

1 **Handling nonlinearity in Ensemble Kalman Filter: Experiments with**
2 **the three-variable Lorenz model**

3
4
5
6
7
8
9

Shu-Chih Yang^{1*}, Eugenia Kalnay², and Brian Hunt³

1. Department of Atmospheric Sciences, National Central University

2. Department of Atmospheric and Oceanic Science, University of Maryland

3. Department of Mathematics, University of Maryland

* Corresponding Author: Shu-Chih Yang, Department of Atmospheric Sciences, National Central University. Email: shuchih.yang@atm.ncu.edu.tw.

1 **Abstract**

2 A deterministic Ensemble Kalman Filter (EnKF) with a large enough ensemble is optimal
3 for linear models, since it assumes that a Gaussian analysis ensemble evolves into a
4 Gaussian forecast ensemble. A new type of outer-loop, different from minimizing a cost-
5 function based on the maximum likelihood method, is proposed for EnKF to improve its
6 ability to handle nonlinear evolving dynamics, especially for long assimilation windows.
7 The idea is to increase the observation influence when there is strong nonlinear error
8 growth. For this purpose, the “Running in place” (RIP) algorithm that improves both the
9 ensemble perturbations and the mean within the Local Ensemble Transform Kalman
10 Filter (LETKF) framework is applied to achieve an incremental analysis correction. A
11 “quasi outer-loop” (QOL) method, newly proposed as a simplified version of RIP, aims
12 to improve the ensemble mean so that ensemble perturbations are centered at a more
13 accurate state.

14
15 We first show that in a linear model, Kalman Filter (KF) reusing observations N times, as
16 in KF-RIP, accelerates the spin-up and yields the same long-term analysis as the standard
17 KF, except that the estimated analysis error covariance is N times smaller. We argue that
18 for nonlinear models the use of RIP should be advantageous: first, the linear
19 approximation to the set of possible model states provided by the ensemble is better with
20 a smaller ensemble spread; and second, breaking the analysis increment into smaller steps
21 and recomputing the linear approximation at each step can allow the increments to follow
22 the nonlinear set toward the truth better than a single increment.

23
24 The performance of the LETKF-RIP and LETKF-QOL in the presence of nonlinearities
25 is tested with the Lorenz 3-variable model. The results support the above arguments: RIP
26 and QOL allow the LETKF to use longer assimilation windows with significant
27 improvement of the analysis accuracy during periods of high nonlinear growth. The
28 improvements in accuracy obtained using either RIP or the computationally less
29 expensive QOL are large. For low-frequency observations (every 25 steps, leading to
30 long assimilation windows), and using optimal inflation, the standard LETKF RMS error
31 is 0.68, whereas for the QOL and RIP the RMS errors are 0.47 and 0.35 respectively.

1 This should be compared with the best 4D-Var analysis error of 0.53, obtained using both
2 optimal long assimilation windows (75 steps) and the quasi-static variational analysis.
3 When only one or two of the three variables are observed, the relative improvement of
4 the QOL and RIP algorithms are further increased, suggesting that these methods can be
5 used to handle nonlinearity introduced by data voids. A comparison with other iterative
6 EnKF shows a superior performance of RIP. Based on the success with the QOL and RIP
7 methods, we argue that reusing observations with multi-step analysis increments is
8 advantageous for performing data assimilation with nonlinear models, especially during
9 spin-up.
10

1 **1. Introduction**

2
3 Data assimilation involving nonlinear perturbations becomes problematic in both
4 incremental variational methods, which use linear adjoint models and observation
5 operators to evolve model and observation perturbations, and in Ensemble Kalman Filters
6 (EnKFs), which assume that ensemble perturbations are Gaussian. Nonlinear growth of
7 perturbations can result from strong dynamical instabilities, from observations that have
8 either insufficient sampling frequency, from the use of nonlinear observation operators,
9 or even from model errors. When strong nonlinearities occur, a data assimilation scheme
10 may lose its ability to track the true dynamics, a problem known as “filter divergence”
11 (Miller et al. 1994, Evensen 1992) within Kalman Filter based assimilation schemes
12 (Kalman, 1960). Also, with non-Gaussian ensemble distribution, the EnKF cannot give
13 an optimal solution (Evensen and van Leeuwen, 2000). In this study, we focus on the
14 nonlinearity since it plays a dominant role in the filter failure, despite the fact that the
15 “filter divergence” can also take place in a linear system.

16
17 In the Kalman Filter formulas derived using the best linear unbiased estimation (BLUE),
18 the error covariance is evolved with the tangent linear model. Evensen (1992) pointed
19 out that unbounded error growth could be caused by the use of the tangent linear model
20 due to the lack of nonlinear saturation effects. Ensemble-based Kalman Filters (EnKFs)
21 are less vulnerable to this error growth than either the Kalman Filter or the Extended
22 Kalman Filter because EnKFs have the advantage of using the full nonlinear model that
23 includes the saturation of nonlinear perturbation growth (Evensen, 1994, 1997, Verlaan
24 and Heemink, 2001). Nevertheless, filter divergence can still take place when
25 nonlinearities appear and result in a non-Gaussian distribution of the ensemble (Evensen,
26 1997 and Lawson and Hansen, 2004). As a result, a short assimilation window is required
27 for EnKF to better preserve the Gaussianity of the ensemble. Fertig et al. (2007)
28 compared the 4D-Var and EnKF implemented in the Lorenz 40 variable model (Lorenz,
29 1996). They showed that EnKF did very well with short windows but did not handle well
30 long windows because ensemble perturbations became non-Gaussian. Studies also show
31 that EnKFs characterized by being either deterministic or stochastic (Tippett et al., 2003)

1 may have different ability to handle the nonlinearities or non-Gaussianity. Lawson and
2 Hansen (2004) compared nonlinearities handled by the Ensemble Square Root Filter
3 (EnSRF, Whitaker and Hamill, 2002) and the perturbed observation EnKF (Evensen,
4 1994), representing prototypes of deterministic and stochastic EnKF respectively. They
5 found that, as nonlinearity becomes significant, square root (deterministic) filters break
6 down earlier. They show that the stochastic EnKF effectively does a resampling to get
7 closer to the dominant mode of the *a posteriori* PDF, while the deterministic EnKF
8 cannot do this even though the analysis ensemble from the two EnKF schemes are both
9 linearly transformed from the background ensemble. Not directly affected by the
10 magnitude of the nonlinearity, filter failure can also occur gradually in a system with
11 weak nonlinearity after successive updates with non-Gaussian ensembles, which result in
12 far outliers (Sakov and Oke, 2008). In addition, Leeuweburgh et al. (2005) showed that
13 the deterministic EnSRF tends to introduce non-Gaussianity, but that using a random
14 rotation step in EnSRF could alleviate this. Outliers in the EnSRF systems can also arise
15 from the use of one-sided, non-symmetric solution in the calculation of the square root of
16 a matrix for deriving the analysis perturbations nonlinearity (Sakov and Oke, 2008), but
17 the LETKF uses the symmetric square-root solution (Ott et al., 2004).

18

19 EnKF needs to use short windows, but, in contrast, 4D-Var performs better with
20 longer windows because it is iterated within a window, producing a more accurate
21 nonlinear trajectory under the constraint of the observations. In the incremental 4D-Var
22 framework (Courtier et al. 1994), an outer loop is applied to adjust the nonlinear
23 evolution of the trajectory with the high-resolution nonlinear model with complete
24 physics while the inner loop is used to minimize the incremental cost function with the
25 adjoint model derived from a low-resolution simplified model (Andersson et al. 2005).
26 The 4D-Var outer-loop aims to improve the nonlinear evolution of the model state by
27 improving the sensitivity matrix of the linearization of the observation operator and the
28 background trajectory used for computing the innovation vector. The 4D-Var outer-loop
29 is essential for nonlinear cases in order to improve the linear approximation and derive
30 the correct gradient for minimization the cost function. During the iterations of
31 minimization, the first guess of the initial state (prior state) and its corresponding error

1 statistics are kept constant. An important reason for the failure of EnKF with long
2 windows is that EnKF does not have an outer loop to deal with nonlinear dynamics as the
3 one used in the incremental 3D-Var and 4D-Var (Arlindo DaSilva, pers. comm., 2006,
4 Kalnay et al., 2007a).

5
6 Inspired by the advantages shown by the outer-loop in improving the nonlinear trajectory
7 in 4D-Var (Courtier et al. 1994, Rabier et al. 2000, Andersson et al. 2005), here we
8 introduce a different type of an “outer loop” within the framework of the EnKF and
9 implement it for the Local Ensemble Transform Kalman Filter (LETKF, Hunt et al.
10 2007). The idea is to increase the influence of the observations for cases with nonlinear
11 error growth when the estimation given by the background mean is not reliable and the
12 ensemble perturbations are distorted by the strong nonlinearity. In these cases, the
13 ensemble perturbations cannot represent the uncertainties of the prior mean state, and
14 therefore the observation information cannot be effectively extracted to correct the model
15 state. In these cases, the observation impact in the standard LETKF is clearly suboptimal.
16 A “hard” way to increase the observation influence would be to artificially reduce the
17 observation error before assimilating the observation. A “softer” way is to use the
18 original observation error, and assimilate this observation multiple times so that the total
19 analysis increment is obtained as the sum of multiple smaller increments. Thus, the rule
20 of “using the observations only once” (Ide et al. 1997) is modified by reducing the
21 amplitude of the background error covariance (Section 2). We note that the soft way of
22 using small increments becomes an additional advantage for nonlinear dynamics because
23 the system can remain closer to the true nonlinear attractor with smaller steps.

24
25 The “Running in Place” (RIP) scheme recently developed by Kalnay and Yang (2008,
26 2010) is able to increase the observation influence, and improve the ability of EnKF to
27 deal with nonlinearity. The RIP method updates both the ensemble mean and the
28 ensemble perturbations and was originally proposed to capture the underlying evolving
29 dynamics in order to accelerate the spin-up of EnKF. The observations are repeatedly
30 assimilated and the ensemble spread self-adjusted (reduced). Here we further propose a
31 “quasi outer-loop” (QOL) algorithm, as a simplified version of RIP, aiming to improve

1 the ensemble mean so that ensemble perturbations are centered at a more accurate state.
2 Both the RIP and QOL methods can be applied to improve the LETKF for nonlinear
3 cases like a “cold start” of an EnKF or when the background error statistics suddenly
4 change. An example of cold start is the initialization of regional data assimilation from a
5 global analysis obtained at coarser resolution, thus lacking features that represent the
6 underlying mesoscales. Also, when the model trajectory encounters a rapid change into a
7 new regime with a short transition period (such as in the development of a severe storm),
8 this may result in a sudden change of the background trajectory and error statistics. In
9 these cases, the newly proposed QOL scheme aims to improve the nonlinear trajectory of
10 the ensemble mean while RIP further reduces strong nonlinearities in ensemble
11 perturbations when they do not represent the true dynamical uncertainties.

12

13 We also note that other iterative methods related to Kalman filter have been developed to
14 deal with the nonlinearity or long assimilation window. For example, the iterated Kalman
15 filter (IKF, Bell and Cathey, 1993, Bell, 1994), using an incremental form similar to the
16 incremental variational methods, has been proposed for better handling the nonlinearity
17 than the extended Kalman Filter to consider the nonlinearity of the observation operator.
18 Bell and Cathey (1993) showed that the IKF is equivalent to using the Gauss-Newton
19 method for constructing the maximum likelihood estimate. However, as in the
20 incremental 4D-Var outer loop, the IKF only modifies the sensitivity matrix related to the
21 observation operators, and does not change the initial guess and the corresponding error
22 statistics of the background state during the minimization iterations. For reservoir
23 engineering application, several iteration-based ensemble Kalman Filter/Smoother
24 (IEnKF, Gu and Oliver, 2007, Krymskaya et al. 2009, Li and Reynolds, 2009 and Wang
25 et al., 2010) were recently proposed for parameter estimation and to deal with the
26 nonlinearity arising from the historical matching problem. Among them, Ensemble
27 Randomized Maximum Likelihood (EnRML) by Gu and Oliver (2007) shares the same
28 concept of an outer-loop similar to incremental 4D-Var. In these methods, including the
29 variational methods, the problem of nonlinearity is solved via iterating the linear solution
30 for a cost-function, e.g. finding the optimum with the Gauss-Newton method. With each
31 iteration the model trajectory is improved until the linear approximation of the

1 observation operator (sensitivity matrix) is close enough to the true state and thus the
2 linear solution becomes optimal. When such sensitivity is known at the very first
3 iteration, like in the linear case, the system could yield an optimal update in just one
4 iteration. However, we should note that such framework for outer-looping is more
5 complicated within the EnKF systems because the sensitivities incorporated the error
6 statistics are entangled with the ensemble state. In EnKF we cannot correct the
7 sensitivities (and the ensemble-based estimated covariances) without correcting the
8 model state. For example, the EnRML scheme re-evaluates the sensitivity matrix based
9 on the “re-evolved” ensemble at each iteration.

