Adaptive ensemble Kalman filtering of nonlinear systems

Tyrus Berry and Timothy Sauer *

George Mason University, Fairfax, VA 22030

ABSTRACT

A necessary ingredient of an ensemble Kalman filter is covariance inflation, used to control filter divergence and compensate for model error. There is an ongoing search for inflation tunings that can be learned adaptively. Early in the development of Kalman filtering, (Mehra 1970, 1972) enabled adaptivity in the context of linear dynamics with white noise model errors by showing how to estimate the model error and observation covariances. We propose an adaptive scheme, based on lifting Mehra's idea to the nonlinear case, that recovers the model error and observation noise covariances in simple cases, and in more complicated cases, results in a natural additive inflation that improves state estimation. It can be incorporated into nonlinear filters such as the Extended Kalman Filter (EKF), the Ensemble Kalman filter (EnKF) and their localized versions. We test the adaptive EnKF on a 40-dimensional Lorenz96 model and show the significant improvements in state estimation that are possible. We also discuss the extent to which such an adaptive filter can compensate for model error, and demonstrate the use of localization to reduce ensemble sizes for large problems.

1. Introduction

The Kalman filter is provably optimal for systems where the dynamics and observations are linear with Gaussian noise. If the covariance matrices Q and R of the model error and observation error, respectively, are known, then the standard equations provided originally by Kalman (Kalman 1960) give the maximum likelihood estimate of the current state. If inexact Q and R are used in Kalman's equations, the filter may still give reasonable state estimates, but it will be suboptimal.

In the case of linear dynamics, (Mehra 1970, 1972) showed that the exact Q and R could be reconstructed by auxiliary equations added to the Kalman filter, in the case of Gaussian white noise model error. This advance opened the door for adaptive versions of the filter. More recently, approaches to nonlinear filtering such as the Extended Kalman Filter (EKF) and Ensemble Kalman Filter (EnKF) have been developed (Evensen 1994; Houtekamer and Mitchell 1998; Evensen 2009; Kalnay 2003; Simon 2006). Proposals have been made (Dee 1995; Daley 1992; Li et al. 2009; Anderson 2007) to extend the idea of adaptive filtering to nonlinear filters. However, none of these showed the capability of recovering an arbitrary Q and R simultaneously, even in the simplest nonlinear setting of Gaussian white noise model error and observation error. One goal of this article is to introduce an adaptive scheme in the Gaussian white noise setting that is a natural generalization of Mehra's approach, and that can be implemented on a nonlinear Kalman-type filter. We demonstrate the re-

construction of full, randomly-generated error covariance matrices Q and R in Example 4.1.

The technique we describe builds on the innovation correlation method of Mehra. Significant modifications are required to lift this technique to the nonlinear domain. In particular, we will explicitly address the various stationarity assumptions required by Mehra which are violated for a nonlinear model. As shown by Mehra and later (Daley 1992) and (Dee 1995), the innovation correlations are a blended mixture of observation noise, predictability error (due to state uncertainty), and model error. Disentangling them is nontrivial due to the nonstationarity of the local linearizations characteristic of strongly nonlinear systems. Many previous techniques have been based on averages of the innovation sequence, but our technique will not directly average the innovations.

Instead, at each filter step, the innovations are used, along with locally linearized dynamics and observations, to recover independent estimates of the matrices Q and R. These estimates are then integrated sequentially using a moving average to update the matrices Q_k^{filt} and R_k^{filt} used by the filter at time step k. By treating each innovation separately, using the local linearizations relevant to the current state, we are able to recover the full matrices Q and R when the model error is Gaussian white noise, and to improve state estimates in more complicated settings.

Gaussian white noise model error is an unrealistic assumption for practical applications. Typically, more significant model errors, as well as partial observability, must be

addressed. In this case, we interpret the covariance matrix Q_k^{filt} (see equation 2) as an additive covariance inflation term whose role is to prevent filter divergence and to improve the estimate of the underlying state.

Covariance inflation, originally suggested in (Anderson and Anderson 1999), adjusts the forecast error covariance to compensate for systematic underestimation of the covariance estimates given by the Kalman filter. There is a rapidly growing list of approaches to covariance inflation strategies (Desroziers et al. 2005; Li et al. 2009; Anderson 2008; Constatinescu et al. 2007a,b; Luo and Hoteit 2011, 2012). In (Wang and Bishop 2003; Li et al. 2009; Hamill et al. 2001) techniques are presented for finding inflation factors that implicitly assume a very simple structure for the noise covariance. There are also many techniques based on the covariance of the implied observation errors at each time step (Mitchell and Houtekamer 2000; Anderson 2001; Evensen 2003; Desroziers et al. 2005; Jiang 2007). These were actually anticipated and rejected by Mehra, who called them covariance matching techniques in (Mehra 1972). He showed that these techniques were not likely to converge unless the system noise was already known. Since untangling the observation noise and system noise is the key difficulty of adaptive filtering, we do not wish to assume prior knowledge of either.

At the same time, new theory is emerging that begins to give mathematical support to particular applications, including the recent article (Gonzalez-Tokman and Hunt 2013) that shows convergence of the EnKF in the sense of the existence of a shadowing trajectory under reasonable hyperbolicity assumptions. In their approach, there is no system noise, and perfect knowledge of the (deterministic) model and the observation noise covariance matrix are necessary to make the proof work. Furthermore, optimality is not addressed, which naturally leads to the question of the optimal choice of the covariance matrices used by the filter. We speculate that an appropriate adaptive filter may be the key to finding the optimal covariance matrices.

We begin the next section by demonstrating the importance of knowing the correct Q and R for the performance of a Kalman filter. Next we recall the adaptive filter developed in (Mehra 1970, 1972) that augments the Kalman filter to estimate Q and R in the case of linear dynamics. In Section 3 we describe an adaptive filter that can find Q and R in real time even for nonlinear dynamics and observations, building on the ideas of Mehra. In Section 4 we test the adaptive filter on a 40-dimensional model of (Lorenz 1996) and show the dramatic improvements in filtering that are possible. In addition, we show that this adaptive filter can also compensate for significant model error in the Lorenz96 model. We also propose new ideas to extend our technique to the cases of rank deficient observations and non-additive noise, and discuss an implementation that augments the localized LETKF version of the ensemble Kalman filter.

2. Extensions of the Kalman filter

Kalman filtering (Kalman 1960) is a well-established part of the engineering canon for state and uncertainty quantification. In the linear case with Gaussian noise, Kalman's algorithm is optimal. Since our main interest is the case of nonlinear dynamics, we will use notation that simplifies exposition of two often-cited extensions, the Extended Kalman Filter (EKF) and the Ensemble Kalman Filter (EnKF). For simplicity we will work in the discrete setting, and assume a nonlinear model of the form

$$
\begin{aligned}
x_{k+1} &= f(x_k) + \omega_{k+1} \\
y_{k+1} &= h(x_{k+1}) + \nu_{k+1}
\end{aligned}
\tag{1}
$$

where ω_{k+1} and ν_{k+1} are zero-mean Gaussian noise with covariance matrices Q and R, respectively. The system is given by f, and Q is the covariance of the one-step dynamical noise. The value of an observation is related to x by the function h, with observational noise covariance R. To simplify notation, we assume the covariance matrices are fixed in time, although in later examples we allow Q and R to drift and show that our adaptive scheme will track slow changes.

