Handling Nonlinearity in an Ensemble Kalman Filter: Experiments with the Three-Variable Lorenz Model

SHU-CHIH YANG

Department of Atmospheric Sciences, National Central University, Jhongli, Taiwan

EUGENIA KALNAY

Department of Atmospheric and Oceanic Science, University of Maryland, College Park, College Park, Maryland

BRIAN HUNT

Department of Mathematics, University of Maryland, College Park, College Park, Maryland

(Manuscript received 3 August 2010, in final form 17 January 2012)

ABSTRACT

An ensemble Kalman filter (EnKF) is optimal only for linear models because it assumes Gaussian distributions. A new type of outer loop, different from the one used in 3D and 4D variational data assimilation (Var), is proposed for EnKF to improve its ability to handle nonlinear dynamics, especially for long assimilation windows. The idea of the ''running in place'' (RIP) algorithm is to increase the observation influence by reusing observations when there is strong nonlinear error growth, and thus improve the ensemble mean and perturbations within the local ensemble transform Kalman filter (LETKF) framework. The ''quasi-outer-loop'' (QOL) algorithm, proposed here as a simplified version of RIP, aims to improve the ensemble mean so that ensemble perturbations are centered at a more accurate state.

The performances of LETKF–RIP and LETKF–QOL in the presence of nonlinearities are tested with the three-variable Lorenz model. Results show that RIP and QOL allow LETKF to use longer assimilation windows with significant improvement of the analysis accuracy during periods of high nonlinear growth. For low-frequency observations (every 25 time steps, leading to long assimilation windows), and using the optimal inflation, the standard LETKF RMS error is 0.68, whereas for QOL and RIP the RMS errors are 0.47 and 0.35, respectively. This can be compared to the best 4D-Var analysis error of 0.53, obtained by using both the optimal long assimilation windows (75 time steps) and quasi-static variational analysis.

1. Introduction

Data assimilation involving nonlinear perturbations becomes problematic in both incremental variational methods, which use linear adjoint models and observation operators to evolve model and observation perturbations, and ensemble Kalman filters (EnKFs), which assume that ensemble perturbations are Gaussian. The nonlinear growth of perturbations can result from strong dynamical instabilities, observations that have an insufficient sampling frequency, the use of nonlinear observation operators, or even from model errors. When strong nonlinearities occur, a data assimilation scheme may lose its ability to track the true dynamics, a problem known as ''filter divergence'' (Miller et al. 1994; Evensen 1992) within Kalman filter–based assimilation schemes (Kalman 1960). Also, with a non-Gaussian ensemble distribution, EnKF cannot give an optimal solution (Evensen and van Leeuwen 2000). In this study, we focus on the nonlinearity since it plays a dominant role in the filter failure, despite the fact that the filter divergence can also take place in a linear system.

In the Kalman filter formulas derived using the best linear unbiased estimation (BLUE), the error covariance is evolved with the tangent linear model. Evensen (1992)

Corresponding author address: Shu-Chih Yang, Department of Atmospheric Sciences, National Central University, Jhongli 32001, Taiwan.
E-mail: shuchih.yang@atm.ncu.edu.tw

DOI: 10.1175/MWR-D-11-00313.1

pointed out that unbounded error growth could be caused by the use of the tangent linear model because of the lack of nonlinear saturation effects. EnKFs are less vulnerable to this error growth than either the Kalman filter or the extended Kalman filter because EnKFs have the advantage of using the full nonlinear model that includes the saturation of nonlinear perturbation growth (Evensen 1994, 1997; Verlaan and Heemink 2001). Nevertheless, filter divergence can still take place when nonlinearities appear and result in a non-Gaussian distribution of the ensemble (Evensen 1997; Lawson and Hansen 2004). As a result, a short assimilation window is required for EnKF to better preserve the Gaussianity of the ensemble. Fertig et al. (2007) compared four-dimensional variational data assimilation (4D-Var) and the EnKF implemented in the Lorenz 40-variable model (Lorenz 1996). They showed that the EnKF did very well with short windows but did not handle long windows well because the ensemble perturbations became non-Gaussian. Studies also show that EnKFs characterized by being either deterministic or stochastic (Tippett et al. 2003) may have different abilities in handling nonlinearities and non-Gaussianity. Lawson and Hansen (2004) compared nonlinearities handled by the ensemble square root filter (EnSRF; Whitaker and Hamill 2002) and the perturbed observation EnKF (Burgers et al. 1998; Houtekamer and Mitchell 1998; Evensen 2003), representing prototypes of deterministic and stochastic EnKFs, respectively. They found that as nonlinearity becomes significant, square root (deterministic) filters break down earlier. They showed that the stochastic EnKF effectively does a resampling to get closer to the dominant mode of the a posteriori probability distribution function (PDF), while the deterministic EnKF cannot do this even though the analysis ensembles from these two EnKF schemes are both linearly transformed from the background ensemble. Not directly affected by the magnitude of the nonlinearity, the filter failure can also occur gradually in a system with weak nonlinearity after successive updates with non-Gaussian ensembles, which results in far outliers (Sakov and Oke 2008). In addition, Leeuwenburgh et al. (2005) showed that the deterministic EnSRF tends to introduce non-Gaussianity, but that using a random rotation step in EnSRF could alleviate this. Outliers in the EnSRF systems can also arise from the use of a one-sided, nonsymmetric solution in the calculation of the square root of a matrix for deriving the analysis perturbations (Sakov and Oke 2008), but the LETKF uses the symmetric square root solution (Ott et al. 2004; Hunt et al. 2007).

EnKFs need to use short windows, but in contrast, 4D-Var performs better with longer windows because it is iterated within a window, producing a more accurate nonlinear trajectory under the constraint of the observations. In the incremental 4D-Var framework (Courtier et al. 1994), an outer loop is applied to adjust the nonlinear evolution of the trajectory with the high-resolution nonlinear model with complete physics, while the inner loop is used to minimize the incremental cost function with the adjoint model derived from a low-resolution simplified model (Andersson et al. 2005). The 4D-Var outer loop aims to improve the nonlinear evolution of the model state by improving the sensitivity matrix of the linearization of the observation operator and the background trajectory used for computing the innovation vector. The 4D-Var outer loop is essential for nonlinear cases to improve the linear approximation and derive the correct gradient for minimizing the cost function. During the iterations of minimization, the first guess of the initial state (prior state) and its corresponding error statistics are kept constant. An important reason for the failure of EnKFs with long windows is that EnKFs do not have an outer loop to deal with nonlinear dynamics like the incremental 3D-Var and 4D-Var frameworks do (A. DaSilva 2006, personal communication; Kalnay et al. 2007a).

Inspired by the advantages shown by the outer loop in improving the nonlinear trajectory in 4D-Var (Courtier et al. 1994; Rabier et al. 2000; Andersson et al. 2005), here we introduce a different type of ''outer loop'' within the EnKF framework and implement it for the local ensemble transform Kalman filter (LETKF; Hunt et al. 2007). The idea is to increase the influence of the observations for cases with nonlinear error growth, when the estimation given by the background mean is not reliable and the ensemble perturbations are distorted by strong nonlinearities. In these cases, the ensemble perturbations cannot represent the uncertainties of the prior state, and therefore the observation information cannot be effectively extracted to correct the model state. In these cases, the observation impact in the standard LETKF is clearly suboptimal. A ''hard'' way to increase the observation influence would be to artificially reduce the observation error before assimilating the observation. A ''soft'' way is to use the original observation error, and assimilate this observation multiple times so that the total analysis increment is obtained as the sum of multiple smaller increments. Thus, the rule of ''using the observations only once'' (Ide et al. 1997) is modified by reducing the amplitude of the background error covariance (section 2). We note that the soft way of using small increments becomes an additional advantage for nonlinear dynamics because the system can remain closer to the true nonlinear attractor with smaller steps.

The ''running in place'' (RIP) scheme recently developed by Kalnay and Yang (2008, 2010) is able to increase the observation influence and improve the ability of EnKF to deal with nonlinearity. The RIP method updates both the ensemble mean and the ensemble perturbations and

was originally proposed to capture the underlying evolving dynamics to accelerate the spinup of EnKFs. The observations are repeatedly assimilated and the ensemble spread is self-adjusted (reduced). Here we further propose a ''quasi-outer-loop'' (QOL) algorithm as a simplified version of the RIP, aiming to improve the ensemble mean so that ensemble perturbations are centered at a more accurate state. Both the RIP and the QOL methods can be applied to improve the LETKF scheme for nonlinear cases, like a ''cold start'' of an EnKF or when the background error statistics suddenly change. An example of a cold start is the initialization of regional data assimilation from a global analysis obtained at coarser resolution, thus lacking features that represent the underlying mesoscale evolution. Also, when the model trajectory encounters a rapid change into a new regime with a short transition period (such as in the development of a severe storm), this may result in a sudden change of the background trajectory and error statistics. In these cases the newly proposed QOL scheme aims to improve the nonlinear trajectory of the ensemble mean while the RIP further reduces strong nonlinearities in ensemble perturbations when they do not represent the true dynamical uncertainties.

We also note that other iterative methods related to the Kalman filter have been developed to deal with nonlinearities or long assimilation windows. For example, the iterated Kalman filter (IKF; Bell and Cathey 1993; Bell 1994), using an incremental form similar to the incremental variational methods, has been proposed for handling nonlinearities better than the extended Kalman filter by accounting for the nonlinearity of the observation operator. Bell and Cathey (1993) showed that IKF is equivalent to using the Gauss–Newton method for constructing the maximum likelihood estimate. However, as in the incremental 4D-Var outer loop, the IKF only modifies the sensitivity matrix related to the observation operators and does not change the initial guess and the corresponding error statistics of the background state during the iterations for minimization. For reservoir engineering applications, several iteration-based ensemble Kalman filters (IEnKF)–smoothers (Gu and Oliver 2007; Krymskaya et al. 2009; Li and Reynolds 2009; Wang et al. 2010) were recently proposed for parameter estimation and to deal with the nonlinearity arising from the historical matching problem. Among them, the ensemble randomized maximum likelihood (EnRML) by Gu and Oliver (2007) shares the concept of an outer loop similar to the incremental 4D-Var. In these methods, including the variational methods, the problem of nonlinearity is solved by iterating the linear solution for a cost function (e.g., finding the optimum with the Gauss–Newton method). At each iteration, the model trajectory is improved until the linear approximation of the observation operator (sensitivity matrix) is close enough to the true state, and the linear solution becomes optimal. When such a sensitivity is known at the very first iteration, like in the linear case, the system could yield an optimal update with just one iteration. However, we should note that such a framework for outer looping is more complicated within the EnKF systems because the sensitivities incorporated with the error statistics are entangled with the ensemble state. In the EnKF, we cannot correct the sensitivities (and the ensemble-based estimated covariances) without correcting the model state. For example, the EnRML scheme reevaluates the sensitivity matrix based on the ''re-evolved'' ensemble at each iteration.