10
11 In the RIP (Kalnay and Yang, 2011) and proposed QOL schemes, by contrast, we take
12 advantage of a "no-cost" smoother for the ETKF/LETKF (Kalnay et al., 2007b) to
13 improve not the sensitivity, but the full initial ensemble (RIP), or just the initial ensemble
14 mean (QOL), based on the knowledge about the state from the "future" observations
15 within the window. We will show that by using observations multiple times, RIP/QOL
16 reduce the spin-up and attain the same optimal analysis as the regular Kalman Filter even
17 in the case of a linear system. In a nonlinear system they have the additional advantage of
18 correcting the ensemble incrementally following a trajectory closer to the true dynamics.
19 We will show that this improves the analysis beyond what can be obtained under the
20 framework of the maximum likelihood method with Gauss-Newton method (e.g. the 4-
21 DVar outer-loop or the EnRML).

22
23 In this study, we first compare in Section 2 the standard KF with KF-RIP for a
24 simple linear model, showing that KF-RIP (or QOL) reach the same asymptotic analysis
25 accuracy as KF but substantially accelerate the spin-up, while reducing the ensemble
26 spread (estimated error variance). Assimilation experiments testing the impact of the
27 QOL and RIP on nonlinearity are carried out with the Lorenz 3-variable model (Lorenz,
28 1963), described in Section 3, together with the setup for the observing system simulation
29 experiments. Section 4 introduces the standard LETKF, the RIP and QOL methods within
30 the LETKF framework. In Section 5, the ability to handle nonlinearities of RIP and QOL,
31 which modify the prior ensemble, are compared with EnRML, which, like the standard

1 4D-Var outer loop, does not modify the prior state. Finally, Section 6 contains a summary
2 and discussion of the theoretical framework for RIP and QOL.

3 4 **2. Basic RIP in a linear model** 5

6 In this section, we show that if the number of iterations of RIP per analysis cycle is
7 constant over time, then in the linear scenario where the Kalman Filter (KF) is optimal,
8 using RIP produces the same analysis means after spin-up as not using RIP. The main
9 differences are that with RIP, the estimated error variance will be systematically smaller,
10 and the total analysis increment in a given analysis cycle is applied as the sum of multiple
11 smaller increments. The full RIP algorithm implemented in the LETKF framework is
12 discussed in more detail in section 3. Here, the basic RIP is presented in the framework of
13 Kalman Filter with linear dynamics.

14
15 With linear dynamics, the RIP algorithm is the same as assimilating observations multiple
16 times, and the background at each iteration is provided by the analysis derived from the
17 previous iteration. We acknowledge that if the background covariance used in a data
18 assimilation procedure is commensurate with the typical error in the background state,
19 then using observations more than once in a given assimilation cycle can lead to
20 overfitting the observations. However, in a procedure based on the Kalman filter, if all of
21 the observations are used more than once, then the background and analysis covariances
22 computed by the filter will, after spin-up, underestimate the background and analysis
23 errors. Furthermore, in the linear scenario in which the Kalman Filter is optimal, if each
24 observation is used the same number of times, the resulting analyses will in the long run
25 be essentially the same as if each observation was used once.

26
27 We apply a basic RIP scheme to a simple linear model with linear error growth and
28 compare its results to the solution of optimal Kalman Filter (KF). The simple linear
29 model has the following governing equation for a scalar x

$$30 \quad x_n = M(x_{n-1}) = Cx_{n-1} \quad (1)$$

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

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

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

4

5 The Kalman Filter analysis and estimated error variance are

6
$$x_{a,n} = x_{b,n} + \left(\frac{\sigma_{b,n}^2}{\sigma_{b,n}^2 + \sigma_o^2} \right) (y_n - x_{b,n}) \quad (3)$$

7 and

8
$$\sigma_{a,n}^2 = \frac{\sigma_{b,n}^2 \sigma_o^2}{\sigma_{b,n}^2 + \sigma_o^2} \quad (4)$$

9 where $x_{a,n}$, $x_{b,n}$, and y_n are respectively the values of analysis, background and
10 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
11 that the observation operator is the identity in this example.

12

13 In the RIP algorithm with N iterations, the background state at each iteration i is evolved
14 from the “smoothed” analysis (Eq. 5) at t_{n-1} that has already assimilated the observation at
15 later time t_n , indicated with a tilde:

16
$$\tilde{x}_{a,n-1}^i = \tilde{x}_{a,n-1}^{i-1} + \left[\frac{(\tilde{\sigma}_{a,n-1}^{i-1})^2}{(\tilde{\sigma}_{a,n-1}^{i-1})^2 + \sigma_o^2} \right] (y_n - \tilde{x}_{a,n-1}^{i-1}) \quad (5)$$

17
$$x_{b,n}^i = M_{t_{n-1} \rightarrow t_n} [\tilde{x}_{a,n-1}^i] \quad (6)$$

18 Similarly, the “smoothed” analysis error variance, $(\tilde{\sigma}_{a,n-1}^i)^2$, at t_{n-1} is modified because of
19 the observation at t_n , shown as Eq. (7). It is evolved and provides the background error
20 variance at iteration i .

21
$$(\tilde{\sigma}_{a,n-1}^i)^2 = \frac{(\tilde{\sigma}_{a,n-1}^{i-1})^2 \sigma_o^2}{(\tilde{\sigma}_{a,n-1}^{i-1})^2 + \sigma_o^2} \quad (7)$$

22
$$(\sigma_{b,n}^i)^2 = G_{t_{n-1} \rightarrow t_n} [(\tilde{\sigma}_{a,n-1}^i)^2] \quad (8)$$

23 KF formulas are used again to re-assimilate the observation y_n and to estimate the
24 analysis error variance.

$$x_{a,n}^i = x_{b,n}^i + \left[\frac{(\sigma_{b,n}^i)^2}{(\sigma_{b,n}^i)^2 + \sigma_o^2} \right] (y_n - x_{b,n}^i) \quad (9)$$

$$(\sigma_{a,n}^j)^2 = \frac{(\sigma_{a,n}^j)^2 \sigma_o^2}{(\sigma_{b,n}^j)^2 + \sigma_o^2} \quad (10)$$

For the first iteration, $\tilde{x}_{a,n-1}^1 = x_{a,n-1}^N$ and $\tilde{\sigma}_{a,n-1}^1 = \sigma_{a,n-1}^N$, the final analysis and error variance from RIP for the previous time interval obtained at t_{n-1} . If there is only one iteration, $x_{a,n}^1$ and $\sigma_{a,n}^1$ give the same analysis as the KF solution.

In the following perfect model experiment, we choose $C = 1.25$ but the results are rather insensitive to these choices. 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 and error variance are 30 and 5, respectively, but the results derived with this simple model are not sensitive to the initial values either. We choose the initial analysis far from the truth to illustrate how RIP recovers more quickly from poorly specified initial conditions. After spin-up, 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 of KF, with RIP using 2 or 10 iterations. With one additional iteration (KF is considered to be the first iteration), the RIP analysis reach an error of 0.5877, smaller than the one derived with the optimal KF, and the estimated analysis error variance is 0.18, half as large as the error variance estimated with the KF. The results also suggest that RIP has a time-smoothing effect on the analysis error from averaging the observation error at t_{n-1} and t_n . With 10 iterations, the RIP analysis error increases to 0.6069 and the estimated error variance is 0.036, one tenth of the KF estimation. Such increase in error is due to the increasing correlation of the observation and the background errors in RIP iterations, resulting from the earlier influence of the later observation. We also note that RIP shows a clear advantage in accelerating the spin-up (Figure 1b) that increases with the number of iterations (Kalnay and Yang, 2010). As discussed in the introduction, if the number of RIP iterations N per analysis cycle is constant, the estimated analysis error variance in the long term is N times smaller than the

1 analysis error variance estimated from KF. We should note that the estimated error
2 variance in RIP does not represent the analysis error, but that it is essential for RIP to use
3 the smaller estimated variance in order to perform N analysis increments with optimal
4 results.

5
6 With more iterations, the RIP analysis approximates the KF analysis, as shown in Figure
7 (1c). Although RIP after the second iteration does not fulfill anymore the assumption that
8 observation and background errors are uncorrelated, the results with linear dynamics
9 confirm that the RIP algorithm assimilating observations multiple times essentially gives
10 the same KF analysis accuracy. KF-RIP shows the best performance with just two
11 iterations (regular KF + one additional RIP iteration), when background and observation
12 errors are least correlated for cases with iteration number larger than two.

13
14 Figure 1(b) shows that the estimated analysis error variance from KF-RIP becomes
15 smaller with the number of iterations. More importantly, the estimated variance is the
16 same as the one that would be computed from KF with higher assumed observation
17 accuracy; specifically, with the observation error variance σ_o^2 replaced by σ_o^2/N . KF
18 performed with an observation variance divided by N can be viewed as a “hard” way to
19 increase the observation influence since the analysis correction is derived all at once. The
20 KF-RIP can then be viewed as a “softer” way to achieve same analysis correction but
21 with multi-step analysis corrections. We also argue that the RIP algorithm is
22 advantageous in nonlinear cases since the analysis increment is obtained in small steps
23 and thus is better able to follow the true nonlinear dynamics.

24 25 **3. RIP and QOL algorithms for the Local Ensemble Transform Kalman Filter** 26 **(LETKF)**

27 In Section 2, the KF-RIP could be carried out by using explicitly future observations.
28 With LETKF, future information can be also be used earlier by means of a “no-cost”
29 smoother. In this section we describe the LETKF ensemble Kalman filter, the “no-cost”
30 smoother valid at the beginning of the assimilation window, and apply the smoother to
31 derive the RIP and QOL algorithms for the LETKF.

1

2 **3.a The Local Ensemble Transform Kalman Filter (LETKF)**

3

4 The LETKF scheme (Hunt et al., 2007) performs an analysis locally in space using local
 5 information, including the background state (short-range forecasts) and observations. The
 6 analysis correction by the LETKF is calculated within the space spanned by the local
 7 ensemble. Here we describe the 3-dimensional LETKF (i.e., with all observations
 8 available at the end of the assimilation window). The 4-dimensional extension of the
 9 LETKF (where observations are available throughout the assimilation window) is
 10 straightforward (Hunt et al., 2004, 2007). We note that in this study no localization is
 11 required due to the small dimension of the Lorenz 3-variable model and therefore, the
 12 variables denote the globally defined model states or the observations. Without
 13 localization, the LETKF is formally equivalent to the Ensemble Transform Kalman Filter
 14 (ETKF, Bishop et al., 2001) but with centered spherical simplex ensemble (Wang et al.
 15 2004, Ott et al., 2004). Since the notation used in both methods is different, here we
 16 follow Hunt et al. (2007).

17

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

23

$$24 \quad \mathbf{X}_n^a = \mathbf{X}_n^b \mathbf{W}_n^a. \quad (11)$$

25 Here $\mathbf{X}_n^b = [\delta \mathbf{x}_n^{b,1} | \dots | \delta \mathbf{x}_n^{b,K}]$ is the matrix of the background perturbations whose
 26 columns are the vectors of ensemble member perturbations (deviations from the mean):
 27 $\delta \mathbf{x}_n^{b,k} = \mathbf{x}_n^{b,k} - \bar{\mathbf{x}}_n^b$, where $\mathbf{x}_n^{b,k}$ is the k^{th} background ensemble member and $\bar{\mathbf{x}}_n^b$ is the
 28 background ensemble mean. The analysis perturbations are represented in a similar way.

