# **Observation Quality Control with a Robust Ensemble Kalman Filter**

SOOJIN ROH

Department of Statistics, Texas A&M University, College Station, Texas

MARC G. GENTON

CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

MIKYOUNG JUN

Department of Statistics, Texas A&M University, College Station, Texas

ISTVAN SZUNYOGH

Department of Atmospheric Sciences, Texas A&M University, College Station, Texas

#### **IBRAHIM HOTEIT**

CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

(Manuscript received 8 March 2013, in final form 15 July 2013)

### ABSTRACT

Current ensemble-based Kalman filter (EnKF) algorithms are not robust to gross observation errors caused by technical or human errors during the data collection process. In this paper, the authors consider two types of gross observational errors, additive statistical outliers and innovation outliers, and introduce a method to make EnKF robust to gross observation errors. Using both a one-dimensional linear system of dynamics and a 40-variable Lorenz model, the performance of the proposed robust ensemble Kalman filter (REnKF) was tested and it was found that the new approach greatly improves the performance of the filter in the presence of gross observation errors and leads to only a modest loss of accuracy with clean, outlier-free, observations.

## 1. Introduction

In data assimilation, the process of detecting and accounting for observation errors that are statistical outliers is called *quality control* (QC; e.g., Daley 1991). An operational numerical weather prediction system may employ multiple layers of QC. For instance, observations with implausible values are usually rejected even before they enter the data assimilation process. We refer to the algorithms used for such rejection decisions as *offline QC* algorithms. The fact that an observation passes the offline QC procedures does not guarantee that it is not a statistical outlier, however. For instance, an error in a highly accurate observation can be a statistical outlier

DOI: 10.1175/MWR-D-13-00091.1

when the error has a large representativeness error component. Such errors have to be dealt with by the data assimilation algorithm. We refer to the QC procedures that are part of the data assimilation algorithms as *online QC* algorithms.

Online QC algorithms detect observation errors that are statistical outliers by examining the difference between the observation and the prediction of the observation by the background. This difference is called the innovation. For instance, a simple online QC can be implemented by rejecting the observations for which the absolute value of the innovation is larger than a prescribed threshold. Another approach, which is more desirable from a theoretical point of view, is to employ *robust statistics* in the formulation of the state-update step of the data assimilation scheme (e.g., Huber 1981; Hampel 1968; Maronna et al. 2006). In particular, the presumed probability distribution of the observation

Corresponding author address: Marc G. Genton, CEMSE Division, KAUST, Thuwal 23955-6900, Saudi Arabia. E-mail: marc.genton@kaust.edu.sa

errors can be modified such that the update step can anticipate errors that would be considered statistical outliers if the observation errors were strictly Gaussian. The practical challenge posed by this approach is to find a modification of the prescribed probability distribution function, which leads to a data assimilation algorithm that can be implemented in practice.

An operational online QC algorithm using robust observation error statistics (Anderson and Järvinen 1999) was first introduced by the European Centre for Medium-Range Weather Forecasts (ECMWF). The general idea of this approach was to define the probability distribution of the observation errors as the sum of two probability distributions: a normal distribution representing the "normal" observation errors and another distribution representing the "gross" observation errors. This approach was originally proposed as an offline QC procedure by Ingleby and Lorenc (1993), but the variational framework made its integration into the data assimilation scheme possible. The formulation of the algorithm by Anderson and Järvinen (1999) became known as variational QC (Var-QC). In the latest operational version of Var-QC, called the *Huber norm QC* (Tavolato and Isaksen 2010), the probability of medium and large observation errors decreases linearly making it faster than a Gaussian distribution but slower than a uniform distribution.

A wide variety of robust filtering schemes has been proposed in the mathematical statistics literature in the past decades. In particular, Meinhold and Singpurwalla (1989) replaced the normality assumption with fat-tailed distributions such as the t distribution, whereas Naveau et al. (2005) considered a skewed version of the normal distribution. West (1981, 1983, 1984) suggested a method for robust sequential approximate Bayesian estimation. Fahrmeir and Kaufmann (1991) and Fahrmeir and Kunstler (1999) offered posterior mode estimation and penalized likelihood smoothing in robust state-space models. Kassam and Poor (1985) discussed the minimax approach for the design of robust filters for signal processing. Schick and Mitter (1994) derived a first-order approximation for the conditional prior distribution of the state. Ershov and Liptser (1978), Stockinger and Dutter (1987), Martin and Raftery (1987), Birmiwal and Shen (1993), and Birmiwal and Papantoni-Kazakos (1994) also proposed robust filtering schemes that were resistant to outliers.

Recently, Ruckdeschel (2010) proposed a robust Kalman filter in the setting of time-discrete linear Euclidean state-space models with an extension to hidden Markov models, which is optimal in the sense of minimax mean-squared errors. He used the Huberization method but investigated its performance only on a one-dimensional linear system. Luo and Hoteit (2011) employed the  $H_{\infty}$  filter to make ensemble Kalman filters (EnKF) robust enough to gross background errors. The  $H_{\infty}$  filter minimizes the maximum of a cost function different from the minimum variance used in the Kalman filter. They demonstrated their approach on both a onedimensional linear and a multidimensional nonlinear model. Calvet et al. (2012) introduced an impact function that quantified the sensitivity of the state distribution and proposed a filter with a bounded impact function.

