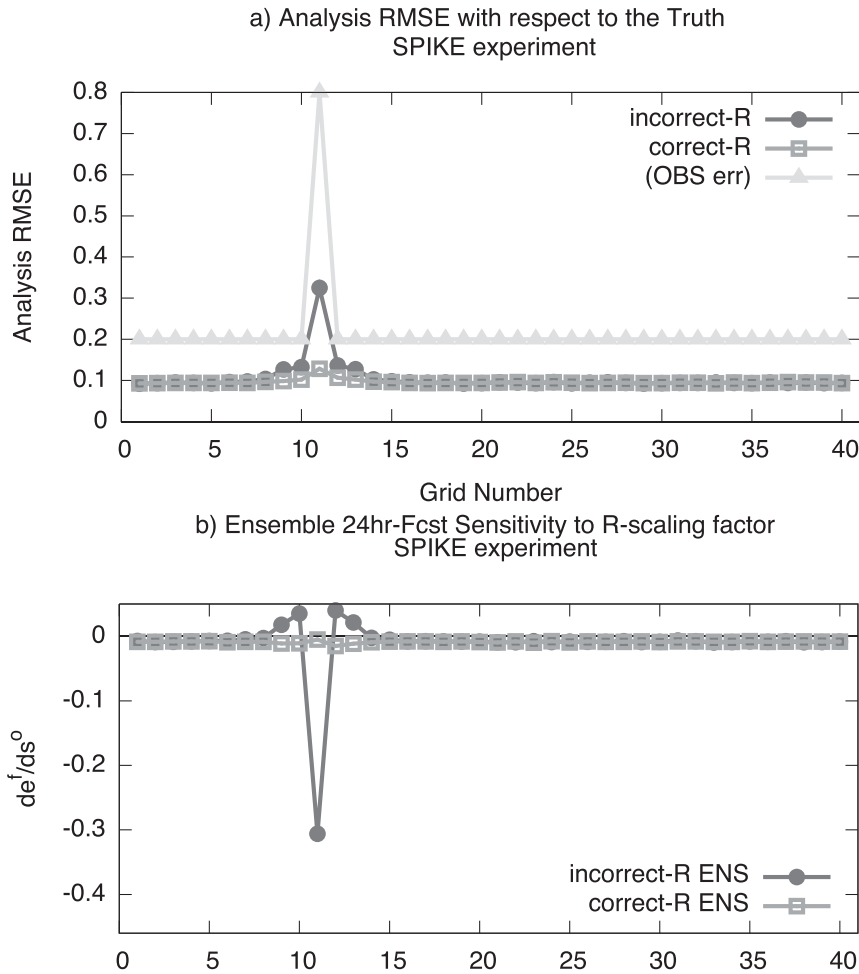


FIG. 1. (a) Analysis errors verified against the truth for the SPIKE experiment displayed as a function of grid number. The filled circles and open squares show the analysis errors, respectively, for the incorrect-**R** and correct-**R** runs. As a reference, observation errors verified against the truth are also shown by the triangles. (b) Ensemble-based 24-h forecast sensitivity to the **R**-scaling factors for the incorrect-**R** (filled circles) and correct-**R** (open squares) runs of the SPIKE experiment.

circles), EFSR successfully diagnoses large negative sensitivity at the 11th grid point where the observation error variance was intentionally made large (note that a negative sensitivity means that the forecast error would decrease by inflating the prescribed observation error variance, i.e., the current prescribed observation error variance is too small). On the other hand, for the correct-**R** run (open squares), the EFSR shows almost flat zero sensitivity, which means that there are no false alarms”.

It is interesting to note that, in the incorrect-**R** run (filled circles), despite the fact that the observation error variances for the observations near the “bad” observation are correctly specified, the EFSR diagnoses positive sensitivity, which tells us that we should decrease the

observation error variances for them. Our interpretation for this is as follows:

The sensitivity gradient $\partial e_{i|0}^f / \partial s_i^o$, being a partial derivative, tells us how, for each index i , a small displacement in s_i^o from unity would change the forecast error $e_{i|0}^f$ if the prescribed error variances for other observations are kept unchanged. Thus, if there is an observation that makes the forecast worse, then we can make the forecast better by giving higher credence to the adjacent, more accurate observations.

This raises one concern: the FSR methods may not give a reliable diagnostic if accurate and inaccurate observations are located close to each other. This concern is addressed in the next experiment.

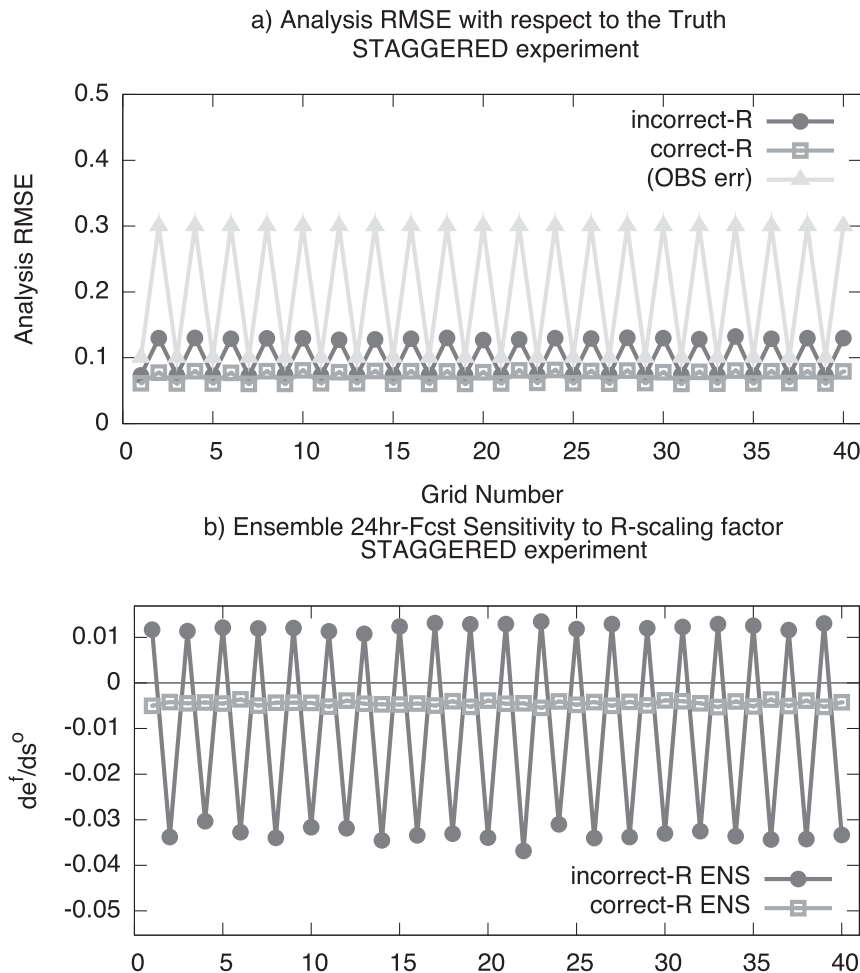


FIG. 2. As in Fig. 1, but for the STAGGERED experiment.