29 The analysis weight perturbation matrix, \mathbf{W}_n^a , is obtained from

$$30 \quad \mathbf{W}_n^a = [(K-1)\hat{\mathbf{P}}_n^a]^{1/2}, \quad (12)$$

1 where

$$2 \quad \hat{\mathbf{P}}_n^a = [(K-1)\mathbf{I}/\rho + \mathbf{Y}_n^{bT} \mathbf{R}^{-1} \mathbf{Y}_n^{bT}]^{-1} \quad (13)$$

3 is the analysis error covariance in the ensemble space, and $\mathbf{Y}_n^b = [\delta\mathbf{y}_n^{b,1} | \dots | \delta\mathbf{y}_n^{b,K}]$ is the
 4 matrix of background perturbations in observation space, $\delta\mathbf{y}_n^{b,k} = h(\mathbf{x}_n^{b,k}) - \overline{h(\mathbf{x}_n^{b,k})}$. \mathbf{R} is
 5 the observation error covariance, $h(\bullet)$ is the observation operator that converts a variable
 6 from model to observation space and ρ is the multiplicative covariance inflation, which in
 7 Eq (13) has essentially the same effect as inflating the amplitude the background error
 8 covariance (Hunt et al. 2007). We note that the nonlinear observation operator $h(\bullet)$ is also
 9 an important source of nonlinearity for observations like satellite radiances or
 10 scatterometer backscatter. With the LETKF, such nonlinear effects are partially taken into
 11 account by using the full nonlinear observation operator, in contrast to variational
 12 methods where the adjoints of the linearized observation operators (Jacobians) are
 13 required in order to minimize the cost function.

14

15 The ensemble mean analysis at time t_n is obtained from

16

$$17 \quad \bar{\mathbf{x}}_n^a = \mathbf{X}_n^b \bar{\mathbf{w}}_n^a + \bar{\mathbf{x}}_n^b \quad (14)$$

18 where

$$19 \quad \bar{\mathbf{w}}_n^a = \hat{\mathbf{P}}_n^a \mathbf{Y}_n^{bT} \mathbf{R}^{-1} (\mathbf{y}_n^o - \bar{\mathbf{y}}_n^b). \quad (15)$$

20

21 In Eq. (15), \mathbf{y}_n^o and $\bar{\mathbf{y}}_n^b$ are the vectors of observations and of the background mean state
 22 in the observation space. The difference between them, $\mathbf{d}_n = \mathbf{y}_n^o - \bar{\mathbf{y}}_n^b$, is referred to as the
 23 innovation vector. As pointed out by Ott et al. (2004) and Hunt et al., (2007) the use of a
 24 symmetric square root algorithm in Eq. (12) ensures that of all possible square roots,
 25 \mathbf{W}^a is the matrix closest to the identity, and that it varies smoothly in space and time. The
 26 accuracy of the LETKF analysis depends on the accuracy of both the matrix of weights
 27 \mathbf{W}^a associated with the flow dependent errors (or “errors of the day”) and of the vector of
 28 weights $\bar{\mathbf{w}}^a$ used to estimate the mean analysis correction. The properties of these
 29 weights are important for the performance of the LETKF. Yang et al. (2009a) showed

1 that the weights vary on larger spatial scales than either the analyses or the analysis
2 increments, so that interpolating the weights derived from a coarse resolution LETKF can
3 still retain the accuracy of the full resolution analysis.

5 **3.b The no-cost smoother**

6
7 In Eqs. (11) and (14), the weights linearly combine background ensemble trajectories in
8 such a way that they are closest to the true atmosphere at the analysis time t_n . The linear
9 combination of ensemble trajectories is also a model trajectory, and if it is close to the
10 truth at the analysis time it should also be close to the truth throughout the assimilation
11 window ($t_{n-1} - t_n$), at least to the extent that we can neglect nonlinearities and model
12 errors. Therefore the same weights obtained at the analysis time should be valid
13 throughout the assimilation window. This allows constructing a “no-cost” smoother for
14 certain types of EnKF (including the LETKF) that improves the analysis state at the
15 beginning of the assimilation window from observations obtained later within the
16 window, without the need of an adjoint model (Kalnay et al. 2007b, Yang et al. 2009b,
17 Kalnay and Yang, 2010, footnote 1). Such ensemble-based smoother for asynchronous
18 assimilation is applicable to any EnKF scheme that expresses the updated ensemble via
19 right-multiplied ensemble transforms as in Eq. (11) (Sakov and Oke, 2008), depending on
20 how nonlinearity affects the time correlation between the model states. Thus, the form of
21 the no-cost smoother can be compared with the ensemble Kalman smoother (EnKS) used
22 in Evensen and van Leeuwen (2000) and Evensen (2003, Appendix D) based on the
23 solution of the stochastic (perturbed observations) EnKF. Specifically, the no-cost
24 ETKF/LETKF smoother has the lagged-1 form of the EnKS in Evensen and van Leeuwen
25 (2000). In both methods, the analysis ensemble is obtained by linearly combining the
26 background ensemble but the ensemble perturbations in the perturbed observation EnKF
27 are updated differently*.

* Given the analysis error covariance, a symmetric square root solution is used in Eq.(12) to ensure the zero mean of the perturbations in LETKF. For perturbed observation EnKF, besides the linear combination of the background perturbations, random Gaussian perturbations are added to the observations \mathbf{Y}_n^o to ensure that the analysis ensemble perturbations have the form $\mathbf{X}_n^a = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{X}_n^b + \mathbf{K}\mathbf{Y}_n^o$, where \mathbf{K} is the

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

$$4 \quad \tilde{\mathbf{x}}_{n-1}^a = \bar{\mathbf{x}}_{n-1}^a + \mathbf{X}_{n-1}^a \bar{\mathbf{w}}_n^a \quad (16a)$$

5 is used to smooth the analysis mean, and

$$6 \quad \tilde{\mathbf{X}}_{n-1}^a = \mathbf{X}_{n-1}^a \mathbf{W}_n^a. \quad (16b)$$

7 is used to smooth the analysis perturbations. More specifically, the background without
8 the information at t_n , is taken from the analysis with the available observation at t_{n-1} . The
9 “tilde” indicates the use of the later information.

10
11 Kalnay and Yang (2010) show that such smoothed analysis is always more accurate than
12 the original LETKF analysis because of its “knowledge” of the observations made later,
13 and, as is the case in other smoothers under the linear assumption, that the forecast from
14 the smoothed analysis valid at t_{n-1} coincides with the EnKF analysis at t_n , i.e.,
15 $M(\tilde{\mathbf{x}}_{n-1}^a) = \bar{\mathbf{x}}_n^a$ and $M(\tilde{\mathbf{X}}_{n-1}^a) = \mathbf{X}_n^a$ (see also Appendix A in Yang et al. 2009b). Sakov et al.
16 (2010) discuss a similar idea of using the weight coefficients for asynchronous data
17 assimilation. Thus, the ensemble-based Kalman smoother and the RIP and QOL methods
18 discussed below could also be applied to any framework of the ensemble-based Kalman
19 Filter that expresses the updated ensemble via right-multiplied ensemble transforms.
20 Although in this study we assimilate observations available at the end of the assimilation
21 window, the RIP and QOL methods are equally applicable if observations are distributed
22 throughout the assimilation window (see Hunt et al., 2004, 2007).

24 **3.c Running in place (RIP) algorithm**

25
26 The linear KF-RIP used in Section 2 is now modified for nonlinear dynamics in the EnKF
27 framework. Because of nonlinearity, EnKF and EnKF-RIP do not give the same results,
28 but, as argued in Section 2, EnKF-RIP is better able to handle nonlinearities due to a
29 smaller ensemble spread and a multi-step analysis increment, i.e., it provides a “softer”

ensemble-based Kalman gain. The EnKS in Evensen et al. (2003) is also a no-cost smoother since no additional adjoint backward integration is required.

1 way to increase the influence of the observations. In the same way that the optimal linear
 2 formulation of EnKF benefits from ad hoc modifications such as covariance inflation,
 3 EnKF-RIP benefits from modifications such as adaptive estimation of the number of
 4 iterations per analysis cycle, and addition of small perturbations.

5
 6 The RIP scheme was originally proposed to accelerate the spin-up of the Ensemble-based
 7 Kalman filter. This spin-up is especially long in the absence of prior information (e.g.
 8 during a cold start) or when the background error statistics suddenly change. With RIP,
 9 the analysis correction and ensemble-based background error covariance are both updated
 10 simultaneously and quickly spin-up towards the corresponding underlying dynamics.
 11 Details of the RIP method are presented in Kalnay and Yang (2008, 2010).

12
 13 The RIP is a more general form of the QOL addressing the nonlinear evolution of all
 14 ensemble members, rather than improving only the mean ensemble state as in the QOL
 15 scheme introduced in the next sub-section.

16
 17 The backbone of the RIP method is the no-cost smoother, which is used multiple times to
 18 adjust the ensemble trajectories within the assimilation window $[t_{n-1}, t_n]$ with the
 19 observations arranged at the end of the window, t_n . During the RIP iterations, the same
 20 set of observations is repeatedly used only if it is estimated that additional useful
 21 information can be extracted. To use the RIP, the standard LETKF (based on the
 22 knowledge of observations made at t_n or before) is applied as the iteration $i=0$, given a
 23 background ensemble initialized from the analysis ensemble at t_{n-1} ($\bar{\mathbf{x}}_{n-1}^{a,0}$ and $\mathbf{X}_{n-1}^{a,0}$) and
 24 deriving weight coefficients at t_n ($\bar{\mathbf{w}}_n^{a,0}$ and $\mathbf{W}_n^{a,0}$), using Eqs. (12), (13) and (15).

25
 26 At the i^{th} iteration, the no-cost smoother is applied so that the smoothed ensemble mean at
 27 t_{n-1} is given by

$$\bar{\mathbf{x}}_{n-1}^{a,i+1} = \bar{\mathbf{x}}_{n-1}^{a,i} + \mathbf{X}_{n-1}^{a,i} \bar{\mathbf{w}}_n^{a,i}, \quad (17)$$

29 and the smoothed perturbations are obtained as

$$\mathbf{X}_{n-1}^{a,i+1} = \mathbf{X}_{n-1}^{a,i} \mathbf{W}_n^{a,i} + \mathbf{E}_{n-1}^{i+1} \quad (18)$$

1 where \mathbf{E}_{n-1}^{i+1} are small random Gaussian perturbations. With the modified analysis
 2 ensemble at t_{n-1} , the nonlinear model is used to integrate all ensemble members forward
 3 to the end of the assimilation window, t_n . The newly evolved ensemble is the new
 4 background ensemble ($\mathbf{x}_n^{b,i+1}$) and the weight coefficients are computed again to obtain
 5 $\bar{\mathbf{w}}_n^{a,i+1}$, $\mathbf{W}_n^{a,i+1}$ based on $\mathbf{y}_n^o - \bar{\mathbf{y}}_n^{b,i+1}$ and $\mathbf{Y}_n^{b,i+1}$ with Eqs. (12), (13) and (15). The new
 6 background ensemble has a better mean (closer to the truth) and perturbations (more
 7 reliable flow-dependent structures) because it is initialized from a more accurate
 8 ensemble. More specifically, \mathbf{y}_n^o (the observation at t_n) has been used i times in $\mathbf{x}_n^{b,i}$ and
 9 $\mathbf{x}_{n-1}^{a,i}$, and $i+1$ times in $\mathbf{x}_n^{a,i}$. At $i=0$, $\mathbf{x}_{n-1}^{a,0}$ is the final analysis ensemble derived from the
 10 previous analysis cycle at t_{n-1} and $\mathbf{x}_n^{b,0}$ are the following forecast. Therefore, they do not
 11 know the observation \mathbf{y}_n^o . With the no-cost smoother, $\mathbf{x}_{n-1}^{a,1}$ then contains the information
 12 about \mathbf{y}_n^o (i.e. use the observation once) and so does the following forecast $\mathbf{x}_n^{b,1}$.
 13

