# Ensemble Kalman filtering with residual nudging

*By* XIAODONG LUO[1]* and IBRAHIM HOTEIT[2], [1]*International Research Institute of Stavanger, Bergen, Norway*; [2]*King Abdullah University of Science and Technology, Thuwal, Saudi Arabia*

## ABSTRACT

Covariance inflation and localisation are two important techniques that are used to improve the performance of the ensemble Kalman filter (EnKF) by (in effect) adjusting the sample covariances of the estimates in the state space. In this work, an additional auxiliary technique, called residual nudging, is proposed to monitor and, if necessary, adjust the residual norms of state estimates in the observation space. In an EnKF with residual nudging, if the residual norm of an analysis is larger than a pre-specified value, then the analysis is replaced by a new one whose residual norm is no larger than a pre-specified value. Otherwise, the analysis is considered as a reasonable estimate and no change is made. A rule for choosing the pre-specified value is suggested. Based on this rule, the corresponding new state estimates are explicitly derived in case of linear observations. Numerical experiments in the 40-dimensional Lorenz 96 model show that introducing residual nudging to an EnKF may improve its accuracy and/or enhance its stability against filter divergence, especially in the small ensemble scenario.

*Keywords: data assimilation, ensemble Kalman filter, residual nudging*

## 1. Introduction

The ensemble Kalman filter (EnKF) (Evensen, 1994; Burgers et al., 1998; Houtekamer and Mitchell, 1998; Anderson, 2001; Bishop et al., 2001; Pham, 2001; Hoteit et al., 2002; Whitaker and Hamill, 2002) is an efficient algorithm for data assimilation in high-dimensional systems. Because of its runtime efficiency and simplicity in implementation, it has received ever-increasing attention from researchers in various fields. In many applications of the EnKF, due to limited computational resources, an EnKF can only be run with an ensemble size much smaller than the dimension of the state space. In such circumstances, problems often arise, noticeably on the quality of the sample covariances, including rank-deficiency, underestimation of the covariance matrices (Whitaker and Hamill, 2002; Sacher and Bartello, 2008), and spuriously large cross-variances between independent (or uncorrelated) state variables (Hamill et al., 2001). To mitigate these problems, it is customary to introduce two auxiliary techniques, covariance inflation (Anderson and Anderson, 1999) and localisation (Hamill et al., 2001), to the EnKF. On the one hand, covariance inflation increases the estimated sample covariances in order to compensate for the effect of underestimation, which also increases the robustness of the EnKF as reported by Luo and Hoteit (2011). On the other hand, covariance localisation introduces a 'distance'-dependent tapering function to the elements of the sample covariances and smoothes out the spuriously large values in them. In addition, covariance localisation also increases the ranks of the sample covariances (Hamill et al., 2009).

Both covariance inflation and localisation are techniques that in effect adjust the sample covariances in the state space. Since data assimilation is a practice of estimation that incorporates information from both the state and observation spaces, it would be natural for one to make use of the information in the observation space to improve the performance of an EnKF.

In this study we propose such an observation-space-based auxiliary technique, called residual nudging, for the EnKF. Here, a 'residual' is a vector in the observation space and is defined as the projection of an analysis mean onto the observation space subtracted from the corresponding observation. In residual nudging, our objective is to make the vector norm of the residual ('residual norm' for short) no larger than a pre-specified value. This is motivated by the observation that if the residual norm is too large, then the corresponding analysis mean is often a poor estimate. In such cases, it is better off to choose as the new estimate a state vector whose residual norm is smaller.

*Corresponding author.
email: xiaodong.luo@iris.no

**1**

The method presented in this work is close to the idea of Van Leeuwen (2010), in which a nudging term is added to the particle filter so that the projections of the particles onto the observation space are drawn closer to the corresponding observation, and the particles themselves are associated with almost equal weights. By doing so, the modified particle filter can achieve a remarkably good performance using only 20 particles in the chaotic 40-dimensional Lorenz-96 (L96) model (Lorenz and Emanuel, 1998), while traditional methods may need thousands of particles (Van Leeuwen, 2010). Other similar, residual-related methods were also found in the literature, for example, see Anderson (2007, 2009) and Song et al. (2010). Anderson (2007, 2009) suggested adaptive covariance inflation schemes in the context of hierarchical ensemble filtering. There the inflation factor $\lambda$ is considered as a random variable (with a presumed initial prior distribution) and in effect adjusts the projection of the background (co)variances onto the observation space.[1] With an incoming observation, the prior distribution is updated to the posterior one based on Bayes' rule, while the residual affects the shape of the posterior distribution of $\lambda$. On the other hand, Song et al. (2010) considered the idea of replacing an existing analysis ensemble member by a new one, in which the residual plays a role in generating the new ensemble member.

Our main purpose is to use residual nudging as a safeguard strategy, with which the projections of state estimates onto the observation space, under suitable conditions, are guaranteed to be within a pre-specified distance to the corresponding observations. We will discuss how to choose the pre-specified distance and construct the (possibly) new state estimates accordingly in case of linear observations. In this work, the ensemble adjustment Kalman filter (EAKF) (Anderson, 2001) is adopted for the purpose of demonstration, while the extension to other filters can be done in a similar way. Through numerical experiments in the L96 model, we show that the EAKF equipped with residual nudging (EAKF-RN) is more robust than the normal EAKF. In addition, the accuracy of the EAKF-RN is comparable to, and sometimes (much) better than, that of the normal EAKF.

This work is organised as follows. Section 2 reviews the filtering step of the EAKF, introduces the concept of residual nudging, and discusses how it can be implemented in the EAKF. Section 3 investigates the effect of residual nudging on the performance of the Kalman filter (KF) in a linear/Gaussian system, which aims to provide some

insights of how residual nudging may affect the behaviour of an already optimal filter. Section 4 extends the investigation to the Lorenz 96 model, in which we examine the performance of the EAKF-RN in various scenarios and compare it with the normal EAKF. Section 5 discusses possible extensions of the current study and concludes the work.

## 2. Ensemble Kalman filtering with residual nudging

Suppose that at the $k$-th assimilation cycle, one has a background ensemble $\mathbf{X}_k^b = \{\mathbf{x}_{k,i}^b\}_{i=1}^n$ with $n$ members. The incoming observation $\mathbf{y}_k^o$ is obtained from the following observation system

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \tag{1}$$

where $\mathbf{H}_k$ is a matrix, and $\mathbf{v}_k$ is the observation noise, with zero mean and covariance $\mathbf{R}_k$. For convenience of discussion, we assume that the dimensions of $\mathbf{X}_k$ and $\mathbf{y}_k$ are $m_x$ and $m_y$, respectively, $m_y \leq m_x$, and $\mathbf{H}_k$ has full row rank.

### 2.1. The filtering step of the EAKF with covariance inflation and localisation

We first summarise the filtering step of the EAKF with both covariance inflation and localisation. For simplicity, here we only consider the scenario with constant covariance inflation and localisation and refer readers to, for example, Anderson (2007, 2009) for the details of adaptive configuration of the EAKF. In the context of EAKF, it is assumed that the covariance $\mathbf{R}_k$ of the observation noise is a diagonal matrix, such that one can assimilate the incoming observation in a serial way. Following Anderson (2007, 2009), we use a single scalar observation to demonstrate the assimilation algorithm in the EAKF. To this end, in this subsection (only) we temporarily assume that the observation vector $\mathbf{y}_k \equiv y_k$ is a scalar random variable, with zero mean and variance $R_k$. The notation of the incoming observation thus becomes $y_k^o$, with the dimension $m_y = 1$. The algorithm description below mainly follows Anderson (2007).

Suppose that the $i$-th ensemble member $\mathbf{x}_{k,i}^b$ of $\mathbf{X}_k^b$ consists of $m_x$ elements $(\mathbf{x}_{k,i}^b)_j (j = 1, \cdots, m_x)$ such that $\mathbf{x}_{k,i}^b = [(\mathbf{x}_{k,i}^b)_1, \cdots, (\mathbf{x}_{k,i}^b)_{m_x}]^T$. Then the sample mean $\hat{\mathbf{x}}_k^b$ of $\mathbf{X}_k^b$ is $\hat{\mathbf{x}}_k^b = \sum_{i=1}^n \mathbf{x}_{k,i}^b / n$. To introduce covariance inflation to the filter, suppose that $\Delta \mathbf{X}_k^b \equiv \{\Delta \mathbf{x}_{k,i}^b : \Delta \mathbf{x}_{k,i}^b = \mathbf{x}_{k,i}^b - \hat{\mathbf{x}}_k^b\}_{i=1}^n$ is the ensemble of deviations with respect to $\mathbf{X}_k^b$ and $\lambda \geq 1$ the inflation factor, then the inflated background ensemble is $\mathbf{X}_k^{inf} \equiv \{\mathbf{x}_{k,i}^{inf} : \mathbf{x}_{k,i}^{inf} = \hat{\mathbf{x}}_k^b + \sqrt{\lambda} \, \Delta \mathbf{x}_{k,i}^b\}_{i=1}^n$ (Anderson, 2007, 2009). With covariance inflation, $\mathbf{X}_k^{inf}$ and $\mathbf{X}_k^b$ have the same mean, but the sample covariance of $\mathbf{X}_k^{inf}$ is $\lambda$ times that

---

[1] In contrast, in residual nudging we are interested in adjusting the projection of the background mean. Comparison and/or combination of these two strategies will be deferred to future investigations.

of $\mathbf{X}_k^b$. In what follows, we do not particularly distinguish background ensembles with and without covariance inflation through different notations. Instead, we always denote the background ensemble by $\mathbf{X}_k^b$, no matter whether it is inflated or not. One can tell whether a background ensemble is inflated by checking the value of $\lambda$; for example, $\lambda = 1$ means no inflation and $\lambda > 1$ means with covariance inflation.

On the other hand, suppose that the projection of $\mathbf{X}_k^b$ onto the observation space is $\mathbf{Y}_k^b = \{y_{k,i}^b : y_{k,i}^b = \mathbf{H}_k \mathbf{x}_{k,i}^b\}_{i=1}^n$, then one can compute the sample mean $\hat{y}_k^b$ and sample variance $\hat{p}_{yy,k}^b$ as

$$
\begin{aligned}
\hat{y}_k^b &= \tfrac{1}{n} \sum_{i=1}^n y_{k,i}^b, \\
\hat{p}_{yy,k}^b &= \tfrac{1}{n-1} \sum_{i=1}^n (y_{k,i}^b - \hat{y}_k^b)^2.
\end{aligned}
\tag{2}
$$

With the incoming observation $y_k^o$, one updates $\hat{y}_k^b$ and $\hat{p}_{yy,k}^b$ to their analysis counterparts, $\hat{y}_k^a$ and $\hat{p}_{yy,k}^a$, respectively, through the following formulae [Anderson, 2007, eqs. (3.2)–(3.3)].

$$
\begin{aligned}
\hat{p}_{yy,k}^a &= [(\hat{p}_{yy,k}^b)^{-1} + R_k^{-1}]^{-1}, \\
\hat{y}_k^a &= \hat{p}_{yy,k}^a [(\hat{p}_{yy,k}^b)^{-1} \hat{y}_k^b + R_k^{-1} y_k^o].
\end{aligned}
\tag{3}
$$

Accordingly, one can update the projection $\mathbf{Y}_k^b$ to its analysis counterpart $\mathbf{Y}_k^a \equiv \{y_{k,i}^a : y_{k,i}^a = y_{k,i}^b + \delta y_{k,i}\}_{i=1}^n$, where the increments $\delta y_{k,i}$ with respect to $y_{k,i}^b$ are given by

$$
\delta y_{k,i} = \sqrt{\frac{\hat{p}_{yy,k}^a}{\hat{p}_{yy,k}^b}} (y_{k,i}^b - \hat{y}_k^b) + \hat{y}_k^a - y_{k,i}^b.
\tag{4}
$$

One can verify that the sample mean and covariance of $\mathbf{Y}_k^a$ are $\hat{y}_k^a$ and $\hat{p}_{yy,k}^a$, respectively. Also note the difference between the concepts of deviations and increments. For distinction, we have used $\Delta$ to denote deviations and $\delta$ to denote increments.

After the above quantities are calculated, one proceeds to update the background ensemble $\mathbf{X}_k^b$ to the analysis of $\mathbf{X}_k^a \equiv \{\mathbf{x}_{k,i}^a : \mathbf{x}_{k,i}^a = \mathbf{x}_{k,i}^b + \delta \mathbf{x}_{k,i}\}_{i=1}^n$, where the increment $\delta \mathbf{x}_{k,i}$ with respect to the $i$-th background ensemble member $\mathbf{x}_{k,i}^b$ is an $m_x$ dimensional vector, that is, $\delta \mathbf{x}_{k,i} = [(\delta \mathbf{x}_{k,i})_1, \cdots, (\delta \mathbf{x}_{k,i})_{m_x}]^T$, where the $j$-th element $(\delta \mathbf{x}_{k,i})_j$ of $\delta \mathbf{x}_{k,i}$ is given by

$$
(\delta \mathbf{x}_{k,i})_j = (\hat{p}_{xy,k}^j / \hat{p}_{yy,k}^b) \delta y_{k,i}, \, j = 1, \cdots, m_x,
\tag{5}
$$

with $\hat{p}_{xy,k}^j$ being the sample cross-variance between all the $j$-th elements of the ensemble members of $\mathbf{X}_k^b$, and the projection ensemble $\mathbf{Y}_k^b = \{y_{k,i}^b\}_{i=1}^n$, that is,

$$
\hat{p}_{xy,k}^j = \frac{1}{n-1} \sum_{i=1}^n [(\mathbf{x}_{k,i}^b)_j - (\hat{\mathbf{x}}_k^b)_j][y_{k,i}^b - \hat{y}_k^b].
\tag{6}
$$

With relatively small ensemble sizes, eq. (6) often results in spuriously large sample cross-variances (Hamill et al., 2001). To tackle this problem, one may introduce covariance localisation (Hamill et al., 2001) to the EAKF, in which the main idea is to multiply $\hat{p}_{xy,k}^j$ in eq. (5) by a 'distance'-dependent tapering coefficient $\eta_{ij} \le 1$ (Anderson, 2007, 2009). We will discuss how to compute $\eta_{ij}$ in the experiments with respect to the L96 model.

After obtaining the analysis ensemble $\mathbf{X}_k^a$, one computes the analysis mean $\hat{\mathbf{x}}_k^a = \sum_{i=1}^n \mathbf{x}_{k,i}^a / n$ (analysis for short) and uses it as the posterior estimate of the system state. Propagating $\mathbf{X}_k^a$ forward through the dynamical model, a background ensemble at the next assimilation time is obtained, and a new assimilation cycle starts, and so on.

### 2.2. Residual nudging

As will be shown later, the EAKF may suffer from filter divergence in certain circumstances, even when it is equipped with both covariance inflation and localisation. To mitigate filter divergence, intuitively one may choose to adjust the estimate $\hat{\mathbf{x}}_k^a$ and move it closer toward the truth $\mathbf{x}_k^{tr}$. In practice, though, $\mathbf{x}_k^{tr}$ is normally unknown, thus it is infeasible to apply this state-space-based strategy. In what follows, we introduce a similar, but observation-space-based strategy, in which the main idea is to monitor, and, if necessary, adjust the residual norm of the estimate. For this reason, we refer to this strategy as residual nudging.