d. The STAGGERED experiment

The STAGGERED experiment is designed to assess whether the FSR diagnostics are robust to cases where observations with different magnitudes of error are located close to each other. The true observation error variances are 0.1^2 and 0.3^2 , respectively, for odd- and even-numbered grid points. In the incorrect-**R** run, they are all prescribed as 0.2^2 ; we should thus reduce–increase the error variances at odd–even grid points. The design of the STAGGERED experiment is very similar to the one performed by [Daescu and Todling \(2010\)](#), who sought to validate their AFSR diagnostics. The precise setup is not identical, but the incorrect-**R** run in our STAGGERED experiment and that in one of their experiments, which they named DAS-1, are similar in that the same Lorenz '96 model is used, that the accurate and less accurate observations are placed next to each other, and that the DA system prescribes a constant observation error variance to both the accurate and less accurate observations.

[Figure 2a](#) shows the analysis RMSE verified against the truth for the STAGGERED experiment. Both incorrect-**R** (filled circles) and correct-**R** (open squares) runs are successful in the sense that the analysis is more accurate than the observations. In the incorrect-**R** run, the analyses are substantially more accurate on the odd grids where observations are more accurate; on the other hand, in the correct-**R** run, the difference in the quality of analysis between the odd and even grids is much smaller. The fact that correctly specifying the observation error variances markedly improves the analysis leads us to expect that FSR diagnostics should give us correct guidance on how to tune the **R**.

The EFSR-based forecast sensitivity to scaling factors of **R** is shown in [Fig. 2b](#). Despite our concern that FSR diagnostics might not work well if accurate and inaccurate observations are located close to each other (see the previous subsection), EFSR turned out to be successful: in the incorrect-**R** run, it shows clear positive and negative sensitivity on the odd and even grids, which

indicates that we should decrease the observation error variances on the odd grids where the prescribed observation error variances (0.2^2) are larger than their actual values (0.1^2) and the opposite is true for the even grids. Similar results are reported by [Daescu and Todling \(2010\)](#) for their DAS-1 experiment (their Fig. 1b). On the other hand, in the correct- \mathbf{R} run, EFSR shows almost identical sensitivity to the odd and even grids, suggesting that we should inflate \mathbf{R} with a constant factor for all observations or, equivalently, that we should reduce the background inflation factor [Eq. (16)]. These results indicate that the FSR diagnostics are successful.

e. Adaptive online estimation of \mathbf{R} and inflation factor guided by EFSR

EFSR diagnostics allows us to estimate the gradient of the scalar forecast error $e_{t|0}^f$ with respect to the prescribed observation error covariance \mathbf{R} and the background covariance inflation factor. This gradient information can be combined with a gradient-based iterative algorithm, such as the steepest-descent method, to solve the optimization problem whose goal is to minimize $e_{t|0}^f$ by adjusting the prescribed \mathbf{R} and the

inflation factor. Naively performing such an optimization would require repeating many times the analysis and forecast for the same analysis time each time updating \mathbf{R} and the inflation factor, which would be computationally unfeasible. Alternatively, we can spread the iterations over assimilation cycles, assuming that the changes in covariance parameters that (would have) improved t -hour forecast from the analysis of t hours ago should also improve the current analysis. Such an adaptive tuning algorithm has already been proposed and proven to be successful by [Shaw and Daescu \(2017\)](#) within the context of adjoint-based sensitivity for model error bias and covariance parameters within weak-constraint 4D-Var. A difficulty in combining the sensitivity diagnostics and an iterative optimization scheme is in how best to determine the step size. A larger step size may achieve faster convergence but at the risk of overcorrection that may harm the analysis. In this sense, a smaller step size (i.e., slowly adjusting the \mathbf{R} and the inflation parameter) is safer and preferable, although convergence may be slow. Here, we explore a simple algorithm with a pre-determined constant step size that can be schematically summarized as follows:

$s_i^o := 1.0, i = 1, \dots, K$	Initialize observation error scaling factors
$\alpha := \alpha^{\text{default}} (= 1.25)$	Initialize multiplicative covariance inflation factor
do $n = 1, \dots$	Loop over cycles
$\mathbf{R} := \text{scale_oberr}(\mathbf{R}^{\text{default}}, s_i^o)$	Apply scaling to \mathbf{R} by Eq. (11)
$\mathbf{X}_{t_n t_{n-1}}^f = \alpha[\mathbf{x}_{t_n t_{n-1}}^{f(1)} - \bar{\mathbf{x}}_{t_n t_{n-1}}^f, \dots, \mathbf{x}_{t_n t_{n-1}}^{f(K)} - \bar{\mathbf{x}}_{t_n t_{n-1}}^f]$	Apply covariance inflation
$[\mathbf{x}_{t_n}^{a(1)}, \dots, \mathbf{x}_{t_n}^{a(K)}] := \text{EnKF}(\bar{\mathbf{x}}_{t_n t_{n-1}}^f, \mathbf{X}_{t_n t_{n-1}}^f, \mathbf{y}_{t_n}^o, \mathbf{R})$	Analysis update
$\mathbf{x}_{t_{n+1} t_n}^{f(i)} := M(\mathbf{x}_{t_n}^{a(i)}), i = 1, \dots, K$	Compute the first guess for the next cycle
Compute $\frac{\partial e_{t_n t_n-24}^f}{\partial s_i^o}$ for $i = 1, \dots, K$	Compute EFSR by Eqs. (12), (8), and (4) using the current analysis $\bar{\mathbf{x}}_{t_n}^u$ as the verifying truth $\bar{\mathbf{x}}_{t_n}^v$
$\frac{\partial e_{t_n t_n-24}^f}{\partial s^b} := -\sum_i \frac{\partial e_{t_n t_n-24}^f}{\partial s_i^o}$	
if $\left[\sum_i \gamma \left(\frac{\partial e_{t_n t_n-24}^f}{\partial s_i^o} \right)^2 > \epsilon e_{t_n t_n-24}^f \right]$ then	Update the scaling factors if the steepest-descent update with step-size γ is estimated to improve the scalar forecast error at least by a factor of ϵ
$s_i^o := s_i^o - \gamma \frac{\partial e_{t_n t_n-24}^f}{\partial s_i^o}$	
end if	