14 The random perturbations, \mathbf{E}_{n-1}^{i+1} , are added in Eq. (18) because, under linear conditions,
 15 integrating the smoothed analysis from t_{n-1} to t_n will end up with the same analysis
 16 ensemble previously derived. Adding these very small random Gaussian perturbations
 17 before the forward integration of the smoothed analysis ensemble not only avoids
 18 evolving into the same ensemble, but also has the advantage of exploring new growing
 19 directions that may have been missed by the ensemble (Kalnay et al., 2007a). Corazza et
 20 al. (2002) have shown that adding random perturbations on the ensemble members can
 21 “refresh” the space spanned by the ensemble into new growing directions that may be
 22 missed otherwise, so that the forecast errors project better into the slightly perturbed
 23 ensemble space. In the EnKF, the analysis corrections are made within the space spanned
 24 by the ensemble. Under nonlinear condition and by slightly perturbing the ensemble
 25 space, we can generate a (slightly) different evolved ensemble and have better chance to
 26 correct the errors within the newly growing error subspace, given exactly the same
 27 observations. Adding small random perturbations at each iteration also has the effect of
 28 slightly reducing the time correlation between the analysis and the observation errors
 29 (Craig Bishop, personal communication, 2009).
 30

1 The criterion of when to stop the iterations is based on the reduction of the forecast misfit
2 to the observations (measured by the RMS of the differences between the observations
3 and the forecasts $\mathbf{RMS_OMF}_i$) and scaled by the observation error, σ_o . The procedure of
4 modifying and re-evolving the ensemble at t_{n-1} is repeated, for as long as the relative
5 reduction of the misfit ε is greater than a threshold ε_s (Eq. 19). If the reduction of the fit is
6 smaller than the chosen threshold ε_s , it indicates little useful additional information can be
7 extracted from the same set of observations without seriously overfitting them.

$$\varepsilon = \frac{\mathbf{RMS_OMF}_{i-1} - \mathbf{RMS_OMF}_i}{\sigma_o} > \varepsilon_s \quad (19)$$

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

19 In the experiments with the Lorenz model, iterations of the RIP scheme are continued as
20 long as $\varepsilon > 0.001$ in Eq. (19) but the maximum number of iterations allowed is set to 10.
21 The criteria for stopping may depend on the density of observations. Experimental results
22 suggest that more iterations with a stricter threshold are required to optimize the
23 performance of the RIP method with fewer observations (see Section 4.d). Although the
24 success of the RIP scheme was demonstrated in Kalnay and Yang (2010), we should note
25 that the computational cost of the RIP scheme is relatively high since all the ensemble
26 members have to be integrated using the nonlinear model. In the next sub-section, the
27 QOL scheme is introduced as a simplified version of the RIP scheme where only the
28 ensemble mean is integrated within the assimilation window, and only one or two
29 iterations are allowed.

3.d Quasi Outer-Loop (QOL) algorithm

The main purpose of the 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 the RIP. The steps for performing the 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 the QOL, given the analysis ensemble $(\bar{\mathbf{x}}_{n-1}^{a,0}$ and $\mathbf{X}_{n-1}^{a,0})$ at t_{n-1} and weight coefficients for correcting the mean state $(\bar{\mathbf{w}}_n^{a,0})$ at t_n . The QOL uses the weights $\bar{\mathbf{w}}_n^{a,i}$ obtained at the end of the window to update the analysis ensemble mean at t_{n-1} :

$$\bar{\mathbf{x}}_{n-1}^{a,i+1} = \bar{\mathbf{x}}_{n-1}^{a,i} + \mathbf{X}_{n-1}^{a,0} \bar{\mathbf{w}}_n^{a,i} \quad (20)$$

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 the QOL. After applying Eq. (20), the nonlinear model is used to evolve only the ensemble mean trajectory: $\bar{\mathbf{x}}_n^{b,i+1} = M(\bar{\mathbf{x}}_{n-1}^{a,i+1})$. The innovation vector is now updated to $\mathbf{y}_n^o - h(\bar{\mathbf{x}}_n^{b,i+1})$.

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 previous iteration, and, as in RIP, small perturbations are added:

$$\mathbf{X}_n^{b,i+1} = \mathbf{X}_n^{a,i} + \mathbf{E}_n^{i+1} \quad (21)$$

As in Eq (18), small random Gaussian perturbations \mathbf{E}_n^{i+1} are added so that $\mathbf{X}_n^{b,i+1}$ is not identical to $\mathbf{X}_n^{a,i}$, and as noted before, this also reduces slightly the dependence of the analysis on the observations. We then define the weights corresponding to the updated innovation vector as in Eq. (4):

$$\bar{\mathbf{w}}_n^{a,i+1} = \hat{\mathbf{P}}_n^{a,i+1} (\mathbf{Y}_n^{b,i+1})^T \mathbf{R}^{-1} (\mathbf{y}_n^o - h(\bar{\mathbf{x}}_n^{b,i+1})) \quad (22)$$

1 Eq. (19) is again used to determine whether another QOL iteration should be applied,
2 with a criterion of $\varepsilon > 0.01$. Otherwise, $\bar{\mathbf{x}}_n^{a,i+1}$ and $\mathbf{X}_n^{a,i+1}$ are the final analysis ensemble
3 mean and perturbations and serve as the initial conditions for the next analysis cycle. The
4 results are less sensitive to the value of ε than RIP, since only one or two outer-loop
5 iterations are allowed. If the criterion is not satisfied, the QOL becomes the same as the
6 standard LETKF with zero iteration.

7

8 Jazwinski (1970) had already suggested a very similar idea of an outer loop allowing the
9 ensemble to be recentered on a more accurate nonlinear solution for the Extended
10 Kalman Filter (p. 276, footnote 3). This method was then introduced as the IKF in Bell
11 and Cathey (1993) to linearize the observation operator with respect to a more accurate
12 state so that the linearity assumption is less likely to be violated. The incremental 4D-Var
13 (Courtier et al., 1994) and other iterative EnKFs (e.g. Gao and Reynolds, 2009, Gu and
14 Oliver, 2007) related to the Gauss-Newton method share the same concept of outer-
15 looping. An important difference between these methods and the RIP/QOL is that we
16 modify the background error covariance at each iteration in order to reflect the
17 improvement of the mean state (Eq. 21), whereas in 4D-Var, IKF and IEnKFs the
18 background state (first guess) and the corresponding statistics are kept constant.

19

20 **4. Results with the Lorenz three-Variable model**

21

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

27

28 **4.a Lorenz 3-variable model and experiment settings**

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

30

1 $\frac{dx}{dt} = \sigma(y - x); \quad \frac{dy}{dt} = rx - y - xz; \quad \frac{dz}{dt} = xy - bz$

2

3 The standard values for the parameters in this system used here, $\sigma = 10$, $r = 28$ and
4 $b = 8/3$, result in a chaotic behavior with two regimes (the famous butterfly pattern). The
5 model is integrated here using the fourth-order Runge-Kutta method with a time step of
6 0.01. With exponential instability of the model solution and abrupt regime changes, this
7 model has been widely used to demonstrate issues related to predictability (e.g., Palmer,
8 1993) or to evaluate the feasibility of an assimilation scheme (e.g., Miller et al., 1994,
9 Evensen, 1997, Yang et al. 2006, Kalnay et al. 2007a). Evans et al. (2004) showed that
10 the growth rate of perturbations with this model depends on the state where the
11 perturbation resides, and that large perturbation growth rate is observed in the last orbit
12 before the model trajectory changes regimes. These are also the locations where the
13 ensemble Kalman filter is observed to diverge (see Miller et al., 1994, Evensen, 1997
14 who used an adaptive time scheme integrator and an ensemble of 1000 members,),
15 because strong nonlinear perturbation growth results in poor estimates of the state by the
16 ensemble and causes the filter to fail.

17

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

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31

4.b Nonlinearity with a long assimilation window

With the same model and assimilation settings (section 4.a), Kalnay et al. (2007a) compared the analysis accuracy between fully optimized LETKF and 4D-Var schemes. Their results (reproduced in Table 1) show that the performance of LETKF with 3 ensemble members is comparable to 4D-Var when using a short assimilation window (8 time steps or 0.08 time units) for which 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 more accurate than the LETKF analysis, but only if the 4D-Var quasi-static variational method of solution (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 the LETKF is hindered by the fact that the nonlinear perturbations are no longer Gaussian.

Table 1 compares the analysis accuracy of 4D-Var, the standard LETKF, the LETKF with QOL, and the LETKF with RIP, 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. 8 and 11) have zero mean with standard deviations of 0.0004 and 0.0001 respectively. All the LETKF-related schemes use $K=3$ ensemble members. 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 the ensemble size is necessarily small in practical applications. In Table 1, 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 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

1 obtained by Kalnay et al. (2007a). Their study showed that with an observation frequency
2 of 25 time steps, the optimal length of the assimilation window for 4D-Var is a window
3 of 75 time steps for this model, and that for such long windows the optimized 4D-Var is
4 significantly more accurate than the optimized LETKF.

5
6 The LETKF-QOL improves the performance of the standard LETKF to an error level
7 now better than the optimized 4D-Var. The improvement is especially significant for the
8 longer window with nonlinear perturbations, suggesting that the LETKF-QOL is much
9 better in handling the nonlinearity of the trajectory for long assimilation windows. We
10 also note that the optimal multiplicative covariance inflation is much reduced from 22%
11 with the standard LETKF to 8% with the QOL, indicating that the ensemble-based
12 background error covariance can better represent the dynamic uncertainties. RIP shows a
13 performance similar to the QOL with the linear window, for which little further
14 improvement can be gained. With the longer nonlinear window, however, the
15 improvement obtained using the RIP is significantly larger than with the QOL, and the
16 RMS analysis error is further reduced from 0.47 to 0.35 with only 4.7% multiplicative
17 covariance inflation. The RIP out-performs the QOL because it also enables improving
18 the analysis ensemble perturbations as well as the mean at the beginning of each
19 assimilation window. We note that, on the average, eight iterations are required but when
20 the dynamics become very nonlinear, as during a regime transition, RIP may take as
21 many as 10 iterations to fulfill the criterion $\varepsilon \leq 0.001$. We note that the RIP analysis (with
22 3 ensemble members and about eight iterations) is still more accurate than the standard
23 LETKF with 24 ensemble members.

24
25 Table 1 confirms the advantage of QOL for dealing with long assimilation windows and
26 that RIP is even more advantageous. Figure 2 shows the true time evolution of the
27 variable y (black line), the analysis/forecast ensemble mean state from the standard
28 LETKF (blue lines), and the ensemble mean from the LETKF with the QOL (red lines) in
29 a typical example of a dynamical instability leading to nonlinearity and filter divergence.
30 In Figure 2, the dots are the analyses, i.e., the initial states of the forecasts. Such filter
31 divergence is found to occasionally occur in the standard LETKF even if the ensemble

1 size is increased (here $K=3$). As shown in Figure 2, the standard LETKF (in blue) fails to
2 track the regime transition and gives wrong information about whether a regime change
3 actually takes place or whether the system stays close to the borderline state. For
4 example, starting at time=17.0, the analysis mean states (blue dots) become less and less
5 accurate and the corresponding forecasts are completely off the true trajectory even
6 though the observations are still being assimilated. Only when the background happens to
7 be again close to the true dynamics (given by the observations) near time=24.5, is the true
8 trajectory re-captured, and the LETKF gets back onto the right regime solution. With the
9 QOL, by contrast, the evolution of the trajectory is much improved and stays on the
10 correct regime, following the true trajectory very well, and suggesting that the quasi
11 outer-loops have a clear advantage during unstable periods. When getting close to the
12 boundary between regimes (Evans et al., 2004), nonlinear perturbation growth quickly
13 degrades the forecast trajectory for the standard LETKF and for the QOL before the
14 second iteration. Particularly for the assimilation window between time=17.25 and 17.5
15 (between point C and D), both forecasts quickly get into the wrong regime. Even with the
16 observation at time=17.5, the analysis obtained from the standard LETKF is still in the
17 wrong regime, causing a further worse forecast and leading to filter divergence. By
18 comparison, the QOL corrects the trajectory because it makes use of the observation *at*
19 *the end of the assimilation window*. Thus, the analysis mean, serving as the initial state as
20 the next analysis cycle (the red dot at Point D) is close to the truth and is able to stay in
21 the right regime.