In the RIP (Kalnay and Yang 2010) and proposed QOL schemes, by contrast, we take advantage of a ''no-cost'' smoother for ETKF/LETKF (Kalnay et al. 2007b) to improve not the sensitivity but the full initial ensemble (RIP) or just the initial ensemble mean (QOL), based on the knowledge about the state from the ''future'' observations within the window. We will show that by using observations multiple times, RIP/QOL reduces the spinup and attains the same optimal analysis as the regular Kalman filter even in the case of a linear system. In a nonlinear system there is the additional advantage of correcting the ensemble incrementally following a trajectory closer to the true dynamics. We will show that this improves the analysis beyond what can be obtained under the framework of the maximum likelihood method with the Gauss–Newton method (e.g., the 4D-Var outer loop or EnRML).

The paper is organized as follows: in section 2 the KF-RIP method is compared with the standard KF for a simple linear model. Section 3 introduces the standard LETKF and the RIP and QOL methods within the LETKF framework. Assimilation experiments testing the impact of QOL and RIP on nonlinearity with the Lorenz three-variable model (Lorenz 1963) are presented in section 4. In section 5, the ability to handle the nonlinearities of RIP and QOL, which modify the prior ensemble, is compared with EnRML, which like the standard 4D-Var outer loop does not modify the prior state. Finally, section 6 contains a summary and discussion of the theoretical framework for RIP and QOL.

2. Basic RIP in a linear model

The full RIP algorithm implemented with the LETKF framework is discussed in detail in section 3. Here the basic RIP with a constant number of iterations per analysis cycle is presented in the framework of the Kalman filter with linear dynamics, for which KF is optimal. We show that with a linear model, using RIP produces the same analysis means after spinup as not using RIP. The main differences are that with RIP the estimated error

variance is systematically smaller, the total analysis increment in a given analysis cycle is applied as the sum of multiple smaller increments, and the spinup time is reduced.

With linear dynamics, the RIP algorithm is the same as assimilating observations multiple times, and the background at each iteration is provided by the analysis derived from the previous iteration. We acknowledge that if the background error covariance used in a data assimilation procedure is commensurate with the typical error in the background state, then using observations more than once in a given assimilation cycle underestimates the observation error and can lead to overfitting the observations. However, in a procedure based on the Kalman filter, if all observations are used more than once in all assimilation cycles, then the background and analysis error covariances are also underestimated after spinup. As a result, in the linear scenario, if each observation is used the same number of times, the resulting analyses will in the long run be essentially the same as in the optimal KF solution, where each observation is used only once (P. Sakov 2010, personal communication).

We apply a basic RIP scheme to a simple linear model with linear error growth and compare its results to the optimal solution of the Kalman filter. The simple linear model has the following governing equation for a scalar x:

$$x_n = M(x_{n-1}) = Cx_{n-1}, \tag{1}$$

where n indicates the time step. The error variance growth is also linear:

$$\sigma_n^2 = G(\sigma_{n-1}^2) = C^2\sigma_{n-1}^2, \tag{2}$$

where C is a constant. In the following, we will use the operators M and G to denote the forward integration for the model state and error variance, respectively.

The Kalman filter analysis and estimated error variance are

$$x_{a,n} = x_{b,n} + \left(\frac{\sigma_{b,n}^2}{\sigma_{b,n}^2 + \sigma_o^2}\right)(y_n - x_{b,n}) \tag{3}$$

and

$$\sigma_{a,n}^2 = \frac{\sigma_{b,n}^2\sigma_o^2}{\sigma_{b,n}^2 + \sigma_o^2}, \tag{4}$$

where $x_{a,n}$, $x_{b,n}$, and y_n are, respectively, the values of the analysis, background, and observation at time t_n and $\sigma_{a,n}^2$, $\sigma_{b,n}^2$, and σ_o^2 are the corresponding error variances. We note that the observation operator is the identity in this example.

In the RIP algorithm (Kalnay and Yang 2010; see section 3c) with N iterations, the background state at each iteration i is evolved from the "smoothed" analysis at t_{n-1} that has already assimilated the observation at a later time t_n. With linear dynamics, the forecast from the smoothed analysis at t_{n-1} coincides with the KF analysis at t_n (Yang et al. 2009a). Therefore, the KF–RIP procedure with linear dynamics can be reduced to repeatedly assimilating observations N times at t_n without involving the smoothed analysis at t_{n-1}. With RIP, the analysis at the ith iteration is

$$\tilde{x}_{a,n}^i = \tilde{x}_{a,n}^{i-1} + \left[\frac{(\tilde{\sigma}_{a,n}^{i-1})^2}{(\tilde{\sigma}_{a,n}^{i-1})^2 + \sigma_o^2}\right](y_n - \tilde{x}_{a,n}^{i-1}). \tag{5}$$

Here, the tilde is used to indicate that the observation has been used more than once.

Similarly, the estimated analysis error variance, $(\tilde{\sigma}_{a,n}^i)^2$, at t_n is

$$(\tilde{\sigma}_{a,n}^i)^2 = \frac{(\tilde{\sigma}_{a,n}^{i-1})^2\sigma_o^2}{(\tilde{\sigma}_{a,n}^{i-1})^2 + \sigma_o^2}. \tag{6}$$

The iteration starts with $\tilde{x}_{a,n}^0 = x_{b,n}$ and $(\tilde{\sigma}_{a,n}^0)^2 = (\sigma_{b,n})^2$. If there is only one iteration, $\tilde{x}_{a,n}^1$ and $(\tilde{\sigma}_{a,n}^1)^2$ give the same analysis and error variance as the KF solution.

In the following perfect model experiment, we choose $C = 1.25$ but the results are rather insensitive to this choice. The truth starts from $x_0 = 0$, and an observation is created at every time step with an error variance of 1. The initial conditions at $t = 0$ for the analysis value and error variance are 30 and 5, respectively, but the results derived with this simple model are not sensitive to these initial values either. We choose the initial analysis far from the truth to illustrate how RIP recovers more quickly from the poorly specified initial conditions. After spinup, which takes less than 20 analysis cycles, the KF analysis has a mean accuracy of 0.6175, and the mean estimated analysis error variance is 0.36.

Figure 1a compares the results from KF and RIP using either 2 or 10 iterations. With one additional iteration (KF is considered to be the first iteration), the RIP analysis reaches the same accuracy as the optimal KF analysis after the spinup period (defined as the first 15 analysis cycles), and the estimated analysis error variance is 0.18, half as large as the error variance estimated with KF. With 10 iterations, the RIP analysis accuracy also reaches the same value, but the estimated error variance is 0.036, one-tenth of the KF estimation. Figure 1b shows that the estimated analysis error variance from KF-RIP converges from the large value and becomes smaller with the number of iterations. The estimated

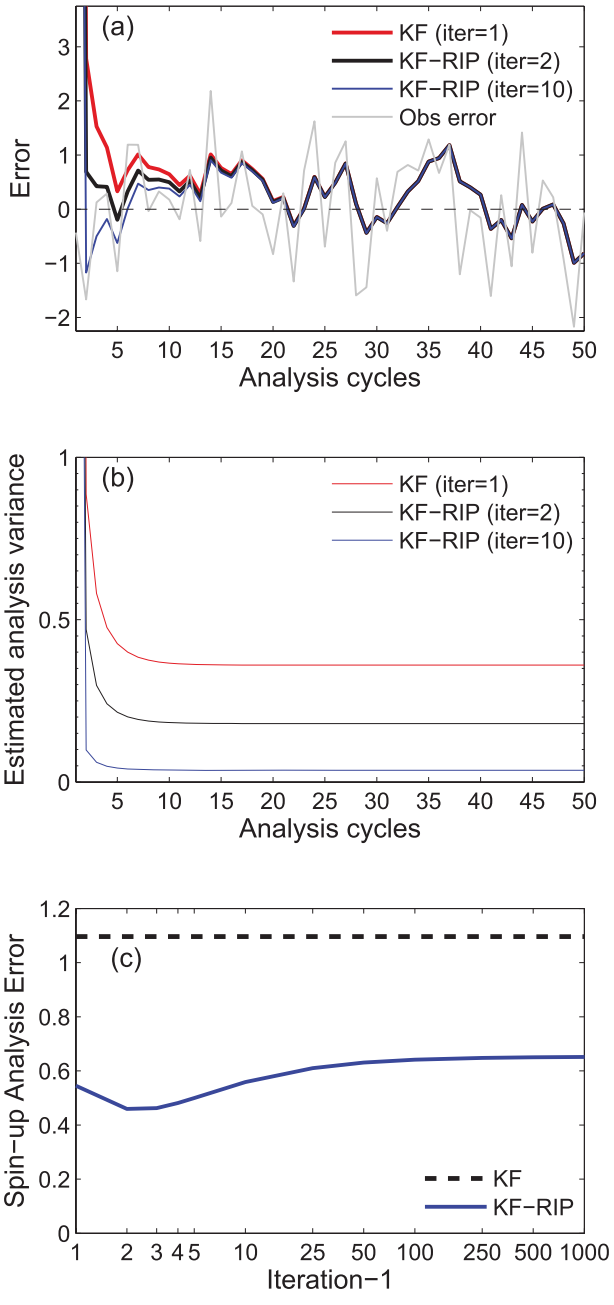

FIG. 1. (a) Analysis error of KF and KF–RIP with 2 and 10 iterations and Gaussian observation errors with $\sigma_o = 1$. (b) Estimated analysis error variance of KF and KF–RIP with 2 and 10 iterations. (c) RMS analysis error from KF and KF–RIP computed with different iteration numbers and averaged for the spinup period (the first 15 analysis cycles).

error variance is the same as would be computed from KF with higher assumed observation accuracy, with the observation error variance σ_o^2 replaced by σ_o^2/N.

As shown in Fig. 1a, the RIP analysis with linear dynamics is not sensitive to the number of iterations beyond the spinup. Although RIP after the second iteration no longer fulfills the assumption that the observation and background errors are uncorrelated, the results with linear dynamics confirm that the RIP algorithm assimilating observations multiple times gives the same KF analysis accuracy. We also note that RIP shows a clear advantage in accelerating the spinup (the first 15 analysis cycles in Fig. 1a) when increasing the number of iterations (Kalnay and Yang 2010) from 1 to 10. Figure 1c shows that for this simple model the optimal number of iterations during spinup (first 15 cycles) is three–four, but the advantage of KF-RIP in reducing spinup error is very apparent even with just two iterations.

Based on the discussions above, we conclude that the Kalman filter performed with an observation error variance divided by N can be viewed as a ''hard'' way to increase the observation influence since the analysis correction is computed all at once. KF–RIP can then be viewed as a ''softer'' way to achieve the same analysis correction but with multistep analysis corrections. We also argue that the RIP algorithm should be advantageous in nonlinear cases since the analysis increment is obtained in small steps and thus is better able to follow the true nonlinear dynamics.