if [any ($s_i^o < 0$)] cancel the above update

$$\text{if } \left(\frac{\partial e_{t_n|t_n-24}^f}{\partial s^b} \times 0.1 > \varepsilon e_{t_n|t_n-24}^f \right) \alpha := \alpha \times \sqrt{0.9}$$

$$\text{if } \left(\frac{\partial e_{t_n|t_n-24}^f}{\partial s^b} \times 0.1 < -\varepsilon e_{t_n|t_n-24}^f \right) \alpha := \alpha / \sqrt{0.9}$$

if ($\alpha < 1.0$) cancel the above update

end do

But undo the above update if any of the scaling factors would become negative

Update inflation factor if increasing or decreasing it by 10% is estimated to improve the scalar forecast error at least by a factor of ε

But undo the above update if it results in covariance deflation

This algorithm does not update the observation error variance scaling factors s_i^o and the background error covariance inflation factor α , if the estimated relative forecast improvement by their update is below a chosen error level ε . This condition serves as a kind of convergence criterion and is found to be important to suppress oscillatory fluctuations (successive overshoots and undershoots) that degrade analysis performance.

We implemented the above adaptive algorithm and experimented with many choices of γ and ε , assimilating the observations that we used in the SPIKE experiment. As in the incorrect-**R** run, we initialized the prescribed observation error variance by $\sigma_j^o = 0.2^2$ for all observations and the covariance inflation factor by $\alpha = 1.25$. We run the regular DA cycles using the initial **R** and α for the first 240 cycles (60 days in dimensional time) before starting the adaptive adjustment process.

The results from the experiment with $\gamma = 0.5$ and $\varepsilon = 0.01$ are shown in Fig. 3. The estimated observation error standard deviations σ_j^o at the end of 10-yr cycles (dark triangles in Fig. 3a) successfully exhibited a spikelike pattern that peaks at the inaccurate observations at the 11th grid. The analysis RMSEs verified against the truth for this adaptive experiment (dark filled circles in Fig. 3a) are around ~ 0.03 on any of the grids, which is much smaller than ~ 0.1 (cf. Fig. 1a, lighter gray open squares) seen in the correct-**R** run of the SPIKE experiment with fixed **R** and the inflation parameter. Time series of the estimated observation error standard deviations at each grid point (Fig. 3b) show that they slowly converge to the nearly steady state shown in Fig. 3a. The covariance inflation factor α quickly decreases from 1.25 to 1.0125 within ~ 100 cycles (not shown).

Different choices of the parameters γ and ε yielded results that have different estimates of **R** and the speed of convergence, but all the tested combinations ($\gamma = 1.0, 0.5$, and 0.1 ; $\varepsilon = 0.05, 0.01$, and 0.005) were successful in that a larger observation error variance is assigned to the 11th grid than to the other grids, and in that they result in

analysis RMSEs that are smaller than in the non-adaptive correct-**R** run shown in Fig. 1a by the open squares.

4. Experiments with a real DA system: System description and experimental setup

We implemented and tested EFSR on the NCEP's quasi-operational global NWP system designed to test the new proactive quality control (Hotta et al. 2017). The system is based on the operational suite that had been operational until January 2015 but with the reduced horizontal resolutions of T254 for the deterministic runs and T126 for the ensemble (as opposed to the operational T574 and T254). In this two-way interactive hybrid DA system, the variational Gridpoint Statistical Interpolation analysis system (GSI) incorporates flow-dependent background ensemble covariance from the EnKF first-guess perturbations to produce a deterministic analysis, and the EnKF analysis ensemble is recentered on the deterministic analysis thus produced. The weights given to the static and ensemble parts of the covariance in the hybrid GSI are 25% and 75%, respectively. The EnKF part of the DA system, which, in the operational suite, is the serial ensemble square root filter (EnSRF) of Whitaker and Hamill (2002), is replaced with the LETKF. The ensemble size is 80 and both localization and inflation are applied to the covariance. We note that the localization and inflation parameters used in this experiment are tuned for the operational higher-resolution system and thus may be suboptimal for our system. Nevertheless, our system worked well and without any problems.

We remark that in a hybrid DA system, the Kalman gain assumed in EFSR and EFSO computation differs from what is actually used in the hybrid analysis. Because of this inconsistency, EFSR and EFSO may not correctly estimate the sensitivities on **R** or observations of the forecast initialized from the hybrid analysis; Hotta et al. (2017) nevertheless confirmed that the EFSO-estimated total