EnKFs have been successfully implemented in highly complex operational prediction models in the atmospheric and oceanic sciences. They are Monte Carlo approximations of the traditional Kalman filter (KF; Kalman 1960) and use ensembles of forecasts to estimate the mean and covariance of the presumed normal distribution of the background. Similar to KF, EnKFs are not robust enough to gross errors in the estimate of the background mean or the observation (e.g., Schlee et al. 1967). The main goal of this paper is to design an EnKF scheme that is robust to observation errors that are statistical outliers. Harlim and Hunt (2007) and Luo and Hoteit (2011) made EnKF robust to unexpectedly large background errors. Here, we propose to make EnKF robust to gross observation errors by Huberization, a procedure that can be implemented on any EnKF scheme.

The rest of the paper is organized as follows. Section 2 first illustrates the effects of gross observation errors on the performance of EnKF; then, it describes our proposed approach to cope with such errors. Section 3 demonstrates the effectiveness of our approach for a one-dimensional linear system, while section 4 shows the results for the 40-variable Lorenz model. Finally, section 5 summarizes the main results of the paper.

### 2. A robust ensemble Kalman filter

## a. Ensemble Kalman filters

Let  $\mathbf{x}_t \in \mathbb{R}^n$  be a finite-dimensional representation of the state of the atmosphere at time *t*, and

$$\mathbf{x}_t = \mathcal{M}(\mathbf{x}_{t-1}) \tag{1}$$

be a model for the evolution of the state between discrete times with a fixed interval. For the sake of simplicity, we assume that the observations of the state are taken at discrete times, for which the model solution is available, and that the functional relationship between the state and the vector of observations,  $\mathbf{y}_t \in \mathbb{R}^p$ , at time *t* is

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \boldsymbol{\epsilon}_t. \tag{2}$$

Here,  $\mathbf{H}_t \in \mathbb{R}^{p \times n}$  is the *observation operator* and the random variable  $\boldsymbol{\epsilon}_t \in \mathbb{R}^p$  is the observation error, which is assumed to be a zero-mean (Gaussian) process with a known covariance matrix  $\mathbf{R}_t$ . The Kalman filter provides an estimate of the state  $\mathbf{x}_t$  based on the observations taken at the past and the present observation times and on the assumed knowledge (model) of the dynamics.

An EnKF algorithm assumes the availability of an *M*-member ensemble (sample),  $\{\mathbf{x}_t^{b(k)}: k = 1, ..., M\}$ , of a priori state estimates (backgrounds) with random sample errors available at time *t*. This ensemble is called the *background ensemble*. The mean of the background ensemble  $\mathbf{\bar{x}}_t^b$ , which is called the *background*, is our best estimate of the state  $\mathbf{x}_t$  before the assimilation of the observations taken at time *t*. The analysis step of an EnKF generates an *analysis ensemble*,  $\{\mathbf{x}_t^{a(k)}: k = 1, ..., M\}$ , such that its mean  $\mathbf{\bar{x}}_t^a$ , called the *analysis*, satisfies

$$\overline{\mathbf{x}}_{t}^{a} = \overline{\mathbf{x}}_{t}^{b} + \mathbf{K}_{t}(\mathbf{y}_{t} - \mathbf{H}_{t}\overline{\mathbf{x}}_{t}^{b}), \qquad (3)$$

while the ensemble-based estimate of the *analysis error* covariance matrix  $\mathbf{P}_t^a \in \mathbb{R}^{n \times n}$ , which is defined by the sample mean covariance matrix for the ensemble, satisfies either

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_t^b, \qquad (4)$$

without the perturbed observations, or

$$\mathbf{P}_t^a = (\mathbf{I} - \mathbf{K}_t \mathbf{H}_t) \mathbf{P}_t^b + O(M^{-1/2}), \qquad (5)$$

with the perturbed observations. In Eqs. (3), (4), and (5), the *Kalman gain matrix*  $\mathbf{K}_{t} \in \mathbb{R}^{n \times p}$  is given by

$$\mathbf{K}_{t} = \mathbf{P}_{t}^{b} \mathbf{H}_{t}^{\mathrm{T}} (\mathbf{H}_{t} \mathbf{P}_{t}^{b} \mathbf{H}_{t}^{\mathrm{T}} + \mathbf{R}_{t})^{-1}, \qquad (6)$$

and the ensemble-based estimate of the *background error covariance matrix*  $\mathbf{P}_t^b \in \mathbb{R}^{n \times n}$  is provided by the sample covariance matrix for the background ensemble. In our numerical experiments, we use the method of perturbed observations (Houtekamer and Mitchell 1998; Burgers et al. 1998) to obtain the analysis ensemble. In this technique,  $\mathbf{P}_t^a$  satisfies Eq. (5). The analysis process at time *t* is completed by the forecast step of the EnKF, in which the model dynamics are applied to each member of the analysis ensemble to obtain the members of the background ensemble for the next observation time, t + 1.

The components of the vector of differences  $\mathbf{y}_t - \mathbf{H}_t \mathbf{\bar{x}}_t^b$ between the observations and their predicted values are called *innovations* (each innovation describes the discrepancy between an observation and its predicted value). In addition, the components of the change in the state estimate  $\mathbf{\bar{x}}_t^a - \mathbf{\bar{x}}_t^b$ , due to the assimilation of the



FIG. 1. Plot of the true states (solid line) and the traditional EnKF (dashed line) as a function of time *t* for a one-dimensional linear system with (top) additive outliers  $\xi_t = 5$  and (bottom) innovations outliers with  $\alpha = 0.2$  and  $k_t = 25$ . The occurrences of outliers are marked with open circles.

observations included in  $\mathbf{y}_t$ , are called *analysis increments*. The role of the Kalman gain matrix  $\mathbf{K}_t$  is to map the innovations into analysis increments. According to Eq. (6), the Kalman gain accounts for the observation errors based on the prescribed error statistics included in  $\mathbf{R}_t$ . It thus has no information about the errors in a particular observation or the magnitude of a particular innovation. Since the analysis increments are unbounded functions of the innovations, a large innovation due to a gross (outlier) observation error can cause a large degradation in the accuracy of the state estimate.