3. The RIP and QOL algorithms for LETKF

Different from the linear KF–RIP procedure illustrated in section 2, there are two other important steps in the ensemble-based KF–RIP scheme for nonlinear cases including the use of a no-cost smoother and being able to improve the nonlinear trajectory by re-evolving it with the full nonlinear model. With LETKF, future information can be used earlier by means of the no-cost smoother. In this section we describe LETKF and the no-cost smoother valid at the beginning of the assimilation window, and apply the smoother to derive the RIP and QOL algorithms for LETKF.

a. LETKF

The LETKF scheme (Hunt et al. 2007) performs an analysis locally in space using local information, including the background state (short-range forecasts) and observations. The analysis correction by LETKF is calculated within the space spanned by the local ensemble. Here we describe the three-dimensional LETKF (i.e., with all observations available at the end of the assimilation window). The four-dimensional extension of LETKF (where observations are available throughout the assimilation window) is straightforward (Hunt et al. 2004, 2007). We note that in this study no localization is required because of the small dimension of the Lorenz three-variable model, and therefore, the variables denote the globally

defined model states or the observations. Without localization, LETKF is formally equivalent to the ensemble transform Kalman filter (Bishop et al. 2001) but with a centered spherical simplex ensemble (Wang et al. 2004; Ott et al. 2004). Since the notation used in the two methods is different, here we follow Hunt et al. (2007).

At the analysis time, the LETKF computes a weight matrix to linearly combine the background ensemble members so that the mean and error covariance of the analysis ensemble agree with the Kalman filter formulas. With K ensemble members, the analysis ensemble perturbations at analysis time t_n are computed with a right-multiplied transform of the background perturbation matrix:

$$\mathbf{X}_n^a = \mathbf{X}_n^b \mathbf{W}_n^a. \tag{7}$$

Here, $\mathbf{X}_n^b = (\delta\mathbf{x}_n^{b,1} | \ldots . | \delta\mathbf{x}_n^{b,K})$ is the matrix of background perturbations whose columns are the vectors of ensemble member perturbations (deviations from the mean), and $\delta\mathbf{x}_n^{b,k} = \mathbf{x}_n^{b,k} - \overline{\mathbf{x}}_n^b$, where $\mathbf{x}_n^{b,k}$ is the kth background ensemble member and $\overline{\mathbf{x}}_n^b$ is the background ensemble mean. The analysis perturbations are represented in a similar way. The analysis weight perturbation matrix $\mathbf{W}_n^a$ is obtained from

$$\mathbf{W}_n^a = [(K-1)\hat{\mathbf{P}}_n^a]^{1/2}, \tag{8}$$

where

$$\hat{\mathbf{P}}_n^a = [(K-1)\mathbf{I}/\rho + \mathbf{Y}_n^{b\mathrm{T}} \mathbf{R}^{-1} \mathbf{Y}_n^b]^{-1} \tag{9}$$

is the analysis error covariance in the ensemble space, $\mathbf{Y}_n^b = (\delta\mathbf{y}_n^{b,1} | \ldots . | \delta\mathbf{y}_n^{b,K})$ is the matrix of background perturbations in the observation space, $\delta\mathbf{y}_n^{b,k} = h(\mathbf{x}_n^{b,k}) - h(\mathbf{x}_n^{b,k})$, $\mathbf{R}$ is the observation error covariance, $h(\cdot)$ is the observation operator that converts a variable from the model to the observation space, and ρ is the multiplicative covariance inflation, which in Eq. (9) has essentially the same effect as inflating the amplitude of the background error covariance (Hunt et al. 2007). We note that the nonlinear observation operator $h(\cdot)$ is also an important source of nonlinearities for observations like satellite radiances or scatterometer backscatter. With LETKF, such nonlinear effects are partially taken into account by using the nonlinear observation operator, in contrast to variational methods where the adjoints of the linearized observation operators (Jacobians) are required to minimize the cost function.

The ensemble-mean analysis at time t_n is obtained from

$$\overline{\mathbf{x}}_n^a = \mathbf{X}_n^b \overline{\mathbf{w}}_n^a + \overline{\mathbf{x}}_n^b, \tag{10}$$

where

$$\overline{\mathbf{w}}_n^a = \hat{\mathbf{P}}_n^a \mathbf{Y}_n^{b\mathrm{T}} \mathbf{R}^{-1} (\mathbf{y}_n^o - \overline{\mathbf{y}}_n^b). \tag{11}$$

In Eq. (11), $\mathbf{y}_n^o$ and $\overline{\mathbf{y}}_n^b$ are vectors of the observations and the background mean state in observation space. The difference between them, $\mathbf{d}_n = \mathbf{y}_n^o - \overline{\mathbf{y}}_n^b$, is referred to as the innovation vector. As pointed out by Ott et al. (2004) and Hunt et al. (2007), the use of a symmetric square root algorithm in Eq. (8) ensures that among all possible square roots, $\mathbf{W}^a$ is the matrix closest to the identity, and that it varies smoothly in space and time. The accuracy of the LETKF analysis depends on the accuracy of both the weight matrix $\mathbf{W}^a$, associated with the flow-dependent errors (or "errors of the day"), and the weight vector $\overline{\mathbf{w}}^a$, used to estimate the mean analysis correction. The properties of these weights are important for the performance of LETKF. Yang et al. (2009b) showed that the weights vary on larger spatial scales than those of either the analyses or the analysis increments, so that interpolating the weights derived from a coarse-resolution LETKF can still retain the accuracy of the full-resolution LETKF analysis.

b. The no-cost smoother

In Eqs. (7) and (10), the weights linearly combine the background ensemble trajectories in such a way that they are closest to the true atmosphere at the analysis time t_n. A linear combination of ensemble trajectories is also an approximate model trajectory, and if it is close to the truth at the analysis time it should also be close to the truth throughout the assimilation window ($t_{n-1} - t_n$), at least to the extent that we can neglect nonlinearities and model errors. Therefore, the same weights obtained at the analysis time should be valid throughout the assimilation window. This allows for constructing a no-cost smoother for certain types of EnKFs (including LETKF) that improves the analysis state at the beginning of the assimilation window from observations obtained later within the window, without the need of an adjoint model (Kalnay et al. 2007b; Yang et al. 2009a; Kalnay and Yang 2010, their first footnote). Such ensemble-based smoothers for asynchronous assimilation are applicable to any EnKF scheme that expresses the updated ensemble via right-multiplied ensemble transforms as in Eq. (7), depending on how nonlinearity affects the time correlation between the model states. Thus, the form of the no-cost smoother can be compared with the ensemble Kalman smoother (EnKS) used in Evensen and van Leeuwen (2000) and Evensen (2003, his appendix D) based on the solution of the stochastic (perturbed observations) EnKF. Specifically, the no-cost ETKF/LETKF smoother has the lagged-1 form of EnKS in Evensen and van Leeuwen

(2000). In both methods the analysis ensemble is obtained by linearly combining the background ensemble, but the ensemble perturbations in the perturbed observation EnKF are updated differently.

In the no-cost smoother, we apply the weights derived at the end of the window t_n to the analysis ensemble derived at the previous analysis time t_{n-1} so that

$$\tilde{\mathbf{x}}^a_{n-1} = \overline{\mathbf{x}}^a_{n-1} + \mathbf{X}^a_{n-1}\overline{\mathbf{w}}^a_n \tag{12a}$$

is used to smooth the analysis mean, and

$$\tilde{\mathbf{X}}^a_{n-1} = \mathbf{X}^a_{n-1}\mathbf{W}^a_n \tag{12b}$$

is used to smooth the analysis perturbations. More specifically, the background is taken from the analysis derived with the available observation at t_{n-1} and does not know the information at t_n. The tilde indicates the use of the later information at t_n.

Kalnay and Yang (2010) show that such a smoothed analysis is always more accurate than the original LETKF analysis because of its ''knowledge'' of the observations made later. As in the case of other smoothers under the linear assumption, $M(\tilde{\mathbf{x}}^a_{n-1}) = \overline{\mathbf{x}}^a_n$ and $M(\tilde{\mathbf{X}}^a_{n-1}) = \mathbf{X}^a_n$ because the forecast from the smoothed analysis valid at t_{n-1} coincides with the EnKF analysis at t_n (see appendix A in Yang et al. 2009a). Thus, the ensemble-based Kalman smoother and the RIP and QOL methods discussed below could also be applied to any framework of the ensemble-based Kalman filter that expresses the updated ensemble via right-multiplied ensemble transforms. Although in this study we assimilate observations available at the end of the assimilation window, the RIP and QOL methods are equally applicable if observations are distributed throughout the assimilation window (see Hunt et al. 2004, 2007).

c. RIP algorithm

The linear KF–RIP used in section 2 is now modified for nonlinear dynamics in the EnKF framework. Because of nonlinearities, EnKF and EnKF–RIP do not give the same results but, as argued in section 2, EnKF-RIP is better able to handle nonlinearities because of the use of a smaller ensemble spread and a multistep analysis increment (i.e., it provides a softer way to increase the influence of observations). In the same way that the optimal linear formulation of EnKF benefits from ad hoc modifications such as covariance inflation, EnKF–RIP benefits from modifications such as adaptive estimation of the number of iterations per analysis cycle and the addition of small perturbations.

The RIP scheme was originally proposed to accelerate the spinup of the ensemble-based Kalman filter. This spinup is especially long in the absence of prior information (e.g., during a cold start) or when the background error statistics suddenly change. With RIP, the analysis correction and ensemble-based background error covariance are both updated simultaneously and quickly spin up toward the corresponding underlying dynamics. Details of the RIP method are presented in Kalnay and Yang (2008, 2010).

The RIP method is a more general form of the QOL method addressing the nonlinear evolution of all ensemble members, rather than improving only the mean ensemble state as in the QOL scheme introduced in the next subsection.

The backbone of the RIP method is the no-cost smoother, which is used multiple times to adjust the ensemble trajectories within the assimilation window $[t_{n-1}, t_n]$ using the observations arranged at the end of the window t_n. During the RIP iterations, the same set of observations is repeatedly used only if it is estimated that additional useful information can be extracted. To use RIP, the standard LETKF (based on the knowledge of observations made at t_n or before) is applied as iteration $i = 0$. During this iteration, the weight coefficients $\overline{\mathbf{w}}^{a,0}_n$ and $\mathbf{W}^{a,0}_n$ are derived at t_n, using Eqs. (8), (9), and (11).