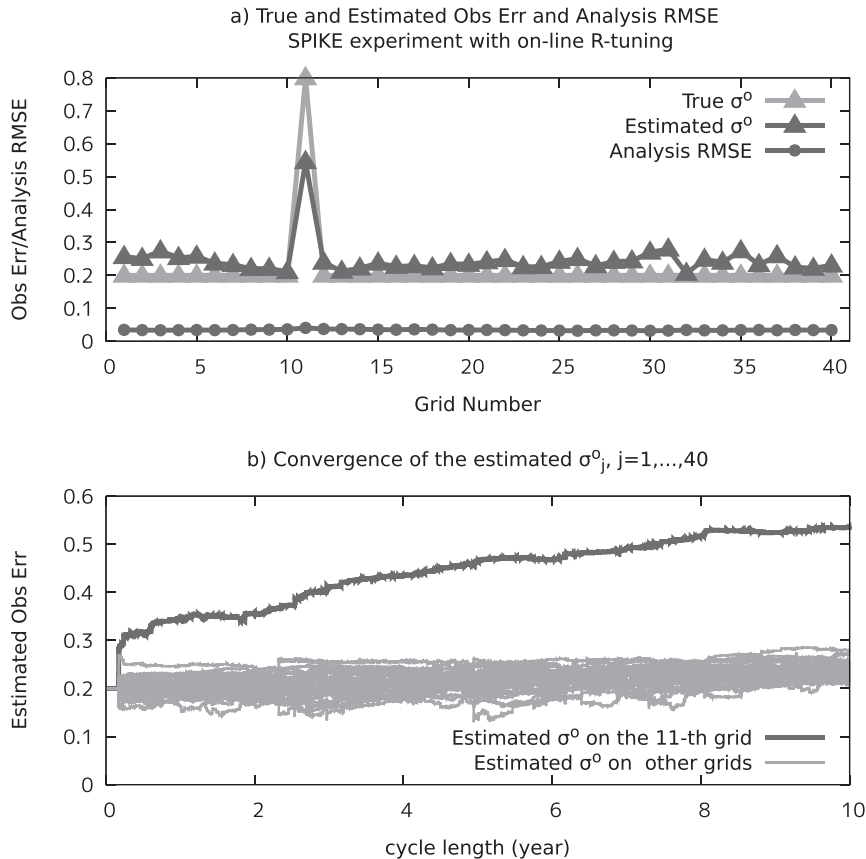


FIG. 3. Results from an adaptive online \mathbf{R} -estimation experiment with parameters $\gamma = 0.5$ and $\varepsilon = 0.01$. (a) True (triangles in lighter gray) and estimated (triangles in darker gray) observation error standard deviations, and the analysis RMSEs verified against the truth (filled circles), displayed as a function of grid number. (b) Time series of the estimated observation error standard deviations at every grid point. The thick line in darker gray that grows with time represents the observation error standard deviation at the 11th grid.

forecast error reductions were in good agreement with the actual forecast error reductions (see insets in their Fig. 1), presumably because of the large weight (75%) given to the ensemble part of the background error covariance within the variational GSI. Validity of using EFSR or EFSO within a hybrid DA system depends on its specific configuration and, as such, should be assessed on a system-to-system basis.

The experiment is performed with 6-hourly cycles for the 31-day period from 8 January to 8 February 2012, with a 7-day spinup period from 1 to 7 January 2012. All observations that were assimilated in the operational system during this period are also assimilated in our experiments. Observations are grouped into different types as in Ota et al. (2013) and Hotta et al. (2017). The EFSO impacts and EFSR-based forecast sensitivities to \mathbf{R} -scaling factors [Eq. (12)] are computed for each observation type and then averaged over the 31-day period. For both EFSO and EFSR computation, the dry

and moist total energy norms defined by Ehrendorfer et al. (1999) are used, with forecast lead times of 6 and 24 h. The precise definition of the norm is given in Eq. (9) of Ota et al. (2013). The same localization function as in the LETKF analysis is used to compute EFSO and EFSR but with the localization scheme that includes horizontal advection as proposed by Ota et al. (2013). In a hybrid DA system there are two choices for the verifying truth [\mathbf{x}_t^v in Eq. (2)]; here, the analysis from the variational part is used, but we have confirmed that the results are very similar when instead the LETKF mean analysis is used as \mathbf{x}_t^v .

5. Experiments with a real DA system: Results

Figure 4 shows the EFSR-based forecast sensitivities to \mathbf{R} -scaling factors for different observation types measured with moist or dry norm and 6- or 24-h lead time. In interpreting this figure, note that positive

Forecast sensitivity to observation error variance scaling factor

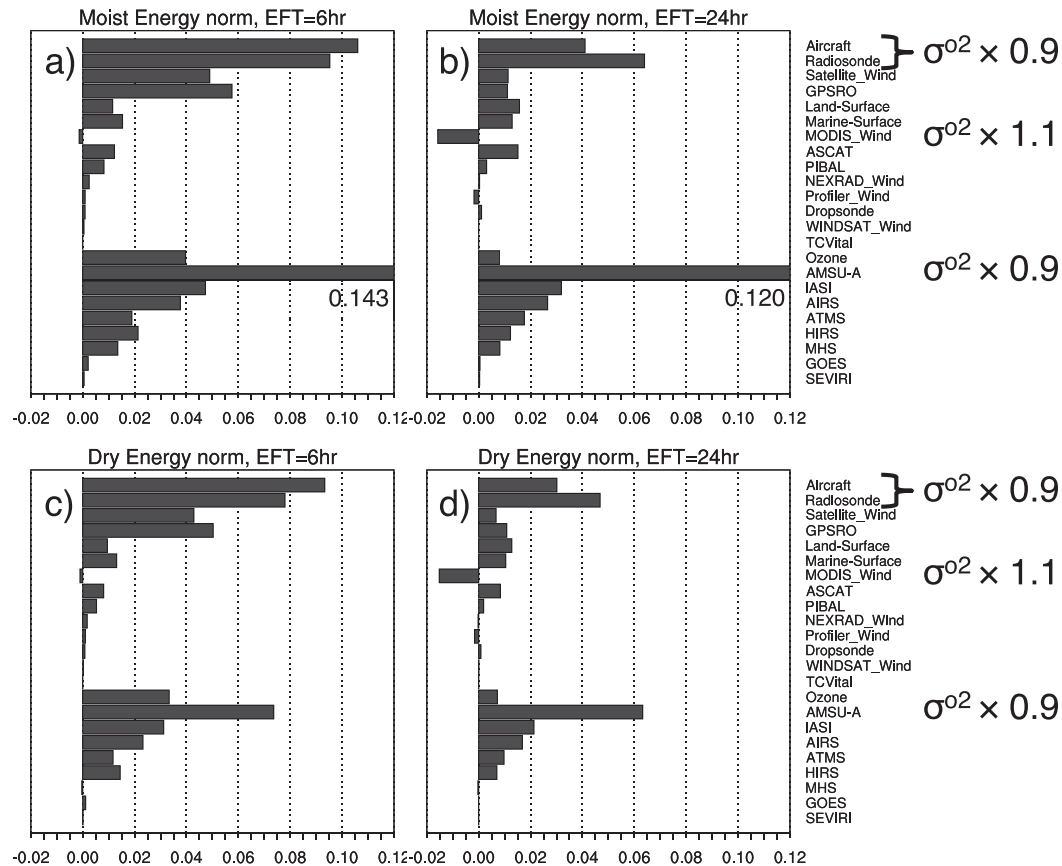


FIG. 4. Forecast sensitivity to observation error scaling factors for different types of observations measured with the (a),(b) moist or (c),(d) dry total energy norms, at (a),(c) 6- or (b),(d) 24-h lead time averaged over the 31-day period. The units are J kg^{-1} . The observation types whose error variances are scaled in the **R**-sensitivity experiment are labeled with either $\sigma^2 \times 0.9$ or $\sigma^2 \times 1.1$. The sensitivities for AMSU-A in (a) and (b) are, respectively, 0.143 and 0.120 J kg^{-1} , as indicated by the numbers below each bar.

sensitivity of an observation type means that the forecast error will increase by increasing the corresponding prescribed observation error variance and thus we should reduce the **R**. The fact that sensitivities are positive for all types except Moderate Resolution Imaging Spectroradiometer (MODIS) winds, for any combinations of the norm and lead time, suggests that the background covariance inflation factor applied in this experiment is insufficient (section 2c) or that the localization length is shorter than optimal because in that case the weights given to the observations that are away from the analyzed grid are excessively down weighted, resulting in overconfidence on the background.