22

23 **4.c Ensemble distribution**

24

25 This section we evaluate the ensemble of the standard LETKF, QOL and RIP with the
26 statistical moments to depict the shape of the ensemble distribution, including the
27 standard deviation and kurtosis of the ensemble. Also, to exhibit statistical significance,
28 results in the following are shown with a large ensemble member for $K=24$ and both
29 QOL and RIP uses random perturbations with an amplitude of 0.003. Filter divergence
30 (i.e., the analysis stays in the wrong regime for several cycles) still occurs occasionally
31 with the standard LETKF with 24 ensemble members while the other two schemes are

1 always in the right regime even with $K=3$. The large ensemble size just helps the standard
2 LETKF to recover from filter divergence more quickly so that it can begin to track the
3 true dynamics again.

4
5 First, the standard deviation of the ensemble represents the spread of the ensemble
6 distribution. Table 2 shows the mean RMS error and ensemble spread (standard
7 deviation) of the background and analysis ensemble from three schemes. Results show
8 that on the average the ensemble spread and mean state error are quite comparable in the
9 standard LETKF, but the mean state is the least accurate among the three schemes. When
10 the standard LETKF performs poorly, the spread is severely underestimated. Among the
11 three schemes, RIP shows a most accurate analysis mean but with a small ensemble
12 spread. This agrees with what we illustrated in Section 2 that a small ensemble spread is a
13 necessary characteristic of RIP in order to perform the assimilation in a “soft” way
14 assimilating the same observation repeatedly, and giving the optimal analysis. Given the
15 success of RIP, we argue that the small ensemble spread from RIP is essential for
16 performing incremental analysis correction instead of representing the uncertainty of the
17 mean state. As for QOL, the analysis ensemble spread is only slightly smaller than the
18 RMS error, reflecting the fact that only one or two iterations are used.

19
20 For a linear model, if the analysis ensemble has a Gaussian distribution, the forecast
21 ensemble will also be Gaussian. If the model is nonlinear and the assimilation window is
22 long, ensemble perturbations grow nonlinearly, and as discussed in the introduction, the
23 presence of strong nonlinearities will distort the distribution of the ensemble making it
24 non-Gaussian, so that the Kalman Filter formulation ceases to be optimal. To investigate
25 whether using the QOL improves Gaussianity, kurtosis is used to measure the “flatness”
26 of the LETKF ensemble distribution. In a Gaussian distribution, kurtosis is equal to zero.
27 Large positive kurtosis represents a non-Gaussian flatter distribution with heavy tails, i.e.
28 with many outliers. Negative kurtosis corresponds to a peaked distribution narrower than
29 the Gaussian.

30

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

15

16 Without adding the random perturbations, the RMS analysis errors become slightly larger
17 for both schemes (0.49 to 0.51 with QOL and 0.33 to 0.37 with RIP). Comparing the
18 results with and without the \mathbf{E} term for both schemes, results show that the random
19 perturbations have contributed positively to the Gaussianity of the ensemble distribution:
20 the IQRs of the kurtosis become smaller, especially for the RIP method, since for the
21 QOL only up to two iterations are allowed. Results from Table 3 also imply that the
22 random perturbations with Gaussian statistics have the function of modifying the
23 ensemble distribution into a more Gaussian shape so that fewer outlier members occur. A
24 similar stochastic effect from the random perturbations regulating the ensemble
25 distribution has been discussed for a stochastic EnKF (Lawson and Hansen, 2004).

26

27 **4.d Impact of incomplete observations**

28

29 So far, observations were made available for all three variables, x , y and z , so that the
30 influence of error covariances between variables is not dominant. We further consider the
31 impact of nonlinearity due to sparse observations. For this purpose, instead of

1 assimilating observations in x , y and z , we only observe subsets of these observations,
2 both for the linear (8 time-step) and non-linear (25 time-step) windows (Table 4). We
3 note that with fewer observations available, the optimal threshold used in the RIP was
4 reduced to 0.0001, indicating that more iterations are required to make the model states
5 pull closer to the observations.

6
7 Yang et al. (2006) pointed out that in the Lorenz 3-variable model, the y variable provides
8 the most effective observations to synchronize the model trajectory with the true states. In
9 agreement with this observation, when only two observations are assimilated every 8 time
10 steps with the standard LETKF (the first row in Table 4a), the analysis accuracy with
11 observations (x, y) or (y, z) is comparable with the analysis with all three observations
12 assimilated, while the one assimilating observations (x, z) is much less accurate. With the
13 25 time-step window, the analysis error with observations (x, z) is even larger than the
14 observation error. As a consequence, with fewer observations and a longer assimilation
15 window, it becomes more difficult to keep the model trajectories close to the truth, and
16 the errors for the non-observed variables are large enough to quickly amplify the
17 nonlinear effect.

18
19 From the last row in Table 4(a) and (b), with only one or two observations, the relative
20 improvement of the RIP method with respect to the standard LETKF is even larger than
21 when all three observations are available. For the nonlinear window and incomplete
22 observations, the RIP method has at least 40% smaller errors than the standard LETKF.
23 The RIP analysis with two observations is more accurate than the standard LETKF
24 analysis with all three observations. Observations of only x and z also become more
25 effective with the RIP method. For the 8-steps linear window, the improvement with the
26 RIP method is most significant when the observation y is not available because the
27 available observations are less effective than y . Even when observing z alone, it is still
28 possible to constrain the model trajectory. The result implies that the RIP method can be
29 used to help overcome nonlinearity problems exhibited in data void regions.

30
31 With the nonlinear window, the performance of the QOL is always in between the RIP

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

11
12 Among all the cases, assimilating observations y and z with the QOL and RIP methods
13 significantly improves the Gaussianity of the ensemble compared with the standard
14 LETKF. The nonlinear growth of the errors can be best controlled with these two
15 observations because the nonlinear terms appear only in the governing equations for y
16 and z variables. The ensemble distribution of the non-observed x variable also becomes
17 more Gaussian: the median values of the kurtosis are 1.06, 0.35 and 0.24, for the standard
18 LETKF, QOL and RIP methods respectively, and the corresponding IQR values are 7.69,
19 4.06 and 1.49. This implies that the dynamical adjustment of RIP and QOL is useful to
20 constrain the perturbations growth for unobserved variable or data-void region.

21

22 **4.e Impact of the choice of the threshold**

23 The threshold (ϵ) used in RIP that determines when to stop the iterations and avoid over-
24 fitting the observations. Figure 3 shows the RIP performance with different criteria as the
25 threshold for cases with all 3 observations and only 2 observations (x and z), still under
26 the limitation of a maximum of 10 iterations. Results from Figure 3 indicate that the
27 performance of RIP has reached the asymptotic level at $\epsilon > 0.001$. In addition, even with
28 a large threshold like $\epsilon > 5$ (only one or no iteration is used) RIP always provides
29 improvement over the standard LETKF and the performance is close to the QOL. In
30 contrast, RIP with two observations is more sensitive to the criterion and the best
31 performance is obtained at $\epsilon > 10^{-5}$ (always using close to 10 iterations). A threshold that

1 is too small and results too many iterations will produce some degradation in the analysis
2 accuracy due to over-fitting the observations. For example, if 10 iterations are always
3 used in RIP for assimilating all observations, the RMS error degrades from 0.35 to 0.39.
4 The results also suggest that most of the corrections take place at the first few iterations
5 and RIP with small ensemble size can always perform well with a conservative criterion
6 that stops iterating rather early.

7
8 We note again that the QOL is less sensitive to the threshold since in practice we only
9 allow for a second iteration when the evolution of the background state is under strong
10 nonlinear instability.

11 12 **4.f Impact of the forecast/observation error correlation**

13
14 We note that after the first iteration, reusing the observation in QOL and RIP, introduces
15 a correlation of the forecast and observation errors. Since this correlation is not accounted
16 for in RIP and QOL, this error correlation should reduce their accuracy. To estimate the
17 impact of neglecting the correlation between the observation and background errors, we
18 have artificially eliminated the correlation by repeating the experiments in Table 1 but
19 introducing new, independent observation errors at every new iteration (something that
20 could not be done in practice). There was indeed a reduction of analysis error in RIP and
21 QOL, but it is small compared to the large improvement obtained with both methods over
22 the standard LETKF. For the infrequent observation case (assimilation window of 25
23 steps) the QOL analysis errors without error correlation were reduced from 0.47 to 0.38,
24 and for RIP from 0.35 to 0.26 respectively. The fact that both RIP and QOL have
25 achieved excellent results in reducing the analysis errors, both for the Lorenz 63 model
26 and with much larger systems with real observations (Yang et al. 2011 and Penny et al.
27 2011) indicates that neglecting the correlation between the observation and background
28 errors does less harm to the assimilation with RIP/QOL than the strong nonlinear
29 perturbation growth does to the standard LETKF. Nevertheless, the results imply that the
30 performance RIP/QOL could be further improved by accounting for this correlation of
31 errors effect of correlated errors.

5. Comparisons between iterative EnKFs 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), also designed for handling nonlinear data assimilation. The EnRML was originally proposed for solving the 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, determined by the parameter β_i . For this comparison, the formulas of EnRML are modified to derive the optimal initial state within each window as in 4D-Var, without including 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. 23) is defined to assimilate observations arranged at the end of the assimilation window, t_n .

$$J(\mathbf{x}_{n-1}) = \frac{1}{2} [\mathbf{x}_{n-1} - \mathbf{x}_{b,n-1}]^T \mathbf{P}_{b,n-1}^{-1} [\mathbf{x}_{n-1} - \mathbf{x}_{b,n-1}] + \frac{1}{2} [H(M[\mathbf{x}_{n-1}]) - \mathbf{y}_{o,n}]^T \mathbf{R}^{-1} [H(M[\mathbf{x}_{n-1}]) - \mathbf{y}_{o,n}] \quad (23)$$

The optimal estimation, \mathbf{x}_{n-1} , at the beginning of the window t_{n-1} is obtained by minimizing Eq (23).

In Eq. (23), is $\mathbf{x}_{b,n-1}$ the first guess at 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 observations are perturbed following Evensen (1994):

The implementation of the EnRML algorithm is as follows:

1. The cost-function is minimized for each ensemble member (k), with perturbed observations, giving the new iteration $i+1$ of the window initial condition:

$$J(\mathbf{x}_{n-1}^k) = \frac{1}{2} [\mathbf{x}_{n-1}^k - \mathbf{x}_{bn-1}^k]^T \mathbf{P}_{bn-1}^{-1} [\mathbf{x}_{n-1}^k - \mathbf{x}_{bn-1}^k] + \frac{1}{2} [H(\mathbf{x}_n^k) - \mathbf{y}_{on}^k]^T \mathbf{R}^{-1} [H(\mathbf{x}_n^k) - \mathbf{y}_{on}^k]$$

$$\mathbf{x}_{n-1}^{k,i+1} = \beta_i \mathbf{x}_{bn-1}^k + (1 - \beta_i) \mathbf{x}_{n-1}^{k,i} - \beta_i \left(\frac{1}{K-1} \right) \mathbf{X}_{bn-1}^k (\mathbf{Y}_{bn}^k)^T \left(\mathbf{R} + \left(\frac{1}{K-1} \right) \mathbf{Y}_{bn}^k (\mathbf{Y}_{bn}^k)^T \right) [\mathbf{x}_n^{k,i} - \mathbf{y}_{on}^{k,i} - (\mathbf{x}_n^{k,i} - \mathbf{x}_{bn}^k)]$$

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

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

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

Recall that with only 3 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 3 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 and RIP, which now have errors of 0.49 and 0.33 respectively.

Running EnRML with 24 ensemble members is successful, with β_i starting at 0.5 and with an average of 7 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 RIP with 3 ensemble members.

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

8
9 The statistical behavior of the ensemble perturbations among QOL, RIP and EnRML are
10 also examined. Results show that the Gaussianity of the EnRML is not significantly
11 different from the one with RIP (figure not shown). However, we observe that the spread
12 of the EnRML ensemble represents well the uncertainty of the mean state when the
13 EnRML analysis is very accurate; in comparison, the ensemble spread from RIP is
14 smaller for these cases because, as discussed in section 2, RIP has adjusted to the re-use
15 of the observations by reducing the ensemble spread.