We are given a multivariate time series of observations, with the goal of estimating the state x as a function of time. The Kalman filter follows an estimated state x_k^a and an estimated state covariance matrix P_k^a. Given the estimates from the previous step, the Kalman update first creates forecast estimates x_{k+1}^f and P_{k+1}^f using the model, and then updates the forecast estimate using the observation y_{k+1}. The goal of nonlinear Kalman filtering is to correctly interpret and implement the linear Kalman equations

$$
\begin{aligned}
x_{k+1}^f &= F_k x_k^a \\
y_{k+1}^f &= H_{k+1} x_{k+1}^f \\
P_{k+1}^f &= F_k P_k^a F_k^T + Q_k^{\text{filt}} \\
P_{k+1}^y &= H_{k+1} P_{k+1}^f H_{k+1}^T + R_k^{\text{filt}} \\
K_{k+1} &= P_{k+1}^f H_{k+1}^T (P_{k+1}^y)^{-1} \\
P_{k+1}^a &= (I - K_{k+1} H_{k+1}) P_{k+1}^f \\
x_{k+1}^a &= x_{k+1}^f + K_{k+1} \left(y_{k+1} - y_{k+1}^f \right).
\end{aligned}
\tag{2}
$$

where P_{k+1}^y is the covariance of the observation, and y_{k+1}^f is the forecast observation implied by the previous state estimate. The matrices Q_k^{filt} and R_k^{filt} should be chosen to equal Q and R, respectively, in the linear case; how to choose them in general is the central issue of our investigation. The EKF extends the Kalman filter to systems of the form (1) by explicitly computing linearizations F_k and H_{k+1} from the dynamics. In the EnKF, these quan-
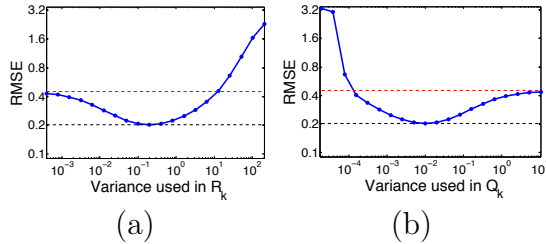
FIG. 1. Results of EnKF on a Lorenz96 data set (see Section 4) with 400 different combinations of diagonal Q_k^{filt} and R_k^{filt} matrices. RMS error was computed by comparing the filter output to the time series without observation noise. The correct Q and R (used to generate the simulated data) were diagonal matrices with entries $Q_{ii} = 0.01$ and $R_{ii} = 0.2$, respectively. The RMS error of the signal prior to filtering was 0.44 (shown as red dotted line) the RMSE of the optimal filter using $Q_k^{\text{filt}} = Q$ and $R_k^{\text{filt}} = R$ was 0.20 (shown as black dotted line). In (a) we show the effect of varying R_k^{filt} when $Q_k^{\text{filt}} = Q$ and in (b) the effect of varying Q_k^{filt} when $R_k^{\text{filt}} = R$.

tities are computed more indirectly, using ensembles; we give more complete details in Section 4.

Filter performance depends strongly on the covariances Q_k^{filt} and R_k^{filt} that are used in the algorithm. In fact, for linear problems the algorithm is provably optimal only when the true covariances $Q_k^{\text{filt}} = Q$ and $R_k^{\text{filt}} = R$ are used. For nonlinear problems, even for Gaussian white noise model error, using the exact $Q_k^{\text{filt}} = Q$ can lead to filter instability (see Figure 5(c)-(d)). This is because the local linearizations F_k and H_{k+1} introduce an additional model error which may systematically underestimate P_k^f. More generally, the model error in applications is typically not Gaussian white noise, and Q_k^{filt} must be interpreted as an additive inflation which attempts to compensate for the covariance structure of the model error.

In Fig. 1 we illustrate the effect of various suboptimal Q_k^{filt} and R_k^{filt} on a nonlinear problem by running an EnKF on a 40-site Lorenz96 model (Lorenz 1996; Lorenz and Emanuel 1998). (Full details of the model are given in Section 4.) In the limit of large Q_k^{filt} or small R_k^{filt} we find that the RMSE of filtered state estimates approach the RMSE of the unfiltered signal. Intuitively, when Q_k^{filt} is large and R_k^{filt} is small the filter has no confidence in the forecast relative to the observation and the filter simply returns the observation. For a filter to be nontrivial we must use a smaller Q_k^{filt} and a larger R_k^{filt}. However, as shown in Fig. 1, when the former is too small or the latter is too large, the RMSE of the filtered signals can actually be higher than the RMSE of the original noisy signals. In one extreme the filter becomes trivial, and in the other extreme it is possible for the filter to actually degrade the signal. Fig. 1 illustrates that a key to filter performance lies in the choice of Q_k^{filt} and R_k^{filt}.

In Section 3 we present a novel adaptive scheme that augments Eqn. (2) to also update the matrices Q_k^{filt} and R_k^{filt} at each filter step based on the observations. Since the adaptive scheme is a natural generalization of the Kalman update, it can be used in many of the extensions of the Kalman filter for nonlinear problems.

3. An adaptive nonlinear Kalman Filter

For the case of linear dynamics with linear full-rank observation, the adaptive filtering problem was solved in two seminal papers (Mehra 1970, 1972). Mehra considered the innovations of the Kalman filter, which are defined as $\epsilon_k = y_k - y_k^f$ and represent the difference between the observation and the forecast. In his innovation correlation method, he showed that the cross-correlations of the innovations could be used to estimate the true matrices Q and R. Intuitively, this is possible because cross-correlations of innovations will be influenced by the system and observation noise in different ways, which give multiple equations involving Q and R. These differences arise because the perturbations of the state caused by the system noise persist and continue to influence the evolution of the system, whereas each observation contains an independent realization of the observation noise. When enough observations are collected, the resulting system can be solved for the true covariance matrices Q and R.

a. Adaptive Kalman filter for linear dynamics.

In the case of linear dynamics with linear full-rank observations, the forecast covariance matrix P^f and the Kalman gain K have a constant steady state solution. Mehra shows that for an optimal filter we have the following relationship between the expectations of the cross-correlations of the innovations and the matrices involved in the Kalman filter:

$$\begin{aligned} \Gamma_0 &\equiv \mathbb{E}[\epsilon_k \epsilon_k^T] = HP^f H^T + R \\ \Gamma_1 &\equiv \mathbb{E}[\epsilon_{k+1} \epsilon_k^T] = HFP^f H^T - HFK\Gamma_0. \end{aligned}$$

Now consider the case when the Kalman filter is not optimal, meaning that the filter uses matrices Q_k^{filt} and R_k^{filt} that are not equal to the true values Q and R. In this case we must consider the issue of filter convergence. We say that the filter converges if the limit of the state estimate covariance matrices exists and is given by

$$M \equiv \lim_{k \to \infty} P_k^f = \lim_{k \to \infty} \mathbb{E}[(x_k - x_k^f)(x_k - x_k^f)^T]$$

where P_k^f is the estimate of the uncertainty in the current state produced by the Kalman filter at time step k. The filter is called non-optimal because the state estimate distribution will have higher variance than the optimal filter, as measured by the trace of M. As shown in (Mehra 1970, 1972), for the non-optimal filter, M is still the covariance

matrix of the state estimate as long as the above limit exists. Moreover, M satisfies an algebraic Riccati equation given by

$$M = F\left[(I - K_\infty H)M(I - K_\infty H)^T + K_\infty R K_\infty^T\right]F^T + Q \tag{3}$$

where K_∞ is the limiting gain of the non-optimal filter defined by Q^{filt} and R^{filt}. Note that the matrices Q and R in (3) are the true, unknown covariance matrices.