We can also observe from Fig. 4 that, among all the observation types, aircraft, radiosonde, and Advanced Microwave Sounding Unit A (AMSU-A) exhibit higher sensitivities than the others, and that MODIS winds

show negative sensitivity. This feature is consistently seen in any combination of the lead times and the error norms.

To assess the validity of the EFSR diagnostics described above, we performed an **R**-sensitivity experiment in which the observation error variances of aircraft, radiosonde, and AMSU-A are reduced by a factor of 0.9 and that of MODIS winds is increased by a factor of 1.1. We chose these scaling factors, which are close to 1, because EFSR gives an estimate on how forecast errors would change by an *infinitesimally small* perturbation to the **R**. Using this new **R** both in the EnKF and the variational analysis, we reran the cycling experiment for the entire period, including the spinup, and used the last 31 days for verification.

If our EFSR diagnostics are valid, the use of the new **R** should enable us to make better use of the observations,

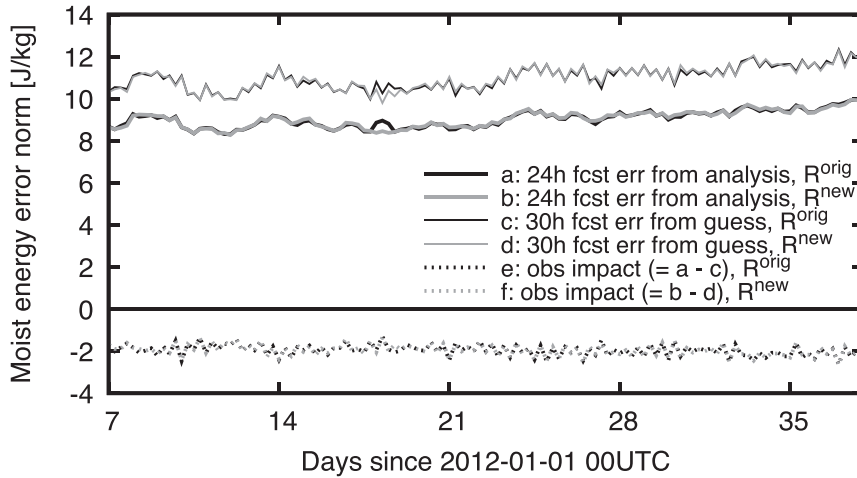


FIG. 5. Time series of (line a, thick black solid line) the error of the 24-h forecast from analysis $e_{24|0}^f$ computed for the control experiment with the original observation error covariance; (line b, thick gray solid line) as in line a, but computed for the \mathbf{R} -sensitivity experiment; (line c, thin black solid line) as in line a, but for the 30-h forecast from the analysis at the previous cycle $e_{24|-6}^f$; (line d, thin gray solid line) as in line b, but for $e_{24|-6}^f$; (line e, dotted black line) the forecast error reduction owing to the assimilation of observations for the control experiment (i.e., the difference between the lines a and c); and (line f, dotted gray line) as in line e, but for the \mathbf{R} -sensitivity experiment (i.e., the difference between the lines b and d). For all of the above, the errors are measured with the moist total energy norm.

thereby improving our analyses and forecasts. In our experiment, however, such forecast improvement by the use of the new \mathbf{R} was not detected (Fig. 5). There was a short episode of 24-h forecast improvement from 18 to 19 January [cf. the black and gray thick solid lines (lines a and b)]; this period incidentally includes the “dropouts” that Hotta et al. (2017) identified as cases 8 and 9], but the overall improvement was not statistically significant. The 1-month average of the difference of 24-h forecast errors measured with the moist total energy norm before and after the modification of \mathbf{R} (i.e., the difference between lines a and b) was only -0.008 J kg^{-1} , which is much smaller than the standard deviation 0.092 J kg^{-1} of the paired difference (which gives $p > 0.27$, the lower bound estimated by assuming that all samples are independent). Similarly, no statistically significant changes were detected for the 30-h forecast errors initialized from the first guess [black and gray thin solid lines (lines c and d)] or the forecast error reduction by the assimilation of observations [black and gray dotted lines (lines e and f)].

The fact that no statistically significant differences were found between the two cycled experiments may seem disappointing, but this was in fact an expected outcome because the estimated 24-h forecast improvement by the scaling of \mathbf{R} , computed as $\sum_i (\partial e_{24|0}^f / \partial s_i^o) \delta s_i^o$, where $i \in \{\text{aircraft, radiosonde, MODIS winds, AMSU-A}\}$, is only 0.03 J kg^{-1} , while the standard deviation of

the 24-h forecast error $e_{24|0}^f$ was 0.39 J kg^{-1} . This suggests that step sizes δs_i^o that are much larger than our choices of ± 0.1 should have been chosen in order for the forecast improvement to stand out over the natural variability, but choosing larger δs_i^o has the risk of potentially invalidating the EFSR estimation. The EFSR results computed using the data from the cycled experiment with renewed \mathbf{R} (Fig. 6) differ very little from those of the original cycle (Fig. 4b), suggesting that the renewed \mathbf{R} is still very far from nearing convergence. Further effort is necessary to demonstrate the effectiveness of the EFSR approach to improve forecasts in an operational environment, and an adaptive approach, such as the one illustrated in section 3e for the toy system, would be a promising option.