16
17 In summary, EnRML derives the optimal state and its uncertainty through minimizing a
18 cost-function. Given a poor prior state and/or non-representative prior error statistics, the
19 ability to fit the model solution to observations through the minimization of a cost-
20 function is limited. RIP instead aims to improve the prior state and corresponding error
21 structure so that the observations can be better used: at every iteration, RIP, solves a
22 different cost-function that is closer to being quadratic since the prior state is improved.
23 Since in the RIP ensemble perturbations are repeatedly rescaled, RIP has an ensemble
24 spread smaller than the analysis error but consistent with the multiple use of the
25 observations.

26 27 **6. Summary and discussion**

28
29 A new type of outer-loop, different from the incremental minimization of a cost-function
30 based on the maximum likelihood method, is proposed for the Ensemble Kalman Filter
31 (EnKF) framework. It improves the EnKF ability to handle the strongly nonlinear

1 evolving dynamics that can take place in long assimilation windows. Unlike the
2 variational outer loops, which do not modify the prior state, RIP and its simpler version,
3 QOL, take advantage of the information from “future observations” to improve the initial
4 ensemble in the assimilation window, and use the observations more than once.

5
6 In Section 2 we show with a simple linear model that after the initial spin-up, using RIP
7 with a constant number of iterations per analysis cycle produces essentially the same
8 analysis means as the optimal Kalman Filter. The main differences are that with RIP, the
9 size of the ensemble squared spread is reduced by N , the number of iterations per cycle,
10 and the total analysis increment in a given analysis cycle is applied as the sum of multiple
11 smaller increments. We argue that in a nonlinear scenario, these differences can be an
12 advantage: first, the linear approximation to the set of possible model states provided by
13 the ensemble is better with a smaller ensemble spread; and second, breaking the analysis
14 increment into smaller steps and recomputing the linear approximation at each step can
15 allow the increments to follow the nonlinear set toward the truth better than a single
16 increment.

17
18 When there is strongly nonlinear error growth, the extraction of information from
19 observations in the standard EnKF is less than optimal. Therefore, it is important to
20 increase the influence of the observations on the analysis. RIP does this by improving
21 both the ensemble mean and the perturbations incrementally, reusing the observations N
22 times, rather than reducing the observation error covariance by N . The QOL method,
23 newly proposed as a simplified version of RIP, aims to improve just the ensemble mean
24 so that ensemble perturbations are centered at a more accurate state. Both the RIP and
25 QOL schemes are based on the application of a no-cost smoother to improve the analysis
26 at the beginning of the assimilation window and thus the nonlinear trajectory of the
27 ensemble based on information provided by “future” observations available within the
28 assimilation window. QOL and RIP were implemented here within the LETKF
29 framework to examine their ability to handle the nonlinearities of the dynamics, but they
30 could be applied to any EnKF where the analysis perturbation matrix is obtained from the

1 background perturbation matrix through a right multiplication by a matrix of weights,
2 such as the ETKF or LETKF.

3
4 In a nonlinear scenario, EnKF and EnKF-RIP will give different results, but as we argued
5 above, EnKF-RIP may handle nonlinearities better due to smaller ensemble spread and a
6 multi-step analysis increment. Just as Ensemble Kalman Filters benefit from ad hoc
7 modifications (such as covariance inflation) from the optimal linear formulation, RIP
8 benefits from modifications such as adaptively varying the number of iterations per
9 analysis cycle, and adding small perturbations. In this article we use such modifications
10 of the basic RIP approach described above, but we emphasize that the basic approach
11 provides most of the improvement we find in our numerical experiments. These
12 modifications require tuning two parameters in order to optimize the performance of both
13 RIP and QOL. The first is a threshold for the improvement of the fit of the forecasts to
14 the observations, and the second is the size of additive random perturbations. For the RIP,
15 the threshold determines when to stop the iterations. Experimental results suggest that for
16 smaller ensembles, a larger threshold (with fewer iterations) would be appropriate to
17 avoid overfitting the observations. Kalnay and Yang (2010) tested LETFK-RIP to
18 accelerate the spin-up in a quasi-geostrophic model, and used the threshold to estimate
19 whether the spin-up phase was over, and then stop the RIP iterations and return the
20 system to the standard LETKF. The QOL is much less sensitive to the threshold because
21 only one or two iterations are used when the model trajectory undergoes nonlinear
22 instabilities. Since most of the improvements are attained in the first few iterations, both
23 RIP and QOL should work with a conservative approach using only few iterations.
24 Results from both schemes show that the analysis accuracy is slightly improved by
25 adding small random perturbations that also play a role in regulating the Gaussianity of
26 the ensemble distribution.

27
28 The performance of the LETKF with the QOL and RIP schemes is tested with the Lorenz
29 3-variable model. Results show that QOL and RIP allow the LETKF to use longer
30 assimilation windows with significant improvements of the analysis accuracy, especially
31 during periods of highly nonlinear growth. For low-frequency observations (every 25

1 steps, leading to long assimilation windows), and using optimal inflation, the standard
2 LETKF RMS error is 0.68, whereas for the QOL and RIP the RMS errors are 0.47 and
3 0.35 respectively. This can be compared with the best 4D-Var analysis error of 0.53,
4 obtained using both optimal long assimilation windows (75 steps) and the quasi-static
5 variational analysis (Pires et al., 1996). As discussed in Section 2, RIP has a small
6 ensemble spread due to the incremental analysis correction, while the standard LETKF,
7 with an ensemble spread comparable to the uncertainty of the mean state, computes the
8 analysis increment only once, but with worse accuracy during strong nonlinear growth.
9 When only one or two of the 3 variables are observed, the improvements of the QOL and
10 RIP algorithms with respect to the standard LETKF are further enhanced, suggesting that
11 these methods can be used to handle nonlinearity introduced by data voids.

12

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

20

21 The original outer-loop used in variational data assimilation in incremental form does not
22 change the background state (first guess) in the outer loop: it only improves the nonlinear
23 trajectory of the model used to compute the innovations and the state used to linearize the
24 observation operator (Courtier, 1994). By contrast, the QOL uses the no-cost smoother
25 and the future observations within the assimilation window to improve the ensemble
26 mean and re-center the smoothed perturbations around this improved mean. RIP reruns
27 the whole ensemble, so that it not only improves the initial ensemble mean but also the
28 perturbations. This makes the QOL and RIP quite different from the variational outer
29 loop and seems to violate the basic rule that “data should be used once and then
30 discarded” (Ide et al, 1997). In this study, we provide a theoretical explanation for the
31 success of RIP based on linear dynamics and show that repeatedly assimilating same

1 observations is a “soft” way to increasing the observation influence. Most importantly,
2 such multi-step analysis increment becomes advantageous for nonlinear cases since the
3 updated state can better follow the true dynamics. In addition, we note that the smaller
4 ensemble spread is a characteristic of RIP, allowing multi-step analysis corrections with
5 small increments, and is associated with the fact that the observations are used multiple
6 times. If the ensemble needs to be inflated for the purpose of obtaining enough spread for
7 the long-term forecasts, one can then either use fewer RIP iterations or inflate the
8 ensemble with an adaptive inflation scheme (e.g. Miyoshi, 2011).

9
10 Since the performance of the QOL and RIP on the 3-variable Lorenz model and several
11 other small models is excellent, they are now being tested within more realistic and
12 dynamically complex models in order to investigate the potential impact on improving
13 the nonlinearity of the background trajectory. Applications to typhoon regional data
14 assimilation (Yang et al., 2011) and global ocean assimilation (Penny, 2011) also have
15 demonstrated very encouraging results.

16 17 **Acknowledgements**

18 We are very grateful for the insights provided by Anna Trevisan, Arlindo da Silva and
19 Fuqing Zhang. Takemasa Miyoshi and Kayo Ide and other members of the University of
20 Maryland Weather and Chaos Group, Ross Hoffman (AER) and Xiang-Yu Huang
21 (NCAR Data Assimilation Test Center), made very helpful suggestions. We also
22 appreciate the comments and suggestions of Pavel Sakov and two other reviewers. Shu-
23 Chih Yang is sponsored by Taiwan National Science Council grant 97-2111-m-008-25,
24 98-2111-m-008-014 and the NCU Development Program for the Top-Ranked University
25 sponsored by the Ministry of Education. Eugenia Kalnay acknowledges support from
26 NASA grants NNX08AD40G, NN07AM97G, and DOE grant DEFG0207ER64437.

27
28

1 **6. References**

- 2 Andersson, E., M. Fisher, E. Hólm, L. Isaksen, G. Radnóti, and Y. Trémolet, 2005: Will
3 the 4D-Var approach be defeated by nonlinearity? *ECMWF Tech. Memo.* , **479**,
4 pp 26.
- 5 Bell, B.M. and Cathey, F.W. 1993. The iterated Kalman filter update as a Gauss-Newton
6 method. *IEEE Transactions on Automatic Control* **38** (2): 294–297.
7 DOI:10.1109/9.250476.
- 8 ----- 1994: The iterated Kalman smoother as a Gauss-Newton method. *SIAM J.*
9 *Optimization*, **4**, 626-636.
- 10 Bishop, C. H., Etherton, B. J. and Majumdar, S. 2001. Adaptive sampling with the
11 ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*
12 **129**, 420–436.
- 13 Corazza, M., E. Kalnay, D. J. Patil, R. Morss, M. Cai, I. Szunyogh, B. R. hunt, E. Ott and
14 J. Yorke, 2002: Use of the breeding technique to estimate the structure of the
15 analysis “errors of the day”. *Nonlinear Processes in Geophysics*, **10**, 233- 243.
- 16 Courtier, P., J. N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational
17 implementation of 4DVAR, using an incremental approach. *Quart. J. Roy.*
18 *Meteor. Soc.*, **120**,1367-1387.
- 19 Evans, E., N. Bhatti, J. Kinney, L. Pann, M. Pena, S-C Yang, E. Kalnay and J. Hansen,
20 2004: RISE: Undergraduates Find That Regime Changes in Lorenz's Model are
21 Predictable. *Bull. Amer. Meteor. Soc.*, **85**, 520-524.
- 22 Evensen, G., 1992: Using the extended kalman filter with a multilayer quasi-geostrophic
23 ocean model. *J. Geophys. Res.*, **97** (C11), 17 905–17 924.
- 24 -----, 1994: Sequential data assimilation with a nonlinear quasi-geo strophic model
25 using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99** (C5),
26 10 143–10 162.
- 27 -----, 1997: Advanced data assimilation for strongly nonlinear dynamics. *Mon.*
28 *Wea. Rev.*, **125**, 1342–1354.
- 29 ----- and P. J. van Leeuwen, 2000: An ensemble Kalman Smoother for nonlinear
30 dynamics. *Mon. Wea. Rev.*, **128**, 1852-1867.
- 31 -----, 2003: The ensemble Kalman filter: theoretical formulation and practical

1 implementation. *Ocean Dyn.* **53**, 343–367.

2 Fertig, E., Harlim J., and B. R. Hunt, 2007: A comparative study of 4D-VAR and a 4D
3 Ensemble Kalman Filter: Perfect Model Simulations with Lorenz-96. *Tellus A*, **59**,
4 96-101.

5 Gu, Y. and D. S. Oliver, 2007: An iterative ensemble Kalman filter for multiphase fluid
6 flow data assimilation. *SPE journal*, 437-446.

7 Hamill, T.M., 2000: Interpretation of rank histograms for verifying ensemble forecasts.
8 *Mon. Wea. Rev.*, **129**, 550-560.

9 Hunt, B.R., E. Kalnay, E.J. Kostelich, E. Ott, D.J. Patil, T. Sauer, I. Szunyogh, J.A.
10 Yorke, and A.V. Zimin, 2004: Four-Dimensional Ensemble Kalman Filtering.
11 *Tellus* 56A, 273-277.

12 -----, E. Kostelich, I. Szunyogh, 2007: Efficient data assimilation for
13 spatiotemporal chaos: a Local Ensemble Transform Kalman Filter. *Physica D*,
14 **230**, 112-126.

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

17 Jazwinski, A.H., 1970. *Stochastic processes and filtering theory*. Academic Press, 376
18 pp.

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

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

23 -----, H. Li, T. Miyoshi, S.-C. Yang and J. Ballabrera, 2007b: Response to the
24 Discussion on “4D-Var or EnKF?” by Nils Gustaffson. *Tellus A*, 59, 778-780.