Motivated by the appearance of the true covariance matrices in the above equations, Mehra gave the following procedure for finding these matrices and hence the optimal filter. After running a non-optimal filter long enough that the expectations in Γ_0 and Γ_1 have converged, the true Q and R can be estimated by solving the equations

$$\begin{aligned}
M &= (HF)^{-1}(\Gamma_1 + HFK_\infty\Gamma_0)H^{-T} \\
R &= \Gamma_0 - HMH^T \\
Q &= M - F\left[(I - K_\infty H)M(I - K_\infty H)^T + K_\infty R K_\infty^T\right]F^T.
\end{aligned} \tag{4}$$

Clearly this method requires H to be invertible, and when the observation had low rank Mehra could not recover Q with his method. With a more complicated procedure he was still able to find the optimal Kalman gain matrix K even when he could not recover Q. However, this procedure used the fact that the optimal gain is constant for a linear model.

b. Extension to nonlinear dynamics.

Our goal is to apply this fundamental idea to nonlinear problems. Unfortunately, while the technique of Mehra has the correct basic idea of examining the time correlations of the innovation sequence, there are many assumptions that fail in the nonlinear case. First, the innovation sequence is no longer stationary, and thus the expectations for Γ_0 and Γ_1 are no longer well-defined. Second, the matrices H and F (interpreted as local linearizations) are no longer fixed and the limiting values of K and P^f no longer exist, and therefore all of these matrices must be estimated at each filter step. Third, Mehra was able to avoid explicitly finding Q in the case of a rank deficient observation by using the limiting K, which does not exist in the nonlinear case. For nonlinear dynamics the optimal Kalman gain is not fixed, making it more natural to estimate Q directly. When the observations are rank-deficient, parameterization of the matrix Q will be necessary.

Consequently, the principal issue with lifting Mehra's technique to a nonlinear model is that the local linear dynamics are changing with each time step. Even in the case of Gaussian white noise model error, the only quantities which may be assumed fixed over time in the nonlinear problem are the covariance matrices Q and R. With this insight, we solve Mehra's equations (4) at each filter step and update Q_k^{filt} and R_k^{filt} at each step with a exponentially weighted moving average. Thus our iteration becomes

$$\begin{aligned}
P_k^e &= (H_{k+1}F_k)^{-1}(\epsilon_{k+1}\epsilon_k^T + H_{k+1}F_k K_k \epsilon_k \epsilon_k^T)H_k^{-T} \\
Q_k^e &= P_k^e - F_{k-1}P_{k-1}^a F_{k-1}^T \\
R_k^e &= \epsilon_k \epsilon_k^T - H_k P_k^f H_k^T \\
Q_{k+1}^{\text{filt}} &= Q_k^{\text{filt}} + \delta(Q_k^e - Q_k^{\text{filt}}) \\
R_{k+1}^{\text{filt}} &= R_k^{\text{filt}} + \delta(R_k^e - R_k^{\text{filt}})
\end{aligned} \tag{5}$$

where δ must be chosen small enough to smooth the moving average. Note that these equations naturally augment the Kalman update in that they use the observation to update the noise covariances. This iteration is straightforward to apply with the Extended Kalman Filter, because H_k and F_k are explicitly known. For the Ensemble Kalman Filter, these quantities must be estimated from the ensembles. This extra step is explained in detail in Section 4.

To explain the motivation behind our method, first consider the following idealized scenario. We assume that the unknown state vector x_k is evolved forward by a known time varying linear transformation $x_{k+1} = F_k x_k + \omega_{k+1}$ and observed as $y_{k+1} = H_{k+1}x_{k+1} + \nu_{k+1}$. Furthermore, we assume that ω_{k+1} and ν_{k+1} are stationary white noise random variables which are independent of the state, time, and each other, and that $\mathbb{E}[\omega\omega^T] = Q$ and $\mathbb{E}[\nu\nu^T] = R$. In this scenario, we can express the innovation as

$$\epsilon_k = y_k - y_k^f = H_k(x_k - x_k^f) + \nu_k.$$

If we were able to average over all the possible realizations of the random variables x_k, x_k^f, and ν_k (conditioned on all previous observations) at the fixed time k we would have

$$\begin{aligned}
\mathbb{E}[\epsilon_k \epsilon_k^T] &= \mathbb{E}[H_k(x_k - x_k^f)(x_k - x_k^f)^T H_k^T] + \mathbb{E}[\nu_k \nu_k^T] \\
&= H_k P_k^f H_k^T + R.
\end{aligned}$$

However, the most important realization is that only the last expectation in this equation can be replaced by a time average. This is because unlike ν_k, which are independent, identically distributed random variables, the distribution P_k^f is changing with time and thus each innovation is drawn from a different distribution. In our idealized scenario, the non-stationarity of ϵ_k arises from the fact that the dynamics are time varying, and more generally for nonlinear problems this same effect will occur due to inhomogeneity of the local linearizations.

Since we have no way to compute the expectation $\mathbb{E}[\epsilon_k \epsilon_k^T]$, we instead compute the matrix

$$\begin{aligned}
R_k^e &= \epsilon_k \epsilon_k^T - H_k P_k^f H_k^T \\
&= \nu_k \nu_k^T + H_k(x_k - x_k^f)(x_k - x_k^f)^T H_k^T - H_k P_k^f H_k^T \\
&+ \nu_k(x_k - x_k^f)^T H_k^T + H_k(x_k - x_k^f)\nu_k^T.
\end{aligned}$$

Note that the last two terms have expected value of zero (since $\mathbb{E}[\nu_k] = 0$ for all k), and for each fixed k the difference

$$H_k(x_k - x_k^f)(x_k - x_k^f)^T H_k^T - H_k P_k^f H_k^T$$

4

has expected value given by the zero matrix. Since these terms have expected value of zero for each k, we can now replace the average over realizations of ν_k with an average over k since ν_k is assumed to be stationary. Thus we have

$$\lim_{K \to \infty} \frac{1}{K} \sum_{k=1}^{K} R_k^e = \mathbb{E}_k[R_k^e] = \mathbb{E}_k[\nu_k \nu_k^T] = \mathbb{E}_{\nu_k}[\nu_k \nu_k^T] = R$$

where $\mathbb{E}_k$ denotes an average over time and $\mathbb{E}_{\nu_k}$ denotes an average over possible realizations of the random variable ν_k. This motivates our definition of R_k^e as the empirical estimate of the matrix R based on a single step of the filter. Our method is to first recover the stationary component and then average over time instead of averaging the innovations. Thus the equation for R_k is simply a moving average of the estimates R_k^e. Of course, in real applications, the perturbation ν_k will not usually be stationary. However, our method is still advantageous since the matrix $\epsilon_k \epsilon_k^T$ is largely influenced by $H_k P_k^f H_k^T$ and thus by subtracting these matrices we expect to improve the stationarity of the sequence R_k^e.

A similar argument motivates our choice of P_k^e and Q_k^e as the empirical estimates of the forecast and model error covariances respectively. First, we continue the expansion of the k-th innovation as

$$
\begin{aligned}
\epsilon_{k+1} &= y_{k+1} - y_{k+1}^f = H_{k+1}(x_{k+1} - x_{k+1}^f) + \nu_{k+1} \\
&= H_{k+1}(F_k x_k + \omega_{k+1} - F_k x_k^a) + \nu_{k+1} \\
&= H_{k+1}(F_k x_k - F_k(x_k^f + K_k \epsilon_k)) + H_{k+1}\omega_{k+1} + \nu_{k+1} \\
&= H_{k+1}F_k(x_k - x_k^f) \\
&\quad - H_{k+1}F_k K_k \epsilon_k + H_{k+1}\omega_{k+1} + \nu_{k+1}.
\end{aligned}
$$

By eliminating terms which have mean zero we find the following expression for the cross covariance

$$\epsilon_{k+1}\epsilon_k^T = H_{k+1}F_k(x_k - x_k^f)(x_k - x_k^f)^T H_k^T - H_{k+1}F_k K_k \epsilon_k \epsilon_k^T.$$

Note that the expected value of $(x_k - x_k^f)(x_k - x_k^f)^T$ is the forecast covariance matrix P_k^f given by the filter. Solving the above equation for $(x_k - x_k^f)(x_k - x_k^f)^T$ gives an empirical estimate of the forecast covariance from the innovations. This motivates our definition of P_k^e, the empirical forecast covariance, which we find by solving

$$H_{k+1}F_k P_k^e H_k^T = \epsilon_{k+1}\epsilon_k^T + H_{k+1}F_k K_k \epsilon_k \epsilon_k^T.$$

It is tempting to use the empirical forecast to adjust the filter forecast P_k^f, however this is infeasible. The empirical forecast is extremely sensitive to the realization of the noise and since the forecast covariance is not stationary for nonlinear problems there is no way to average these empirical estimates. Instead, we isolate a stationary model error covariance which can be averaged over time by separating the predictability error from the forecast error.