### b. The effects of observation outliers

We consider two common types of observation outliers: additive outliers (AO) and innovations outliers (IO; Fox 1972; Genton 2003; Genton and Lucas 2003, 2005). In an AO model, we observe

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \boldsymbol{\xi}_t + \boldsymbol{\epsilon}_t, \tag{7}$$

where  $\boldsymbol{\xi}_t \in \mathbb{R}^p$  is a vector of unknown outlying values. It is assumed that only a few components of  $\boldsymbol{\xi}_t$  are different from zero. In an IO model, the observation error is assumed to be a contaminated multivariate Gaussian distribution

$$\boldsymbol{\epsilon}_t \sim (1-\alpha)N_p(\boldsymbol{0}, \boldsymbol{\mathsf{R}}_t) + \alpha N_p(\boldsymbol{0}, k_t \times \boldsymbol{\mathsf{R}}_t), \qquad (8)$$

where  $0 < \alpha < 1$ ,  $k_t > 1$ , and  $N_p$  denotes the *p*-variate Gaussian distribution. That is, the observation errors have a zero mean and a probability  $1 - \alpha$  of coming from a normal distribution with covariance matrix **R**<sub>t</sub> and a



FIG. 2. Bias vs efficiency for the EnKF and two REnKFs for a one-dimensional linear system for t = 30, 31, 32, 33, 34, and 35. The additive outliers with  $\xi_t = 8$  occur at times t = 31, 32, and 33.

(usually small) probability  $\alpha$  of coming from a normal distribution with higher variances  $k_t \times \mathbf{R}_t$ . The value of  $k_t$  is assumed to be unknown. The additive outlier model corresponds to a situation where some of the observations are affected by a strong observation bias, whereas the innovations outlier model corresponds to a situation where there is an  $100 \times \alpha$  percent chance that the observation error variance is larger than the prescribed value given by  $\mathbf{R}_t$ .

We illustrate the effects of the type of outlier on EnKF analyses with the help of a one-dimensional linear system

$$x_t = x_{t-1} + e_t \tag{9}$$

and the observation equation

$$y_t = x_t + \epsilon_t, \tag{10}$$

where  $e_t$  and  $\epsilon_t$  are zero-mean Gaussian processes with unit variance. The results shown in Fig. 1 were obtained by using the traditional EnKF algorithm to obtain the analysis ensemble that satisfies Eqs. (3) and (5) for Eqs. (9) and (10). We assimilate observations at every time step. The outliers occur at the times where the errors are marked by open circles. The top panel shows the results for the AO model, with  $\xi_t = 5$  for the outliers, while the bottom panel shows the results for the IO



FIG. 3. As in Fig. 2, but for bias vs radius.

model, with  $\alpha = 0.2$  and  $k_t = 25$  for the outliers. The accuracy of the state estimates are clearly degraded at the time steps where the outliers are present in either outlier model.

## c. A robust ensemble Kalman filter

The detrimental effect of the outliers on the EnKF state estimate can be reduced by decreasing the magnitude of those components of the innovation vector that have unusually large absolute values. This can be done by defining an upper bound for the allowable absolute value of the innovations. When the magnitude of an innovation is found to be larger than the prescribed upper bound, the magnitude of the innovation can be clipped at the upper bound. To be precise, the innovation  $\delta y$  is left unchanged if  $-c < \delta y < c$  for some c > 0 and clipped at -c if  $\delta y < -c$  and at *c* if  $\delta y > c$ . This componentwise clipping of the innovation is called *Huberization*, and the tunable parameter *c* is called the *clipping height*.

The *Huberized analysis*  $\hat{\mathbf{x}}^a$  can be written as

$$\hat{\mathbf{x}}_{t}^{a} = \overline{\mathbf{x}}_{t}^{b} + \mathbf{K}_{t} G_{\mathbf{c}} (\mathbf{y}_{t} - \mathbf{H}_{t} \overline{\mathbf{x}}_{t}^{b}), \qquad (11)$$

where for any  $\mathbf{c} \in \mathbb{R}^p_+$  and  $\mathbf{u} \in \mathbb{R}^p$ , the *Huber* function  $G_{\mathbf{c}}(\mathbf{u})$  is defined by (i = 1, ..., p)



FIG. 4. The true states, EnKF, and two REnKFs with efficiencies  $\delta = 0.99$  and 0.7 for a one-dimensional linear system. The additive outliers with  $\xi_t = 8$  occur at times t = 31, 32, and 33.

$$[G_{\mathbf{c}}(\mathbf{u})]_i = \begin{cases} u_i, & \text{if } |u_i| < c_i, \\ c_i, & \text{if } u_i \ge c_i, \\ -c_i, & \text{if } u_i \le -c_i. \end{cases}$$
(12)

Here,  $c_i$  and  $u_i$  are the *i*th elements of **c** and **u**, respectively. The observation is clipped componentwisely by the clipping height of the same dimension. When Huberization achieves its goal of reducing the contamination of the prescribed distribution of the observation errors, the observation error covariance matrix **R**<sub>*i*</sub> provides a better representation of the observation error covariance. Hence, we do not modify any entries of **R**<sub>*i*</sub>.