By definition, the residual with respect to the analysis mean $\hat{\mathbf{x}}_k^a$ is $\hat{\mathbf{r}}_k^a \equiv \mathbf{H}_k \hat{\mathbf{x}}_k^a - \mathbf{y}_k^o$. We also define the 2-norm of a vector $\mathbf{z}$ as

$$
||\mathbf{z}||_2 \equiv \sqrt{\mathbf{z}^T \mathbf{z}}.
\tag{7}
$$

The objective in residual nudging is the following. We accept $\hat{\mathbf{x}}_k^a$ as a reasonable estimate if its residual norm $||\hat{\mathbf{r}}_k^a||_2$ is no larger than a pre-specified value, for instance, $\beta \sqrt{\text{trace}(\mathbf{R}_k)}$, with $\beta > 0$ being called the noise level coefficient hereafter (the reason in choosing this pre-specified value is explained later). Otherwise, we consider $\hat{\mathbf{x}}_k^a$ a poor estimate, and thus find a replacement, for instance, $\tilde{\mathbf{x}}_k^a$, based on the estimate $\hat{\mathbf{x}}_k^a$ and the observation $\mathbf{y}_k^o$ so that the residual norm of $\tilde{\mathbf{x}}_k^a$ is no larger than $\beta \sqrt{\text{trace}(\mathbf{R}_k)}$. To this end, we stress that the assumption $m_y \le m_x$ may be necessary in certain cases (see the later discussion). In this work, we focus on the cases with $m_y \le m_x$, which is true for many geophysical problems.

The objective of residual nudging can be achieved as follows. First of all, we compute a scalar $c_k \in [0, 1]$, called the fraction coefficient hereafter [cf. eq. (9a) later for the reason], according to the formula

$$
c_k = \min(1, \beta \sqrt{\text{trace}(\mathbf{R}_k)} / ||\hat{\mathbf{r}}_k^a||_2),
\tag{8}
$$

where the function min(a,b) finds the minimum between the scalars $a$ and $b$. The rationale behind eq. (8) is this: if $\| \widehat{\mathbf{r}}_k^a \|_2 > \beta \sqrt{\text{trace}(\mathbf{R}_k)}$, then we need to multiply $\| \widehat{\mathbf{r}}_k^a \|_2$ by a coefficient $c_k < 1$ to reduce $\| \widehat{\mathbf{r}}_k^a \|_2$ to the pre-specified value. Otherwise, we do nothing and keep $\| \widehat{\mathbf{r}}_k^a \|_2$ as it is, which is equivalent to multiplying $\| \widehat{\mathbf{r}}_k^a \|_2$ by $c_k = 1$.

Next, we construct a new estimate $\widetilde{\mathbf{x}}_k^a$ by letting

$$\widetilde{\mathbf{x}}_k^a = c_k \, \widehat{\mathbf{x}}_k^a + (1 - c_k) \, \mathbf{x}_k^o, \tag{9a}$$

$$\mathbf{x}_k^o = \mathbf{H}_k^T (\mathbf{H}_k \mathbf{H}_k^T)^{-1} \mathbf{y}_k^o. \tag{9b}$$

The term $\mathbf{H}_k^T (\mathbf{H}_k \mathbf{H}_k^T)^{-1}$ in eq. (9b) is the Moore-Penrose generalised inverse of $\mathbf{H}_k$, such that $\mathbf{x}_k^o$ in eq. (9b) provides a least-square solution for the equation $\mathbf{H}_k \mathbf{x} = \mathbf{y}_k^o$ (Engl et al., 2000, chapter 2). We refer to $\mathbf{x}_k^o$ as the observation inversion hereafter. With eq. (9), the new residual $\widetilde{\mathbf{r}}_k^a = \mathbf{H}_k \widetilde{\mathbf{x}}_k^a - \mathbf{y}_k^o = c_k \widehat{\mathbf{r}}_k^a$ so that $\| \widetilde{\mathbf{r}}_k^a \|_2 = c_k \| \widehat{\mathbf{r}}_k^a \|_2 \leq \beta \sqrt{\text{trace}(\mathbf{R}_k)}$ according to eq. (8).

In residual nudging we only attempt to adjust the analysis mean $\widehat{\mathbf{x}}_k^a$ of the EAKF, but not its covariance. To this end, let the analysis ensemble be $\mathbf{X}_k^a = \{ \mathbf{x}_{k,i}^a : \mathbf{x}_{k,i}^a = \widehat{\mathbf{x}}_k^a + \Delta \mathbf{x}_{k,i}^a \}_{i=1}^n$, where the deviations $\Delta \mathbf{x}_{k,i}^a = \mathbf{x}_{k,i}^a - \widehat{\mathbf{x}}_k^a$. We then replace the original analysis mean $\widehat{\mathbf{x}}_k^a$ by $\widetilde{\mathbf{x}}_k^a$ and change the analysis ensemble to $\widetilde{\mathbf{X}}_k^a = \{ \widetilde{\mathbf{x}}_{k,i}^a : \widetilde{\mathbf{x}}_{k,i}^a = \widetilde{\mathbf{x}}_k^a + \Delta \mathbf{x}_{k,i}^a \}_{i=1}^n$. Therefore, in comparison with the normal EAKF, the EAKF with residual nudging (EAKF-RN for short) just has additional steps in eqs. (8) and (9), while all the other procedures remain the same. In doing so, residual nudging is compatible with both covariance inflation and localisation.

## 2.3. Discussion

Choosing the pre-specified value in the form of $\beta \sqrt{\text{trace}(\mathbf{R}_k)}$ is motivated by the following consideration. Let $\mathbf{x}_k^{tr}$ be the truth such that $\mathbf{y}_k^o = \mathbf{H}_k \mathbf{x}_k^{tr} + \mathbf{v}_k$. Then $\widetilde{\mathbf{r}}_k^a = \mathbf{H}_k \widetilde{\mathbf{x}}_k^a - \mathbf{y}_k^o = \mathbf{H}_k (\widetilde{\mathbf{x}}_k^a - \mathbf{x}_k^{tr}) - \mathbf{v}_k$ and by the triangle inequality

$$\| \widetilde{\mathbf{r}}_k^a \|_2 \leq \| \mathbf{H}_k (\widetilde{\mathbf{x}}_k^a - \mathbf{x}_k^{tr}) \|_2 + \| \mathbf{v}_k \|_2. \tag{10}$$

For a reasonably good estimate $\widetilde{\mathbf{x}}_k^a$, we expect that the magnitude of $\mathbf{H}_k \widetilde{\mathbf{x}}_k^a - \mathbf{H}_k \mathbf{x}_k^{tr}$ should not substantially exceed the observation noise level. On the other hand, we have $(\mathbb{E} \| \mathbf{v}_k \|_2)^2 \leq \mathbb{E} \| \mathbf{v}_k \|_2^2 = \text{trace}(\mathbb{E}(\mathbf{v}_k \mathbf{v}_k^T)) = \text{trace}(\mathbf{R}_k)$, thus the expectation $\mathbb{E} \| \mathbf{v}_k \|_2$ of the norm of the observation noise is (at most) in the order of $\sqrt{\text{trace}(\mathbf{R}_k)}$. One may thus use $\sqrt{\text{trace}(\mathbf{R}_k)}$ to characterise the noise level. By requiring that a reasonably good estimate has $\| \mathbf{H}_k (\widetilde{\mathbf{x}}_k^a - \mathbf{x}_k^{tr}) \|_2$ in the order of $\sqrt{\text{trace}(\mathbf{R}_k)}$ (or less), one comes to the choice in the form of $\beta \sqrt{\text{trace}(\mathbf{R}_k)}$. The criterion in choosing the above threshold is very similar to that in certain quality control algorithms (called check of plausibility, see, for example, Gandin, 1988, for a survey), in which one is

assumed to have prior knowledge about, for instance, the mean $\bar{y}_s$ and variance $\sigma_s$ of a scalar observation $y_s$. In quality control, $y_s$ is often assumed to be a Gaussian random variable so that for a measured observation $y_s^o$, if the ratio $|y_s^o - \bar{y}_s| / \sigma_s$ is too large, then $y_s^o$ is discarded, or is at least deemed suspected (Gandin, 1988). The main differences between residual nudging and quality control are the following. While quality control checks the plausibility of an incoming observation, residual nudging checks the plausibility of a state estimate and suggests a replacement if the original state estimate does not pass the test. Moreover, as long as the 2-norm is used, the expectation $\mathbb{E} \| \mathbf{v}_k \|_2^2$ is always trace$(\mathbf{R}_k)$, independent of the distribution of $\mathbf{v}_k$. This independence, on the one hand, implies that the inequality in eq. (10), hence the threshold $\beta \sqrt{\text{trace}(\mathbf{R}_k)}$, holds without requiring the knowledge of the distribution of $\mathbf{H}_k (\widetilde{\mathbf{x}}_k^a - \mathbf{x}_k^{tr})$. On the other hand, the absence of the knowledge of the distribution means that less statistical information is gained in choosing the threshold $\beta \sqrt{\text{trace}(\mathbf{R}_k)}$. For instance, one may not be able to assign a statistical meaning to $\beta \sqrt{\text{trace}(\mathbf{R}_k)}$, nor obtain a confidence (or significance) level in accepting (or rejecting) a state estimate. Finally, it is also possible for one to adopt another distance metric, for example, the 1- or $\infty$-norm, for which the inequality in eq. (10) still holds. In such circumstances, the expectation, $\mathbb{E} \| \mathbf{v} \|_1^2$ or $\mathbb{E} \| \mathbf{v}_k \|_\infty^2$, may not be equal to trace$(\mathbf{R}_k)$ anymore so that one may need to choose a threshold different from $\beta \sqrt{\text{trace}(\mathbf{R}_k)}$. Despite the stated differences, we expect that residual nudging can be used in conjunction with observation quality control, although this is not pursed in the current study.

Even though the noise level coefficient $\beta$ in residual nudging is chosen to be time-invariant, the resulting fraction coefficient $c_k$, in general, changes with time according to eq. (8). The coefficient $\beta$ affects how the new analysis $\widetilde{\mathbf{x}}_k^a$ combines the original one $\widehat{\mathbf{x}}_k^a$ and the observation inversion $\mathbf{x}_k^o$. This can be seen from eqs. (8) and (9a). Because $c_k \in [0, 1]$, the new analysis $\widetilde{\mathbf{x}}_k^a$ in eq. (9a) is a convex combination of $\widehat{\mathbf{x}}_k^a$ and $\mathbf{x}_k^o$, that is, an estimate somewhere in-between the original estimate $\widehat{\mathbf{x}}_k^a$ and the observation inversion $\mathbf{x}_k^o$, depending on the value of $c_k$. If one chooses a large value for $\beta$, or if for a fixed $\beta$ the original residual norm $\widehat{\mathbf{r}}_k^a$ is sufficiently small, then the fraction coefficient $c_k \to 1$ according to eq. (8), thus $\widetilde{\mathbf{x}}_k^a \to \widehat{\mathbf{x}}_k^a$ according to eq. (9a). Therefore, $\widetilde{\mathbf{x}}_k^a$ will be a good estimate if $\widehat{\mathbf{x}}_k^a$ is so, but may not be able to achieve a good estimation accuracy when $\widehat{\mathbf{x}}_k^a$ itself is poor. On the other hand, if one chooses a very small value for $\beta$, or if for a fixed $\beta$ the original residual norm $\widehat{\mathbf{r}}_k^a \to +\infty$ (e.g. with filter divergence), then $c_k \to 0$ and $\widetilde{\mathbf{x}}_k^a \to \mathbf{x}_k^o$. In this case, the estimate $\widetilde{\mathbf{x}}_k^a$ is calculated mainly based on the information content of the observation $\mathbf{y}_k^o$ and may result in a relatively poor accuracy. This is

largely because of (1) the presence of the observation noise $\mathbf{v}_k$ in eq. (1), and (2) the ignorance of the prior knowledge of the model dynamics. As a result, pushing the projection of state estimates very close to noisy observations may have some negative consequences. For instance, in geophysical applications, dynamical balances of the numerical models may not be honoured so that the estimation errors may be relatively large. However, using $\mathbf{x}_k^o$ as the estimate may be a relatively safe (although conservative) strategy against filter divergence. In the sense of the above discussion, the choice of $\beta$ reflects the extent to which one wants to achieve the trade-off between a filter's potential accuracy and stability against divergence. This point is further demonstrated through some further experiments.

Some numerical issues related to the computation of the observation inversion $\mathbf{x}_k^o$ are discussed in order. One is the existence and uniqueness of the observation inversion. Under the assumptions that $m_y \leq m_x$ and that $\mathbf{H}_k$ is of full row rank, the observation inversion, as a solution of the equation $\mathbf{H}_k \mathbf{x} = \mathbf{y}_k^o$, does exist (Meyer, 2001, chapter 4). Finding a concrete solution, however, is in general an under-determined problem; hence, the solution is not unique unless $m_y = m_x$. This point can be seen as follows. When $m_y < m_x$, the null space $\mathbb{S}^N$ of $\mathbf{H}_k$ contains non-zero elements; that is, there exist elements $\mathbf{x}_n \in \mathbb{S}^N$, $\mathbf{x}_n \neq \mathbf{0}$, such that $\mathbf{H}_k \mathbf{x}_n = \mathbf{0}$ (Meyer, 2001, chapter 4). As a result, given an observation inversion $\mathbf{x}_k^o$, $\mathbf{x}_k^o + \mathbf{x}_n$ is also a solution of the equation $\mathbf{H}_k \mathbf{x} = \mathbf{y}_k^o$ for any $\mathbf{x}_n \in \mathbb{S}^N$. Therefore, which solution one should take is an open problem in practice. In the context of state estimation, it is desirable to choose a solution that is close to the truth $\mathbf{x}_k^{tr}$, which, unfortunately, is infeasible without the knowledge of $\mathbf{x}_k^{tr}$. As a trade-off, one may choose as a solution some estimate that possesses certain properties. The Moore-Penrose generalised inverse $\mathbf{x}_k^o$ given in eq. (9b) is such a choice, which is the unique, and 'best-approximate', solution in the sense that it has the minimum 2-norm among all least-squares solutions (Engl et al., 2000, Theorem 2.5).

It is also worth mentioning what may happen if our assumptions, that $m_y \leq m_x$ and that $\mathbf{H}_k$ is of full row rank, are not valid. In the former case, with $m_y > m_x$, the equation $\mathbf{H}_k \mathbf{x} = \mathbf{y}_k^o$ is over-determined, meaning that there may be no solution that solves the equation exactly. One may still obtain an approximate solution by recasting the problem of solving the linear equation as a linear least-squares problem, which yields the unique, least-squares solution in the form of $\mathbf{x}_k^o = (\mathbf{H}_k^T \mathbf{H}_k)^{-1} \mathbf{H}_k^T \mathbf{y}_k^o$, similar to (but different from) eq. (9b). Because $\mathbf{H}_k \mathbf{x}_k^o - \mathbf{y}_k^o$ may not be $\mathbf{0}$, in general, one may thus not be able to find a new estimate $\widetilde{\mathbf{x}}_k^a$ with a sufficiently small (e.g. zero) residual. Therefore, the inequality $\|\widetilde{\mathbf{r}}_k^a\|_2 \leq \beta \sqrt{\text{trace}(\mathbf{R}_k)}$ may not hold for some sufficiently small $\beta$. This restriction is consistent with the nature of over-determined problems (i.e. no exact solution).