In our idealized scenario, the forecast error $P_k^f = P_k^p + Q$ is the sum of the predictability error P_k^p and the model error Q which we wish to estimate. We can estimate the predictability error using the dynamics and the analysis covariance from the previous step of the Kalman filter as $P_k^p \approx F_{k-1}P_{k-1}^a F_{k-1}^T$. Finally, we estimate the model error covariance by

$$Q_k^e = P_k^e - F_{k-1}P_{k-1}^a F_{k-1}^T.$$

As with R_k^e, in our example of Gaussian white noise model error, this estimate of Q_k^e is stationary and thus can be averaged over time in order to estimate the true model error covariance Q. While in real applications the actual form of the forecast error is more complicated, a significant component of the forecast error will often be given by the predictability error estimated by the filter. Thus it is natural to remove this known non-stationary term before attempting to estimate the stationary component of the model error.

If the true values of Q and R are known to be constant in time, then a cumulative average can be used, however this would take much longer to converge. The exponentially weighted moving average allows the estimates of Q_k^{filt} and R_k^{filt} to adjust to slow or sporadic changes in the underlying noise covariances. A large delta will allow fast changes in the estimates but results in Q_k^{filt} and R_k varying significantly. We define $\tau = 1/\delta$ to be the stationarity time scale of our adaptive filter. Any changes in the underlying values of Q and R are assumed to take place on a time scale sufficiently greater than τ. In the next section, we will show (see Figure 2) that our iteration can find complicated covariance structures while achieving reductions in RMS error.

c. Compensating for rank-deficient observations.

For many problems, H_k or $H_{k+1}F_k$ will not be invertible because the number of observations per step m is less than the number of elements n in the state. In this case, the above algorithm cannot hope to reconstruct the entire matrix Q. However, we can still estimate a simplified covariance structure by parameterizing $Q_k^e = \sum q_p Q_p$ as a linear combination of fixed matrices Q_p. This parameterization was first suggested by Bélanger in (Bélanger 1974). To impose this restriction we first set

$$C_k = \epsilon_{k+1}\epsilon_k^T + H_{k+1}F_k K_k \epsilon_k \epsilon_k^T - H_{k+1}F_k F_{k-1}P_{k-1}^a F_{k-1}^T H_k^T$$

and note that we need to solve $H_{k+1}F_k Q_k^e H_k^T = C_k$. Thus we simply need to find the vector q with values q_p that minimize the Frobenius norm

$$\left\| C_k - \sum_p q_p H_{k+1}F_k Q_p H_k^T \right\|_{\text{F}}.$$

To solve this we simply vectorize all the matrices involved. Let $\text{vec}(C_k)$ denote the vector made by concatenating the columns of C_k. We are looking for the least squares solution of

$$A_k q = \sum_p q_p \text{vec}(H_{k+1} F_k Q_p H_k^T) \approx \text{vec}(C_k)$$

where the pth column of A_k is $\text{vec}(H_{k+1} F_k Q_p H_k^T)$. We can then find the least squares solution $q = A_k^\dagger \text{vec}(C_k)$ and form the estimated matrix Q_k^e.

In the applications section we will consider two particular parameterizations of Q_k^e. The first is simply a diagonal parameterization using n matrices $(Q_p) = E_{pp}$, where E_{ij} is the elementary matrix whose only nonzero entry is 1 in the ij position. The second parameterization is a block constant structure which will allow us to solve for Q_k^e in the case of a sparse observation. For the block constant structure we choose a number b of blocks which divides n and then we form b^2 matrices $\{Q_{(p,r)}\}_{p,r=1}^b$ where $(Q_{(p,r)}) = \sum_{l,m=1}^{n/b} E_{pn/b+l,rn/b+m}$. Thus each matrix $Q_{(p,r)}$ consists of a $n/b \times n/b$ submatrix which is all ones, and they sum to a matrix whose entries are all ones.

Note that it is very important to choose matrices Q_p which complement the observation. For example, if the observations are sparse then the matrix H_k will have rows which are all zero. The block constant parameterization naturally interpolates from nearby sites to fill the unobserved entries.

4. Application to the Lorenz model

In this section we demonstrate the adaptive EnKF by applying it to the Lorenz96 model (Lorenz 1996; Lorenz and Emanuel 1998), a nonlinear system which is known to have chaotic attractors and thus provides a suitably challenging testbed. We will show that our method recovers the correct covariance matrices for the observation and system noise. Next we will address the role of the stationarity parameter $\tau = 1/\delta$ and demonstrate how this parameter allows our adaptive EnKF to track changing covariances. Finally, we demonstrate the ability of the adaptive EnKF to compensate for model error by automatically tuning the system noise.

The Lorenz96 model is an N-dimensional ODE given by

$$\frac{dx^i}{dt} = -x^{i-2}x^{i-1} + x^{i-1}x^{i+1} - x^i + F \qquad (6)$$

where $x = [x^1(t), \ldots, x^N(t)]$ is a vector in $\mathbb{R}^N$ and the superscript on x^i (considered modulo N) refers to the i-th vector coordinate. The system is chaotic for parameters such as $N = 40$ and $F = 8$, the parameters used in the forthcoming examples. To realize the Lorenz96 model as a system of the form (1), we consider $x = x_k$ to be the state in $\mathbb{R}^N$ at time step k. We define $f(x_k)$ to be the result of integrating the system (6) with initial condition x_k for a single time step of length $\Delta t = 0.05$. For simplicity we took the observation function h to be the identity function except in Example 3 below, where we examine a lower dimensional observation. When simulating the system (1) to generate data for our examples we generated the noise vector ω_{k+1} by multiplying a vector of N independent standard normal random numbers by the symmetric positive definite square root of the system noise covariance matrix Q. Similarly, we generated ν_{k+1} using the square root of the observation noise covariance matrix R.

In the following examples we will demonstrate how the proposed adaptive EnKF is able to improve state estimation for the Lorenz96 model. In all of these examples we will use the root mean squared error (RMSE) between the estimated state and the true state as a measure of filter performance. The RMSE will always be caluculated over all $N = 40$ observation sites of the Lorenz96 system. In order to show how the error evolves as the filter runs, we will often plot RMSE against filter steps by averaging over a window of 1000 filter steps centered at the indicated filter step. Since we are using our algorithm to estimate the matrices Q and R we will also be interested in how close our estimates Q_k^{filt} and R_k^{filt} are. Since we are interested in recovering the entire matrices Q and R we use RMSE to measure the error $Q - Q_k^{\text{filt}}$ and $R - R_k^{\text{filt}}$. The RMSE in this case refers to the square root of the average of the squared errors over all the entries in the matrices.