A simple alternative to Huberization for handling observation error outliers is to discard the suspect observations from the data assimilation process. In fact, this is the online QC approach that has been employed by EnKF algorithms in weather prediction models (e.g., Szunyogh et al. 2008). In the simple numerical examples given here, we discard the observation if  $|\delta y| > c$  for a prescribed c. In these applications, the prescribed smallest magnitude of the innovation that triggers a rejection of the observation depends on the magnitude of the ensemble-based estimate of the background error variance at the observation location (the related entry of  $\mathbf{H}_{t}\mathbf{P}_{t}^{b}\mathbf{H}_{t}^{T}$ ) and/or the variance of the observation error (the related diagonal element of  $\mathbf{R}_t$ ). Because this approach is based on discarding the observation rather than reducing the contamination from the observation error, the entries of  $\mathbf{R}_t$  that are related to the discarded observation must also be removed.



FIG. 5. As in Fig. 4, but for radii r = 0.0001 and 0.01 rather than efficiencies.

#### d. Choosing parameter **c**

The tunable parameter of both strategies to handle the outlier observation errors, which were described in section 2c, is the *p*-dimensional vector  $\mathbf{c} \in \mathbb{R}^p$ . An ideal choice for  $\mathbf{c}$  would remove the contamination from the observation error or lead to the rejection of the observation without making any change in the state estimates of clean, outlier-free, observations. While such an ideal choice for  $\mathbf{c}$  usually does not exist, we can define a measure of our tolerance for degradation in the accuracy of the state estimates for clean observations.

One measure of tolerance can be defined by introducing the notion of *relative efficiency*. The relative efficiency of two algorithms to estimate the state is defined by the ratio of the variance of the error in the two estimates they provide. The relative efficiency of EnKF with and without online QC,

$$\delta = \frac{E|\mathbf{x}_t - \overline{\mathbf{x}}_t^a|_{\mathrm{id}}^2}{E|\mathbf{x}_t - \hat{\mathbf{x}}_t^a|_{\mathrm{id}}^2},\tag{13}$$

falls into the interval  $\delta \in (0, 1]$ . Here,  $|\cdot|$  denotes the Euclidean norm and the subscript id indicates that the norm is to be computed for clean, outlier-free, observations. If no quality control was applied (the components of **c** were set to infinity), then the relative efficiency would be  $\delta = 1$ . Equivalently, achieving a perfect relative efficiency,  $\delta = 1$ , would require choosing  $\mathbf{c} = \infty$ . The lower the value of  $\delta$  we accept, the lower the values we can choose for the components of **c**. A common choice for the relative efficiency is  $\delta = 0.95$  (e.g., Huber 1981).



FIG. 6. As in Fig. 2, but with innovations outliers with  $k_t = 25$  occurring at times t = 31, 32, and 33 with probability of contamination  $\alpha = 0.2$ .

To use the relative efficiency as a criterion for the selection of **c**, we have to find a practical approach to computing the components of **c** for a given value of  $\delta$ . In Kalman filtering, the variance of the analysis error is usually estimated by the trace of  $\mathbf{P}_t^a$  given by Eq. (4). While this approach would provide a simple formula for the numerator in Eq. (13), the denominator could not be written with the help of Eq. (4) because  $G_{\mathbf{c}}(\mathbf{u})$  is a nonlinear function of the innovation. It cannot thus be absorbed into the Kalman gain matrix. Hence, after dropping the subscript *t* that denotes the time, the only alternative left is to substitute  $\overline{\mathbf{x}}^a$  from Eq. (3) and  $\hat{\mathbf{x}}^a$  from Eq. (11) into Eq. (13), which yields

$$\delta = \frac{E|\mathbf{x} - \overline{\mathbf{x}}^b - \mathbf{K}(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^b)|_{\mathrm{id}}^2}{E|\mathbf{x} - \overline{\mathbf{x}}^b - \mathbf{K}G_{\mathbf{c}}(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^b)|_{\mathrm{id}}^2}$$
$$= \frac{E|(\mathbf{I} - \mathbf{K}\mathbf{H})(\mathbf{x} - \overline{\mathbf{x}}^b) + \boldsymbol{\epsilon}|_{\mathrm{id}}^2}{E|(\mathbf{x} - \overline{\mathbf{x}}^b) - \mathbf{K}G_{\mathbf{c}}[\mathbf{H}(\mathbf{x} - \overline{\mathbf{x}}^b) + \boldsymbol{\epsilon}]|_{\mathrm{id}}^2}.$$
(14)

The second equality comes from the observation Eq. (2). The *i*th component  $c_i$  of the clipping height **c** is obtained by using

$$E|\mathbf{x} - \overline{\mathbf{x}}^{a}|_{\mathrm{id}}^{2} = E|\mathbf{x} - \overline{\mathbf{x}}^{b} - (\mathbf{K})_{i}(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^{b})_{i}|_{\mathrm{id}}^{2} \quad (15)$$



FIG. 7. As in Fig. 3, but with innovations outliers with  $k_t = 25$  occurring at times t = 31, 32, and 33 with a probability of contamination  $\alpha = 0.2$ .