Consistent with the insignificant differences in the forecast error reductions ($\Delta e = e_{24|0}^f - e_{24|-6}^f$) before and after the modification of \mathbf{R} (lines e and f in Fig. 5), the total EFSO impacts (i.e., sum of the EFSO impacts of all the assimilated observations) resulted in no statistically significant difference from the modification of \mathbf{R} (mean and standard deviation of the paired difference being 0.0252 and 0.172 J kg^{-1} , respectively, with effective sample size at most 124, giving the lower bound of p value $p > 0.11$). Interestingly, however, the renewal of \mathbf{R} did change how the total observational impact is distributed over different types of observations. Figure 7 compares the 1-month averages of the

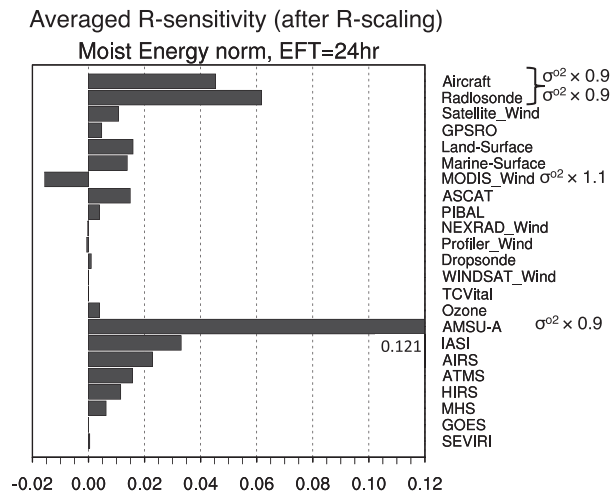


FIG. 6. As in Fig. 4b, but computed for the data from the **R**-sensitivity experiment. The sensitivity for AMSU-A was 0.121 J kg^{-1} , as indicated by the number below the corresponding bar.

EFSO impacts from each observation type before (upper bars with darker shading) and after (lower bars with lighter shading) the **R** scaling. The error bars represent the 95% confidence interval computed with Welch's *t* test for paired differences.

From Fig. 7 we can observe the following features. First, the EFSO impacts from aircraft, radiosonde, and AMSU-A all were statistically significantly increased by reducing their **R**. Second, the EFSO impact from the MODIS winds had no statistically significant increase or decrease; this is understandable given that the EFSO impact from the MODIS winds is almost neutral. Third, the EFSO impact from the Infrared Atmospheric Sounding Interferometer (IASI) is also increased, although its observation error variance was not changed; this third feature is not easy to interpret but perhaps we should not trust this result too much since its statistical significance is small compared to the increase found for aircraft, radiosonde, and AMSU-A. Finally, the EFSO impact from the Global Positioning System Radio Occultation (GPS-RO) is somewhat decreased. We remark that increases in (E)FSO impacts in general do not imply improvements in the analyses or forecasts; in fact, recent adjoint-based impact studies by Hoover and Langland (2017) and Kim et al. (2017) showed that forecast improvements tend to accompany *decreases* rather than increases in FSO impacts since the error norm of the forecast from the background (e_{t-6}^f) tends to improve (i.e., decrease) more than the error norm of the forecast from the analysis (e_{t0}^f) does.

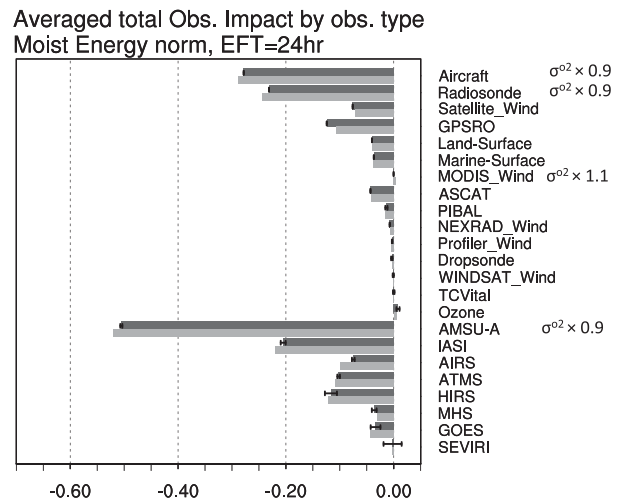


FIG. 7. Forecast sensitivity to observation impact before (top bars) and after (bottom bars) the observation error variance scaling, evaluated for different types of observations measured over the moist total energy norm and 24-h lead time averaged over the 31-day period. The error bars on the top bars represent the confidence intervals at the 95% level computed by a *t* test for the differences among the paired data (see text for details). The units are J kg^{-1} . As in Fig. 4, the observation types whose error variances are scaled in the **R**-sensitivity experiment are labeled with either $\sigma^2 \times 0.9$ or $\sigma^2 \times 1.1$.

6. Conclusions

The observation error covariance matrix **R** is an external parameter to a DA system that is often accompanied by significant uncertainty. Despite its uncertainty, the choice of **R** is known to have a significant impact on analysis and forecast accuracy and thus has been subject to extensive manual tuning by operational NWP centers. A method for objectively estimating **R** has thus been sought after, and several such methods have already been proposed. In this study, we adapted adjoint-based **R**-sensitivity diagnostics of Daescu (2008) to an EnKF using the approach of Kalnay et al. (2012), who formulated an ensemble version of FSO. Our ensemble equivalent of Daescu's method, which we call EFSR, can be implemented with only a minor code modification to EFSO and can be computed along with EFSO at a very small additional computational cost. We tested the proposed EFSR on a toy system using the Lorenz '96 model and confirmed that EFSR can effectively detect misspecifications of **R**, and that an adaptive online adjustment of **R** and the inflation factor guided by EFSR can significantly improve the estimation of **R**, thereby significantly improving the quality of analysis. We then applied EFSR to NCEP's quasi-operational global NWP system. The **R**-sensitivity

experiment guided by the EFSR diagnostics in which the \mathbf{R} for the selected four observation types are inflated or deflated by small factors (1.1 or 0.9), however, did not result in statistically significant improvements in the forecasts, which suggests that more effort is necessary to demonstrate the effectiveness of the EFSR approach in an operational environment. The neutral impact of our EFSR-guided \mathbf{R} tuning can be explained by the smallness of our chosen step size. The use of the adaptive online adjustment approach that we outlined in section 3e is an attractive alternative that we plan to explore.

The main focus of this paper was on verifying the validity of the EFSR formulation. As such, our tuning effort to improve the operational system was only preliminary. In our experiment with the NCEP's quasi-operational system, we examined the \mathbf{R} sensitivity at the highest aggregation level (grouping of observations into "nonradiance" report types and "radiance" sensors). To further tune the system, finer aggregations based, for example, on the elements/channels/altitudes for each type/sensor, should yield more informative diagnostics, as have been done by Cardinali and Healy (2014) for GPS-RO based on the altitudes and Lupu et al. (2015) for IASI based on the channels.