The EnKF uses an ensemble of state vectors to represent the current state information. In examples 4.1 through 4.4 below we will use the unscented version of the EnKF Julier et al. (2000, 2004). For simplicity, we do not propagate the mean ($\kappa = 0$ in Simon (2006)) or use any scaling. When implementing an EnKF, the ensemble can be integrated forward in time and observed using the nonlinear equations (1) without any explicit linearization. Our augmented Kalman update requires the local linearizations F_k and H_{k+1}. While these are assumed available in the extended Kalman filter, for an EnKF these linear transformations must be found indirectly from the ensembles. Let E_k^a be a matrix containing the ensemble perturbation vectors (the centered ensemble vectors) which represents the prior state estimate and let E_{k+1}^f be the forecast perturbation ensemble which results from applying the nonlinear dynamics to the prior ensemble. We define F_k to be the linear transformation that best approximates the transformation from the prior perturbation ensemble to the forecast perturbation ensemble. Similarly, to define H_{k+1} we use the inflated forecast perturbation ensemble E_{k+1}^x and the ensemble E_{k+1}^y which results from applying the nonlinear observation to forecast ensemble. To summarize, in the context of an ensemble Kalman filter, we define F_k and

H_{k+1} by

$$F_k = E_{k+1}^f (E_k^a)^\dagger$$
$$H_{k+1} = E_{k+1}^y (E_{k+1}^x)^\dagger \qquad (7)$$

where † indicates the pseudo-inverse. There are many methods of inflating the forecast perturbation ensemble from E_{k+1}^f. In the examples below we always use the additive inflation factor Q_k^{filt} by finding the covariance matrix P_{k+1}^f of the forecast ensemble, forming $P_{k+1}^x = P_{k+1}^f + Q_k^{\text{filt}}$, and then taking the positive definite square root of P_{k+1}^x to form the inflated perturbation ensemble E_{k+1}^x.

Example 4.1. Finding the noise covariances. In the first example, both the Q and R matrices were generated randomly (constrained to be symmetric and positive definite) and the discrete-time Lorenz96 model was simulated for twenty thousand time steps. We then initialized diagonal matrices Q_k^{filt} and R_k^{filt} as shown in Fig. 2, and applied the standard EnKF to the simulated data. The windowed RMS error is shown in the red curve of Figure 2(d). Note that the error initially decreases but then quickly stabilizes as the forecast covariance given by the EnKF converges to its limiting behavior.

Next we applied the adaptive EnKF on the same simulated data using Q_k^{filt} and R_k^{filt} as our initial guess for the covariance matrices. In Figure 2(a)-(c) we compare the true Q matrix to the initial diagonal Q_k^{filt} and the final Q_k^{filt} estimate produced by the adaptive EnKF. Here, the adaptive EnKF recovers the complex covariance structure of the system noise in a 40-dimensional system. Moreover, the resulting RMS error, shown by the blue curve in Figure 2(d), shows a considerable improvement. This example shows that for Gaussian white noise model and observation errors, our adaptive scheme can be used with an EnKF to recover a randomly generated covariance structure even for a strongly nonlinear model.

Fig. 2 shows that the improvement in state estimation builds gradually as the adaptive EnKF converges to the true values of the covariance matrices Q and R. The speed of this convergence is determined by the parameter τ, which also determines the accuracy of the final estimates of the covariance matrices. The role of the stationarity constant τ will be explored in the next example.

Example 4.2. Tracking changing noise levels. To demonstrate the role of $\tau = 1/\delta$ and to illustrate the automatic nature of the adaptive EnKF, we consider a Lorenz96 system where both Q and R are multiples of the identity matrix, with R constant and Q varying in time. The trace of Q (normalized by the state dimension $N = 40$) is shown as the black curve in Figure 3(a) where it is compared to the initial Q (shown in red) and the estimates of Q produced by the adaptive EnKF with $\tau = 500$ (shown in
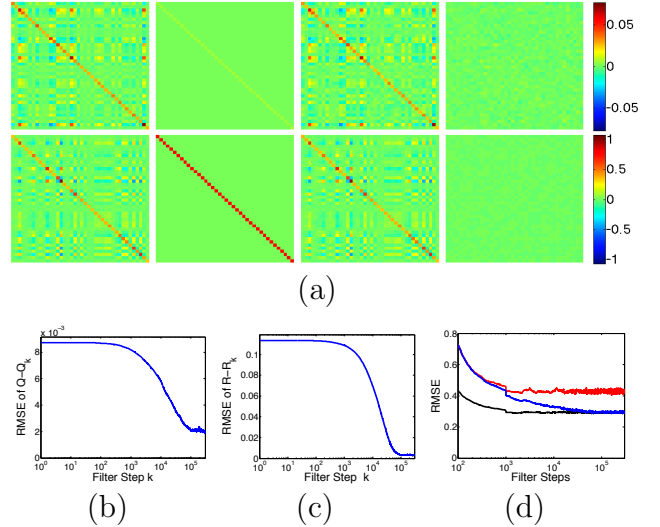

(a)

(b)   (c)   (d)

FIG. 2. We show the long term performance of the adaptive EnKF by simulating Lorenz96 for 300000 steps and running the adaptive EnKF with stationarity $\tau = 20000$. (a) First row, left to right: true Q matrix used in the Lorenz96 simulation, the initial guess for Q_k^{filt} provided to the adaptive filter, the final Q_k^{filt} estimated by the adaptive filter, and the final matrix difference $Q - Q_k^{\text{filt}}$. The second row shows the corresponding matrices for R; (b) RMSE of $Q - Q_k^{\text{filt}}$ as Q_k^{filt} is estimated by the filter; (c) RMSE of $R - R_k^{\text{filt}}$ as R_k^{filt} is estimated by the filter; (d) comparison of windowed RMSE vs. number of filter steps for the conventional EnKF run with the true Q and R (black, lower trace), and the conventional EnKF run with the initial guess matrices (red, upper trace), and our adaptive EnKF initialized with the guess matrices (blue, middle trace).

green) and $\tau = 2000$ (shown in blue). Note that when τ is smaller the adaptive filter can move more quickly to track the changes of the system, however, when the system stabilizes the larger value of τ is more stable and gives a better approximation of the stable value of Q.

Next we examine the effect of τ on the RMS error of the state estimates. Figure 3(b) shows that leaving the value of Q_k^{filt} equal to the initial value of Q leads to a large increase in RMSE for the state estimate, while the adaptive EnKF can track the changes in Q. We can naturally compare our adaptive EnKF to an "oracle" EnKF which is given the exact value of Q at each point in time; this is the best case scenario represented by the black curves in Figure 3. Again we see that a small τ results in a smaller peak deviation from the oracle, but the higher stationarity constant τ tracks the oracle better when the underlying Q is not changing quickly. Thus the parameter τ trades adaptivity (lower τ) for accuracy in the estimate (higher τ).

Example 4.3. Sparse observation. In this example we examine the effect of a low-dimensional observation on

7

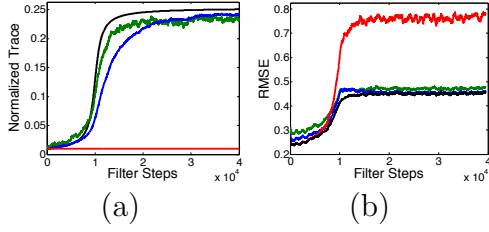(a)   (b)

FIG. 3. A Lorenz96 data set with slowly varying Q is produced by defining the system noise covariance matrix Q_k as a multiple of the identity matrix, with the multiple changing in time. (a) Trace of Q_k (black) and Q_k^{filt} for the EnKF (red) compared to the trace of Q_k^{filt} (normalized by $N = 40$) for the Adaptive EnKF at stationarity levels $\tau = 500$ (green) and $\tau = 2000$ (blue). (b) Comparison of the RMSE in state estimation for the EnKF with fixed Q_k^{filt} (red) and the Adaptive EnKF with stationarity $\tau = 500$ (green) and $\tau = 2000$ (blue). The black curve represents an oracle EnKF which is given the correct covariance matrix Q at each point in time.

the adaptive EnKF. As explained in Section 3, Mehra uses the stationarity of the Kalman filter for linear problems to build a special adaptive filter for problems with rank deficient observations. However, our current version of the adaptive EnKF cannot find the full Q matrix when the observation has lower dimensionality than the state vector (possible solutions to this open problem are considered in Section 5). In Section 3 we presented a special algorithm that parameterized the Q matrix as a linear combination of fixed matrices reducing the required dimension of the observation..