for the computation of the numerator and

$$E|\mathbf{x} - \hat{\mathbf{x}}^{a}|_{\mathrm{id}}^{2} = E|\mathbf{x} - \overline{\mathbf{x}}^{b} - (\mathbf{K})_{i}G_{\mathbf{c}_{i}}[(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^{b})_{i}]|_{\mathrm{id}}^{2}$$
(16)

for the computation of the denominator. In the above,  $(\mathbf{K})_i$  is the *i*th column of the Kalman gain matrix  $\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^b = \mathbf{H}(\mathbf{x} - \overline{\mathbf{x}}^b) + \boldsymbol{\epsilon}, \ \mathbf{x} - \overline{\mathbf{x}}^b \sim N_n(\mathbf{0}, \mathbf{P}^b)$  and  $\boldsymbol{\epsilon} \sim N_p(\mathbf{0}, \mathbf{R})$ . In Eqs. (15) and (16), the means are then computed using a Monte Carlo approach, sampling from these Gaussian distributions. The  $c_i$  used to clip the *i*th innovation is chosen as if the analysis process consisted of assimilating the *i*th observation only. The selected clipping heights vary according to whether we clip the observation to  $\mathbf{c}$  by Huberizing or to  $\mathbf{0}$  by deleting. There is a precedent for this criterion of selecting a clipping height for robust Kalman filters (Ruckdeschel 2010). In Ruckdeschel (2010), however, a one-dimensional clipping height is selected to clip the norm of the multi-dimensional observations for the multidimensional case. When outliers occur at few variables, the norm may not be changed much by these few outliers and therefore the outliers may not be clipped. The variables where the outliers do not occur should be evenly clipped once

the one-dimensional clipping height is smaller than the norm. A multidimensional clipping height that we propose clips the elements that are considered to have outliers.

Another criterion is to select  $c_i$  such that

$$(1-r)E[|(\mathbf{y} - \mathbf{H}\overline{\mathbf{x}}^b)_i|_{\mathrm{id}} - c_i]_+ = rc_i \qquad (17)$$

for a given radius  $r \in (0, 1)$ . Here,  $(x)_+ = |x| \times \max(x/|x|, 0)$ . The radius *r* is a proportion of the amount of clipping in the innovation. The clipping heights are the same for either type of clipping because this criterion does not depend on how we clip innovations. A smaller radius provides a larger clipping height and fewer clipping outliers. This radius criterion has been used to select a clipping height in the robust Kalman filter scheme (Ruckdeschel 2010).

The important issues in selecting the clipping height **c** are the computational complexity of the sample covariance matrices. First, a small ensemble size may produce inaccurate estimates of the covariance matrices (Whitaker and Hamill 2002). Another is that doing the Monte Carlo integration method to choose the clipping height **c** for all time steps is time consuming. To increase the accuracy of the covariance matrices and save computation time, we may use  $\lim \mathbf{P}_t^b$  for one common clipping height **c** to use at every time step in case we can obtain the limit, instead of using  $\mathbf{P}_t^b$  at each time t. If we let  $\mathbf{P}_{\infty}$  be the unknown  $n \times n$  true covariance matrix at  $t = \infty$ , then we have  $\mathbf{P}_{\infty} = \lim_{M \to \infty, t \to \infty} \mathbf{P}_{t}^{b}$ . When M is sufficiently large, we can assume that  $\lim_{t \to \infty} \mathbf{P}_t^b \approx \mathbf{P}_{\infty}$ . We show in the next sections that the sample covariance matrix converges to its limit in a one-dimensional linear system and that the average of the sample covariance matrix is used as an alternative to a limit in our multidimensional nonlinear system.

#### 3. A one-dimensional linear system

To illustrate the effect of outliers in a one-dimensional linear system, we assume that the system equation and the observation equation are given by Eqs. (9) and (10), respectively. These simple equations have been used by Meinhold and Singpurwalla (1983). The estimate of the classical ensemble Kalman filter [Eq. (3)] becomes

$$\overline{x}_{t}^{a} = \overline{x}_{t}^{b} + P_{t}^{b} (P_{t}^{b} + 1)^{-1} (y_{t} - \overline{x}_{t}^{b}), \qquad (18)$$

given that the observation error variance is 1.

We investigate the performance of the robust ensemble Kalman filter (REnKF) for this system using 20member ensembles and a variance inflation factor of 1.1. A limit of the sample variance of the ensembles  $P_t^b$  of 1.63 is used to determine the clipping height *c*. We use



FIG. 8. The average sample ensemble covariances between variable 21 and other variables using 10 000 ensemble members from t = 101 to 300.

500 replications for graphical representations with boxplots (Tukey 1970). The efficiencies  $\delta = 0.99$ , 0.95, 0.9, 0.8, and 0.7, respectively, correspond to the clipping heights c = 4.25, 2.64, 2.19, 1.60, and 1.21 when we Huberize the observations. The same efficiencies correspond to the clipping heights c = 6.02, 4.80, 4.40, 3.71, and 3.21 when we discard the observations. The radii r =0.0001, 0.001, 0.003, 0.005, and 0.01 respectively correspond to the clipping heights c = 5.20, 4.24, 3.77, 3.48, and 3.14 when we Huberize or discard the observations.