It does not necessarily mean that residual nudging cannot be applied to an over-determined problem but instead implies that the noise level coefficient $\beta$ should entail a lower bound that may be larger than 0.

In the latter case, without loss of generality, suppose that $m_y \leq m_x$ and $\mathbf{H}_k$ is not of full row rank, then the matrix product $\mathbf{H}_k \mathbf{H}_k^T$ is singular so that it may be numerically unstable to compute its inverse. In such circumstances, one needs to employ a certain regularisation technique to obtain an approximate, but stable, solution. For instance, one may adopt the Tikhonov regularisation (Engl et al., 2000, chapter 4) so that the solution in eq. (9b) becomes $\mathbf{x}_k^o = \mathbf{H}_k^T (\mathbf{H}_k \mathbf{H}_k^T + \alpha \mathbf{I})^{-1} \mathbf{y}_k^o$, where $\alpha$ is the regularisation parameter chosen according to a certain criterion. The observation inversion in eq. (9b) can be treated as a special case of the Tikhonov regularisation solution with $\alpha = 0$, while the concept of residual nudging is also applicable to the general cases with $\alpha \neq 0$ following our deduction in section 2.2.[2] In this sense, the state estimate of the EAKF-RN can be considered as a hybrid of the original EAKF estimate and the (regularised) least-squares solution of the equation $\mathbf{H}_k \mathbf{x} = \mathbf{y}_k^o$. This point of view opens up many other possibilities, given the various types of regularisation techniques in the literature (see, e.g., Engl et al. 2000).

The computation of the matrix product $\mathbf{H}_k^T (\mathbf{H}_k \mathbf{H}_k^T)^{-1}$ is a non-trivial issue in large-scale problems and is worthy of further discussion.[3] In general cases where the observation operator $\mathbf{H}_k$ is time varying, the computational cost is comparable to that in evaluating the Kalman gain. In terms of numerical computations, one possible choice is to apply QR factorisation (Meyer, 2001, chapter 5) to $\mathbf{H}_k^T$ such that $\mathbf{H}_k^T$ is factorised as the product of an orthogonal, $m_x \times m_x$ matrix $\mathbf{Q}$ and an upper-triangular, $m_x \times m_y$ matrix $\mathbf{U}$, where for notational convenience we drop the time index $k$ in these matrices. Note that $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T \mathbf{Q} = \mathbf{I}_{m_x}$, and $\mathbf{U} = [\mathbf{U}_{m_y}^T, \mathbf{0}_{(m_x - m_y)m_y}^T]^T$, with $\mathbf{I}_{m_x}$ being the $m_x$-dimensional identity matrix, $\mathbf{0}_{(m_x - m_y)m_y}$ the $(m_x - m_y) \times m_y$ zero matrix, and $\mathbf{U}_m$ a non-singular, upper-triangular, $m_y \times m_x$ matrix in which all elements below the main diagonal are zero. With some algebra, it can be shown that the product $\mathbf{H}_k^T (\mathbf{H}_k \mathbf{H}_k^T)^{-1} = \mathbf{Q} [\mathbf{U}_{m_y}^{-1}, \mathbf{0}_{(m_x - m_y)m_y}]^T = \mathbf{Q}_{m_x m_y} (\mathbf{U}_{m_y}^{-1})^T$, where $\mathbf{Q}_{m_x m_y}$ is a matrix that is comprised of the first $m_y$ columns of $\mathbf{Q}$, and the inverse $\mathbf{U}_{m_y}^{-1}$ of the upper-triangular matrix

---

[2]In general cases with $\alpha \neq 0$, it can be shown that a sufficient condition to achieve residual nudging is, for example, $c_k (\|\widehat{\mathbf{r}}_k^a\|_2 - \|\mathbf{y}_k^o\|_2) \leq \beta \sqrt{\text{trace}(\mathbf{R}_k)} - \|\mathbf{y}_k^o\|_2$, with the (possibly new estimate $\widetilde{\mathbf{x}}_k^a$ again given by eq. (9a).

[3]For the experiments to be presented later, since the dimensions of the dynamical models are relatively low, we choose to directly compute the matrix product $\mathbf{H}_k^T (\mathbf{H}_k \mathbf{H}_k^T)^{-1}$. The matrix inversion $(\mathbf{H}_k \mathbf{H}_k^T)^{-1}$ is done through the MATLAB (R2011b) built-in function INV.

$\mathbf{U}_{m_y}$ can be computed element-by-element in a recursive way (called back substitution, Meyer 2001, chapter 5). In certain circumstances, further reduction of computational cost and/or storage can be achieved, for instance, when $\mathbf{H}_k$ is sparse (Meyer, 2001, chapter 5) or when $\mathbf{H}_k$ is time invariant, for example, in a static observation network. In the latter case, one only needs to evaluate the product $\mathbf{H}_k^T (\mathbf{H}_k \mathbf{H}_k^T)^{-1}$ once and for all.

## 3. Numerical results in a linear scalar system

Here we use a scalar, first-order autoregressive (AR1) model driven by Gaussian white noise, to investigate the performance of the KF (Kalman, 1960), and that of the KF with residual nudging (KF-RN), in which residual nudging is introduced to the posterior estimate of the KF in the same way as in the EAKF. The motivation in conducting this experiment is as follows. With linear and Gaussian observations, the KF provides the optimal estimate in the sense of, for instance, minimum variance (MV) (Jazwinski, 1970). Therefore, we use the KF estimate as the reference to examine the behaviour of the KF-RN under different settings, which reveals how residual nudging may affect the performance of the KF.

The scalar AR1 model is given by

$$x_{k+1} = 0.9x_k + u_k, \tag{11}$$

where $u_k$ represents the dynamical noise and follows the Gaussian distribution with zero mean and variance 1 and is thus denoted by $u_k \sim N(u_k : 0, 1)$. The observation model is described by

$$y_k = x_k + v_k, \tag{12}$$

where $v_k \sim N(v_k : 0, 1)$ is the observation noise and is uncorrelated with $u_k$.

In the experiment, we integrate the AR1 model forward for 10 000 steps (integration steps hereafter), with the initial value randomly drawn from the Gaussian distribution $N(0, 1)$, and the associated initial prior variance being 1. The true states (truth) $\{x_k\}_{i=1}^{10000}$ are obtained by drawing samples of dynamical noise from the distribution $N(0, 1)$, adding them to $x_k$ to obtain $x_{k+1}$ at the next integration step, and so on. The synthetic observations $y_k^o$ are obtained by adding to model states $x_k$ samples of observation noise from the distribution $N(0, 1)$. For convenience of comparison, we generate and store synthetic observations at every integration step. However, we choose to assimilate them for every $S_a$ integration steps, with $S_a \in \{1, 2, 4, 8\}$, in order to investigate the impact of $S_a$ on filter performance. In doing so, data assimilation with different $S_a$, or other experiment settings (e.g. the noise level coefficient $\beta$ in the KF-RN), will have identical observations at the same integration steps. For convenience, hereafter we may sometimes use the

concept 'assimilation step', with one assimilation step equal to $S_a$ integration steps. In addition, we may also call $S_a$ the assimilation step when it causes no confusion.

In the KF-RN, we also choose to vary the noise level coefficient $\beta$, with $\beta \in \{0.01, 0.05, 0.1, 0.5, 1, 2, 3, 4, 6, 8, 10\}$, in order to investigate its effect on filter performance. In order to reduce statistical fluctuations, we repeat the experiment 20 times, each time with randomly drawn initial value, samples of dynamical and observation noise (so that the truth and the corresponding observations are produced at random). Except for the introduction of residual nudging, the KF-RN has the same configurations and experiment settings as the KF.

We use the average root mean squared error (average RMSE) to measure the accuracy of a filter estimate. For an $m_x$-dimensional system, the RMSE $e_k$ of an estimate $\hat{\mathbf{x}}_k = [\hat{x}_{k,1}, \cdots, \hat{x}_{k,m_x}]^T$ with respect to the true state vector $\mathbf{x}_k^{tr} = [x_{k,1}^{tr}, \cdots, x_{k,m_x}^{tr}]^T$ at time instant $k$ is defined as

$$e_k = \| \hat{\mathbf{x}}_k - \mathbf{x}_k^{tr} \|_2 / \sqrt{m_x}. \tag{13}$$

The average RMSE $\hat{e}_k$ at time instant $k$ over $M$ repetitions of the same experiment is thus defined as $\hat{e}_k = \sum_{j=1}^{M} e_k^j / M (M = 20$ in our setting), where $e_k^j$ denotes the RMSE at time instant $k$ in the $j$-th repetition of the experiment. We also define the time mean RMSE $\hat{e}$ as the average of $\hat{e}_k$ over the assimilation time window with $N$ integration steps, that is, $\hat{e} = \sum_{k=1}^{N} \hat{e}_k / N (N = 10\,000$ here).

We also use the spread to measure the estimated uncertainty associated with an estimation. To this end, let $\hat{\mathbf{P}}_k$ be the estimated covariance matrix with respect to the estimate $\hat{\mathbf{x}}_k$. Then the spread $s_k$ at time instant $k$ is defined as

$$s_k = \sqrt{\text{trace}(\hat{\mathbf{P}}_k)/m_x}. \tag{14}$$

Table 1. Time mean RMSEs and spreads of the KF, and the minimum time mean RMSEs (over different $\beta$) of the KF-RN, in the AR1 model with different $S_a$

|  | $S_a =$ | | | |
|---|---|---|---|---|
|  | 1 | 2 | 4 | 8 |
| KF | | | | |
|   RMSE | 0.6184 | 0.8260 | 1.0592 | 1.2997 |
|   Spread | 0.7729 | 1.0413 | 1.3419 | 1.8241 |
| KF-RN | | | | |
|   Min RMSE | 0.6183 | 0.8259 | 1.0592 | 1.2997 |
|   Achieved at | $\beta = 2$ | $\beta = 2$ | $\beta \geq 2$ | $\beta \geq 2$ |

The KF and KF-RN have identical time mean spreads; therefore, only those of the KF are presented. In the bottom row, we also report the ranges of $\beta$ in which the minimum time mean RMSEs of the KF-RN are achieved.

The average spread $\hat{s}_k$ and the time mean (average) spread $\hat{s}$ are defined in a way similar to their counterparts with respect to the RMSE.

Table 1 reports the time mean RMSEs and spreads of the KF at different assimilation steps $S_a$. The time mean RMSE of the KF grows as $S_a$ increases, indicating that the performance of the KF deteriorates as the assimilation frequency decreases. The time mean spread of the KF exhibits a similar tendency as $S_a$ increases. However, the time mean spread tends to be larger than the time mean RMSE, indicating that the corresponding variance is over-estimated.

Figure 1 shows the time mean RMSEs of the KF-RN (dash–dot lines marked by diamonds), as functions of the noise level coefficient $\beta$, at different assimilation steps $S_a$. Given the different orders of magnitudes of $\beta$, we adopt the logarithmic scale for the $x$-axes. Then for comparison, we also plot the time mean RMSEs of the KF (solid lines) at each $S_a$. Since the time mean RMSEs of the KF are independent of the choice of $\beta$, they are horizontal lines in the plots. However, the choice of $\beta$ does influence the performance of the KF-RN. As shown in all of the plots of Fig. 1, if one adopts a small $\beta$, for instance, $\beta = 0.01$, for the KF-RN, then the resulting time mean RMSE is higher than that of the KF. This is because such a choice may

force the KF-RN to rely excessively on the observations when updating the prior estimates, such that the information contents in the prior estimates are largely ignored. As $\beta$ grows, the time mean RMSE of the KF-RN decreases and eventually converges to that of the KF when $\beta$ is sufficiently large, for instance, $\beta \geq 3$. These results are consistent with our expectation of the behaviour of a filter equipped with residual nudging, as has been discussed above in section 2.3.

It is also of interest to gain some insights of the behaviour of the fraction coefficients $c_k$ in the KF-RN with different $\beta$. To this end, Fig. 2 plots two sample time series of $c_k$ in the KF-RN with $\beta = 0.1$ (upper left panel), and $\beta = 1$ (lower left panel), respectively, together with their corresponding histograms (right panels). For convenience of visualisation, the assimilation time window is shortened to 1000 steps (with the observations assimilated for every four steps). At $\beta = 0.1$, $c_k$ tends to be relatively small, with the mean value being 0.4213 and the median 0.3027. Among the 250 $c_k$ values, 210 of them are less than 1, meaning that residual nudging is effective at those steps. A histogram of $c_k$ is also shown on the upper right panel. There it indicates that $c_k$ distributes like a U-shape, with relatively large proportions of $c_k$ taking values that are less than 0.2, or equal to 1. On the other hand, at $\beta = 1$, $c_k$ tends to remain close to 1, with
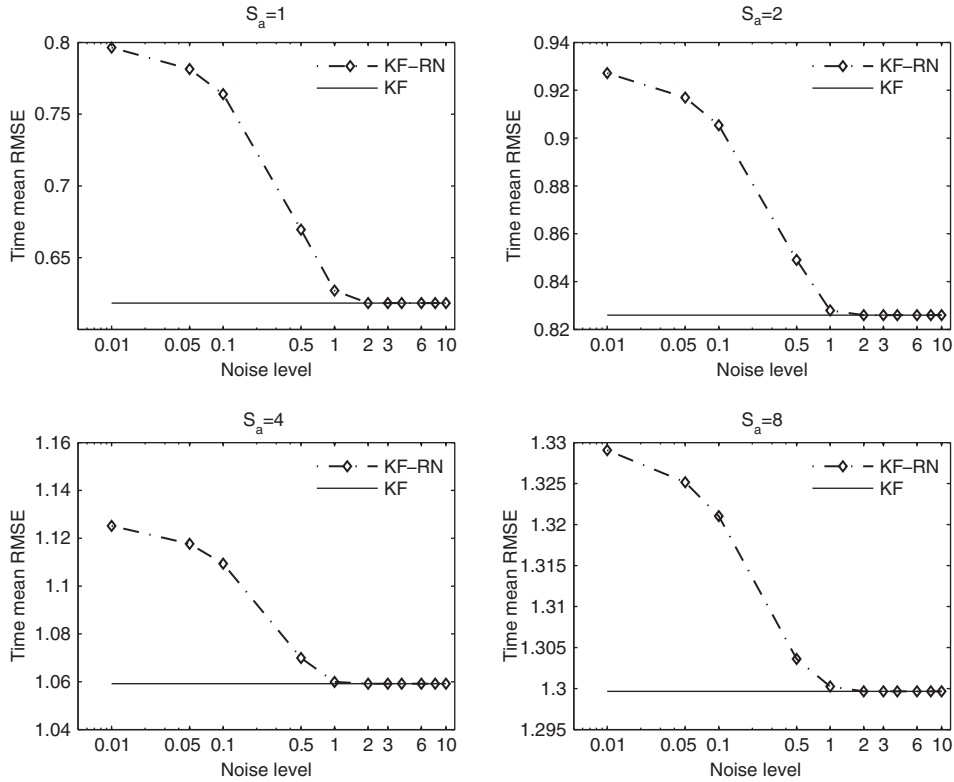