To demonstrate this form of the adaptive EnKF, we use a sparse observation. We first observe 20 sites equally spaced among the 40 total sites (below we will consider observing only 10 sites). In this example the observation y_k is 20-dimensional while the state vector x_k is 40-dimensional, giving a rank-deficient observation. We note that because the observation is sparse the linearization, H, of the observation will have rows which are all zeros. Solving for Q_k^e requires inverting H so we cannot use the diagonal parameterization of Q_k^{filt}, and instead we use the block constant parameterization which allows the Q_k^{filt} values to be interpolated from nearby sites. In order to check that the adaptive EnKF can still find the correct covariance matrices we simulated 100000 steps of the discretized Lorenz96 system observing only 20 evenly spaced sites. We take the true Q to have a block constant structure, and we take the true R to be a randomly generated symmetric positive definite matrix as shown in Figure 4. Since our algorithm will impose a block constant structure on Q_k^{filt}, we chose a block constant matrix for Q so that we could confirm that the correct entry values were recovered.

To test the block constant parameterization of the adaptive EnKF we chose initial matrices Q_k^{filt} and R_k^{filt} which

(a)

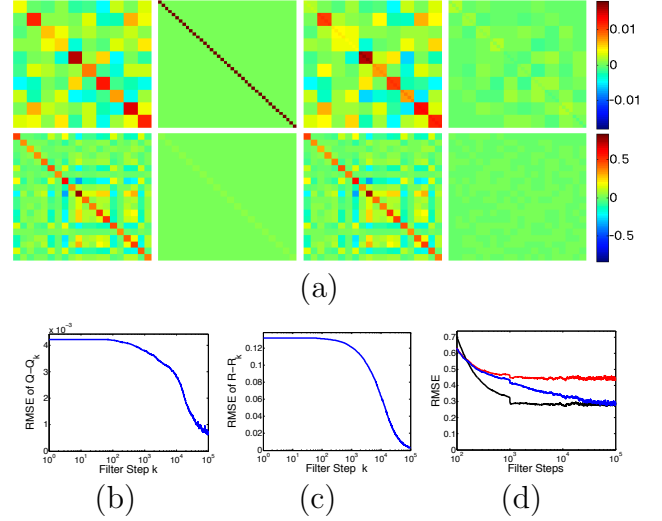(b)   (c)   (d)

FIG. 4. We apply the adaptive EnKF with a sparse observation by only observing every other site (20 total observed sites) of a Lorenz96 simulation with 100000 steps. The Q_k^{filt} matrix is assumed to be constant on 4x4 sub-matrices and the true Q used in the simulation is given the same block structure. (a) First row, left to right: true Q matrix used in the Lorenz96 simulation, the initial guess for Q_k^{filt} provided to the adaptive filter, the final Q_k^{filt} estimated by the adaptive filter, and the final matrix difference $Q - Q_k^{\text{filt}}$. The second row shows the corresponding matrices for R; (b) RMSE of $Q - Q_k^{\text{filt}}$ as Q_k^{filt} is estimated by the filter; (c) RMSE of $R - R_k^{\text{filt}}$ as R_k^{filt} is estimated by the filter; (d) comparison of windowed RMSE vs. number of filter steps for the conventional EnKF run with the true Q and R (black, lower trace), and the conventional EnKF run with the initial guess matrices (red, upper trace), and our adaptive EnKF initialized with the guess matrices (blue, middle trace).

were diagonal with entries 0.02 and 0.05 respectively. We used a large stationarity of $\tau = 15000$ which requires longer to converge but gives a better final approximation of Q and R, this is why the Lorenz96 system was run for 100000 steps. Such a large stationarity and long simulation was chosen to illustrate the long term behavior of the adaptive EnKF. The estimates of Q_k^{filt} were parameterized with a 10×10 block constant structure ($b = 10$ using the method of section 3). In Figure 4 we compare the true Q and R matrices to the initial guesses and the final estimates of our adaptive EnKF. This example shows that, even in the case of a rank deficient observation, the adaptive EnKF can recover an arbitrary observation noise covariance matrix and a parameterized system noise covariance matrix.

Observations in real applications can be very sparse, so we now consider the case when only 10 evenly spaced sites are observed. Such a low dimensional observation makes state estimation very difficult as shown by the increased RMSE in Figure 5 compared to Figure 4. Even more interesting is that we now observe filter divergence, where

the state estimate trajectory completely loses track of the true trajectory. Intuitively this is much more likely when fewer sites are observed, and the effect is shown in Figure 5(c) by sporadic periods of very high RMSE. Filter divergence occurs when the variance of Q_k^{filt} is small, implying overconfidence in the state estimate and the resulting forecast. In Figure 5 this is clearly shown since filter divergence occurs only when the true matrix Q is used in the filter, whereas both the initial guess and the final estimate produced by the adaptive EnKF are *in*flated. The fact that the filter diverges when provided with the true Q and R matrices illustrates how this example pushes the boundaries of the EnKF. However, our adaptive EnKF produces an inflated version of the true Q matrix as shown in Figure 5 which not only avoids filter divergence, but also significantly reduces RMSE relative to the initial diagonal guess matrices. This example illustrates that the breakdown of the assumptions of the EnKF (local linearizations, Gaussian distributions) can lead to assimilation error even when the nonlinear dynamics are known. In the presence of this error, our adaptive EnKF must be interpreted as an inflation scheme and we judge it by its performance in terms of RMSE rather than recovery of the underlying Q.

Example 4.4. Compensating for model error. The goal of this example is to show that the covariance of the system noise is a type of additive inflation and thus is a natural place to compensate for model error. Intuitively, increasing Q_k^{filt} increases the gain, effectively placing more confidence in the observation and less on our forecast estimate. Thus, when we are less confident in our model, we would naturally want to increase Q_k^{filt}. Following this line of thought, it seems natural that if Q_k^{filt} is sub-optimally small, the model errors would manifest themselves in poor state estimates and hence large innovations. In this example, the adaptive EnKF automatically inflates Q_k^{filt} based on the observed innovations, leading to significantly improved filter performance.

To illustrate this effect, we fixed the model used by the adaptive EnKF to be the Lorenz96 model from (6) with $N = 40$ and $F = 8$, and changed the system generating the data. The sample trajectory of the Lorenz96 system was created with 20000 time steps. For the first 10000 steps we set $F = 8$. We then chose $N = 40$ fixed random values of F^i, chosen from a distribution with mean 8 and standard deviation 4. These new values were used for the last 10000 steps of the simulation. Thus, when running our filter we would have the correct model for the first 10000 steps but a significantly incorrect model for the last 10000 steps.