To see the impact of additive outliers, we suppose that the additive outliers with  $\xi_t = 8$  are present in the data at t = 31, 32, and 33. Figure 2 shows the boxplots of the bias versus efficiency  $\delta$ . Figure 3 shows the boxplots of the bias versus radius r. As the clipping value c decreases, that is, as the efficiency decreases or as the radius increases, the bias of the robust estimators shrinks, whereas the error variance decreases to a point but then increases again. The chosen clipping heights are in the range where the error variance keeps increasing. The bias of the Huberizing filter decreases to zero slower than that of the discarding filter, but the error variance increases slower than that of the discarding filter. The bias starts to recover from t = 34 when the outliers disappear. Figure 4 shows the trajectories of the true state, the traditional ensemble Kalman filter, and two robust ensemble Kalman filters with efficiencies  $\delta = 0.99$ and 0.7. Both robust ensemble Kalman filters have smaller jumps in the state estimation errors at the times of the outliers than the traditional ensemble Kalman filter has. At efficiency  $\delta = 0.99$ , the discarding filter



FIG. 9. Bias vs efficiency of the EnKF and two REnKFs for variable 11 of the Lorenz model for t = 70-78. The additive outliers  $\xi_t = 10$  occur at variables 11, 12, and 13 at times t = 71, 72, and 73.

removes the jump entirely, coinciding with a bias of zero, but at efficiency  $\delta = 0.7$ , its estimation is inaccurate, coinciding with the big error variance shown in Fig. 2. Figure 5 shows the trajectories of the true state, the traditional ensemble Kalman filter, and the two robust ensemble Kalman filters with efficiencies  $\delta = 0.99$  and 0.7 and radii r = 0.0001 and 0.01. For r = 0.01, the discarding filter is more precise than the Huberizing filter at t = 31, 32, and 33, but it is more imprecise in the absence of outliers from t = 10 to 20. It agrees that the larger the radius, the smaller bias and the larger error variance shown in Fig. 3. To examine the effect of innovations outliers, we suppose that the innovation outliers with  $k_t = 25$  occur at time t = 31, 32, and 33. Figure 6 shows the boxplots of the bias versus efficiency  $\delta$ . Figure 7 shows the boxplots of the bias versus radius *r*. The bias stays at zero for all filters because the innovations outliers are set to have zero means. In terms of the error variance, the robust ensemble Kalman filters have increasing error variance as the efficiency  $\delta$  decreases or as the radius *r* increases. At t = 31, 32, and 33, the traditional ensemble Kalman filter has the largest error variance. The efficiency  $\delta$  gives a smaller error variance than the radius *r* gives because it



FIG. 10. As in Fig. 9, but for bias vs radius rather than efficiencies.

has a larger clipping value compared to the radius and as such does not clip much. The Huberization is better than getting rid of observations in terms of the error variance. At times with no outliers, the robust ensemble Kalman filter, however, has a larger error variance than the traditional ensemble Kalman filter.

### 4. A multidimensional nonlinear system

# a. The Lorenz model

We modify the 40-variable nonlinear dynamical system of Lorenz and Emanuel (1998) by adding a random model error term  $dw_i$ . Then, the model equation is given by

$$dx_{i} = [(x_{i+1} - x_{i-2})x_{i-1} - x_{i} + F]dt + dw_{i},$$
  

$$i = 1, \dots, 40,$$
(19)

where  $dw_i$  is a scalar from a Gaussian distribution with a zero mean and variance of 0.05, F = 8, and the boundary conditions are assumed to be periodic. We use a fourth-order stochastic Runge–Kutta scheme with a time step of 0.05 nondimensional units to integrate the model. The background ensemble members are initialized from random fields and integrated for 500 steps. Each state variable is observed directly, and observations having uncorrelated errors are assimilated at every time step. The observation equation follows





FIG. 11. The true values, EnKF, and two REnKFs with efficiency  $\delta = 0.9999$  and 0.98 for variable 11 of the Lorenz model. The additive outliers with  $\xi_t = 10$  occur at variables 11, 12, and 13 at times t = 71, 72, and 73.

$$(y_{t,1},\ldots,y_{t,40})^{\mathrm{T}} := \mathbf{y}_t = \mathbf{x}_t + \boldsymbol{\epsilon}_t, \qquad (20)$$

where  $\boldsymbol{\epsilon}_t$  is zero-mean white noise with variance **R** =  $0.05 \times I_{40}$ , and  $I_{40}$  is the identity matrix of size 40. The model is integrated for 190 time steps, and the first 100 time steps are discarded. The 20-member ensembles are used, and a localization constant of 15 and ensemble inflation factor of 1.07 are used following Whitaker and Hamill (2002). Experiments were conducted using the EnKF and REnKF with perturbed observations.

### b. Choice of the clipping height for the Lorenz model

We discuss how to choose the clipping height c and investigate the behavior of the robust ensemble Kalman filter for the Lorenz model. We use the average of the sample background covariance matrix  $\mathbf{P}_t^b$  from t = 101 to 300 of  $M = 10\,000$  ensemble members to select a 40-dimensional clipping height vector **c** based on a Monte Carlo integration method. Figure 8 illustrates the 21st column of the averaged sample background covariance matrix. The sample background covariance matrices were computed by running the model forward and assimilating the observations.

Since the dynamics of the model, distribution of the observations, and observation error statistics are homogenous, all components of the clipping height vector **c** have similar values. The radii r = 0.0001, 0.0005, 0.001,0.01, and 0.05 respectively correspond to the clipping heights 3.30, 2.89, 2.7, 2, and 1.44. The efficiencies  $\delta =$ 0.9999, 0.999, 0.99, 0.985, and 0.98 respectively correspond to the clipping heights 2.45, 1.62, 0.55, 0.32, and



FIG. 12. As in Fig.11, but for r = 0.0001 and 0.05.