*Fig. 1.* Time mean RMSEs of the KF and the KF-RN as functions of the noise level coefficient in the AR1 model, with different $S_a$.
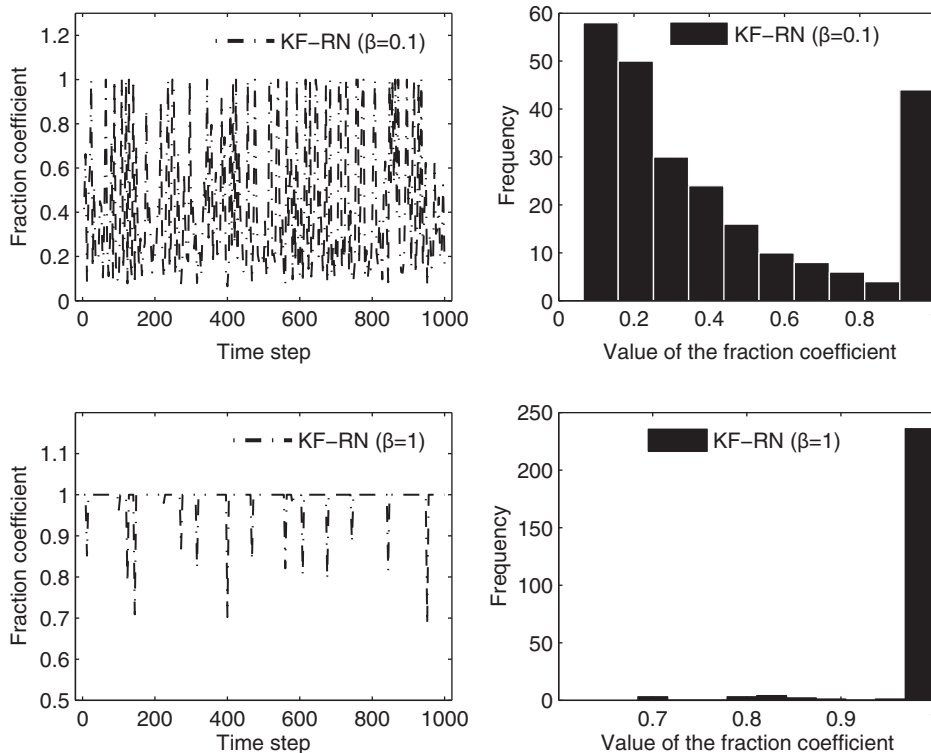
*Fig. 2.* Left panels: Sample time series of the fraction coefficients of the KF-RN with $\beta = 0.1$ (upper) and $\beta = 1$ (lower), respectively. Right panels: The corresponding histograms of the fraction coefficient time series.

the mean being 0.9892 and the median 1, and only 16 out of 250 $c_k$ values are less than 1. These are also manifested in the histogram on the lower right panel, where one can see that $c_k$ largely concentrate on 1.

In Table 1, we report the minimum time mean RMSEs that the KF-RN can achieve by varying the value of $\beta$ at different $S_a$, together with the values of the $\beta$ at which the minima are obtained for specific $S_a$. When $S_a = 1,2$, the minimum time mean RMSEs of the KF-RN, both achieved at $\beta = 2$, are (very) slightly lower than the time mean RMSEs of the KF; the time mean RMSEs of the KF-RN become the same as those of the KF when $\beta \geq 3$. On the other hand, when $S_a = 4,8$, the minimum time mean RMSEs of the KF-RN are identical to the time mean RMSEs of the KF and are obtained when $\beta \geq 2$. The reason that the KF-RN can have lower time mean RMSEs than the 'optimal' KF at $S_a = 1,2$ might be the following. The classic filtering theory states that the KF is optimal under the MV criterion (Jazwinski, 1970); that is, taking the mean of the posterior conditional probability density function (pdf) as the state estimate, the KF has the lowest possible expectation of squared estimation error. Note that here the expectation is taken over all possible values of the truth (i.e. by treating the truth as a random variable). Therefore, in principle one has to repeat the same experi-

ment for a sufficiently large number of times (with randomly drawn truth) in order to verify the performance of the filters under the MV criterion. For computational convenience, though, we only repeat the experiment 20 times. Thus in our opinion the slight out-performance of the KF-RN might be largely attributed to statistical fluctuations.

In Table 1 we do not present the time mean spreads of the KF-RN because they are in fact identical to those of the KF. This is because in the KF, the forecast and update of the (estimated) covariance matrix of the system state are not influenced by the mean estimate of the system state (Jazwinski, 1970). Since residual nudging only changes the estimate of the system state (if necessary) and nothing else, it is expected that the KF and KF-RN share the same covariance matrix. This point, however, is not necessarily true in the context of ensemble filtering in a non-linear system. For instance, if the dynamical model is non-linear, then the background covariance at the next assimilation time is affected by the analysis mean at the current time, such that two analysis ensembles with different sample (analysis) means but identical sample (analysis) covariance may result in different sample (background) means and covariances at the next assimilation time.

The above results suggest that it may not be very meaningful to introduce residual nudging to a Bayesian filter that already performs well. In practice, though, due to the existence of various sources of uncertainties (Anderson, 2007; Luo and Hoteit, 2011), a Bayesian filter is often suboptimal and is even likely to suffer from divergence (Schlee et al., 1967). In such circumstances, instead of only looking into the accuracy of a filter, it may also be desirable to take the stability of the filter into account. Through the experiments below, we show that equipping the EAKF with residual nudging can not only help improve its stability but also achieve a filter accuracy that is comparable to, sometimes even (much) better than, that of the normal EAKF, especially in the small ensemble scenario.

## 4. Numerical results in the 40-dimensional L96 model

### 4.1. Experiment settings

Here we use the 40-dimensional Lorenz-96 (L96) model (Lorenz and Emanuel, 1998) as the test bed. The governing equations of the L96 model are given by

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})x_{i-1} - x_i + F, \ i = 1, \cdots, 40. \quad (15)$$

The quadratic terms simulate advection, the linear term represents internal dissipation and $F$ acts as the external forcing term (Lorenz, 1996). Throughout this work, we choose $F = 8$ unless otherwise stated. For consistency, we define $x_{-1} = x_{39}$, $x_0 = x_{40}$ and $x_{41} = x_1$ in eq. (15) and construct the state vector $\mathbf{x} \equiv [x_1, x_2, \cdots, x_{40}]^T$.

We use the fourth-order Runge-Kutta method to integrate (and discretise) the system from time 0 to 75, with a constant integration step of 0.05. To avoid the transition effect, we discard the trajectory between 0 and 25 and use the rest for data assimilation. The synthetic observation $\mathbf{y}_k$ is obtained by measuring (with observation noise) every $d$ elements of the state vector $\mathbf{x}_k = [x_{k,1}, x_{k,2}, \cdots, x_{k,40}]^T$ at time instant $k$, that is,

$$\mathbf{y}_k = \mathbf{H}^d \mathbf{x}_k + \mathbf{v}_k, \quad (16)$$

where $\mathbf{H}^d$ is a $(J+1) \times 40$matrix such that $\mathbf{H}^d \mathbf{x}_k = [x_{k,1}, x_{k,1+d}, \cdots, x_{k,1+Jd}]^T$, with $J = \text{floor}(39/d)$ being the largest integer that is less than, or equal to, $39/d$, and $\mathbf{v}_k$is the observation noise following the Gaussian distribution $N(\mathbf{v}_k : \mathbf{0}, \mathbf{I}_{J+1})$, with $\mathbf{I}_{j+1}$ being the $(J+1)$-dimensional identity matrix. The elements $(\mathbf{H}^d)_{pq}$ of the matrix $\mathbf{H}^d$ can be determined as follows.

$$(\mathbf{H}^d)_{pq} = 1 \text{if } q = (p-1)\, d + 1\,, \text{otherwise} (\mathbf{H}^d)_{pq} = 0\,,$$

for $p = 1, \cdots, (J+1), q = 1, \cdots, 40$. In all the experiments below, we generate and store the synthetic observations

at every integration step but assimilate the observations for every four integration steps unless otherwise stated.

The filters in the experiments are configured as follows. To generate an initial background ensemble, we run the L96 model from 0 to 2500 (overall 50 000 integration steps) and compute the temporal mean and covariance of the trajectory.[4] We then assume that the initial state vectors follow the Gaussian distribution with the same mean and covariance and draw a specified number of samples to form the background ensemble. Covariance inflation (Anderson and Anderson, 1999) and localisation (Hamill et al., 2001) are conducted in all the experiments. Concretely, covariance inflation, with the inflation factor $\lambda$, is introduced following the discussion above in section 2.1. Covariance localisation is conducted following Anderson (2007, 2009), which introduces an additional parameter $l_c$, called the length scale (or half-width following Anderson 2007, 2009) hereafter, to the EAKF. The distance $d_{ij}$ between two state variables $x_i$ and $x_j$ are defined as $d_{ij} = \min(|i-j|/40, 1 - |i-j|/40)$, and the corresponding tapering coefficient $\eta_{ij}$ [cf. the text below eq. (6)] is determined by the fifth-order polynomial function $\xi(d_{ij}, l_c)$ in Gaspari and Cohn (1999) with half-width $l_c$. For $d_{ij} < 2l_c$, one has $0 < \eta_{ij} \leq 1$, and $\eta_{ij} = 0$ otherwise. With both covariance inflation and localisation, the performance of the normal EAKF is in general comparable to the established results with respect to the L96 model under similar experiment setting, see, for example, Hunt et al. (2004) and Fertig et al. (2007).

To reduce statistical fluctuations, we repeat each experiment below for 20 times, each time with randomly drawn initial state vector, initial background ensembles and observations. Except for the introduction of residual nudging, in all experiments the normal EAKF and the EAKF-RN have identical configurations and experiment settings.

### 4.2. Experiment results

4.2.1. Results with different observation operators. Here we consider four different observation operators $\mathbf{H}^d$, with $d = 1, 2, 4, 8$, respectively. For convenience, we refer to them as the full, 1/2, 1/4 and 1/8 observation scenarios, respectively. The concrete configurations of the normal EAKF and the EAKF-RN are as follows. In both filters, the ensemble size is fixed to be 20. The half-width $l_c$ of

---

[4]Let $\{\mathbf{x}_k\}_{k=1}^N$ be a set of state vectors at different time instants which form a state trajectory from time instant 1 to $N$. Then the temporal mean and covariance of the trajectory are taken as the sample mean and covariance of the set $\{\mathbf{x}_k\}_{k=1}^N$, respectively.
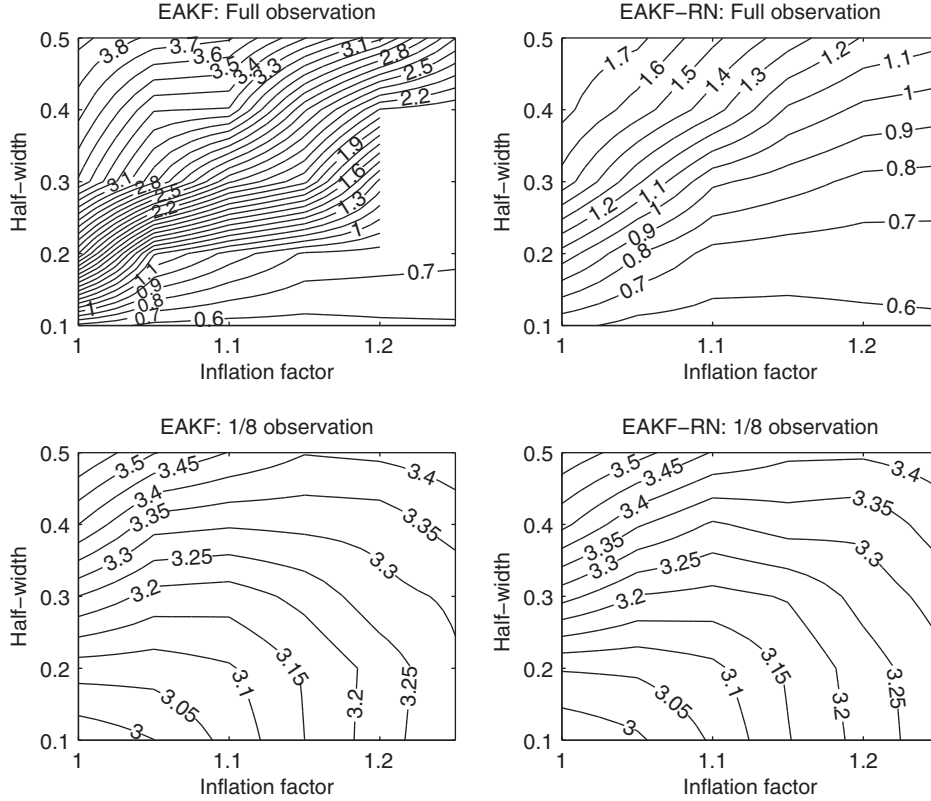
*Fig. 3.* Time mean RMSEs of the normal EAKF and the EAKF-RN, as functions of inflation factor and half-width, in the full and 1/8 observation scenarios.

covariance localisation increases from 0.1 to 0.5, each time with an even increment of 0.1. For convenience, we denote this setting by $l_c \in \{0.1 : 0.1 : 0.5\}$. Similar notations will be frequently used later. The inflation factor $\lambda \in \{1 : 0.05 : 1.25\}$, and the noise level coefficient $\beta = 2$ in the EAKF-RN.

The upper panels of Fig. 3 shows the contour plots of the time mean RMSEs of the normal EAKF (left) and that of the EAKF-RN (right), in the full observation scenario, as functions of the inflation factor $\lambda$ and the half-width $l_c$. Given a fixed $\lambda$, the time mean RMSEs of both the EAKF and EAKF-RN tend to increase as the half-width $l_c$ increases. On the other hand, given a fixed $l_c$, when $l_c = 0.1$, the time mean RMSEs of both filters exhibit the U-turn behaviour, that is, the time mean RMSEs tend to decrease as $\lambda$ grows, until it reaches a certain value (1.10 for both filters). After that, the time mean RMSEs will increase instead as $\lambda$ grows further. However, when $l_c > 0.1$, the time mean RMSEs of both filters tend to decrease as $\lambda$ increases within the range of tested $\lambda$. The normal EAKF achieves its minimum time mean RMSE (0.5605) at the point $(l_c = 0.1, \lambda = 1.10)$, and the EAKF-RN also hits its minimum time mean RMSE (0.5586) at the same place. In general, the EAKF and the EAKF-RN have similar

performances at $l_c = 0.1$, but at other places the EAKF-RN may perform substantially better than the EAKF. For instance, at $(l_c = 0.4, \lambda = 1.05)$ the time mean RMSE of the normal EAKF is about 3.3, while that of the EAKF-RN is about 1.6. Moreover, a filter divergence is spotted in the normal EAKF at $(l_c = 0.3, \lambda = 1.25)$ so that the contour plot around this point is empty and indicates no RMSE value. Filter divergence, however, is not observed in the EAKF-RN at the same place. For clarity, here a 'divergence' is identified as an event in which the RMSE of a filter becomes abnormally large. More specifically, the filter is considered divergent in the Lorenz 96 model, if its RMSE at any particular time instant is larger than $10^3$. As mentioned previously, we repeat each experiment 20 times in order to reduce statistical fluctuations. In accordance with this setting, a filter divergence is reported whenever there is at least one (but not necessarily all) divergence(s) out of 20 repetitions.