At the ith iteration, the no-cost smoother is applied so that the smoothed ensemble mean at t_{n-1} is given by

$$\overline{\mathbf{x}}^{a,i+1}_{n-1} = \overline{\mathbf{x}}^{a,i}_{n-1} + \mathbf{X}^{a,i}_{n-1}\overline{\mathbf{w}}^{a,i}_n, \tag{13}$$

and the smoothed perturbations are obtained as

$$\mathbf{X}^{a,i+1}_{n-1} = \mathbf{X}^{a,i}_{n-1}\mathbf{W}^{a,i}_n + \mathbf{E}^{i+1}_{n-1}, \tag{14}$$

where $\mathbf{E}^{i+1}_{n-1}$ are small random Gaussian perturbations. With the modified analysis ensemble at t_{n-1}, the nonlinear model is used to integrate all ensemble members forward to the end of the assimilation window t_n. The newly evolved ensemble is the new background ensemble ($\mathbf{x}^{b,i+1}_n$), and the weight coefficients are computed again to obtain $\overline{\mathbf{w}}^{a,i+1}_n$, $\mathbf{W}^{a,i+1}_n$ based on $\mathbf{y}^o_n - \overline{\mathbf{y}}^{b,i+1}_n$ and $\mathbf{Y}^{b,i+1}_n$ with Eqs. (8), (9), and (11). Equations (7) and (10) are used to compute $\overline{\mathbf{x}}^{a,i+1}_n$ and $\mathbf{X}^{a,i+1}_n$. The new background ensemble has a better mean (closer to the truth) and perturbations (more reliable flow-dependent structures) because it is initialized from a more accurate ensemble. More specifically, $\mathbf{y}^o_n$ (the observation at t_n) has been used i times in $\mathbf{x}^{b,i}_n$ and $\mathbf{x}^{a,i}_{n-1}$, and $i + 1$ times in $\mathbf{x}^{a,i}_n$. At $i = 0$, $\mathbf{x}^{a,0}_{n-1}$ is the final analysis ensemble derived from the previous analysis cycle at t_{n-1} and $\mathbf{x}^{b,0}_n$ is the following forecast. Therefore, they ($\mathbf{x}^{a,0}_{n-1}$, $\mathbf{x}^{b,0}_n$) do not ''know'' the observation $\mathbf{y}^o_n$. With the no-cost smoother, $\mathbf{x}^{a,1}_{n-1}$ then contains the information about $\mathbf{y}^o_n$ (i.e., it uses the observation once) and so does the following forecast $\mathbf{x}^{b,1}_n$.

The random perturbations $\mathbf{E}_{n-1}^{i+1}$ are added in Eq. (14) because, under linear conditions, integrating the smoothed analysis from t_{n-1} to t_n will end up with the same analysis ensemble previously derived. Adding these very small random Gaussian perturbations before the forward integration of the smoothed analysis ensemble avoids it evolving into the same ensemble at t_n, and also has the advantage of exploring new growing directions that may have been missed by the original ensemble (Kalnay et al. 2007a). Corazza et al. (2003) have shown that adding random perturbations onto the ensemble members can ''refresh'' the space spanned by the ensemble into new growing directions, so that the forecast errors project better into the slightly perturbed ensemble space. In EnKFs, the analysis corrections are made within the space spanned by the ensemble. Under nonlinear conditions and by slightly perturbing the ensemble space, we can generate a (slightly) different evolved ensemble and have a better chance to correct the errors within the newly growing error subspace, given exactly the same observations. Adding small random perturbations at each iteration also has the effect of slightly reducing the time correlation between the analysis and the observation errors (C. Bishop 2009, personal communication).

The criterion of when to stop the iterations is based on the reduction of the forecast misfit to the observations (measured by the RMS of the differences between the observations and the forecasts RMS_OMF$_i$) scaled by the observation error σ_o. The procedure of modifying and re-evolving the ensemble at t_{n-1} is repeated as long as the relative reduction of the misfit ε is greater than a threshold ε_s [Eq. (15)]. If the reduction of the fit is smaller than the chosen threshold ε_s, it indicates that little useful additional information can be extracted from the same set of observations:

$$\varepsilon = \frac{\text{RMS\_OMF}_{i-1} - \text{RMS\_OMF}_{i}}{\sigma_o} > \varepsilon_s. \quad (15)$$

Kalnay and Yang (2010) showed that the spinup period needed for LETKF to provide an accurate analysis is very substantially reduced with RIP when the initial ensemble is chosen without using any prior information. Section 2 shows that the reduction of spinup also takes place when using RIP with a linear model and the standard Kalman filter. Although RIP was developed for accelerating spinup, when the error statistics suddenly change because of strong nonlinearities, it can also be applied as a generalized quasi outer loop aimed at improving the whole ensemble, not just the mean.

In the experiments with the Lorenz model, iterations of the RIP scheme are continued as long as $\varepsilon > 0.001$ in Eq. (15) but the maximum number of iterations allowed is set to 10. The criterion for stopping may depend on the density of observations. Experimental results suggest that more iterations with a stricter threshold are required to optimize the performance of the RIP method with fewer observations (see section 4d). Although the success of the RIP scheme was already demonstrated in Kalnay and Yang (2010), we should note that the computational cost of the RIP scheme is relatively high since all the ensemble members have to be integrated using the nonlinear model. In the next subsection, the QOL scheme is introduced as a simplified version of the RIP scheme where only the ensemble mean is integrated within the assimilation window, and only one or two iterations are allowed.

d. QOL algorithm

The main purpose of QOL is to improve the nonlinear trajectory of the ensemble mean and re-center the ensemble without the need to integrate again the whole ensemble as in RIP. The steps for performing QOL in the LETKF framework within the assimilation window $[t_{n-1}, t_n]$ are as follows:

As in RIP, the standard LETKF analysis is used as the iteration $i = 0$ of QOL, given the analysis ensemble ($\overline{\mathbf{x}}_{n-1}^{a,0}$ and $\mathbf{X}_{n-1}^{a,0}$) at t_{n-1} and weight coefficients for correcting the mean state ($\overline{\mathbf{w}}_n^{a,0}$) at t_n. QOL uses the weights $\overline{\mathbf{w}}_n^{a,i}$ obtained at the end of the window to update the analysis ensemble mean at t_{n-1}:

$$\overline{\mathbf{x}}_{n-1}^{a,i+1} = \overline{\mathbf{x}}_{n-1}^{a,i} + \mathbf{X}_{n-1}^{a,0}\overline{\mathbf{w}}_n^{a,i}. \quad (16)$$

Note that $\mathbf{X}_{n-1}^{a,0}$, the matrix of analysis ensemble perturbations from the previous analysis cycle at t_{n-1}, is not modified in QOL. After applying Eq. (16), the nonlinear model is used to evolve only the ensemble mean trajectory: $\overline{\mathbf{x}}_n^{b,i+1} = M(\overline{\mathbf{x}}_{n-1}^{a,i+1})$. The innovation vector is now updated to $\mathbf{y}_n^o - h(\overline{\mathbf{x}}_n^{b,i+1})$.

Without re-evolving the ensemble like in RIP, the matrix of perturbations in QOL still undergoes changes with each iteration because of the change of the mean trajectory. It is given by the analysis ensemble perturbations derived at the previous iteration, and, as in RIP, small perturbations are added:

$$\mathbf{X}_n^{b,i+1} = \mathbf{X}_n^{a,i} + \mathbf{E}_n^{i+1}. \quad (17)$$

As in Eq. (14), small random Gaussian perturbations $\mathbf{E}_n^{i+1}$ are added so that $\mathbf{X}_n^{b,i+1}$ is not identical to $\mathbf{X}_n^{a,i}$. We then define the weights corresponding to the updated innovation vector as in Eq. (11):

$$\overline{\mathbf{w}}_n^{a,i+1} = \hat{\mathbf{P}}_n^{a,i+1}(\mathbf{Y}_n^{b,i+1})^{\mathrm{T}}\mathbf{R}^{-1}[\mathbf{y}_n^o - h(\overline{\mathbf{x}}_n^{b,i+1})]. \quad (18)$$

Similar to RIP, $\overline{\mathbf{x}}_n^{a,i+1}$ and $\mathbf{X}_n^{a,i+1}$ are then computed based on $\mathbf{X}_n^{b,i+1}$, $\mathbf{Y}_n^{b,i+1}$, and $\overline{\mathbf{w}}_n^{a,i+1}$. Equation (15) is again used to determine whether another QOL iteration should be applied a second time, with a criterion of $\varepsilon > 0.01$. Otherwise, $\overline{\mathbf{x}}_n^{a,i+1}$ and $\mathbf{X}_n^{a,i+1}$ are the final analysis ensemble mean and perturbations and serve as the initial conditions for the next analysis cycle. The results are less sensitive to the value of ε than RIP, since only one or two outer-loop iterations are allowed. If the criterion is not satisfied, QOL becomes the same as the standard LETKF with zero iteration.

Jazwinski (1970, p. 276, footnote 3) had already suggested a very similar idea of an outer loop allowing the ensemble to be recentered on a more accurate nonlinear solution for the extended Kalman filter. This method was then introduced as the IKF in Bell and Cathey (1993) to linearize the observation operator with respect to a more accurate state so that the linearity assumption is less likely to be violated. The incremental 4D-Var (Courtier et al. 1994) and other iterative EnKFs (e.g., Li and Reynolds 2009; Gu and Oliver 2007) related to the Gauss–Newton method share the same concept of outer looping. An important difference between these methods and RIP/QOL is that we modify the background error covariance at each iteration to reflect the improvement of the mean state [Eq. (16)] and to achieve multistep corrections, whereas in 4D-Var, IKF, and IEnKFs the background state (first guess) and the corresponding statistics are kept constant.

4. Results with the Lorenz three-variable model

In this section, Observing System Simulation Experiments (OSSEs) are performed for the standard LETKF, LETKF–QOL, and LETKF–RIP schemes with the Lorenz three-variable model. We first discuss the results based on the QOL scheme newly proposed in this study, and then the results from the RIP scheme are viewed as the additional benefit obtained from the use of a more general (and more expensive) form of the QOL scheme.

a. Lorenz three-variable model and experiment settings

The Lorenz (1963) model is a three-variable nonlinear system:

$$\frac{dx}{dt} = \sigma(y - x); \quad \frac{dy}{dt} = rx - y - xz; \quad \frac{dz}{dt} = xy - bz.$$

The standard values for the parameters in this system used here, $\sigma = 10$, $r = 28$, and $b = 8/3$, result in chaotic behavior with two regimes (the famous butterfly pattern). The model is integrated here using the fourth-order Runge–Kutta method with a time step of 0.01. With the exponential instability of the model solution and abrupt regime changes, this model has been widely used to demonstrate issues related to predictability (e.g., Palmer 1993) and to evaluate the feasibility of an assimilation scheme (e.g., Miller et al. 1994; Evensen 1997; Yang et al. 2006; Kalnay et al. 2007a). Evans et al. (2004) showed that the growth rate of perturbations with this model depends on the state where the perturbations reside, and that a large perturbation growth rate is observed in the last orbit before the model trajectory changes regimes. These are also the locations where the filter divergence is observed to occur [see Miller et al. 1994; Evensen 1997 (who used an adaptive time scheme integrator and an ensemble of 1000 members)], because the strong nonlinear perturbation growth results in a poor estimate of the state by the non-Gaussian ensemble and causes the filter to fail.