We first ran a conventional EnKF on this data and found that the RMSE for the last 10000 steps was approximately 180% greater than an oracle EnKF which was given the new parameters F^i. Next we ran our adaptive EnKF
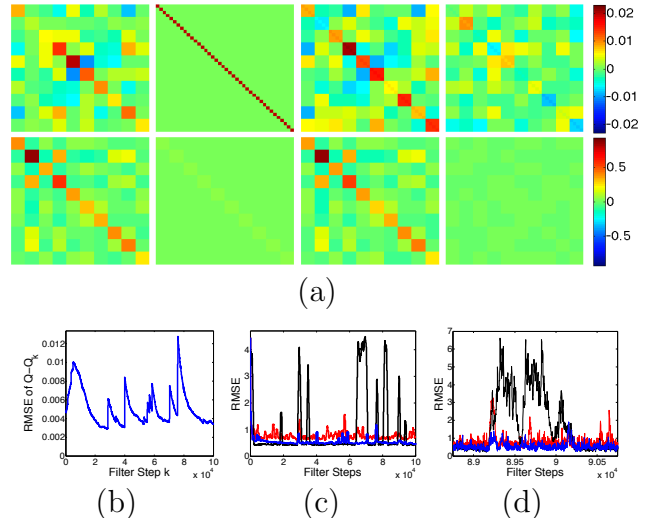

(a)

(b)   (c)   (d)

FIG. 5. We illustrate the effect of extremely sparse observations by only observing every forth site (10 total observed sites) of the Lorenz96 simulation, the Q_k^{filt} matrix is assumed to be constant on 4x4 sub-matrices and the true Q used in the simulation is given the same block structure. (a) First row, left to right: true Q matrix used in the Lorenz96 simulation, the initial guess for Q_k^{filt} provided to the adaptive filter, the final Q_k^{filt} estimated by the adaptive filter, and the final matrix difference $Q - Q_k^{\text{filt}}$. The second row shows the corresponding matrices for R; (b) RMSE of $Q - Q_k^{\text{filt}}$ as the adaptive EnKF is run; (c) comparison of windowed RMSE vs. number of filter steps for the conventional EnKF run with the true Q and R (black, lower trace), and the conventional EnKF run with the initial guess matrices (red, upper trace), and our adaptive EnKF initialized with the guess matrices (blue, middle trace); (d) Enlarged view showing filter divergence, taken from (c). Note that the conventional EnKF occasionally diverges even when provided the true Q and R matrices. The Q_k^{filt} found by the adaptive filter is automatically inflated relative to the true Q which improves filter stability as shown in (c) and (d).

and found that it automatically increased the variance of Q_k^{filt} in proportion to the amount of model error. In Figure 6 we show how the variance of Q_k^{filt} was increased by the adaptive filter; note that the increase in variance is highly correlated with the model error measured as the percentage change in the parameter F^i at the corresponding site. Intuitively, a larger error in the F^i parameter would lead to larger innovations thus inflating the corresponding variance in Q_k^{filt}. However, note that when Q_k^{filt} was restricted to be diagonal the adaptive filter required greater increase (shown in Figure 6(b)) and gave significantly worse state estimates (as shown in Figure 6(c)). Thus, when Q_k^{filt} was not restricted to be diagonal, the adaptive filter was able to find complex new covariance structure introduced by the model error. This shows that the adaptive filter is not simply turning up the noise arbitrarily but is precisely tuning

(a)

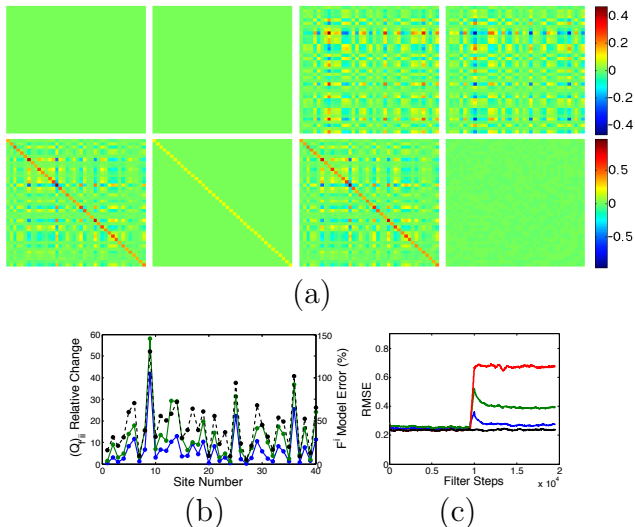(b)   (c)

FIG. 6. For the first 10000 filter steps the model is correct and then the underlying parameters are randomly perturbed at each site. The conventional EnKF is run with the initial true covariances $Q_k^{\text{filt}} = Q$ and $R_k^{\text{filt}} = R$; the adaptive EnKF starts with the same values but it automatically increases the system noise level (Q_k^{filt}) to compensate for the model error resulting in improved RMSE. (a) First row, left to right: true Q matrix used in the Lorenz96 simulation, the initial guess for Q_k^{filt} provided to the adaptive filter, the final Q_k^{filt} estimated by the adaptive filter, and the final matrix difference $Q - Q_k^{\text{filt}}$. The second row shows the corresponding matrices for R; (b) Model error (black, dotted curve) measured as the percent change in the parameter F^i at each site compared to the relative change in the corresponding diagonal entries of Q_k^{filt} found with the adaptive EnKF (blue, solid curve), diagonal (green, solid curve). (c) Results of the adaptive EnKF (blue) compared to conventional EnKF (red) on a Lorenz96 data set in the presence of model error. The green curve is an adaptive EnKF where Q_k^{filt} is forced to be diagonal and the black curve shows the RMSE of an oracle EnKF which is provided with the true underlying parameters F^i for both halves of the simulation.

the covariance of the stochastic process to compensate for the model error. This automatic tuning meant that the final RMSE of our adaptive filter increased less than 15% as a result of the model error.

Example 4.5. Adaptive version of the LETKF.

The goal of this example is to show that our algorithm can naturally integrate into a localization scheme. The unscented version of the EnKF used in the previous examples requires large ensemble sizes, which are impractical for many applications. Localization is a general technique, implemented in various forms, that assumes the state vector has a spatial structure, allowing the Kalman update to be applied locally rather than globally. Some localization schemes have been shown to allow reduced ensemble size.

The local ensemble transform Kalman filter (LETKF)

of (Ott et al. (2004); Hunt et al. (2004, 2007)) is a localization technique that is particularly simple to describe. The local update avoids inverting large covariance matrices, and also implies that we only need enough ensemble members to track the local dynamics for each local region. For simplicity we used the algorithm including the ensemble transform given in Eqn. 41 of (Ott et al. 2004). We performed the local Kalman update at each site using a local region with 11 sites (l=5) and we used a global ensemble with 22 members, compared to 80 ensemble members used in the unscented EnKF.

Building an adaptive version of the LETKF is fairly straightforward. A conventional Kalman update is used to form the local analysis state and covariance estimates. We design an adaptive LETKF by simply applying our iteration, given by Eqn. (5), after the local Kalman update performed in each local region. We will estimate 40 separate pairs of 11×11 matrices $Q_k^{i,\text{filt}}$ and $R_k^{i,\text{filt}}$. This implies that some entries in the full Q_k^{filt} (those far from the diagonal) have no representatives in these local $Q_k^{i,\text{filt}}$. Moreover, the other entries in the full Q_k^{filt} are represented by entries in several $Q_k^{i,\text{filt}}$ matrices. In Fig. 7 we combine the final estimates of $Q_k^{i,\text{filt}}$ into a single 40×40 Q_k^{filt} matrix by averaging all the representatives of each entry. It would also be possible to form this global Q_k^{filt} matrix at each step, however, the LETKF often allows many choices for forming the global results from the local, and our objective here is only to show that our adaptive iteration can be integrated into a localized data assimilation method.