In the 1/2 and 1/4 observation scenarios, there are many cases in which filter divergences are spotted. For this reason, we choose to directly report the assimilation results in Tables 2 and 3, respectively, rather than showing their contour plots as in the full observation scenario. In the 1/2 observation scenario, filter divergences of the normal

*Table 2.* Time mean RMSEs (spreads) of the normal EAKF and the EAKF-RN in the 1/2 observation scenario, as functions of the covariance inflation factor and the half-width of covariance localisation

| | $l_c = 0.1$ | $l_c = 0.2$ | $l_c = 0.3$ | $l_c = 0.4$ | $l_c = 0.5$ |
|---|---|---|---|---|---|
| **EAKF** | | | | | |
| $\lambda = 1.00$ | 1.0721 (0.7049) | Div | Div | Div | Div |
| $\lambda = 1.05$ | 1.0091 (0.7457) | Div | Div | Div | Div |
| $\lambda = 1.10$ | 0.9789 (0.7868) | Div | Div | Div | Div |
| $\lambda = 1.15$ | 0.9662 (0.8209) | Div | Div | Div | Div |
| $\lambda = 1.20$ | 0.9515 (0.8566) | Div | Div | Div | Div |
| $\lambda = 1.25$ | 0.9623 (0.8929) | Div | Div | Div | Div |
| **EAKF-RN** | | | | | |
| $\lambda = 1.00$ | 1.0325 (0.7002) | 1.8256 (0.5697) | 2.1099 (0.5127) | 2.2734 (0.4736) | 2.2964 (0.4579) |
| $\lambda = 1.05$ | 1.0051 (0.7419) | 1.4072 (0.6185) | 1.9879 (0.5644) | 2.1821 (0.5269) | 2.2468 (0.5050) |
| $\lambda = 1.10$ | 0.9598 (0.7842) | 1.2313 (0.6553) | 1.8517 (0.6030) | 2.0342 (0.5699) | 2.1742 (0.5470) |
| $\lambda = 1.15$ | 0.9673 (0.8201) | 1.2024 (0.6870) | 1.6507 (0.6388) | 1.9317 (0.6015) | 2.0953 (0.5845) |
| $\lambda = 1.20$ | 0.9474 (0.8565) | 1.1788 (0.7183) | 1.5776 (0.6680) | 1.9059 (0.6336) | 2.0806 (0.6098) |
| $\lambda = 1.25$ | 0.9650 (0.8935) | 1.1856 (0.7484) | 1.5315 (0.6945) | 1.7778 (0.6603) | 2.0071 (0.6383) |

EAKF, marked by 'Div' in Table 2, are spotted in 24 out of 30 different combinations of $l_c$ and $\lambda$ values (5 $l_c$ values by 6 $\lambda$ values). In contrast, in the EAKF-RN no filter divergence is observed. On the other hand, when there is no filter divergence occurring in either filter, the performance of the EAKF and the EAKF-RN is very close to each other, with the time mean RMSEs of the EAKF-RN slightly lower than those of the EAKF, except at ($l_c = 0.1$, $\lambda = 1.15$) and ($l_c = 0.1$, $\lambda = 1.25$). The situation in the 1/4 observation is similar. As shown in Table 3, the EAKF diverges in 17 out of 30 tested cases, while there is no filter divergence spotted in the EAKF-RN. The performance of the EAKF and the EAKF-RN is close to each other when the EAKF does not diverge.

The lower panels of Fig. 3 show the contour plots of the time mean RMSEs of the normal EAKF (left) and that of the EAKF-RN (right), in the 1/8 observation scenario. In this scenario, no filter divergence is spotted in the EAKF. Overall, the performance of the EAKF and the EAKF-RN is very close to each other, although the EAKF-RN has a slightly lower minimum time mean RMSE (2.9556 achieved at ($l_c = 0.1$, $\lambda = 1$)) than that of the EAKF (2.9619 obtained at the same place).

We then examine the impact of residual nudging on the time mean spreads of the filters in different observation scenarios. For the full and 1/8 observation scenarios, we plot the time mean spreads of the EAKF and the EAKF-RN in Fig. 4; while for the 1/2 and 1/4 observation scenarios, we report them in Tables 2 and 3, in the parentheses after the RMSE values. In all the reported cases in which the EAKF does not diverge, the time mean spreads of the EAKF-RN in general do not significantly

*Table 3.* As in Table 2, except that it is in the 1/4 observation scenario

| | $l_c = 0.1$ | $l_c = 0.2$ | $l_c = 0.3$ | $l_c = 0.4$ | $l_c = 0.5$ |
|---|---|---|---|---|---|
| **EAKF** | | | | | |
| $\lambda = 1.00$ | 2.0685 (1.5730) | Div | Div | Div | Div |
| $\lambda = 1.05$ | 1.9908 (1.7849) | Div | Div | Div | Div |
| $\lambda = 1.10$ | 2.0223 (2.0447) | 2.3014 (1.5640) | Div | Div | Div |
| $\lambda = 1.15$ | 2.0819 (2.3592) | 2.2174 (1.7254) | 2.9502 (1.5820) | Div | Div |
| $\lambda = 1.20$ | 2.1903 (2.6869) | 2.1839 (1.9468) | 2.7534 (1.7191) | Div | Div |
| $\lambda = 1.25$ | 2.3586 (3.0392) | 2.2596 (2.2340) | 2.6413 (1.8780) | Div | Div |
| **EAKF-RN** | | | | | |
| $\lambda = 1.00$ | 2.0840 (1.5689) | 2.6099 (1.1984) | 3.0267 (1.0110) | 3.0453 (0.8703) | 3.0469 (0.7899) |
| $\lambda = 1.05$ | 2.0042 (1.7790) | 2.3341 (1.3762) | 2.8493 (1.1936) | 3.0573 (1.0403) | 3.1015 (0.9618) |
| $\lambda = 1.10$ | 1.9860 (2.0339) | 2.2976 (1.5332) | 2.8154 (1.3484) | 3.0527 (1.2112) | 3.1251 (1.1028) |
| $\lambda = 1.15$ | 2.0766 (2.3648) | 2.2389 (1.7244) | 2.7737 (1.4940) | 3.1247 (1.3341) | 3.2583 (1.2558) |
| $\lambda = 1.20$ | 2.1886 (2.6948) | 2.2312 (1.9710) | 2.6566 (1.6824) | 3.0992 (1.5048) | 3.2340 (1.3674) |
| $\lambda = 1.25$ | 2.3436 (3.0359) | 2.2352 (2.2344) | 2.6168 (1.8427) | 3.0977 (1.6509) | 3.2897 (1.5098) |
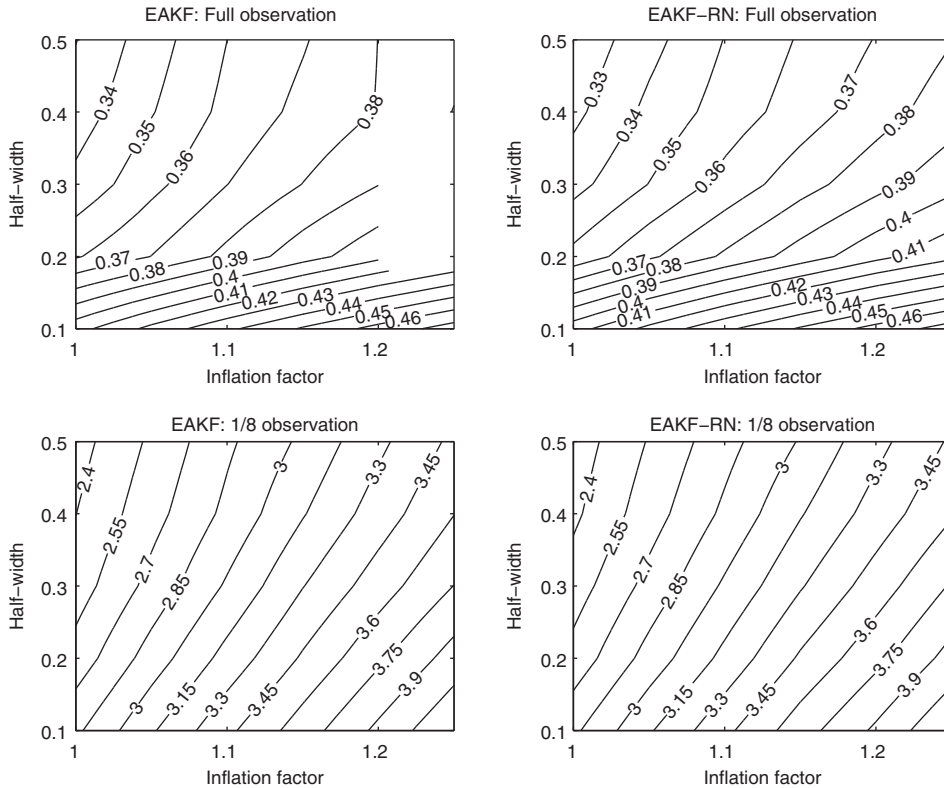
*Fig. 4.* Time mean spreads of the normal EAKF and the EAKF-RN, as functions of inflation factor and half-width, in the full and 1/8 observation scenarios.

deviate from those of the EAKF. In cases that the EAKF does diverge, the EAKF-RN may still maintain positive and finite time mean spreads. The closeness of the time mean spreads of the EAKF and EAKF-RN in the former cases, though, may depend on the experiment settings, for example, the choice of the noise level coefficient $\beta$. However, from our experience, as long as $\beta$ is reasonably large (such as $\beta \geq 2$), the time mean spread of the EAKF-RN often approaches that of the EAKF. For brevity, hereafter we do not report the spread values.

Overall, in both the normal EAKF and the EAKF-RN, their time mean RMSEs tend to increase as the number of elements in an observation decreases. The performance of the EAKF-RN, in terms of time mean RMSE, is in general comparable to, and sometimes (substantially) better than, that of the EAKF. Moreover, the EAKF-RN tends to perform more stable than the EAKF.

*4.2.2. Results with different noise level coefficients.* Next we examine the effect of the noise level coefficient $\beta$ on the performance of the EAKF-RN. The experiment settings are as follows. We conduct the experiments in four observation scenarios as in the previous experiment. The ensemble size of the EAKF-RN is 20. We choose the

noise level coefficient $\beta$ from the sets $\{0\}$, $\{0.02 : 0.02 : 0.1\}$, $\{0.2 : 0.2 : 1\}$ and $\{2,3,4,6,8\}$. The reason to single out $\beta = 0$ will be given soon. Under the above setting, it is infeasible for us to adopt too many combinations of $l_c$ and $\lambda$ as in the previous experiment, either for presentation or computation. Therefore, we only choose two such combinations in the current experiment (similar choices will also be made in subsequent experiments, in which we can only afford to vary some of the parameter values, and have to freeze the rest). In the first combination, we let $l_c = 0.1$ and $\lambda = 1.15$, and in the second, $l_c = 0.3$ and $\lambda = 1.03$. From the previous experiment results, the former choice represents a relatively good filter configuration for the normal EAKF, while the latter a less proper one. We thus use these two configurations to illustrate the effect of residual nudging when the normal EAKF has reasonable/(relatively) poor performance.

Figure 5 depicts the time mean RMSEs of the EAKF-RN as functions of $\beta$ in different observation scenarios, in which the relatively good filter configuration $l_c = 0.1$ and $\lambda = 1.15$ is adopted. Due to different orders of magnitudes of $\beta$, the x-axes are all plotted in the logarithmic scale. For this reason, it is inconvenient to show the results of $\beta = 0$ at $\log 0 (= -\infty)$. Instead, we plot the results at $\beta = 0.005$ and
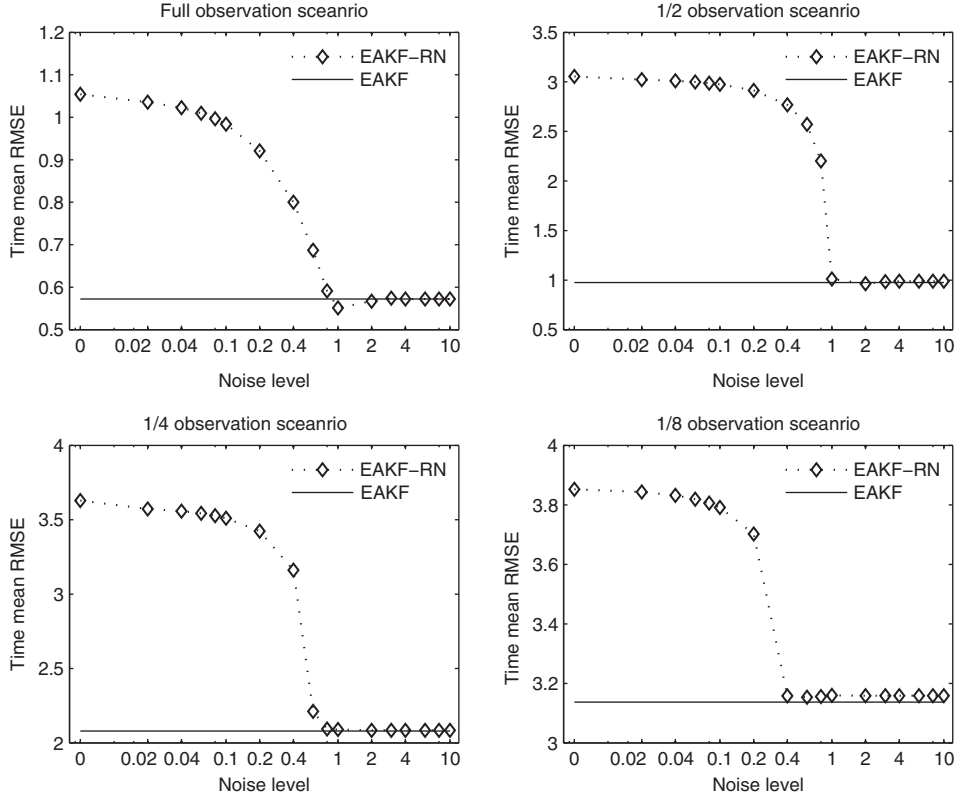
*Fig. 5.* Time mean RMSEs of the normal EAKF and the EAKF-RN as functions of the noise level coefficient in different observation scenarios, with $\lambda = 1.15$ and $l_c = 0.1$.

'artificially' label that point 0. The time mean RMSEs of the normal EAKF are independent of $\beta$ and are plotted as horizontal lines in the relevant subfigures (if no filter divergence in the normal EAKF). In all observation scenarios, the time mean RMSEs of the EAKF-RN are relatively large at small $\beta$ values (for instance, $\beta = 0.02$). As $\beta$ increases, the time mean RMSEs of the EAKF-RN tend to converge to those of the normal EAKF. During the processes of convergence, the minimum time mean RMSE of the EAKF-RN in the full observation scenario is lower than that of the normal EAKF, while the minimum time mean RMSEs of the EAKF-RN in other observation scenarios are either indistinguishable from (in the 1/2 and 1/4 observation scenarios), or slightly higher than (in the 1/8 observation scenario), those of the normal EAKF.