Observing System Simulation Experiments are performed to evaluate the assimilation scheme proposed in this study. The ''truth'' run is obtained from a long integration initialized from (8.0, 0.0, 30.0). After discarding the first 600 time steps of the truth run, observations every 8 or 25 time steps (0.08 and 0.25 time units, respectively) are generated for the three variables by adding Gaussian perturbations with zero mean and variance equal to 2.0 to the true values, and a total of 2000 analysis cycles are then performed. We note that the experimental setup is the same as in Kalnay et al. (2007a), who compared the results using the standard LETKF with those obtained with 4D-Var, with the analysis time defined at the end of the assimilation window. The observation time is the same as the analysis time. Also, the observation error and observing frequency of 25 time steps are the same as in Miller et al. (1994) and Evensen (1997). The initial forecast ensemble is obtained by adding random Gaussian perturbations with a mean equal to 5.0 and variance equal to 1.0 to the truth.

b. Nonlinearity with a long assimilation window

With the same model and assimilation settings (section 4a), Kalnay et al. (2007a) compared the analysis accuracy between the fully optimized LETKF and 4D-Var schemes. Their results (reproduced in Table 1) show that the performance of the LETKF with three ensemble members is comparable to 4D-Var when using a short assimilation window (8 time steps or 0.08 time units) for which the evolutions of the perturbations remain linear and Gaussian. However, when a long assimilation window is used (with 25 time steps or 0.25 time units), the 4D-Var analysis becomes significantly

TABLE 1. The analysis error of the 4D-Var, standard LETKF, LETKF–QOL, and LETKF–RIP schemes; ρ is the optimal multiplicative covariance inflation.

	4D-Var	Standard LETKF (three members)	LETKF–QOL	LETKF–RIP
Observations every 8 time steps (linear window)	0.31(assimilation window = 48)	0.30	0.27	0.27
Observations every 25 time steps (nonlinear window)	0.53(assimilation window = 75)	0.68(ρ = 1.22)	0.47(ρ = 1.08)	0.35(ρ = 1.047)

more accurate than the LETKF analysis, but only if the 4D-Var quasi-static variational method (Pires et al. 1996) is implemented, allowing the use of even longer windows by handling the presence of multiple minima in the cost function. It is clear that the outer loop used in 4D-Var has the advantage of improving the nonlinear trajectory in a long assimilation window, whereas LETKF is hindered by the fact that the nonlinear evolutions of the perturbations are no longer Gaussian.

Table 1 compares the analysis accuracy of 4D-Var, the standard LETKF, LETKF with QOL, and LETKF with RIP schemes, for assimilation windows of 8 and 25 time steps, representing linear and nonlinear behavior, respectively. The random Gaussian perturbations used for the QOL and RIP schemes [in Eqs. (14) and (17)] have zero means with standard deviations of 0.0004 and 0.0007, respectively. All of the LETKF-related schemes use three ensemble members ($K = 3$). This is a small ensemble size compared with the thousands of ensemble members used in other studies such as Evensen (1997) and Evensen and van Leeuwen (2000), but this choice reflects the fact that the ensemble size is necessarily small in practical applications. In Table 1, the analysis accuracy is computed as the RMS analysis error averaged over 2000 analysis cycles (no significant differences were observed when experiments were repeated with 1 000 000 cycles.). For both the 4D-Var and LETKF-related systems, the RMS analysis error is defined at the end of the assimilation window, where the observations are available. The values in the first two columns are the same as those obtained by Kalnay et al. (2007a). Their study showed that with an observation frequency of 25 time steps, the optimal length of the assimilation window for 4D-Var is a window of 75 time steps for this model, and that for such long windows the optimized 4D-Var is significantly more accurate than the optimized LETKF.

LETKF–QOL improves the performance of the LETKF scheme to an error level now better than the optimized 4D-Var. The improvement is especially significant for the longer window when the evolutions of the perturbations become nonlinear. This suggests that LETKF–QOL is much better at handling the nonlinearity of the trajectory for long assimilation windows. We also note that the optimal multiplicative covariance inflation is much reduced from 22% with the standard LETKF to 8% with QOL, indicating that the ensemble-based background error covariance can better represent the dynamic uncertainties. RIP shows a performance similar to QOL with the linear window, for which little further improvement can be gained. With the longer nonlinear window, however, the improvement obtained using RIP is significantly larger than with QOL, and the RMS analysis error is further reduced from 0.47 to 0.35 with only 4.7% multiplicative covariance inflation. RIP outperforms QOL because it also enables improving the analysis ensemble perturbations as well as the mean at the beginning of each assimilation window. We note that, on average, eight iterations are required but when the dynamics become very nonlinear, as during a regime transition, RIP may take as many as 10 iterations to fulfill the criterion $\varepsilon \leq 0.001$. We note that the RIP analysis (with 3 ensemble members and about 8 iterations) is still more accurate than the standard LETKF with 24 ensemble members.

Table 1 confirms that the advantage of QOL is dealing with long assimilation windows and that RIP is even more advantageous. Figure 2 shows the time evolution of the variable y (black line) from the true trajectory and the evolutions initialized from the standard LETKF (blue lines) and from LETKF–QOL (red lines) analysis means. Figure 2 is a typical example of a dynamical nonlinear instability leading to filter divergence. In Fig. 2, the dots are analyses, that is, the initial states of the forecasts. Such filter divergence is found to occasionally occur in the standard LETKF even if the ensemble size is increased (here $K = 3$). As shown in Fig. 2, the standard LETKF (in blue) fails to track the regime transition and gives the wrong information about whether a regime change actually takes place or whether the system stays close to the borderline state. For example, starting at time = 17.0 (point A), the analysis mean states (blue dots) become less and less accurate and the corresponding forecasts are completely off the true trajectory even though the observations are still being assimilated. Only when the background happens to be close to the true dynamics again (given by the observations) near time = 24.5

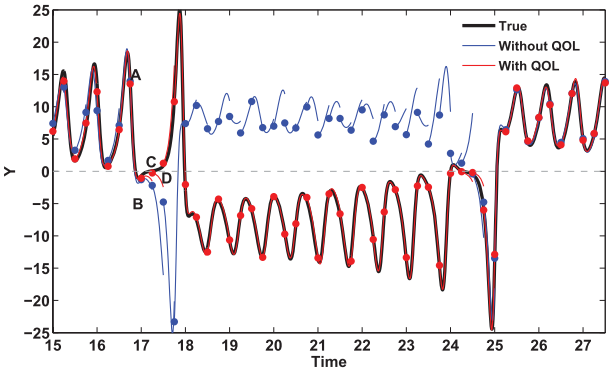

FIG. 2. Time evolution of the variable y during a typical regime transition with strong nonlinear growth between $t = 17$ and $t = 18$. The black line denotes the true trajectory and the blue line denotes the evolution initialized from the standard LETKF analysis mean, marked as the blue dots. The red line is the evolution from the LETKF–QOL analysis mean, marked as the red dots. The points A, B, C, D indicate a regime transition, where the filter divergence occurs with the standard LETKF. Three ensemble members are used in these experiments.

is the true trajectory recaptured, and the standard LETKF gets back onto the right regime solution. With the QOL, by contrast, the evolution of the trajectory is much improved and stays on the correct regime, following the true trajectory very well. This suggests that the quasi outer loops have a clear advantage during unstable periods. When getting close to the boundary between regimes (Evans et al. 2004), the nonlinear perturbation growth quickly degrades the forecast trajectory for the standard LETKF and for QOL before the second iteration. Particularly for the assimilation window between time = 17.25 and 17.5 (between points C and D), both forecasts quickly go into the wrong regime. Even with the observation at time = 17.5, the analysis obtained from the standard LETKF is still in the wrong regime, causing an even worse forecast and leading to filter divergence. By comparison, QOL corrects the trajectory because it makes use of the observation *at the end of the assimilation window*. Thus the analysis mean, serving as the initial state in the next analysis cycle (the red dot at point D), is close to the truth and is able to stay in the correct regime.

c. *Ensemble distribution*

In this section we evaluate the ensemble of the standard LETKF, QOL, and RIP using statistical moments to depict the shape of the ensemble distribution, including the standard deviation and kurtosis of the ensemble. Also, to exhibit statistical significance, results in the following text are shown with a large ensemble (i.e., $K = 24$) and both QOL and RIP use random perturbations with an amplitude of 0.003. Filter divergence (i.e., the analysis stays in the wrong regime for several cycles) still occurs occasionally with the standard LETKF with 24 ensemble members while the analyses of the other two schemes are always in the right regime, even with $K = 3$. The large ensemble size just helps the standard LETKF to recover from filter divergence more quickly so that it can begin to track the true dynamics again.

TABLE 2. The mean RMS error and ensemble spread (standard deviation) of the background and analysis ensemble from the standard LETKF, LETKF–QOL, and LETKF–RIP schemes. Results are obtained with 24 ensemble members.

$K = 24$	Standard LETKF	LETKF–QOL	LETKF–RIP
Background error (spread)	1.32 (1.32)	0.63 (0.66)	0.67 (0.26)
Analysis error (spread)	0.65 (0.63)	0.49 (0.33)	0.33 (0.13)

First, the standard deviation of the ensemble represents the spread of the ensemble distribution. Table 2 shows the mean RMS error and ensemble spread (standard deviation) of the background and analysis ensemble from the three schemes. The results show that, on average, the ensemble spread and mean state errors are reasonably well estimated in the standard LETKF, but its mean state is the least accurate among these three schemes. When the standard LETKF performs poorly, the spread is also severely underestimated. Among the three schemes, RIP shows the most accurate analysis mean but the smallest ensemble spread. This agrees with what we illustrated in section 2—that a small ensemble spread is a necessary characteristic of RIP for performing the assimilation in a soft way by assimilating the same observations repeatedly and for giving the optimal analysis. Given the success of RIP, we argue that the small ensemble spread from RIP is essential for performing incremental analysis correction even though it underestimates the uncertainty of the mean state. As for QOL, the analysis ensemble spread is only slightly smaller than the RMS error, reflecting the fact that only one or two iterations are used.

For a linear model, if the analysis ensemble has a Gaussian distribution, the forecast ensemble will also be Gaussian. If the model is nonlinear and the assimilation window is long, ensemble perturbations grow nonlinearly, and as discussed in the introduction, the presence of strong nonlinearities will distort the distribution of the ensemble and make it non-Gaussian so that the Kalman filter formulation ceases to be optimal. To investigate whether using QOL improves Gaussianity,

TABLE 3. Median and interquartile range of the kurtosis values obtained from the background and analysis ensemble of the standard LETKF, LETKF–QOL, and LETKF–RIP schemes, with and without the random perturbations **E** [in Eqs. (14) and (17)]. The amplitude of the random perturbation is 0.003. The ensemble size is 24, and the kurtosis of a Gaussian distribution is 0.

	Median		Interquartile range (difference between 25% and 75% quantiles)	
	Background ensemble	Analysis ensemble	Background ensemble	Analysis ensemble
Standard LETKF	0.79	0.78	5.96	5.98
LETKF–QOL (with **E**)	0.33	0.32	4.27	4.27
LETKF–QOL (without **E**)	0.47	0.50	4.56	4.78
LETKF–RIP (with **E**)	−0.10	0.45	1.74	1.74
LETKF–RIP (without **E**)	−0.64	0.05	2.16	2.97

kurtosis is used to measure the ''flatness'' of the LETKF ensemble distribution. In a Gaussian distribution, kurtosis is equal to zero. A large positive kurtosis represents a non-Gaussian flatter distribution with heavy tails (i.e., with many outliers). A negative kurtosis corresponds to a peaked distribution narrower than the Gaussian.