One limitation of FSR diagnostics is that they only suggest whether we should reduce or increase each component of \mathbf{R} and not by how much (Daescu and Langland 2013). In this connection, Li et al. (2009a) showed how the consistency diagnostics of Desroziers et al. (2005) can be adapted within the context of EnKF to estimate \mathbf{R} and the covariance inflation factor. Unlike FSR, their diagnostics have the advantage of giving us specific estimates for each component of \mathbf{R} rather than just telling us the direction, but they have the disadvantage of potentially resulting in incorrect estimates if the currently prescribed \mathbf{R} (and \mathbf{B}) values are far from being optimal. Since the two methods are based on different approaches with different assumptions and thus complement each other, we can expect to establish a more robust tuning method by combining the two. In fact, within an adjoint framework, Lupu et al. (2015) successfully improved the \mathbf{R} for IASI by combining Daescu's FSR with the Desroziers method.

In this study, we focused on the forecast sensitivity to observation error variances (i.e., diagonal elements of \mathbf{R}). However, as Eq. (9) shows, EFSR can be easily evaluated for off-diagonal elements of \mathbf{R} as well. A detailed theoretical framework for applying adjoint-based FSR has been recently worked out by Daescu and Langland (2017), and the formalism presented there should be readily applicable to EFSR as well. While

operational NWP systems traditionally assumed uncorrelated observation errors, several NWP centers have operationally implemented interchannel correlations to their \mathbf{R} models for satellite radiances (Bormann et al. 2016; Buehner et al. 2017), following recent theoretical advancements that clarified the benefits of explicitly accounting for correlations in observation errors (e.g., Miyoshi et al. 2013; Weston et al. 2014). This aspect is expected to become even more important in the "big data" era as observing networks become denser and more frequent. In this regard, our EFSR could serve as a useful diagnostic that guides us in building observation error correlation models.

Perhaps our EFSR will prove to be most useful when a new observing system is introduced to a DA system. In such a situation, in addition to assigning optimal \mathbf{R} , designing appropriate quality control (QC) criteria is of prime importance. Lien et al. (2017) proposes a simple and effective procedure to accelerate the development of optimal QC criteria for a new observing system with the aid of EFSO, and EFSR could complement this approach.

Finally, we note that an adjoint-based sensitivity methodology has been recently extended by Shaw and Daescu (2017) to weak-constraint 4D-Var formulation, allowing for the estimation of forecast sensitivity with respect to model-error bias and covariance specification; Shaw and Daescu (2017) also devised an online tuning procedure for these parameters based on the sensitivity guidance. In parallel with weak-constraint 4D-Var, EnKF formulations that explicitly account for model error biases have been proposed (e.g., Baek et al. 2006; Li et al. 2009b), and it will be an interesting future direction to explore the ensemble-based estimation of forecast sensitivity to prescribed parameters related to model errors within the framework of such model-error-aware variants of EnKF.

Acknowledgments. This work is inspired by Dr. Dacian Daescu's talk delivered in 2013 at a seminar of the Weather and Chaos Group of University of Maryland. The manuscript grew out from DH's Ph.D. dissertation, which was supported by Japan Meteorological Agency (JMA) through the Japanese Government Long-term Overseas Fellowship Program. The experiments with the NCEP's quasi-operational system are performed on the Joint Center for Satellite Data Assimilation's (JCSDA) Supercomputer for Satellite Simulations and Data Assimilation Studies [the "S4 supercomputer"; Boukabara et al. (2016)]. We express our gratitude to Dr. Sid Boukabara for his kind support and to Dr. Jim Jung for his guidance on using the Global

Forecast System (GFS) and the GSI on the S4. We thank three anonymous reviews for their insightful comments that significantly improved the manuscript. This work was partially supported by a NOAA/NESDIS/JPSS Proving Ground (PG) and a Risk Reduction (RR) CICS grant (NA14NES4320003) and a JSPS KAKENHI Grant-in-Aid for Research Activity Start-Up (17H07352).