Figure 6 shows the time mean RMSEs of the normal EAKF and the EAKF-RN, with experiment settings similar to those in Fig. 5, except that the covariance localisation and inflation configuration become $l_c = 0.3$ and $\lambda = 1.05$, respectively, which, as will be shown below, makes the normal EAKF perform worse in comparison to the previous case in Fig. 5.

With $l_c = 0.3$ and $\lambda = 1.05$, the resulting EAKF-RN behaves similarly to that with the previous configuration

$l_c = 0.1$ and $\lambda = 1.15$. For the current filter configuration, though, as $\beta$ grows, the time mean RMSEs of the EAKF-RN exhibit clear troughs in all observation scenarios. On the other hand, compared to the previous results in Fig. 5, the performance of the normal EAKF deteriorates in all observation scenarios. Indeed, with the current filter configuration, the normal EAKF may perform (substantially) worse than the EAKF-RN under the same experiment settings, especially if a proper $\beta$ value is chosen for the EAKF-RN. In particular, the normal EAKF diverges in the 1/2 (upper right) and 1/4 (lower left) observation scenarios, while no filter divergence is spotted in the EAKF-RN with $\beta \leq 3$, although the EAKF-RN does diverge in the 1/2 and 1/4 observation scenarios, given $\beta \geq 4$. This suggests that one may increase the stability of the EAKF-RN against filter divergence by decreasing the value of $\beta$ so that $c_k$ is closer to 0 and the observation inversion becomes more influential in eq. (9a), as we have discussed in section 2.3.

It is also worth mentioning the behaviour of the EAKF-RN with small $\beta$ values. As one can see in Figs. 5 and 6, given different filter configurations, the EAKF-RN may behave quite differently at relatively large $\beta$ values. However, as $\beta$ tends to 0, the time mean RMSEs of the
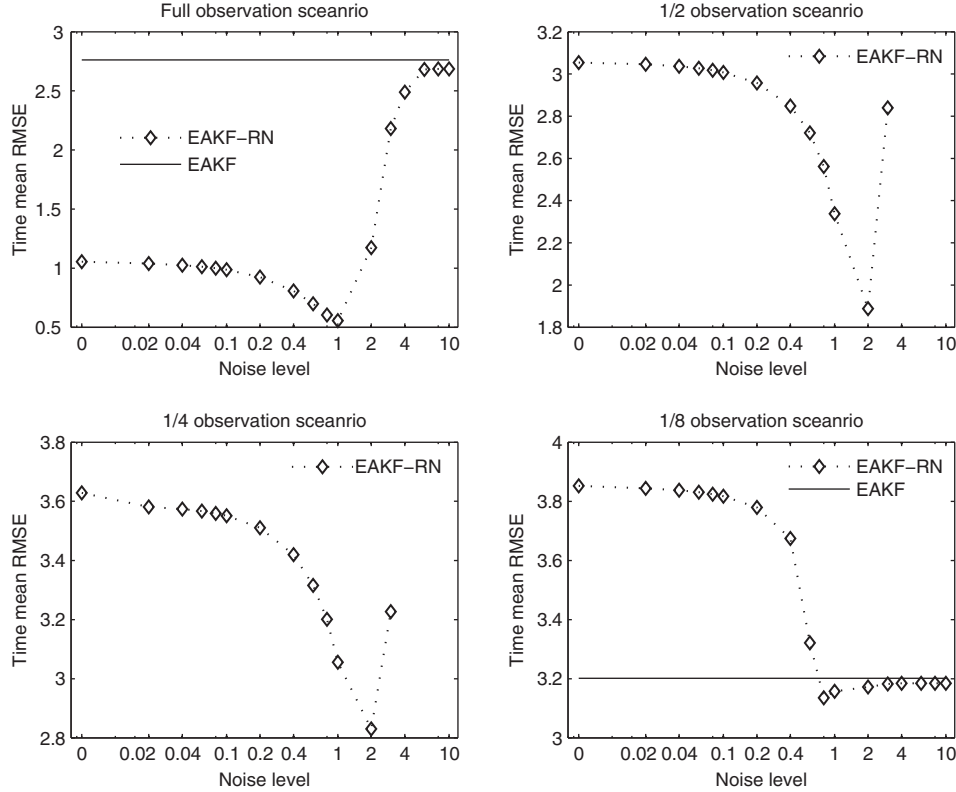
*Fig. 6.* As in Fig. 5, but with $\lambda = 1.05$ and $l_c = 0.3$ for both the filters. Note that in the 1/2 and 1/4 observation scenarios divergences of the normal EAKF are spotted; hence, no horizontal lines are indicated in the corresponding plots. The EAKF-RN also diverges in the 1/2 and 1/4 observation scenarios for $\beta \geq 4$.

EAKF-RN with different configurations tend to converge, despite the different combinations of $l_c$ and $\lambda$. This is because, as $\beta \to 0$, $c_k \to 0$ in eq. (8), the new estimate $\widetilde{\mathbf{x}}_k^a$, according to eq. (9a), approaches the observation inversion $\mathbf{x}_k^o$, which is independent of, for instance, the half-width $l_c$, the inflation factor $\lambda$ and the ensemble size.[5] Since the time mean RMSE continuously depends on $\beta$, it is not surprising to find that in Figs. 5 and 6, the time mean RMSEs of the EAKF-RN with small $\beta$, for instance, at $\beta = 0.02$, are very close to the corresponding values at $\beta = 0$.

More insights of the filters' behaviour may be gained by examining the fraction coefficient $c_k$ in the EAKF-RN. For the relatively good filter configuration ($l_c = 0.1$ and $\lambda = 1.15$), we have seen in Fig. 5 that the EAKF and the EAKF-RN have very close performance, and our experiment results show that $c_k$ mostly concentrate on 1, similar to the situations on the lower panels of Fig. 2 (not reported). Of more interest is the case in which the normal EAKF is less properly configured ($l_c = 0.3$ and $\lambda = 1.05$)

and may suffer from filter divergence. On the upper panels of Fig. 7, we show sample time series of the RMSEs of the normal EAKF and EAKF-RN ($\beta = 2$) in the 1/2 observation scenario. On the upper left panel, the EAKF has an exceptionally large RMSE (in the order of $10^{21}$) at time step $k = 26$ and is thus considered diverged. In contrast, on the upper right panel, the EAKF-RN ($\beta = 2$) has all the RMSEs less than 5 (with the corresponding time mean RMSE being 1.8931), and filter divergence is avoided. The lower left panel shows the time series of the fraction coefficient $c_k$, which has the mean 0.9499 and the median 1. Among 250 $c_k$ values, 78 are less than 1. For reference, a histogram of $c_k$ is plotted on the lower right panel, which confirms that $c_k$ largely concentrate on 1.

In Figure 8, we also examine what happens before the normal EAKF diverges. On the upper panel, we show the time series of the RMSEs of the EAKF (in the solid line with asterisks) and the EAKF-RN ($\beta = 2$, in the dotted line with plus signs). One can see that, at the beginning, for instance, when the time instant $k \leq 15$, the difference between the EAKF and the EAKF-RN is relatively less significant. For $16 \leq k \leq 25$, the difference becomes more obvious. On the middle panel, we report the difference

---

[5]When the observation operator is time-varying, the assimilation step $S_a$ in general has an influence on the observation inversion, as $S_a$ decides when the observations are assimilated.
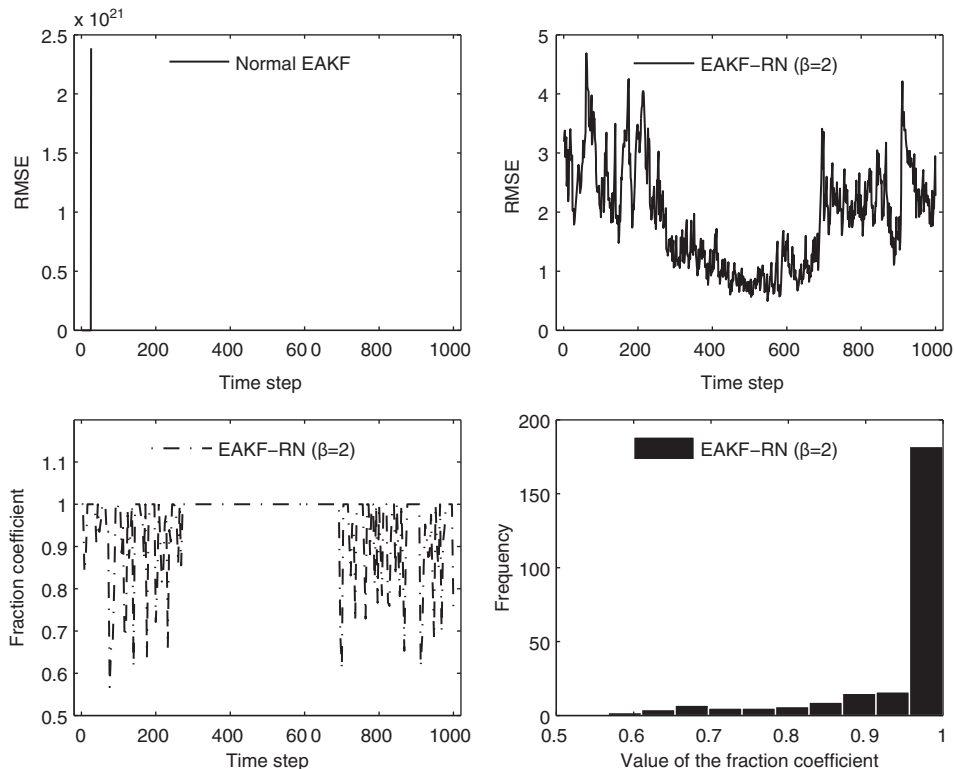
*Fig*. 7. Upper left: sample time series of the RMSE of the normal EAKF in the 1/2 observation scenario; upper right: sample time series of the RMSE of the EAKF-RN ($\beta = 2$) under the same experiment settings as the EAKF; lower left: corresponding fraction coefficient $c_k$ in the EAKF-RN ($\beta = 2$); lower right: corresponding histogram of $c_k$.

between the EAKF and the EAKF-RN ($\beta = 2$), in terms of the RMSE of the EAKF minus that of the EAKF-RN, for $1 \leq k \leq 16$. The reason for not including the RMSE differences at larger time instants is that their amplitudes are relatively large and may make relatively small values indistinguishable from 0, which is not desired for our purpose. On the lower panel, we also show the fraction coefficients $c_k$ of the EAKF-RN ($\beta = 2$) for $1 \leq k \leq 25$. Note the availability of $c_k$ depends on the availability of the incoming observations; therefore, $c_k$ appear for every four steps only. Based on these figures, one may tell what happens to make the EAKF and EAKF-RN behave differently. At time step $k = 4$, there is an incoming observation. However, because $c_4 = 1$, the EAKF and EAKF-RN share identical estimates from $k = 1$ to $k = 7$. At $k = 8$, there is one more incoming observation, and this time $c_8$ is less than 1, meaning that residual nudging is effective, so that there is a (very) small difference spotted between the estimates of the EAKF and EAKF-RN. At $k = 12$, residual nudging is conducted again (but no more for subsequent steps up to $k = 24$), which, together with the previous residual nudging, makes the estimates of the EAKF-RN deviate from those of the EAKF and eventually avoid filter divergence at $k = 26$.

Overall, we have shown that when the normal EAKF is properly configured, the performance of the normal EAKF and the EAKF-RN is in general comparable. However, if the EAKF is not configured properly, then the EAKF-RN may perform (substantially) better than the normal EAKF. For many large scale data assimilation problems, it may be very expensive to conduct an extensive parameter searching in order to configure the EnKF (Anderson, 2007). Should the EnKF be ill-configured, we expect that introducing residual nudging to the EnKF may enhance its performance, in terms of filter accuracy and/or stability against divergence.

*4.2.3. Results with different ensemble sizes.* Here we examine the effect of the ensemble size $n$ on the performance of the normal EAKF and the EAKF-RN. The experiment settings are as follows. We also conduct the experiment in four observation scenarios. The ensemble size $n$ is chosen from the set {2,4,6,8,10,20,40,60,80}. In the experiment, we fix $l_c = 0.1$ and $\lambda = 1.15$ for both the normal EAKF and the EAKF-RN. In the EAKF-RN, we adopt two noise level coefficients, with $\beta$ being 1 and 2, respectively.
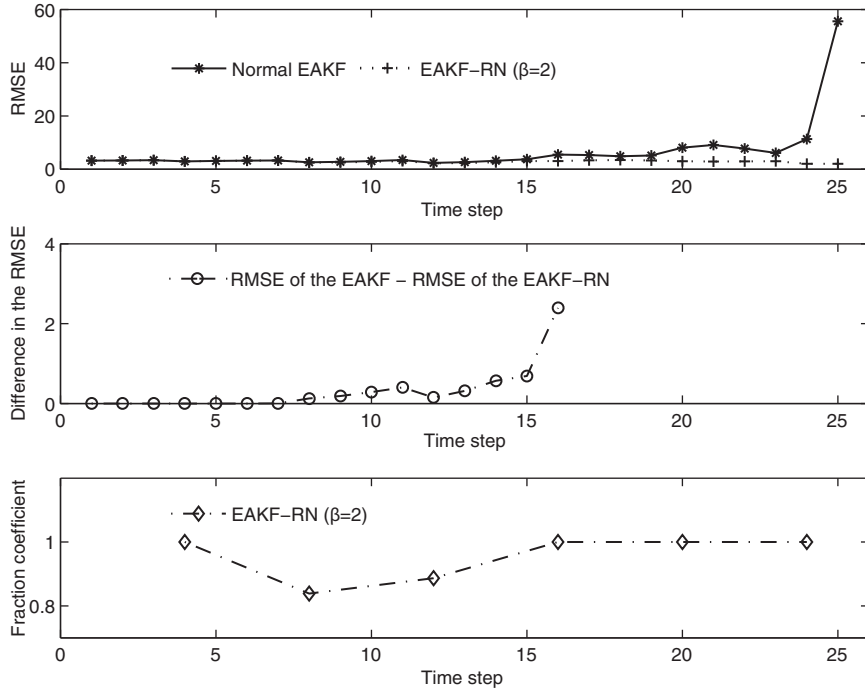
*Fig. 8.* Upper: the RMSE of the EAKF (solid line with asterisks) and EAKF-RN ($\beta = 2$, dotted line with plus signs) between the time instant $k = 1$ and $k = 25$; middle: difference in the RMSE ( = RMSE of the EAKF – RMSE of the EAKF-RN) between $k = 1$ and $k = 16$; lower: the fraction coefficient of the EAKF-RN ($\beta = 2$) between $k = 1$ and $k = 25$.
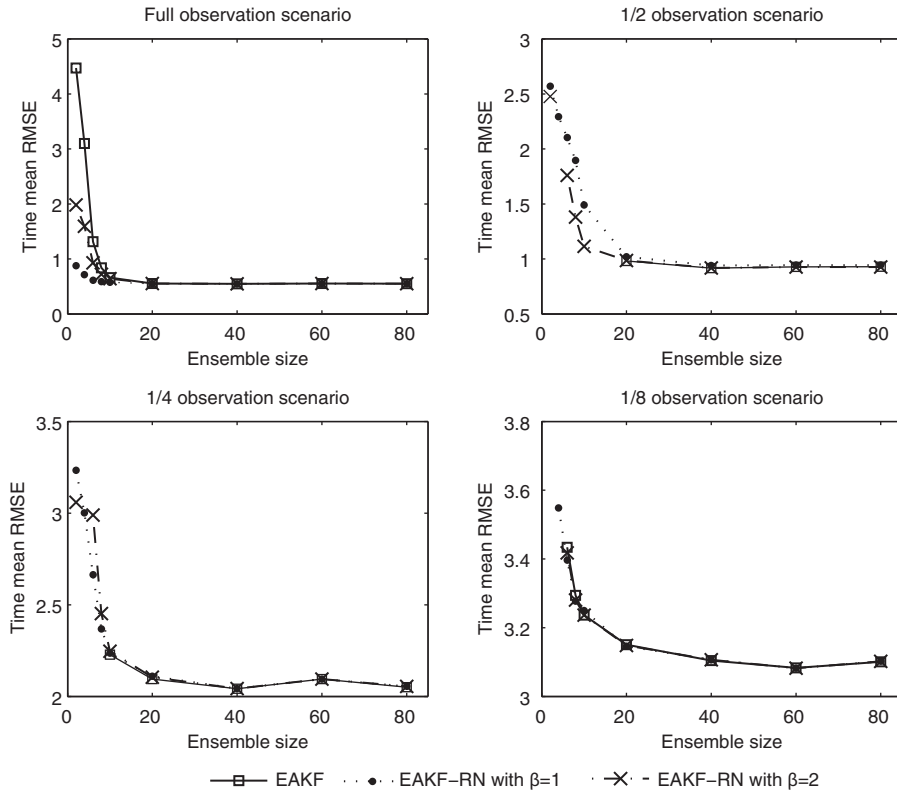