Table 3 compares the median and interquartile range (IQR) of the kurtosis values of the ensemble from the experiments of the QOL and RIP methods with and without the random perturbations. The results show that QOL and RIP have a much more Gaussian distribution than the standard LETKF, for both the background and analysis ensembles: the median is closer to zero with smaller interquartile ranges, compared with the standard LETKF (first row in Table 3). Through nonlinear evolution, strong nonlinearity of the underlying dynamics can still affect the ensemble distribution but has the largest influence on the ensemble from the standard LETKF centered at less accurate states with a larger spread. By keeping the ensemble mean close to the true trajectory, the uncertainty about the dynamical evolution is reduced and the perturbations evolve more linearly with a smaller ensemble spread so that outliers are largely reduced. Therefore, the assumption about Gaussianity is better satisfied with the QOL and RIP schemes. We should note that the standard deviations of the ensemble are smaller for QOL and RIP than for the standard LETKF and that this tends to magnify the impact on kurtosis produced by an outlier.

Without adding random perturbations, the RMS analysis errors become slightly larger for both schemes (from 0.49 to 0.51 with QOL and from 0.33 to 0.37 with RIP). Comparing the results with and without the **E** term for both schemes shows that random perturbations have contributed positively to the Gaussianity of the ensemble distribution. The IQRs of the kurtosis become smaller, especially for the RIP method, since for QOL no more than two iterations are allowed. Results from Table 3 also imply that the random perturbations with Gaussian statistics modify the ensemble distribution into a more Gaussian shape so that fewer outlier members occur. A similar stochastic effect from random perturbations regulating the ensemble distribution has been discussed for a stochastic EnKF (Lawson and Hansen 2004).

d. Impact of incomplete observations

So far, observations were made available for all three variables (x, y, and z) so that the influence of error covariances between variables is not dominant. We further consider the impact of nonlinearity due to sparse observations. For this purpose, instead of assimilating observations in x, y, and z, we only observe subsets of these observations, both for the linear (8 time step) and nonlinear (25 time step) windows (Tables 4 and 5). We note that with fewer observations available, the optimal threshold used in RIP was reduced to 0.0001, indicating that more iterations are required to make the model states pull closer to the observations.

Yang et al. (2006) pointed out that in the Lorenz three-variable model, the y variable provides the most effective observations to synchronize the model trajectory with the true states. In agreement with this observation, when only two observations are assimilated every eight time steps with the standard LETKF (the first row in Table 4), the analysis accuracy with observations (x, y) or (y, z) is comparable to the analysis with all three observations assimilated, while the one assimilating observations (x, z) is much less accurate. With the 25-time-step window, the analysis error with observations (x, z) is even larger than the observation error. As a consequence, with fewer observations and a longer assimilation window, it becomes more difficult to keep the model trajectories close to the truth, and the errors for the nonobserved variables are large enough to quickly amplify the nonlinear effect.

From the last row in Tables 4,5, with only one or two observations, the relative improvement of the RIP method with respect to the standard LETKF is even larger than when all three observations are available.

TABLE 4. The RMS analysis error of the standard LETKF, LETKF–QOL, and LETKF–RIP schemes with the 8-time-step analysis cycles and different combinations of observations. The ensemble size is three in all experiments.

Observations	*x*	*y*	*z*	*xy*	*xz*	*yz*	*xyz*
Standard LETKF	0.88	0.46	4.11	0.35	0.56	0.33	0.3
LETKF–QOL	0.82	0.48	2.73	0.41	0.39	0.30	0.27
LETKF–RIP	0.79	0.43	1.56	0.37	0.38	0.28	0.25

TABLE 5. As in Table 4, but for the 25-time-step analysis cycles.

Observations	*x*	*y*	*z*	*xy*	*xz*	*yz*	*xyz*
Standard LETKF	2.9	1.67	7.16	1.01	1.53	0.78	0.68
LETKF–QOL	1.98	1.23	5.94	0.82	1.16	0.60	0.47
LETKF–RIP	1.57	0.97	3.81	0.56	0.66	0.40	0.35

For the nonlinear window and incomplete observations, the RIP method has at least 40% smaller errors than the standard LETKF. The RIP analysis with two observations is more accurate than the standard LETKF analysis with all three observations. Observations of only x and z also become more effective with the RIP method. For the eight-time-step linear window, the improvement with the RIP method is most significant when the observation y is not available because the available observations are less effective than y. Even when observing z alone, it is still possible to constrain the model trajectory. The result implies that the RIP method can be used to help overcome the problems related to nonlinearity exhibited in data-void regions or unobserved variables.

With the nonlinear window, the performance of QOL is always in between the RIP and standard LETKF methods, sharing the same characteristic effectiveness in assimilating different observations. Regarding the linear window, the improvement is also noticeable for cases without the observation y and particularly evident for cases with the ''least effective'' observation z. This indicates the importance of adjusting the ensemble (and the ensemble-based error covariance) in response to the strong nonlinear instabilities. The ability of EnKF to do multivariate data assimilation and to estimate unobserved variables relies on its linear estimation of the background error covariance, so that in the case of data voids it is particularly important for the perturbations to be small enough to evolve linearly. This helps to explain why the impact of QOL and RIP is even larger when not all observations are available.

Among all the cases, assimilating the observations y and z with the QOL and RIP methods significantly improves the Gaussianity of the ensemble compared with the standard LETKF. The nonlinear growth of the errors can be best controlled with these two observations because the nonlinear terms appear only in the governing equations for the y and z variables. The ensemble distribution of the nonobserved x variable also becomes more Gaussian: the median values of the kurtosis are 1.06, 0.35, and 0.24 for the standard LETKF, QOL, and RIP methods, respectively, and the corresponding IQR values are 7.69, 4.06, and 1.49. This implies that the dynamical adjustment of RIP and QOL is useful to constrain the perturbation growth for unobserved variables or data-void regions.

e. Impact of choice of threshold

The threshold (ε) used in RIP determines when to stop the iterations and reduces overfitting the observations. Figure 3 shows the RIP performance with different threshold values for cases with all three observations and only two observations (x and z), still under the limitation of a maximum of 10 iterations. Results from Fig. 3 indicate that the performance of RIP has reached the asymptotic level at $\varepsilon > 0.001$. In addition, even with a large threshold like $\varepsilon > 5$ (only one or no iteration is used), the RIP always provides improvement over the standard LETKF and the performance is close to that of QOL. In contrast, RIP with two observations is more sensitive to the criterion and the best performance is obtained at $\varepsilon > 10^{-5}$ (always using close to 10 iterations). A threshold that is too small and results in too many iterations will produce some degradation in the analysis accuracy because of overfitting the observations (but see the discussion in section 2). For example, if 10 iterations are always used in RIP for assimilating all observations, the RMS error degrades from 0.35 to 0.39. The results also suggest that most of the corrections take place at the first few iterations and that RIP with a small ensemble size can always perform well with a conservative criterion that stops iterating rather early.

We note again that QOL is less sensitive to the threshold since, in practice, we only allow for a second iteration when the evolution of the background state is under strong nonlinear instability.

f. Impact of the correlation between the forecast and observation errors

We note that after the first iteration, reusing the observations in QOL and RIP introduces a correlation between the forecast and observation errors. Since this correlation is not accounted for in QOL and RIP, this error correlation should reduce QOL's and RIP's accuracy. To estimate the impact of neglecting the correlation between the observation and background errors, we have artificially eliminated the correlation by repeating the

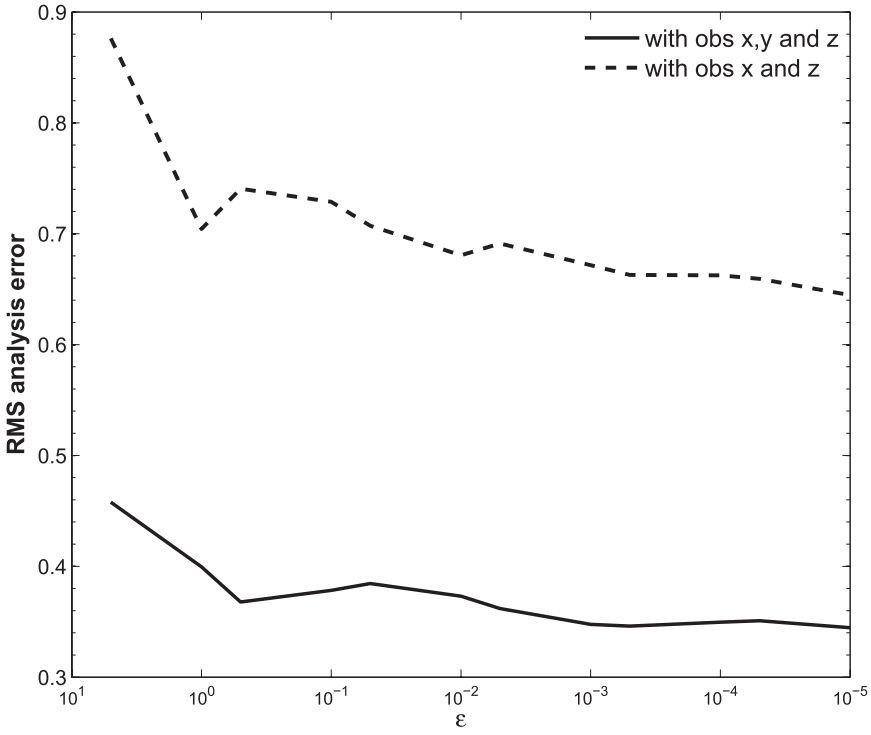

FIG. 3. The RMS analysis error of RIP using different thresholds and either all observations or just two observations. Three ensemble members are used in these experiments.

experiments in Table 1 but introducing new, independent observation errors at every new iteration (something that could not be done in practice). There was indeed a reduction in the analysis error in RIP and QOL, but it is small compared to the large improvement obtained with both methods over the standard LETKF. For the infrequent observation case (assimilation window of 25 steps), the QOL analysis error without the error correlation was reduced from 0.47 to 0.38, and for RIP from 0.35 to 0.26, respectively. The fact that both RIP and QOL have achieved excellent results in reducing the analysis errors for both the Lorenz (1963) model and much larger systems with real observations (Yang et al. 2012; Penny et al. 2011) indicates that neglecting the correlation between the observation and background errors does less harm to the assimilation with RIP/QOL than the strong nonlinear perturbation growth does to the standard LETKF. Nevertheless, the results imply that the performance of RIP/QOL could be further improved by accounting for this effect of correlated errors.

5. Comparisons between iterative EnKF schemes

We now compare the RIP and QOL iterative schemes discussed in section 4 with the ensemble randomized maximum likelihood (EnRML) method proposed by Gu and Oliver (2007), which is also designed for handling nonlinear data assimilation. EnRML was originally proposed for solving historical data matching in the petroleum reservoir flow problem by improving the model parameters. In EnRML, the cost function is minimized with the Gauss–Newton method with a reduced-step adjustment for highly nonlinear cases. For this comparison, the formulas of EnRML are modified to derive the optimal initial state within each window as in 4D-Var, rather than the model parameters as in Gu and Oliver (2007). The modified formulas and implementation of EnRML are briefly discussed below.