25 -----, and S.-C. Yang, 2008: Accelerating the spin-up of Ensemble Kalman
26 Filtering, arXiv:Nonlinear/0806.0180v1.

27 -----, and S-C Yang, 2010: Accelerating the spin-up of Ensemble Kalman Filtering.
28 *Quart. J. Roy. Meteor. Soc.*, 136, 1644-1651.

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

31 Lawson, W. G. and Hansen, J. A. 2004: Implications of stochastic and deterministic filters

1 as ensemble based data assimilation methods in varying regimes of error growth.
2 *Mon. Wea. Rev.*, **132**, 1966–1981.

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

6 Li, G. and A. C. Reynolds, 2009: Iterative ensemble Kalman filters for data
7 assimilation, *SPE Journal*, 496-505.

8 Li, H., E. Kalnay, T. Miyoshi, C. M. Danforth, 2009: Accounting for Model Errors in
9 Ensemble Data Assimilation. *Mon. Wea. Rev.* **137**, 3407-3419.

10 Lorenz, E. N. 1963. Deterministic non-periodic flow. *J. Atmos. Sci.* **20**, 130–141.
11 ----- 1996. Predictability – a problem partly solved. In Proceedings on
12 predictability, held at ECMWF on 4–8 September 1995.

13 Miller, R. N., M. Ghil, and F. Gauthiez,, 1994: Advanced data assimilation in strongly
14 nonlinear dynamical systems. *J. Atmos. Sci.*, **51**, 1037–1055.

15 Miyoshi, T., 2011: The Gaussian Approach to Adaptive Covariance Inflation and Its
16 Implementation with the Local Ensemble Transform Kalman Filter. *Mon. Wea.*
17 *Rev.*, 139, 1519-1535.

18 Ott, E., B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D.
19 J. Patil, and J. A. Yorke, 2004: A local ensemble Kalman filter for atmospheric
20 data assimilation. *Tellus A*, **56**, 415-428.

21 Palmer, T N, 1993: Extended-Range Atmospheric Prediction and the Lorenz model. *Bull*
22 *Amer Meteor. Soc.*, 74, 49-65.

23 Penny, S. 2011: 4D-LETKF Compared with SODA on a Global Ocean Model. The 9th
24 Workshop on Satellite Data Assimilation, University of Maryland. May 24-15,
25 2011. (<http://www.jcsda.noaa.gov/documents/meetings/wkshp2011/dayTwo>
26 [/Penny_S_poster.pdf](http://www.jcsda.noaa.gov/documents/meetings/wkshp2011/dayTwo/Penny_S_poster.pdf))

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

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

1 1143-1170.

2 Sakov, P., G. Evensen, and L. Bertino 2010: Asynchronous data assimilation with the
3 EnKF. *Tellus* 62A, 24-29.

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

6 Tippett, M. K., J. L. Anderson, C. H. Bishop, T. M. Hamill, and J. S. Whitaker, 2003:
7 Ensemble Square Root Filters. *Mon. Wea. Rev.*, **131**, 1485-1490.

8 Trevisan, A., M. D’Isidoro and O. Talagrand, 2010: Four-dimensional variational
9 assimilation in the unstable subspace and the optimal subspace dimension. *Quart.*
10 *J. Roy. Meteor. Soc.*, **136**, 487-496.

11 Verlaan, M. and A. W. Heemink, 2001: Nonlinearity in Data Assimilation Applications:
12 A Practical Method for Analysis, *Mon. Wea. Rev.*, **129**, 1578–1589.

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

16 Wang, Y., G. Li and A.C. Reynolds, 2010: Estimation of Depths of Fluid Contacts by
17 History Matching Using Iterative Ensemble Kalman Smoothers, *SPE Journal* (in
18 press).

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

21 Yang, S-C , E. Kalnay, B. Hunt, N. Bowler, 2009a: Weight interpolation for efficient data
22 assimilation with the Local Ensemble Transform Kalman Filter, *Quart. J. Roy.*
23 *Meteor. Soc.*, **135**, 251-262.

24 -----, M. Corazza, A. Carrassi, E. Kalnay, and T. Miyoshi, 2009b: Comparison of
25 ensemble-based and variational-based data assimilation schemes in a quasi-
26 geostrophic model. *Mon. Wea. Rev.*, **137**, 639-709.

27 -----, E. Kalnay and T. Miyoshi, 2011: "Running in place" with the WRF-LETKF
28 for typhoon assimilation and prediction. 15th Symposium on Integrated
29 Observing and Assimilation Systems for the Atmosphere, Oceans, and Land
30 Surface, 2011 AMS meeting in Seattle, WA (extended abstract available at
31 <http://ams.confex.com/ams/91Annual/webprogram/Paper187577.html>).

1 Zupanski, M., S. J. Fletcher, I. M. Navon, B. Uzunoglu, R. P. Heikes, D. A. Randall, R.
2 D. Ringler4, D. Daescu, 2006: Initiation of ensemble data assimilation. *Tellus A*,
3 **58**, 159-170.
4
5

1 **List of Tables**

2 **Table 1** The analysis error of 4D-Var, LETKF, LETKF with QOL and LETKF with RIP.
3 ρ is the optimal multiplicative covariance inflation.

4

5 **Table 2** The RMS analysis error and ensemble spread of the standard LETKF, LETKF
6 with QOL and LETKF with RIP. Results are obtained with 24 ensemble members.

7

8 **Table 3** Median and interquartile range of the Kurtosis values obtained from the
9 background and analysis ensemble of standard LETKF, QOL and RIP methods, with and
10 without the random perturbations E. The amplitude of the random perturbation is 0.003.

11

12 **Table 4(a)** The RMS analysis error of LETKF, LETKF with QOL and LETKF with RIP
13 with the 8 time-step analysis cycles and different observations. The ensemble size is three
14 in all experiments.

15 **Table 4(b)** Same as Table 4(a), except for the 25 time-step analysis cycles.

16

17

18

19

1 **List of Figures**

2 **Figure 1 (a)** Analysis error of KF, KF-RIP with 2 and 10 iterations and observation error,
3 (b) same as (a) except for the estimated error variance and (c) averaged RMS analysis
4 error from KF and KF-RIP with different iteration numbers

5

6 **Figure 2** Time evolution of the variable y during a typical regime transition with strong
7 nonlinear growth between $t=17$ and $t=18$. The black line denotes the true trajectory and
8 the blue line denotes the evolution initializing from the standard LETKF analysis mean,
9 marked as the blue dots. The red lines are the evolution from the analyses with the QOL
10 marked as red dots. The points A, B, C, D indicate a regime transition, where the filter
11 divergence occurs with the standard LETKF. Three ensemble members are used in these
12 experiments.

13

14 **Figure 3** RMS analysis error of RIP with different criteria as the threshold with three
15 observations and with two observations. Three ensemble members are used in these
16 experiments.

17

18 **Figure 4 (a)** The RMS error from different analyses from the 800th to 1200th analysis
19 cycles and (b) analysis of the y variable from the truth and different analyses.

20

21

1 **Table 1** The analysis error of 4D-Var, LETKF, LETKF with QOL and LETKF with RIP.
 2 ρ is the optimal multiplicative covariance inflation.

	4D-Var	LETKF (3 members)	LETKF + QOL	LETKF + RIP
Obs every 8 time steps (linear window)	0.31	0.30	0.27	0.27
Obs every 25 time steps (nonlinear window)	0.53 (assim window=75)	0.68 ($\rho = 1.22$)	0.47 ($\rho = 1.08$)	0.35 ($\rho = 1.047$)

3
 4
 5
 6
 7
 8

1

2 **Table 2** The RMS analysis error and ensemble spread of the standard LETKF, LETKF
3 with QOL and LETKF with RIP. Results are obtained with 24 ensemble members.

4

K=24	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)

5

1

2 **Table 3** Median and interquartile range of the Kurtosis values obtained from the
3 background and analysis ensemble of standard LETKF, QOL and RIP methods, with and
4 without the random perturbations **E**. The amplitude of the random perturbation is 0.003.
5

	Median		Interquartile range (difference between 25% and 70% quantiles)	
	Background ensemble	Analysis ensemble	Background ensemble	Analysis ensemble
Standard LETKF	0.79	0.78	5.96	5.98
QOL (with E)	0.33	0.32	4.27	4.27
QOL (without E)	0.47	0.50	4.56	4.78
RIP (with E)	-0.10	0.45	1.74	1.74
RIP (without E)	-0.64	0.05	2.16	2.97

6

7

1

2 **Table 4(a)** The RMS analysis error of LETKF, LETKF with QOL and LETKF with RIP
3 with the 8 time-step analysis cycles and different observations. The ensemble size is three
4 in all experiments.
5

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

6

7

8 **Table 4(b)** Same as Table 4(a), except for the 25 time-step analysis cycles.

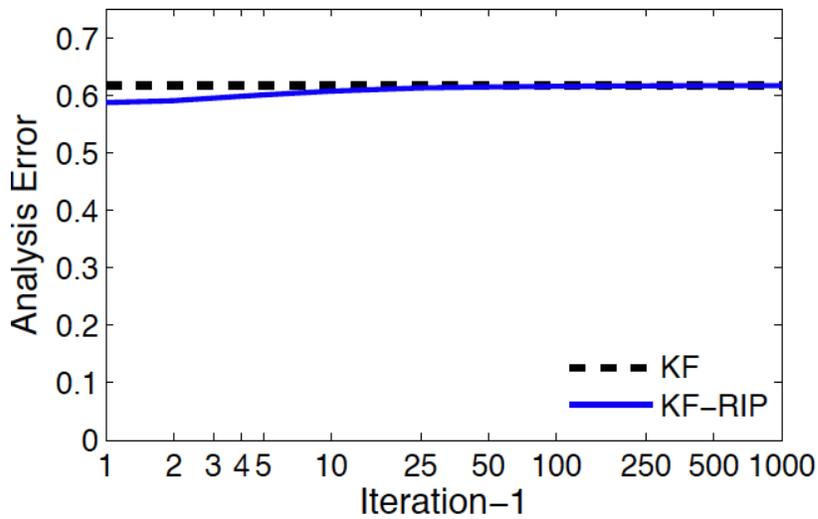
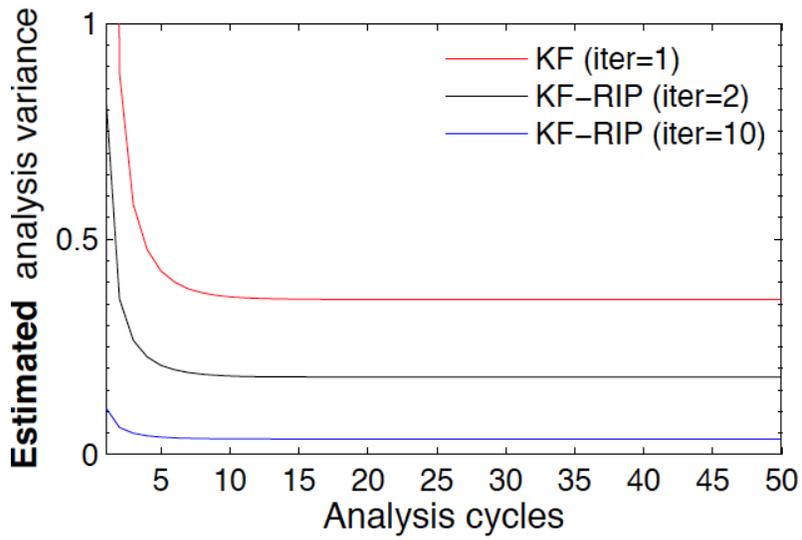
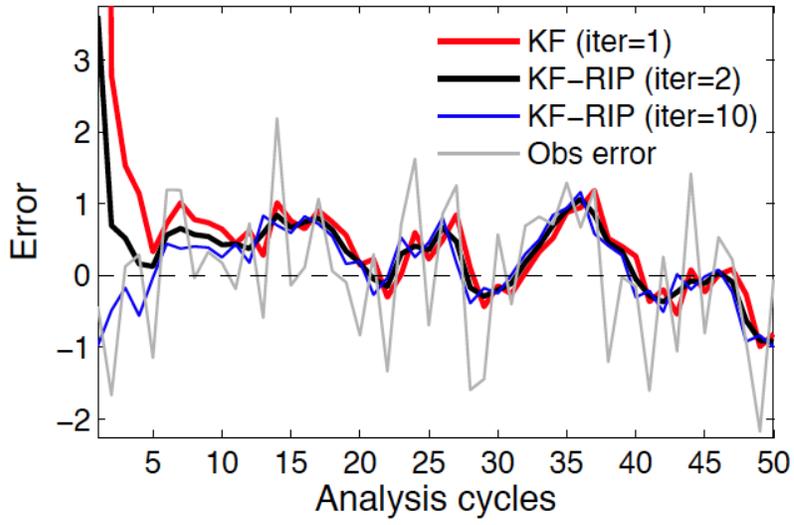
Observations	x	y	z	xy	xz	yz	xyz
Standard LETKF	2.9	1.67	7.16	1.01	1.53	0.78	0.68
QOL	1.98	1.23	5.94	0.82	1.16	0.60	0.47
RIP	1.57	0.97	3.81	0.56	0.66	0.40	0.35