0.16 when we Huberize observations, and they respectively correspond to the clipping heights 3.8, 3, 1.8, 1.43, and 1.06 when we discard observations. We use 200 replications for graphical representations with boxplots.

# c. The effects of outliers

To see the effect of additive outliers in the Lorenz model, we assume that additive outliers with  $\xi_t = 10$ occur for neighboring variables 11, 12, and 13 at t = 71, 72, and 73. Figure 9 shows the boxplots of the bias versus efficiency in the presence of additive outliers with  $\xi_t = 10$ for the Lorenz model. Figure 10 shows the boxplots of the bias versus radius in the presence of additive outliers with  $\xi_t = 10$ . As the clipping value decreases, that is, as the radius increases or as the efficiency decreases, the bias for the robust filters gets closer to zero, similar to the behavior observed in the one-dimensional linear system. The discarding filter forces the bias to go to zero faster than the Huberization filter. The error variance decreases to a point but then it increases again as the clipping value c decreases. The explanation for this behavior is that a proper clipping height truncates observations safely, but a too small clipping height clips observations too much: in the expression

$$\hat{\mathbf{x}}_{t}^{a} = \overline{\mathbf{x}}_{t}^{b} + \mathbf{K}_{t} G_{\mathbf{c}} (\mathbf{y}_{t} + \boldsymbol{\xi}_{t} - \mathbf{H}_{t} \overline{\mathbf{x}}_{t}^{b}), \qquad (21)$$

the clipping function  $G_{\mathbf{c}}$  cuts, in addition to  $\boldsymbol{\xi}_t$ , a significant portion of  $\mathbf{y}_t - \mathbf{H}_t \mathbf{\bar{x}}_t^b$ . Such aggressive clipping can lead to zero bias at the expense of a large error variance of the state estimates. However, the rate of change of the error variance is different for the two robust ensemble Kalman filters. The discarding filter is more aggressive



FIG. 13. As in Fig. 9, but for innovations outliers with  $k_t = 100$  occur at variables 11, 12, and 13 at times t = 71, 72, and 73 with a probability of contamination  $\alpha = 0.2$ .

and its error variance therefore increases faster than the Huberization filter. The bias for all filters increases from t = 71 to 73 because the outliers are carried over and not perfectly removed. The bias starts to recover from t = 74 when outliers do not occur. Figure 11 shows the time evolution of a component of the state vector for the true state, traditional ensemble Kalman filter, and two robust ensemble Kalman filters with efficiencies  $\delta = 0.9999$  and 0.98. Coinciding with the large error variance in the boxplots, the estimation with efficiency 0.98 is imprecise because the maximum relative efficiency that the robust ensemble Kalman filter can achieve for each observation

is 0.9747. Figure 12 shows the time evolution of a component of the state vector for the true state, traditional ensemble Kalman filter, and two robust ensemble Kalman filters with radii r = 0.0001 and 0.05. At the same radius, 0.05, the Huberization filter is more accurate than the discarding filter.

To investigate the effect of innovations outliers, we assume that the observation error comes from white noise with extreme variance at variables 11, 12, and 13 at t = 71, 72, and 73. Figure 13 shows the boxplots of the bias versus efficiency in the presence of innovations outliers with  $k_t = 100$  in the Lorenz model. Figure 14 shows



FIG. 14. As in Fig. 13, but for bias vs radius.

the boxplots of the bias versus radius in the presence of the same innovations outliers. For r > 0 and  $\delta < 1$ , the bias stays at zero but the error variance decreases to a certain point and then increases again as the clipping height decreases, and at t = 70 when no outliers occur, both robust ensemble Kalman filters experience a loss of accuracy.

## 5. Discussion

We proposed a robust ensemble Kalman filter for the robust estimation of the state of a spatiotemporal dynamical system in the presence of observational outliers. We applied this robust ensemble Kalman filter to a one-dimensional linear system and a multidimensional nonlinear system. Using this filtering technique, which is based on the Huberization method, the negative effects of the outliers on the state estimates can be greatly reduced. The clipping values were selected using the efficiency and radius criteria. We compared the results of the robust ensemble Kalman filter with those from the classical ensemble Kalman filter. We also compared the robust ensemble Kalman filter based on the Huberization filter, which pulls the outliers back to  $\mathbf{c}$  or  $-\mathbf{c}$ , and the robust ensemble Kalman filter, which discards outliers. We found that compared to the conventional EnKF, the robust ensemble Kalman filter reduced the bias in the state estimates at the expense of increasing the error variance. The increase of the error variance differed depending on the filtering method. The Huberization filter was found to perform better than the discarding filter for the examples given in the paper, which may be because the model we used gives the true state. The robust ensemble Kalman filter is efficient with simple models, and we plan to test it in realistic ocean and atmospheric systems.

Finding the proper clipping values for a data assimilation system that assimilates many types of observations using a complex model is expected to be a labor intensive process. There is no reason to believe, however, that the process would be more challenging or would require more work than determining the parameters of the quality-control procedures currently used in operational numerical weather prediction. In fact, the parameters used in the current operational systems should provide invaluable information about the gross errors in the different types of observations, which could be used as guidance for the selection of the clipping values.

Acknowledgments. This work was supported in part by Award KUS-C1-016-04 made by King Abdullah University of Science and Technology (KAUST). Mikyoung Jun's research was also partially supported by NSF Grants DMS-0906532 and DMS-1208421. Istvan Szunyogh acknowledges the support from ONR Grant N000141210785.