REFERENCES

- Anderson, J. L., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **129**, 2884–2903, doi:10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2.
- Baek, S.-J., B. R. Hunt, E. Kalnay, E. Ott, and I. Szunyogh, 2006: Local ensemble Kalman filtering in the presence of model bias. *Tellus*, **58A**, 293–306, doi:10.1111/j.1600-0870.2006.00178.x.
- 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, doi:10.1111/j.1600-0870.2008.00371.x.
- , and —, 2009b: Ensemble covariances adaptively localized with ECO-RAP. Part 2: A strategy for the atmosphere. *Tellus*, **61A**, 97–111, doi:10.1111/j.1600-0870.2008.00372.x.
- Bormann, N., M. Bonavita, R. Dragani, R. Eresmaa, M. Matricardi, and A. McNally, 2016: Enhancing the impact of IASI observations through an updated observation-error covariance matrix. *Quart. J. Roy. Meteor. Soc.*, **142**, 1767–1780, doi:10.1002/qj.2774.
- Boukabara, S. A., and Coauthors, 2016: S4: An O2R/R2O infrastructure for optimizing satellite data utilization in NOAA numerical modeling systems: A step toward bridging the gap between research and operations. *Bull. Amer. Meteor. Soc.*, **97**, 2359–2378, doi:10.1175/BAMS-D-14-00188.1.
- Buehner, M., 2010: Error statistics in data assimilation: Estimation and modelling. *Data Assimilation*, W. Lahoz, B. Khattatov, and R. Menard, Eds., Springer, 93–112.
- , R. McTaggart-Cowan, and S. Heilliette, 2017: An ensemble Kalman filter for numerical weather prediction based on variational data assimilation: VarEnKF. *Mon. Wea. Rev.*, **145**, 617–635, doi:10.1175/MWR-D-16-0106.1.
- Cardinali, C., 2009: Monitoring the observation impact on the short-range forecast. *Quart. J. Roy. Meteor. Soc.*, **135**, 239–250, doi:10.1002/qj.366.
- , and S. Healy, 2014: Impact of GPS radio occultation measurements in the ECMWF system using adjoint-based diagnostics. *Quart. J. Roy. Meteor. Soc.*, **140**, 2315–2320, doi:10.1002/qj.2300.
- Daescu, D. N., 2008: On the sensitivity equations of four-dimensional variational (4D-Var) data assimilation. *Mon. Wea. Rev.*, **136**, 3050–3065, doi:10.1175/2007MWR2382.1.
- , and R. Todling, 2010: Adjoint sensitivity of the model forecast to data assimilation system error covariance parameters. *Quart. J. Roy. Meteor. Soc.*, **136**, 2000–2012, doi:10.1002/qj.693.
- , and R. H. Langland, 2013: Error covariance sensitivity and impact estimation with adjoint 4D-Var: Theoretical aspects and first applications to NAVDAS-AR. *Quart. J. Roy. Meteor. Soc.*, **139**, 226–241, doi:10.1002/qj.1943.
- , and —, 2017: Toward new applications of the adjoint sensitivity tools in data assimilation. *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, S. K. Park and L. Xu, Eds., Vol. III, Springer, 361–382.
- 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, doi:10.1256/qj.05.108.
- Ehrendorfer, M., R. M. Errico, and K. D. Raeder, 1999: Singular-vector perturbation growth in a primitive equation model with moist physics. *J. Atmos. Sci.*, **56**, 1627–1648, doi:10.1175/1520-0469(1999)056<1627:SVPGIA>2.0.CO;2.
- Gasperoni, N. A., and X. Wang, 2015: Adaptive localization for the ensemble-based observation impact estimate using regression confidence factors. *Mon. Wea. Rev.*, **143**, 1981–2000, doi:10.1175/MWR-D-14-00272.1.
- Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111–136, doi:10.3402/tellusa.v38i2.11707.
- Hoover, B. T., and R. H. Langland, 2017: Forecast and observation impact experiments in the Navy Global Environmental Model with assimilation of ECWMF analysis data in the global domain. *J. Meteor. Soc. Japan*, **95**, 2017-023, <https://doi.org/10.2151/jmsj.2017-023>.
- Hotta, D., 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, doi:10.1175/MWR-D-16-0290.1.
- 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, doi:10.1016/j.physd.2006.11.008.
- Ishibashi, T., 2010: Optimization of error covariance matrices and estimation of observation data impact in the JMA global 4D-Var system. *Research Activities in Atmospheric and Oceanic Modelling*, J. Côté, Ed., CAS/JSC Working Group on Numerical Experimentation Rep. 40, https://www.wcrp-climate.org/WGNE/BlueBook/2010/individual-articles/01_Ishibashi_Toshiyuki_ishibashi_optimization_WGNE2010_submit.pdf.
- 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, doi:10.3402/tellusa.v64i0.18462.
- Kim, M., H.-M. Kim, J.-W. Kim, S.-M. Kim, C. S. Velden, and B. T. Hoover, 2017: Effect of enhanced satellite-derived atmospheric motion vectors on numerical weather prediction in East Asia using an adjoint-based observation impact method. *Weather Forecasting*, **32**, 579–594, doi:10.1175/WAF-D-16-0061.1.
- Langland, R. H., and N. L. Baker, 2004: Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus*, **56A**, 189–201, doi:10.3402/tellusa.v56i3.14413.
- 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, doi:10.1002/qj.371.
- , —, —, and C. M. Danforth, 2009b: Accounting for model errors in ensemble data assimilation. *Mon. Wea. Rev.*, **137**, 3407–3419, doi:10.1175/2009MWR2766.1.
- , J. Liu, and E. Kalnay, 2010: Correction of ‘Estimating observation impact without adjoint model in an ensemble Kalman filter.’ *Quart. J. Roy. Meteor. Soc.*, **136**, 1652–1654, doi:10.1002/qj.658.

- Lien, G.-Y., D. Hotta, E. Kalnay, and T. Miyoshi, 2017: Accelerating assimilation development for new observing systems using EFSO. *Nonlinear Processes Geophys.*, <https://doi.org/10.5194/npg-2017-45>, in press.
- Liu, J., and E. Kalnay, 2008: Estimating observation impact without adjoint model in an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **134**, 1327–1335, doi:10.1002/qj.280.
- Lorenz, E. N., 1996: Predictability: A problem partly solved. *Proc. Seminar on Predictability*, Vol. 1, Reading, United Kingdom, ECMWF, 1–18.
- , 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](https://doi.org/10.1175/1520-0469(1998)055<0399:OSFSWO>2.0.CO;2).
- Lupu, C., C. Cardinali, and A. P. McNally, 2015: Adjoint-based forecast sensitivity applied to observation-error variance tuning. *Quart. J. Roy. Meteor. Soc.*, **141**, 3157–3165, doi:10.1002/qj.2599.
- Miyoshi, T., E. Kalnay, and H. Li, 2013: Estimating and including observation-error correlations in data assimilation. *Inverse Probl. Sci. Eng.*, **21**, 387–398, doi:10.1080/17415977.2012.712527.
- Ota, Y., J. C. Derber, T. Miyoshi, and E. Kalnay, 2013: Ensemble-based observation impact estimates using the NCEP GFS. *Tellus*, **65A**, 20038, doi:10.3402/tellusa.v65i0.20038.
- Shaw, J. A., and D. N. Daescu, 2017: Sensitivity of the model error parameter specification in weak-constraint four-dimensional variational data assimilation. *J. Comput. Phys.*, **343**, 115–129, doi:10.1016/j.jcp.2017.04.050.
- Sommer, M., and M. Weissmann, 2014: Observation impact in a convective-scale localized ensemble transform Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **140**, 2672–2679, doi:10.1002/qj.2343.
- Talagrand, O., 1999: A posteriori evaluation and verification of the analysis and assimilation algorithms. *Proc. Workshop on Diagnosis of Data Assimilation Systems*, Reading, United Kingdom, ECMWF, 17–28.
- Trémolet, Y., 2008: Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus*, **60A**, 964–978, doi:10.1111/j.1600-0870.2008.00349.x.
- Weston, P. P., W. Bell, and J. R. Eyre, 2014: Accounting for correlated error in the assimilation of high-resolution sounder data. *Quart. J. Roy. Meteor. Soc.*, **140**, 2420–2429, doi:10.1002/qj.2306.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924, doi:10.1175/1520-0493(2002)130<1913:EDAWPO>2.0.CO;2.
- Zhu, Y., and R. Gelaro, 2008: Observation sensitivity calculations using the adjoint of the Gridpoint Statistical Interpolation (GSI). *Mon. Wea. Rev.*, **136**, 335–351, doi:10.1175/MWR3525.1.