9

10

11

12

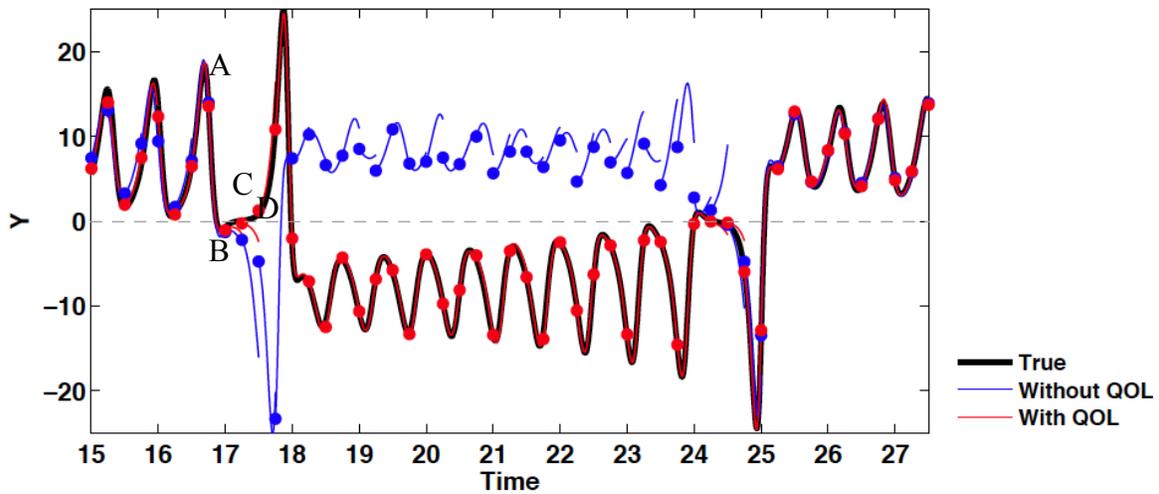


1

2 **Figure 1** (a) Analysis error of KF, KF-RIP with 2 and 10 iterations and observation error,
3 (b) estimated error variance of KF, KF-RIP with 2 and 10 iterations, and (c) averaged
4 RMS analysis error from KF and KF-RIP with different iteration numbers

5

1

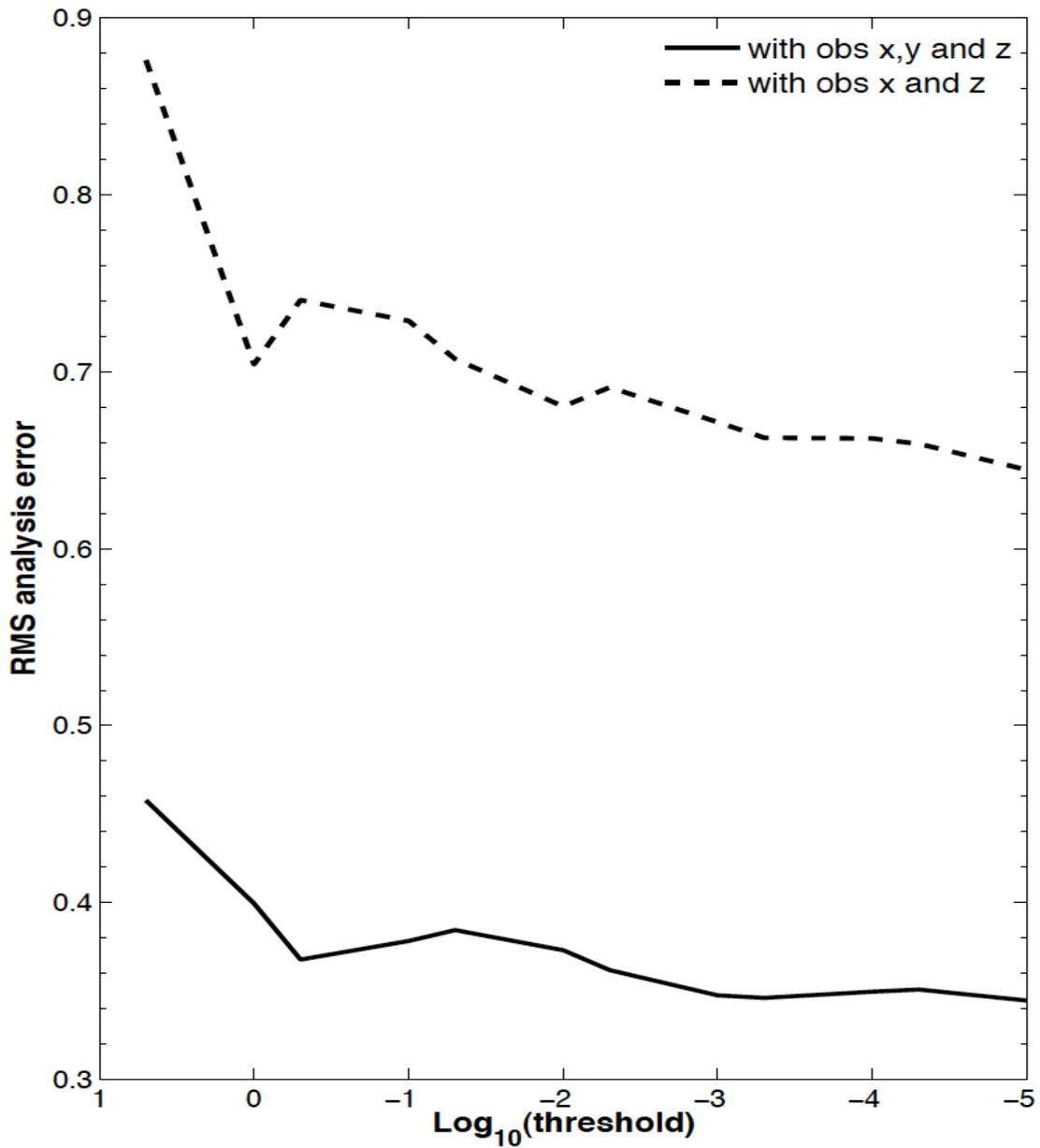


2

3 **Figure 2** Time evolution of the variable y during a typical regime transition with strong
4 nonlinear growth between $t=17$ and $t=18$. The black line denotes the true trajectory and
5 the blue line denotes the evolution initializing from the standard LETKF analysis mean,
6 marked as the blue dots. The red lines are the evolution from the analyses with the QOL
7 marked as red dots. The points A, B, C, D indicate a regime transition, where the filter
8 divergence occurs with the standard LETKF. Three ensemble members are used in these
9 experiments.

10

1



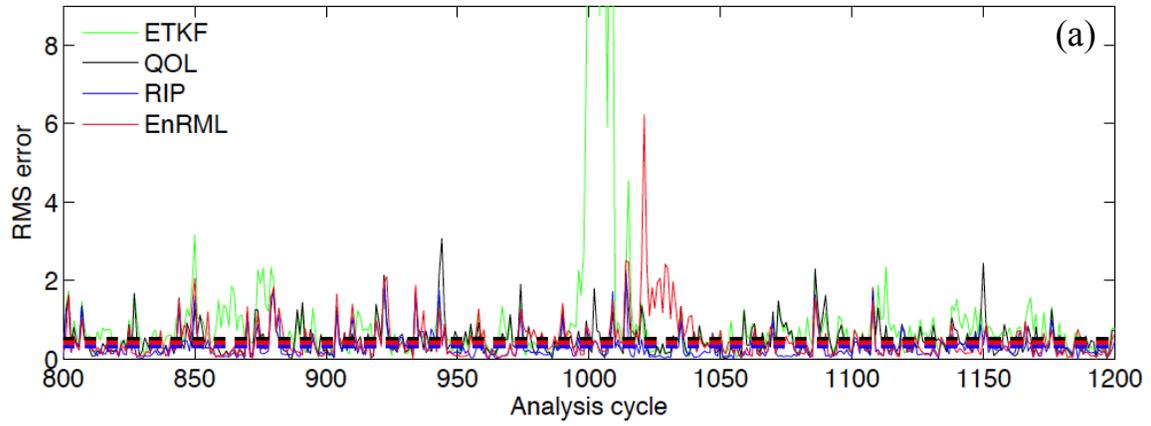
2

3 **Figure 3** RMS analysis error of RIP with different criteria as the threshold with three
4 observations and with two observations. Three ensemble members are used in these
5 experiments.

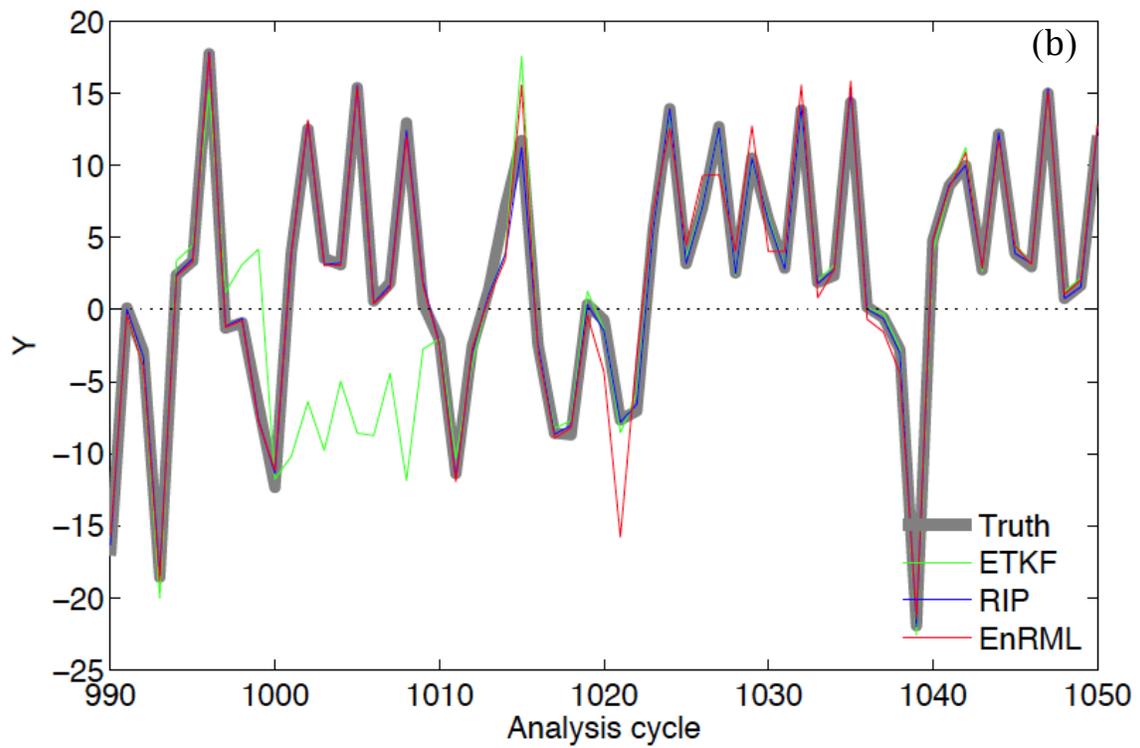
6

7

8



1
2



3
4

5 **Figure 4** (a) The RMS error from different analyses from the 800th to 1200th analysis
6 cycles and (b) analysis of the y variable from the truth and different analyses.