#### REFERENCES

- Anderson, E., and H. Järvinen, 1999: Variational quality control. Quart. J. Roy. Meteor. Soc., 125, 697–722.
- Birmiwal, K., and J. Shen, 1993: Optimal robust filtering. *Stat. Decis.*, **11**, 101–119.
- —, and P. Papantoni-Kazakos, 1994: Outlier resistant prediction for stationary processes. *Stat. Decis.*, **12**, 395–427.
- Burgers, G., P. J. van Leeuwen, and G. Evensen, 1998: Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, **126**, 1719–1724.
- Calvet, L. E., V. Czellar, and E. Ronchetti, cited 2012: Robust filtering. [Available online at http://ssrn.com/abstract=2123477.]
- Daley, R., 1991: Atmospheric Data Analysis. Cambridge Atmospheric and Space Science Series, Cambridge University Press, 457 pp.
- Ershov, A. A., and R. S. Liptser, 1978: Robust Kalman filter in discrete time. *IEEE Trans. Autom. Remote Control*, **39**, 359–367.
- Fahrmeir, L., and H. Kaufmann, 1991: On Kalman filtering, posterior mode estimation and Fisher scoring in dynamic exponential family regression. *Metrika*, 38, 37–60.
- —, and R. Kunstler, 1999: Penalized likelihood smoothing in robust state space models. *Metrika*, **49**, 173–191.
- Fox, A. J., 1972: Outliers in time series. J. Roy. Stat. Soc., B34, 350– 363.
- Genton, M. G., 2003: Breakdown-point for spatially and temporally correlated observations. *Developments in Robust Statistics*, R. Dutter et al., Eds., Springer, 148–159.
- —, and A. Lucas, 2003: Comprehensive definitions of breakdownpoints for independent and dependent observations. J. Roy. Stat. Soc., B65, 81–94.

- —, and —, 2005: Discussion of "Breakdown and groups" by L. Davies and U. Gather. Ann. Stat., 33, 988–993.
- Hampel, F. R., 1968: Contributions to the theory of robust estimation. Ph.D. thesis, University of California.
- Harlim, J., and B. R. Hunt, 2007: A non-Gaussian ensemble filter for assimilating infrequent noisy observations. *Tellus*, 59A, 225–237.
- Houtekamer, P. L., and H. L. Mitchell, 1998: Data assimilation using an ensemble Kalman filter technique. *Mon. Wea. Rev.*, 126, 796–811.
- Huber, P. J., 1981: Robust Statistics. Wiley, 308 pp.
- Ingleby, N. B., and A. C. Lorenc, 1993: Bayesian quality control using multivariate normal distributions. *Quart. J. Roy. Meteor.* Soc., 119, 1195–1225.
- Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. J. Basic Eng., 82, 34–45.
- Kassam, S. A., and H. V. Poor, 1985: Robust techniques for signal processing: A survey. *Proc. IEEE*, **73**, 433–481.
- Lorenz, E. N., and K. A. Emanuel, 1998: Optimal sites for supplementary weather observations: Simulation with a small model. J. Atmos. Sci., 55, 399–414.
- Luo, X., and I. Hoteit, 2011: Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Mon. Wea. Rev.*, **139**, 3938–3953.
- Maronna, A., R. D. Martin, and V. J. Yohai, 2006: *Robust Statistics: Theory and Methods.* Wiley, 436 pp.
- Martin, R. D., and A. E. Raftery, 1987: Robustness, computation and non-Euclidean models. J. Amer. Stat. Assoc., 82, 1044– 1050.
- Meinhold, R. J., and N. D. Singpurwalla, 1983: Understanding the Kalman filter. *Amer. Stat.*, **37**, 123–127.
- —, and —, 1989: Robustification of Kalman filter models. J. Amer. Stat. Assoc., 84, 479–486.
- Naveau, P., M. G. Genton, and X. Shen, 2005: A skewed Kalman filter. J. Multivariate Anal., 95, 382–400.
- Ruckdeschel, P., 2010: Optimally robust Kalman filtering. Berichte des Fraunhofer ITWM 185, 53 pp.
- Schick, I. C., and S. K. Mitter, 1994: Robust recursive estimation in the presence of heavy-tailed observation noise. Ann. Stat., 22, 1045–1080.
- Schlee, F. H., C. J. Standish, and N. F. Toda, 1967: Divergence in the Kalman filter. *Amer. Inst. Aeronaut. Astronaut. J.*, 5, 1114– 1120.
- Stockinger, N., and R. Dutter, 1987: Robust time series analysis: A survey. *Kybernetika*, **23**, 3–88.
- Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, **60**, 113–130.
- Tavolato, C., and L. Isaksen, 2010: Huber norm quality control in the IFS. *ECMWF Newsletter*, No. 122, ECMWF, Reading, United Kingdom, 27–31.
- Tukey, J. W., 1970: Exploratory Data Analysis. Vol. 1. Addison-Wesley, 688 pp.
- West, M., 1981: Robust sequential approximate Bayesian estimation. J. Roy. Stat. Soc., B43, 157–166.
- —, 1983: Generalized linear models: Scale parameters, outlier accommodation and prior distributions. *Bayesian Stat.*, 2, 531– 558.
- —, 1984: Outlier models and prior distributions in Bayesian linear regression. J. Roy. Stat. Soc., B46, 431–439.
- Whitaker, J. S., and T. M. Hamill, 2002: Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, **130**, 1913–1924.