With a similar assimilation setup to the one used in RIP, a cost function [Eq. (19)] is defined to assimilate observations arranged at the end of the assimilation window, t_n:

$$J(\mathbf{x}_{n-1}) = \frac{1}{2}(\mathbf{x}_{n-1} - \mathbf{x}_{b,n-1})^{\mathrm{T}}\mathbf{P}_{b,n-1}^{-1}(\mathbf{x}_{n-1} - \mathbf{x}_{b,n-1}) + \frac{1}{2}\{H[M(\mathbf{x}_{n-1})] - \mathbf{y}_{o,n}\}^{\mathrm{T}}\mathbf{R}^{-1}\{H[M(\mathbf{x}_{n-1})] - \mathbf{y}_{o,n}\}. \quad (19)$$

The optimal estimation, $\mathbf{x}_{n-1}$, at the beginning of the window t_{n-1} is obtained by minimizing Eq. (19).

In Eq. (19), $\mathbf{x}_{b,n-1}$ is the first guess at the initial time t_{n-1} and $\mathbf{P}_{b,n-1}$ is the corresponding error covariance, and the operators, H and M, have the same definition as in section 3. In EnRML, the error covariance matrices are then represented by the ensemble and the observations are perturbed following Evensen (2003).

The implementation of the EnRML algorithm is as follows:

1) The cost function is minimized for each ensemble member (k), with perturbed observations $\mathbf{y}^k_{o,n}$, giving the initial condition for the new iteration $i+1$ of the window:

$$J(\mathbf{x}^k_{n-1}) = \frac{1}{2}(\mathbf{x}^k_{n-1} - \mathbf{x}^k_{b,n-1})^{\mathrm{T}}\mathbf{P}^{-1}_{b,n-1}(\mathbf{x}^k_{n-1} - \mathbf{x}^k_{b,n-1}) + \frac{1}{2}[H(\mathbf{x}^k_n) - \mathbf{y}^k_{o,n}]^{\mathrm{T}}\mathbf{R}^{-1}[H(\mathbf{x}^k_n) - \mathbf{y}^k_{o,n}]$$

$$\mathbf{x}^{k,i+1}_{n-1} = \beta_i\mathbf{x}^k_{b,n-1} + (1-\beta_i)\mathbf{x}^{k,i}_{n-1} - \beta_i\left(\frac{1}{K-1}\right)\mathbf{X}^k_{b,n-1}(\mathbf{Y}^k_{b,n})^{\mathrm{T}}\left[\mathbf{R} + \left(\frac{1}{K-1}\right)\mathbf{Y}^k_{b,n}(\mathbf{Y}^k_{b,n})^{\mathrm{T}}\right]^{-1} \times [H(\mathbf{x}^{k,i}_n) - \mathbf{y}^k_{o,n} - H(\mathbf{x}^{k,i}_n - \mathbf{x}^k_{b,n})]$$

where β_i *is* the parameter for the reduced step adjustment.

2) The misfit between the analysis and the observations is estimated for both $\mathbf{x}^{k,i+1}_{n-1}$ and $\mathbf{x}^{k,i}_{n-1}$:

$$\mathrm{OMF}_i = \frac{1}{2}\{H[M(\mathbf{x}^{k,i}_{n-1})] - \mathbf{y}^k_{o,n}\}^{\mathrm{T}}\mathbf{R}^{-1}\{H[M(\mathbf{x}^{k,i}_{n-1})] - \mathbf{y}^k_{o,n}\}.$$

3) If $\mathrm{OMF}_{i+1} < \mathrm{OMF}_i$, $\mathbf{x}^{k,i+1}_{n-1} = \mathbf{x}^{k,i}_{n-1}$ and β_i is increased; otherwise we keep $\mathbf{x}^{k,i}_{n-1}$ and decrease β_i.
4) Repeat steps 1–3 until one of the following criteria to stop iterating is satisfied:
   (i) $(\mathrm{OMF}_i - \mathrm{OMF}_{i+1})/\mathrm{OMF}_i < 10^{-4}$
   (ii) the number of iterations exceeds the maximum (20).

Recall that with only three ensemble members and a nonlinear assimilation window of 25 steps, QOL and RIP have an RMS analysis error of 0.47 and 0.35, respectively (Table 1). By contrast, with three ensemble members the EnRML algorithm fails because of sampling problems associated with the use of perturbed observations. Since the EnRML scheme needs a large ensemble size, comparisons are then made using 24 ensemble members for assimilating observations arranged at the end of a 25-time-step window. As indicated before, increasing the number of ensemble members from 3 to 24 does not improve the performance of QOL or RIP.

Running EnRML with 24 ensemble members is successful, with β_i starting at 0.5 and with an average of seven iterations needed for minimization. Without the reduced step adjustment the RMS analysis error from EnRML is 1.44, but with the reduced step it becomes 0.41, comparable to the values of 0.33 and 0.49 obtained from RIP and QOL, respectively. Tuning the covariance inflation can slightly improve the performance of EnRML, which attains its best RMS error of 0.39, still larger than that of RIP with three ensemble members.

The RMS error shown in Fig. 4a indicates that the RIP and EnRML methods have a similar performance most of the time, but there are still some situations when the EnRML analysis has catastrophic errors. These usually happen during regime changes or saddle points, as in the example shown in Fig. 4b. These large errors correspond to cases where the minimization of the cost function is trapped and EnRML fails to obtain a reliable initial condition because of an ill-conditioned error covariance matrix based on the background ensemble. In contrast, RIP successfully avoids this problem.

The statistical behavior of the ensemble perturbations among the QOL, RIP, and EnRML schemes is also examined. Results show that the Gaussianity of EnRML is not significantly different from that of RIP (figure not shown). However, we observe that the spread of the EnRML ensemble represents well the uncertainty of the mean state when the EnRML analysis is very accurate; in comparison, the ensemble spread from RIP is smaller for these cases because RIP has adjusted to incrementally use the observations by reducing the ensemble spread, as discussed in section 2.

In summary, EnRML derives the optimal state and its uncertainty through minimizing a cost function. Given a poor prior state and/or nonrepresentative prior error statistics, the ability to fit the model solution to the observations through the minimization of a cost function is limited. RIP instead aims to improve the prior state and corresponding error structure so that the observations can be better used: at every iteration, RIP solves a different cost function that is closer to being quadratic since the prior state is improved. Since the ensemble perturbations are repeatedly rescaled in RIP, it has an ensemble spread smaller than the analysis error but consistent with the multiple uses of observations.

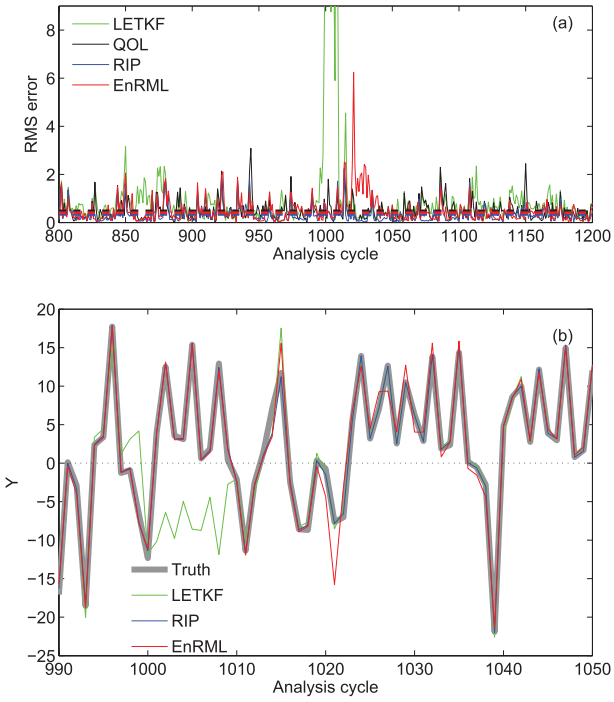

FIG. 4. (a) Comparison of the RMS error from four different analyses from the 800th to 1200th analysis cycle and (b) the values of the *y* variable from the truth and three different analyses from the 990th to 1050th analysis cycle.

6. Summary and discussion

A new type of outer loop, different from the incremental minimization of a cost function based on the maximum likelihood method, is proposed for the ensemble Kalman filter (EnKF) framework. It improves EnKF's ability to handle the strongly nonlinear evolving dynamics that can take place in long assimilation windows. Unlike the variational outer loops, which do not modify the prior state, RIP and its simpler version, QOL, take advantage of the information from "future observations" to improve the initial ensemble in the assimilation window and use the observations incrementally.

In section 2, we show with a simple linear model that after the initial spinup, using RIP with a constant number of iterations per analysis cycle produces the same analysis means as the optimal Kalman filter. The main differences are that with RIP, the estimated analysis error variance is reduced by *N*, the number of iterations per cycle, and the total analysis increment in a given analysis cycle is applied as the sum of multiple smaller increments. We argue that in a nonlinear scenario, the differences introduced by RIP can provide an advantage: first, the linear approximation to the set of possible model states provided by the ensemble is better with a smaller ensemble spread. Second, breaking the analysis increment into smaller steps and recomputing the linear approximation at each step can allow the increments to follow the nonlinear path toward the truth better than a single increment.

When there is strongly nonlinear error growth, the extraction of information from observations in the standard EnKF is less than optimal. RIP extracts the information incrementally, using the observations *N* times to improve both the ensemble mean and the perturbations incrementally, rather than reducing the observation error covariance by *N*. The QOL method, newly proposed as a simplified version of RIP, incrementally adjusts the ensemble mean only so that ensemble perturbations are centered at a more accurate state. Both the RIP and QOL schemes are based on the application of a no-cost smoother, with information provided by future observations available within the assimilation window, to improve the analysis at the beginning of the assimilation window and thus the nonlinear trajectories of the ensemble. QOL and RIP were implemented here within the LETKF framework to examine their ability to handle the nonlinearities of the dynamics, but they could be applied to any EnKF where the analysis perturbation matrix is obtained from the background perturbation matrix through a right multiplication by a matrix of weights, such as ETKF or LETKF.

In a nonlinear scenario, EnKF and EnKF–RIP will give different results, but as we argued above, EnKF–RIP may handle nonlinearities better because of a smaller ensemble spread and a multistep analysis increment. Just as EnKFs benefit from ad hoc modifications (such as covariance inflation) from the optimal linear formulation, EnKF–RIP benefits from modifications such as adaptively varying the number of iterations per analysis cycle and adding small perturbations. In this article we use such modifications of the basic RIP approach described above, but we emphasize that the basic approach provides most of the improvement we find in our numerical experiments. These modifications require tuning two parameters to optimize the performance of both RIP and QOL. The first is a threshold for the improvement of the fit of the forecasts to the observations, and the second is the size of additive random perturbations. For RIP, the threshold determines when to stop the iterations. Experimental results suggest that for smaller ensembles, a larger threshold (with fewer iterations) would be appropriate to avoid the forecasts overfitting the observations. Kalnay and Yang (2010) tested LETKF–RIP to accelerate the spinup in a quasi-geostrophic model, and used the threshold to estimate whether the spinup phase was over, allowing the RIP iterations to stop and return the system to the standard LETKF. QOL is much less sensitive to the threshold

because only one or two iterations are used when the model trajectory undergoes nonlinear instabilities. Since most of the improvements are attained in the first few iterations, both RIP and QOL should work with a conservative approach using only a few iterations. Results from both schemes show that the analysis accuracy is slightly improved by adding small random Gaussian perturbations that also play a role in regulating the Gaussianity of the ensemble distribution.