*Fig. 9.* Time mean RMSEs of the EAKF and the EAKF-RN, as functions of the ensemble size in different observation scenarios.

Figure 9 shows the time mean RMSEs of the normal EAKF (solid lines with squares) and those of the EAKF-RNs with $\beta = 1$ and 2 (dotted lines with bold points, and dash-dotted lines with crosses, respectively), in different observation scenarios. In the full observation scenario, no filter divergence is found for all the ensemble sizes $n$ in either filter. When $n \leq 10$, the EAKF-RN with $\beta = 1$ tends to perform better than the EAKF-RN with $\beta = 2$, while the latter is better than the normal EAKF. This is particularly the case with a relatively small ensemble size, for instance, at $n = 2$. On the other hand, when $n \geq 20$, the time mean RMSEs of the three filters are almost indistinguishable.

In the 1/2 observation scenario, the normal EAKF diverges when $n \leq 10$, so there are no square markers appearing at those $n$ values. The EAKF-RN with $\beta = 2$ appears more robust than the normal EAKF, although there is still a filter divergence spotted at $n = 4$. In contrast, the EAKF-RN with $\beta = 1$ is the most robust filter, which does not diverge for all the tested ensemble sizes. In terms of time mean RMSE, though, when the filters do not diverge, the EAKF-RN with $\beta = 1$ tends to perform worse than the EAKF-RN with $\beta = 2$, while the latter

appears to be indistinguishable from the normal EAKF for $n \geq 20$.

The situations in the 1/4 and 1/8 observation scenarios are similar to that in the 1/2 one. In the 1/4 observation scenario, the normal EAKF diverges for $n \leq 8$, while the EAKF-RN appears to be more robust, except that there is a filter divergence at $n = 4$ for the EAKF-RN with $\beta = 2$. When $n = 2$, the EAKF-RN with $\beta = 2$ performs better than the filter with $\beta = 1$, but at $n = 6$ or 8, the filter with $\beta = 1$ performs better instead. For $n \geq 10$, the performance of all three filters is almost indistinguishable. In the 1/8 observation scenario, the normal EAKF and the EAKF-RN with $\beta = 2$ diverge at $n = 2$ and 4, while the EAKF-RN with $\beta = 1$ diverges only at $n = 2$. For $n = 6$ or 8, the EAKF-RN with $\beta = 1$ has the best performance in terms of time mean RMSE, the EAKF-RN with $\beta = 2$ the second, while the normal EAKF the last. For $n \geq 10$, the performance of the three filters are almost indistinguishable, except that at $n = 10$, the time mean RMSE of the EAKF-RN with $\beta = 1$ is slightly higher than those of the other two filters.

The above results suggest that $n = 20$ appears to be a reasonable ensemble size for the normal EAKF in the
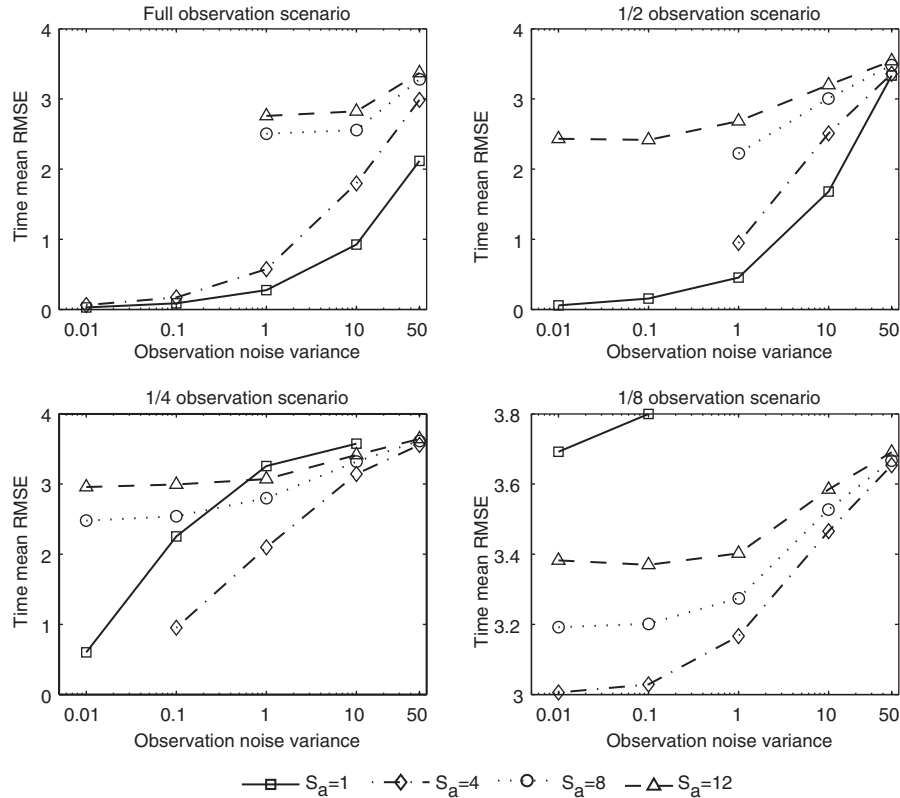


*Fig. 10.* Time mean RMSEs of the normal EAKF, as functions of the assimilation step $S_a$ and the observation noise variance, in different observation scenarios.

L96 model, since in all these four observation scenarios, the performance of the normal EAKF with $n = 20$ is very close to that with larger $n$ values. As the ensemble size $n$ decreases, the normal EAKF becomes more unstable. The performance of the EAKF-RN with $\beta = 1$ and $2$ is almost indistinguishable from the normal EAKF for $n \geq 20$. However, given smaller ensemble sizes, the EAKF-RN tends to perform better than the normal EAKF, in terms of both filter accuracy and stability against filter divergence. In particular, one may enhance the stability of the EAKF-RN by reducing the noise level coefficient $\beta$, since as $\beta \to 0$, the time mean RMSEs of the EAKF-RN in different observation scenarios become independent of the ensemble size $n$ and approach the corresponding values at $\beta = 0$. This property may be of interest in certain circumstances, for instance, those in which, due to practical limitations, one can only afford to run an EnKF with a very small ensemble size, so that filter stability becomes an important factor in consideration.

*4.2.4. Results with different assimilation steps and observation noise variances.* Here we examine the effects of the assimilation step $S_a$ and the observation noise variance on the performance of the normal EAKF and the EAKF-

RN. We assume that the observation noise covariance matrix $\mathbf{R}_k$ is in the form of $\gamma \mathbf{I}$, where $\mathbf{I}$ is the identity matrix with a suitable dimension in different observation scenarios, and $\gamma > 0$ is a real scalar. As a result, the variances of $\mathbf{R}_k$ are $\gamma$ for all variables in an observation vector, while the cross-variances are all zero. The experiment settings are as follows. The ensemble size is 20, $l_c = 0.1$ and $\lambda = 1.15$ for both the normal EAKF and the EAKF-RN. The noise level coefficients $\beta$ is 2 in the EAKF-RN. We conduct the experiment in four different observation scenarios and choose $S_a$ from the set $\{1, 4, 8, 12\}$, and $\gamma$ from the set $\{0.01, 0.1, 1, 10, 50\}$. The relatively large values of $\gamma$, for instance, $\gamma = 10, 50$, are used to represent the scenario in which the quality of the observations is relatively poor. Here we assume that we know the observation noise variance precisely, while in a subsequent experiment, we will consider the case in which the observation noise variance is mis-specified.

Figures 10 and 11 show the time mean RMSEs of the normal EAKF and the EAKF-RN, respectively, in different observation scenarios. In the full observation scenario (upper left panels), for a fixed variance $\gamma$, the time mean RMSEs of both the normal EAKF and the EAKF-RN tend to increase as the assimilation step $S_a$ increases. On the other hand, for a fixed $S_a$, the time mean RMSEs of both filters appear to be monotonically increasing functions of
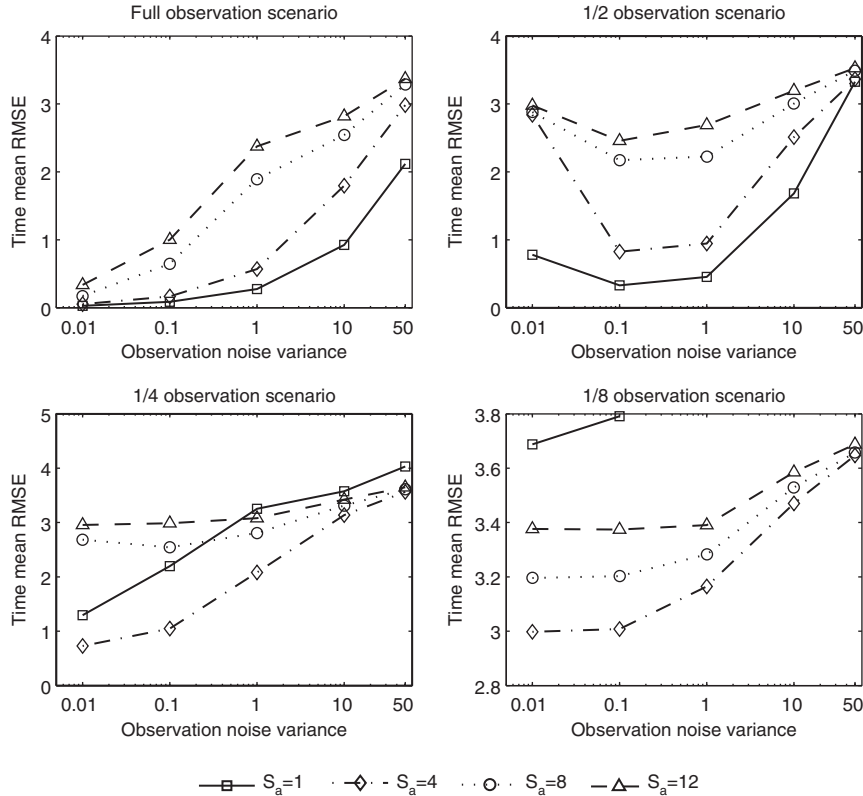


*Fig. 11.*    As in Fig. 10, but for the EAKF-RN with $\beta = 2$.

the variance $\gamma$. With $\gamma = 0.01, 0.1, 1$, the time mean RMSEs of the EAKF-RN tend to be lower than those of the normal EAKF, while with $\gamma = 10, 50$, they are almost indistinguishable, meaning that for relatively poor observation, the normal EAKF and the EAKF-RN have almost the same performance in terms of estimation accuracy, which appears to be also true in other observation scenarios, as will be shown below. In terms of filter stability, for $S_a = 8$ and 12, the normal EAKF diverges at $\gamma = 0.01$ and 0.1, but the EAKF-RN avoids filter divergences at all these places.

In the 1/2 observation scenario (upper right panels), for a fixed variance $\gamma$, the time mean RMSEs of both the normal EAKF and the EAKF-RN also grow as the assimilation step $S_a$ increases. However, for a fixed $S_a$, the time mean RMSEs of the two filters have behaviour different from that in the previous observation scenario. For $S_a = 1$, the time mean RMSE of the normal EAKF is still a monotonically increasing function of $\gamma$; for $S_a = 4, 8$, the normal EAKF diverges at $\gamma = 0.01$ and 0.1 and has monotonically increasing time mean RMSE for $\gamma \geq 1$; for $S_a = 12$, the time mean RMSE of the normal EAKF achieves its minimum at $\gamma = 0.1$ (slightly lower than that at 0.01) and thus exhibits

the U-turn behaviour, a phenomenon that is more visible in the EAKF-RN. Indeed, for all tested $S_a$ values, the time mean RMSEs of the EAKF-RN all have their minima at $\gamma = 0.1$, rather than at $\gamma = 0.01$. The normal EAKF and the EAKF-RN have almost indistinguishable time mean RMSEs for $\gamma \geq 1$. While the normal EAKF tends to perform better than the EAKF-RN at $\gamma = 0.01$ and 0.1 in terms of time mean RMSE, it is more likely to suffer from filter divergence (e.g. at $S_a = 4, 8$). This is an example of the trade-off between filter accuracy and stability, as discussed previously in section 2.3.