The performance of LETKF with the QOL and RIP schemes is tested with the Lorenz three-variable model. Results show that QOL and RIP allow LETKF to use longer assimilation windows with significant improvements of the analysis accuracy, especially during periods of highly nonlinear growth. For low-frequency observations (every 25 time steps, leading to long assimilation windows) and using optimal inflation, the standard LETKF RMS error is 0.68, whereas for QOL and RIP the RMS errors are 0.47 and 0.35, respectively. This can be compared to the best 4D-Var analysis error of 0.53, obtained using both an optimal long assimilation window (75 time steps) and the quasi-static variational analysis (Pires et al. 1996). As discussed in section 2, RIP has a small ensemble spread because of the incremental analysis correction, while the standard LETKF, with an ensemble spread comparable to the uncertainty of the mean state, computes the analysis increment only once but results in worse accuracy during strong nonlinear growth. When only one or two of the three variables are observed, the improvements of the QOL and RIP algorithms with respect to the standard LETKF are further enhanced, suggesting that these methods can be used to handle the nonlinearity introduced by data voids.

Results from RIP and QOL are also compared with the EnRML method, an iterative EnKF with Gauss–Newton minimization with a similar framework to the incremental 4D-Var. With 24 ensemble members, the EnRML performance is in between RIP and QOL. But with only 3 ensemble members, EnRML fails to converge, while both RIP and QOL achieve the same optimal level of performance with 3 members as with 24 members. Moreover, RIP robustly outperforms EnRML at the locations of regime changes and saddle points.

We note that RIP and QOL compensate for inaccuracies of the EnKF due to nonlinearity and non-Gaussianity, and they may not be appropriate for filters that represent non-Gaussian distributions more accurately. With a non-Gaussian background ensemble that undergoes strong nonlinear dynamics, the ensemble from RIP becomes more Gaussian because the nonlinearity of the evolution of the mean state is improved, and because the spread is small. By doing so, RIP reduces the effect of errors in the linear Gaussian approximation used by EnKF. In this sense, RIP is less sensitive to a non-Gaussian distribution than the standard LETKF.

The original outer loop used in variational data assimilation in the incremental form does not change the background state (first guess) in the outer loop; it only improves the nonlinear trajectory of the model used to compute the innovations and the state used to linearize the observation operator (Courtier et al. 1994). In contrast, QOL uses the no-cost smoother and the future observations within the assimilation window to improve the ensemble mean and re-center the smoothed perturbations around this improved mean. RIP reruns the whole ensemble, so that it not only improves the initial ensemble mean but also the perturbations. This makes the QOL and RIP schemes quite different from the variational outer loop and seems to violate the basic rule that ''data should be used once and then discarded'' (Ide et al. 1997). However, as we argued in section 2 for linear models, the background error covariance is underestimated because of multiple uses of past observations. This reduction of the background error covariance balances out the multiple uses of current observations by making the analysis increment smaller for each use of the current observations. Furthermore, using multiple small analysis increments can be advantageous for nonlinear models because the successive increments can track the true dynamics better than one large increment. Although initially the reduced spread of the background ensemble makes it more likely that the truth initially lies outside the ensemble spread, subsequent ensembles should then move toward the truth and better capture the linear dynamics near the truth. Despite the reduced ensemble spread, the structure of the perturbations still represents well the structure of the flow-dependent error uncertainties centered at a more accurate mean state. This is especially true in RIP, where the whole ensemble is re-evolved, improving the structure of the perturbations together with the mean state. Since RIP and QOL underestimate the amplitude of analysis errors, we note that for ensemble forecasting applications the initial amplitudes of the analysis perturbations should be inflated.

The performances of QOL and RIP on the Lorenz three-variable model and several other small models have been shown to be excellent. Applications to typhoon regional data assimilation (Yang et al. 2012) and global ocean assimilation (Penny et al. 2011) also demonstrated very encouraging results, so that QOL and RIP are now being tested within other more realistic and dynamically complex models.

Acknowledgments. We are very grateful for the insights provided by Anna Trevisan, Arlindo da Silva, and

Fuqing Zhang. Takemasa Miyoshi and Kayo Ide and other members of the University of Maryland Weather and Chaos Group, Ross Hoffman (AER), and Xiang-Yu Huang (NCAR Data Assimilation Test Center) made very helpful suggestions. We also appreciate the comments and suggestions of the three reviewers and of the editor, Herschel Mitchell, and Adrienne Norwood's English proofreading. Shu-Chih Yang is sponsored by Taiwan National Science Council Grants 97-2111-m-008-25 and 98-2111-m-008-014, and the NCU Development Program for the Top-Ranked University sponsored by the Ministry of Education. Eugenia Kalnay acknowledges support from NASA Grants NNX08AD40G and NN07AM97G and DOE Grant DEFG0207ER64437.

REFERENCES

Andersson, E., M. Fisher, E. Hólm, L. Isaksen, G. Radnóti, and Y. Trémolet, 2005: Will the 4D-Var approach be defeated by nonlinearity? ECMWF Tech. Memo. 479, 26 pp.

Bell, B. M., 1994: The iterated Kalman smoother as a Gauss-Newton method. *SIAM J. Optim.,* **4,** 626–636.

——, and F. W. Cathey, 1993: The iterated Kalman filter update as a Gauss-Newton method. *IEEE Trans. Autom. Control,* **38,** 294–297.

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

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

Corazza, M., and Coauthors, 2003: Use of the breeding technique to estimate the structure of the analysis ''errors of the day.'' *Nonlinear Processes Geophys.,* **10,** 233–243.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-VAR, using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120,** 1367–1387.

Evans, E., N. Bhatti, L. Pann, J. Kinney, M. Peña, S.-C. Yang, E. Kalnay, and J. Hansen, 2004: RISE: Undergraduates find that regime changes in Lorenz's Model are predictable. *Bull. Amer. Meteor. Soc.,* **85,** 520–524.

Evensen, G., 1992: Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. *J. Geophys. Res.,* **97** (C11), 17 905–17 924.

——, 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.,* **99** (C5), 10 143–10 162.

——, 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon. Wea. Rev.,* **125,** 1342–1354.

——, 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.,* **53,** 343–367.

——, and P. J. van Leeuwen, 2000: An ensemble Kalman smoother for nonlinear dynamics. *Mon. Wea. Rev.,* **128,** 1852–1867.

Fertig, E. J., J. Harlim, and B. R. Hunt, 2007: A comparative study of 4D-VAR and a 4D Ensemble Kalman Filter: Perfect model simulations with Lorenz-96. *Tellus,* **59A,** 96–100.

Gu, Y., and D. S. Oliver, 2007: An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. *Soc. Pet. Eng. J.,* **12,** 438–446.

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

Hunt, B. R., and Coauthors, 2004: Four-dimensional ensemble Kalman filtering. *Tellus,* **56A,** 273–277.

——, E. J. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D,* **230,** 112–126.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan,* **75,** 181–189.

Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory*. Academic Press, 376 pp.

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

Kalnay, E., and S.-C. Yang, 2008: Accelerating the spin-up of Ensemble Kalman Filtering. Arxiv: physics:Nonlinear/0806.0180v1.

——, and ——, 2010: Accelerating the spin-up of Ensemble Kalman Filtering. *Quart. J. Roy. Meteor. Soc.,* **136,** 1644–1651.

——, H. Li, T. Miyoshi, S.-C. Yang, and J. Ballabrera-Poy, 2007a: 4D-Var or Ensemble Kalman Filter? *Tellus,* **59A,** 758–773.

——, ——, ——, ——, and ——, 2007b: Response to the discussion on ''4D-Var or EnKF?'' by Nils Gustaffson. *Tellus,* **59A,** 778–780.

Krymskaya, M. V., R. G. Hanea, and M. Verlaan, 2009: An iterative ensemble Kalman filter for reservoir engineering applications. *Comput. Geosci.,* **13,** 235–244.

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

Leeuwenburgh, O., G. Evensen, and L. Bertino, 2005: The impact of ensemble filter definition on the assimilation of temperature profiles in the tropical Pacific. *Quart. J. Roy. Meteor. Soc.,* **131,** 3291–3300.

Li, G., and A. C. Reynolds, 2009: Iterative ensemble Kalman filters for data assimilation. *Soc. Pet. Eng. J.,* **14,** 496–505.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.,* **20,** 130–141.

——, 1996: Predictability—A problem partly solved. *Proc. Conf. on Predictability,* Reading, United Kingdom, ECMWF.

Miller, R. N., M. Ghil, and F. Gauthiez, 1994: Advanced data assimilation in strongly nonlinear dynamical systems. *J. Atmos. Sci.,* **51,** 1037–1056.

Ott, E., and Coauthors, 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus,* **56A,** 415–428.

Palmer, T. N., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteor. Soc.,* **74,** 49–66.

Penny, S., 2011: Data assimilation of the global ocean using the 4D local ensemble transform Kalman Filter (4D-LETKF) and the Modular Ocean Model (MOM2). Ph.D. thesis, University of Maryland. [Available online at http://hdl.handle.net/1903/11716.]

Pires, C., R. Vautard, and O. Talagrand, 1996: On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus,* **48A,** 96–121.

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.,* **126,** 1143–1170.

Sakov, P., and P. R. Oke, 2008: A deterministic formulation of the ensemble Kalman filter: An alternative to ensemble square root filters. *Tellus,* **60A,** 361–371.

Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003: Ensemble square root filters. *Mon. Wea. Rev.,* **131,** 1485–1490.

Verlaan, M., and A. W. Heemink, 2001: Nonlinearity in data assimilation applications: A practical method for analysis. *Mon. Wea. Rev.,* **129,** 1578–1589.

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

Wang, Y., G. Li, and A. C. Reynolds, 2010: Estimation of depths of fluid contacts by history matching using iterative Ensemble-Kalman smoothers. *Soc. Pet. Eng. J.,* **15,** 509–525.

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

Yang, S.-C., and Coauthors, 2006: Data assimilation as synchronization of truth and model: Experiments with the three-variable Lorenz system. *J. Atmos. Sci.,* **63,** 2340–2354.

——, M. Corazza, A. Carrassi, E. Kalnay, and T. Miyoshi, 2009a: Comparison of local ensemble transform Kalman filter, 3DVAR, and 4DVAR in a quasigeostrophic model. *Mon. Wea. Rev.,* **137,** 693–709.

——, E. Kalnay, B. Hunt, and N. E. Bowler, 2009b: Weight interpolation for efficient data assimilation with the local ensemble transform Kalman filter. *Quart. J. Roy. Meteor. Soc.,* **135,** 251–262.

——, ——, and T. Miyoshi, 2012: Accelerating the EnKF spinup for typhoon assimilation and prediction. *Wea. Forecasting,* in press.