In the 1/4 observation scenario (lower left panels), for a fixed assimilation step $S_a$, the time mean RMSEs of both the normal EAKF and the EAKF-RN again appear to be monotonically increasing as $\gamma$ increases. For a fixed variance $\gamma$, though, the time mean RMSEs of both filters tend to exhibit the U-turn behaviour, in which the minimum time mean RMSE is achieved at $S_a = 4$ (except for the filter divergence in the normal EAKF at $\gamma = 0.01$), rather than at $S_a = 1$. The normal EAKF and the EAKF-RN have almost indistinguishable time mean RMSEs for $\gamma \geq 0.1$. At $\gamma = 0.01$, though, the normal EAKF seems to perform better than the EAKF-RN in terms of time mean
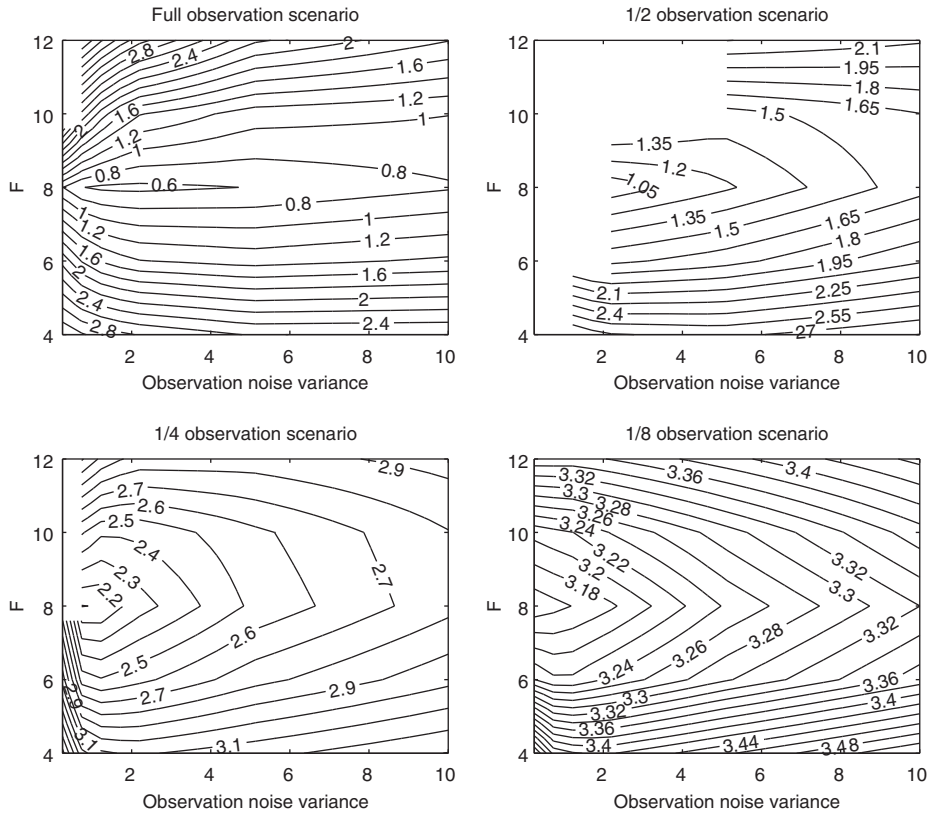


*Fig. 12.* Time mean RMSEs of the EAKF, as functions of the (possibly) mis-specified driving force $F$ and the observation noise variance $\gamma$, in different observation scenarios.
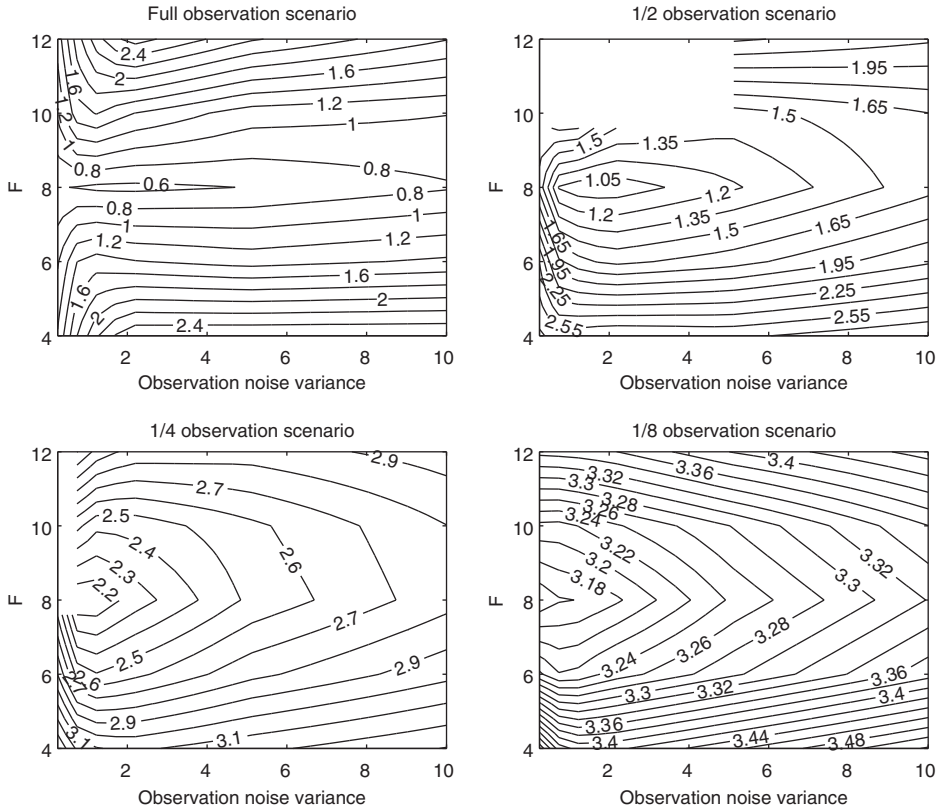
*Fig. 13.*    As in Fig. 12, but for the EAKF-RN with $\beta = 2$.

RMSE. However, filter divergences are spotted at $(S_a = 4, \gamma = 0.01)$ and $(S_a = 1, \gamma = 50)$, which are again avoided in the EAKF-RN.

In the 1/8 observation scenario (lower right panels), the quantitative behaviour of the two filters, as functions of $S_a$ and $\gamma$, is almost the same as that in the 1/4 observation scenario. The main differences are as follows. The time mean RMSEs of the normal EAKF and the EAKF-RN are almost indistinguishable in all tested cases. Filter divergences are spotted at $S_a = 1$, with $\gamma = 1, 10$ and 50, respectively, not only in the normal EAKF, but also in the EAKF-RN. One may, however, avoid these filter divergences in the EAKF-RN by assigning to it a smaller $\beta$, as some of the previous experiment results have suggested.

Overall, the above experiment results are consistent with our discussion in section 2.3. When equipped with residual nudging, the EAKF-RN appears to be more stable than the normal EAKF, although maybe at the cost of some loss of estimation accuracy in certain circumstances (e.g. when with too small $\beta$ values).

### 4.2.5. Results with imperfect models and mis-specified observation error covariances.
Finally, we examine filter performance of the normal EAKF and the EAKF-RN

when they are subject to uncertainties in specifying the forcing term $F$ in eq. (15) and the observation error covariance $\mathbf{R}_k$. We again conduct the experiments in four observation scenarios. The ensemble sizes of both filters are 20. The half-width $l_c$ of covariance localisation is 0.1, and the covariance inflation factor $\lambda$ is 1.15. The true value of $F$ is 8, while the true observation error covariance $\mathbf{R}_k$ is $\mathbf{I}_{20}$. In the experiments, we let the value of $F$ in the (possibly) imperfect model be chosen from the set $\{4,6,8,10,12\}$, and the (possibly) mis-specified covariance $\mathbf{R}_k$ in the form of $\gamma\mathbf{I}_{20}$, with $\gamma \in \{0.25, 0.5, 1, 2, 5, 10\}$.[6] In the EAKF-RN, the noise level coefficient $\beta = 2$.

Figures 12 and 13 show the time mean RMSEs of the normal EAKF and the EAKF-RN, respectively, as functions of the (possibly) mis-specified driving force $F$ and the observation noise variance $\gamma$, in different observation scenarios. In the full observation scenario (upper left panels), for a fixed $\gamma$, the time mean RMSEs of both filters exhibit the U-turn behaviour with respect to $F$, achieving their minima at $F = 8$. This point also appears to

---

[6]The (possibly) mis-specified observation error covariance, in the form of $\gamma\mathbf{I}_{20}$, is used for both background update, as described in section 2.1, and residual nudging through eq. (8).

be valid in other observation scenarios. On the other hand, for a fixed $F$, the behaviour of the filters is very similar to that reported in Figs. 5 and 6, since the role of the (possibly) mis-specified variance $\gamma$ is similar to the observation noise level coefficient $\beta$ (note, though, that $\gamma$ also appears in the computation of the Kalman gain). When $\gamma$ is relatively small (for instance, $\gamma \leq 2$), the EAKF-RN tends to perform better than the normal EAKF in terms of time mean RMSE. Moreover, the normal EAKF diverges at ($F = 12$, $\gamma = 0.25$), while the EAKF-RN avoids the divergence. On the other hand, when $\gamma$ is relatively large (for instance, $\gamma \geq 6$), the EAKF-RN and the normal EAKF have almost indistinguishable performance, not only for the current experiment results but also for those in the other observation scenarios. This is largely because mistakenly over-estimating the variance $\gamma$ has an effect similar to increasing $\beta$ so that the observation inversion in eq. (9a) becomes less influential for state estimation, and the EAKF-RN has almost the same estimate as the normal EAKF.

In the 1/2 observation scenario (upper right panels), when $\gamma$ is relatively small (for instance, $\gamma \leq 1$), the normal EAKF tends to diverge for all $F$. The EAKF-RN avoids filter divergences in some of the areas, though there are still two cases spotted at $F = 12$, with $\gamma = 1$ and $\gamma = 2$, respectively. As $\gamma$ becomes larger, the performance of the normal EAKF and the EAKF-RN is very close to each other, similar to the situation in the full observation scenario. In both the 1/4 and 1/8 observation scenarios (lower panels), there are also almost no differences between the time mean RMSEs of the two filters, although the time mean RMSE of the EAKF-RN appears to be slightly lower than that of the normal EAKF in the 1/4 observation scenario for relatively small $F$ and $\gamma$ (around the lower left corners). Both filters diverge in the 1/4 observation scenario, at ($F = 10$, $\gamma = 0.25$), otherwise neither filter diverges.

## 5. Discussion and conclusion

In this work, we proposed an auxiliary technique, that is, residual nudging, for ensemble Kalman filtering. The main idea of residual nudging is to monitor, and if necessary, adjust the residual norm of a state estimate. In an under-determined state estimation problem, if the residual norm is larger than a pre-specified value, then we reject the estimate and replace it by a new one whose residual norm is equal to the pre-specified value; otherwise we accept the estimate. We discussed how to choose the pre-specified value and demonstrated how one can construct a new state estimate based on the original one and the observation inversion, given a linear observation operator.

Through the numerical experiments in both the scalar AR1 and the Lorenz 96 models, we showed that by choosing a proper noise level coefficient, the EAKF-RN in general works more stable than the normal EAKF, while achieving an accuracy that is often comparable to, sometime even (much) better than, that of the normal EAKF, especially if the normal EAKF is ill-configured. This may occur, for instance, when the EAKF is equipped with improperly chosen covariance inflation factor and/or half-width of covariance localisation, too small ensemble size, and so on. In many data assimilation practices, it may be very expensive to conduct extensive searching for proper inflation factor and/or half-width, or to run a large scale model with too many ensemble members. In such circumstances, we expect that residual nudging may help improve the filter performance, in terms of filter stability, and even accuracy.

We also implemented residual nudging in some other filters, including the stochastic EnKF (Burgers et al., 1998) and the singular evolutive interpolated Kalman filter (SEIK) (Pham, 2001; Hoteit et al., 2002), and observed similar performance improvements (not shown in this work). Since residual nudging only aims to adjust the estimates, we envision that residual nudging can be associated with other data assimilation approaches, including, for instance, the extended KF, the particle filter, and various smoothers. This will be verified elsewhere.

One problem not addressed in this work is the non-linearity of the observation operator. In such circumstances, we conjecture that the rule in choosing the pre-specified value $\beta\sqrt{\text{trace}(\mathbf{R}_k)}$ may still be applicable. However, the construction of new state estimates would become more complicated than eqs. (8) and (9). One possible strategy is to linearise the observation operator or employ more sophisticated methods, such as iterative searching algorithms (see, for example, Gu and Oliver 2007; Lorentzen and Nævdal, 2011), to find new estimates whose residual norms are no larger than $\beta\sqrt{\text{trace}(\mathbf{R}_k)}$. This is another topic that will be investigated in the future.

## 6. Acknowledgements

## References

Anderson, J. L. 2001. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.* **129**, 2884–2903.

Anderson, J. L. 2007. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus.* **59A**(2), 210–224.

Anderson, J. L. 2009. Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus.* **61A**, 72–83.

Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the non-linear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758.

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

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

Engl, H. W., Hanke, M. and Neubauer, A. 2000. *Regularization of Inverse Problems.* Springer, New York.

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

Fertig, E., Harlim, J. and Hunt, B. 2007. A comparative study of 4D-VAR and a 4D ensemble Kalman filter: perfect model simulations with lorenz-96. *Tellus.* **59A**, 96–100.

Gandin, L. S. 1988. Complex quality control of meteorological observations. *Mon. Wea. Rev.* **116**, 1137–1156.

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

Gu, Y. and Oliver, D. 2007. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *SPE Journal.* **12**, 438–446.

Hamill, T. M., Whitaker, J. S. and Snyder, C. 2001. Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter. *Mon. Wea. Rev.* **129**, 2776–2790.

Hamill, T. M., Whitaker, J. S., Anderson, J. L. and Snyder, C. 2009. Comments on 'Sigma-point Kalman filter data assimilation methods for strongly nonlinear systems'. *J. Atmos. Sci.* **66**, 3498–3500.

Hoteit, I., Pham, D. T. and Blum, J. 2002. A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *J. Mar. Syst.* **36**, 101–127.

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

Hunt, B., Kalnay, E., Kostelich, E., Ott, E., Patil, D. and co-authors. Four-dimensional ensemble Kalman filtering. *Tellus.* **56A**(4), 273–277.

Jazwinski, A. H. 1970. *Stochastic Processes and Filtering Theory.* Academic Press, San Diego.

Kalman, R. 1960. A new approach to linear filtering and prediction problems. *Trans. ASME, Ser. D, J. Basic Eng.* **82**, 35–45.

Lorentzen, R. and Nævdal, G. 2011. An iterative ensemble Kalman filter. *IEEE Trans. Automa. Contr.* **56**, 1990–1995.

Lorenz, E. N. 1996. Predictability-a problem partly solved. In: *Predictability* (ed. T. Palmer). ECMWF, Reading, UK, pp. 1–18.

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

Luo, X. and Hoteit, I. 2011. Robust ensemble filtering and its relation to covariance inflation in the ensemble Kalman filter. *Mon. Wea. Rev.* **139**, 3938–3953.

Meyer, C. D. 2001. Matrix analysis and applied linear algebra. SIAM, Philadelphia.

Pham, D. T. 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.* **129**, 1194–1207.

Sacher, W. and Bartello, P. 2008. Sampling errors in ensemble Kalman filtering. Part I: Theory. *Mon. Wea. Rev.* **136**(8), 3035–3049.

Schlee, F. H., Standish, C. J. and Toda, N. F. 1967. Divergence in the Kalman filter. *AIAA Journal.* **5**, 1114–1120.

Song, H., Hoteit, I., Cornuelle, B. and Subramanian, A. 2010. An adaptive approach to mitigate background covariance limitations in the ensemble Kalman filter. *Mon. Wea. Rev.* **138**(7), 2825–2845.

Van Leeuwen, P. J. 2010. Nonlinear data assimilation in geosciences: An extremely efficient particle filter. *Quart. J. Roy. Meteor. Soc.* **136**, 1991–1999.

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