Interestingly, the conventional LETKF requires significant inflation to prevent filter divergence. For example, the noise variances in Fig. 7 are comparable to Example 4.1, however the LETKF often diverges when the diagonal entries of Q_k^{filt} are less than 0.05. So for this example we use $Q_k^{\text{filt}} = Q + (0.1)I_{40}$ and $R_k^{\text{filt}} = R$ as the benchmark for all comparisons, since this represents a filter which was reasonably well-tuned when the true Q and R were both known. To represent that case when the true Q and R are unknown we choose diagonal matrices with the same average covariance as $Q + (0.1)I_{40}$ and R respectively. For the conventional LETKF, we form the local $Q_k^{i,\text{filt}}$ and $R_k^{i,\text{filt}}$ by simply taking the 11×11 sub-matrices of Q_k^{filt} and R_k^{filt} which are given by the local indices. In Fig. 7 we show that the adaptive LETKF significantly inflates both Q_k^{filt} and R_k^{filt}, and it recovers much of the structure of both Q and R, and improves the RMSE, compared to the conventional LETKF using the diagonal Q_k^{filt} and R_k^{filt}.

While this example shows that our iteration can be used to make an adaptive LETKF, we found that the adaptive version required a much higher stationarity ($\tau = 50000$ in Fig. 7). We speculate that this is because local regions can experience specific types of dynamics for extended periods which bias the short term estimates and by setting a
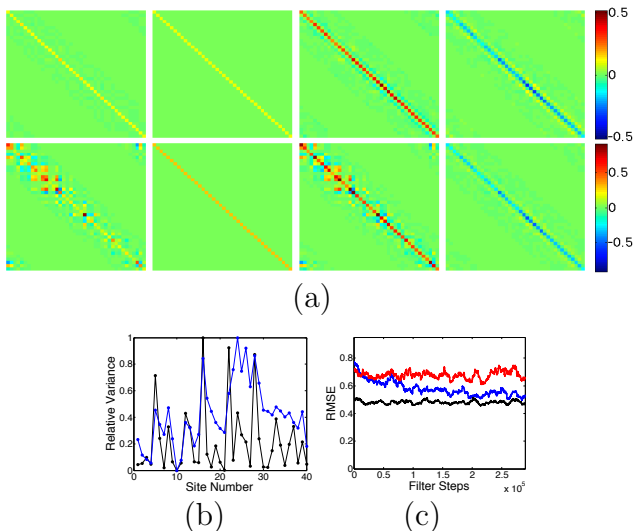
(a)

(b)   (c)

FIG. 7. We compare the conventional and adaptive LETKF on a simulation of 300000 steps of Lorenz96 (a) First row, left to right: benchmark $Q + .1I_{40}$ matrix where Q is was the matrix used in the Lorenz96 simulation, the initial guess for Q_k^{filt} provided to the adaptive filter, the final Q_k^{filt} estimated by the adaptive filter, and the final matrix difference $Q - Q_k^{\text{filt}}$. The second row shows the corresponding matrices for R (leftmost is the true R); (b) the variances from the diagonal entries of the true Q matrix (black, rescaled to range from zero to one) and the those from the final global estimate Q_k^{filt} produced by the adaptive LETKF (blue, rescaled to range from zero to one) (c) Results of the adaptive LETKF (blue) compared to conventional LETKF with the diagonal covariance matrices (red) and the conventional LETKF with the benchmark covariances $Q_k^{\text{filt}} = Q + .1I_{40}$ and $R_k^{\text{filt}} = R$ (black).

large stationarity we remove these biases by averaging over a larger portion of state space. We also found that the LETKF could have large deviations from the true state estimate similar to filter divergence but on shorter time scales. These deviations would cause large changes in the local estimates $Q_k^{i,\text{filt}}$ which required us to impose a limit of on the amount an entry of $Q_k^{i,\text{filt}}$ could change in a single time step of 0.05. It is possible that a better tuned LETKF or an example with lower noise covariances would not require these ad hoc adjustments. Since large inflations are necessary simply to prevent divergence, this adaptive version is a first step towards the important goal of optimizing the inflation.

5. Discussion and outlook

A central feature of the Kalman filter is that the innovations derived from observations are only used to update the mean state estimate, and the covariance update does not incorporate the innovation. This is a remnant of the Kalman filter's linear heritage, in which observation, fore-

cast and analysis are all assumed Gaussian. In any such scenario, when a Gaussian observation is assimilated into a Gaussian prior, the posterior is Gaussian and independent of observation. Equation (2) explicitly shows that the Kalman update of the covariance matrix is independent of the innovation.

Any scheme for adapting the Kalman filter to nonlinear dynamics must consider the possibility of non-Gaussian estimates, which in turn demands a reexamination of this independence. But the stochastic nature of the innovation implies that any attempt to use this information directly will be extremely sensitive to particular noise realizations. In this article, we have introduced an augmented Kalman update in which the innovation is allowed to affect the covariance estimate indirectly, through the filter's estimate Q^{filt} of the system noise. We envision this augmented version to be applicable to any nonlinear generalization of the Kalman filter, such as the EKF and EnKF. Application to the EKF is straightforward, and we have shown in equation (7) how to implement the augmented equations into an ensemble Kalman update.

The resulting adaptive EnKF is an augmentation of the conventional EnKF, to allow the system and observation noise covariance matrices to be automatically estimated. This removes an important practical constraint on filtering for nonlinear problems, since often the true noise covariance matrices are not known a priori. Using incorrect covariance matrices will lead to sub-optimal filter performance, as shown in Section 2, and in extreme cases can even lead to filter divergence. The adaptive EnKF uses the innovation sequence produced by the conventional equations, augmented by our additional equations developed in Section 3, to estimate the noise covariances sequentially. Thus, it is easy to adopt for existing implementations, and as shown in Section 4 it has a significant performance advantage over the conventional version. Moreover, we have shown in Section 4 that the adaptive EnKF can adjust to non-stationary noise covariances and even compensate for significant modeling errors.

The adaptive EnKF introduced here raises several practical and theoretical questions. A practical limitation of our current implementation is the requirement that the rank of the linearized observation H_k is at least the dimension of the state vector x_k. In Section 3 we provide a partial solution which constrains the Q estimation to have a reduced form which must be specified in advance. However, in the context of the theory of embedology, developed in (Sauer et al. 1991), it may be possible to modify our algorithm to recover the full Q matrix from a low-rank observation. Embedology shows that for a generic observation, it is possible to reconstruct the state space using a time-delay embedding. Motivated by this theoretical result, we propose a way to integrate a time-delay reconstruction into the context of a Kalman filter and discuss the remaining

challenges of such an approach.

In order to recover the full system noise covariance matrix in the case of rank deficient observation we propose an augmented observation formed by concatenating several iterations of the dynamics,

$$\hat{y}_k^f = \hat{h}(x_k^f) = (h(x_k^f), h \circ f(x_k^f), ..., h \circ f^M(x_k^f))$$

which will be compared to time-delayed observation vector $\hat{y}_k = (y_k, y_{k+1}, ..., y_{k+M})$. For a generic observation h, when the system is near an attractor of box counting dimension N, our augmented observation $\tilde{h}$ is generically invertible for $M > 2N$ (Sauer et al. 1991). We believe that this augmented observation will not only solve the rank deficient observation problem, but may also improve the stability of the Kalman filter by including more observed information in each Kalman update. However, a challenging consequence of this technique is that the augmented observation is influenced by both the dynamical noise and the observation noise realizations. Thus extracting the noise covariances from the resulting innovations may be non-trivial.

An important remaining theoretical question is to develop proofs that our estimated Q and R converge to the correct values. For R this is a straightforward claim, however the interpretation of Q for nonlinear dynamics is more complex. This is because the Kalman update equations assume that at each step the forecast error distribution is Gaussian. Even if the initial prior is assumed to be Gaussian, the nonlinear dynamics will generally change the distribution, making this assumption false. Until recently it was not even proved that the EnKF tracked the true trajectory of the system. It is likely that the choice of Q_k^{filt} for an EnKF will need to account for both the true system noise Q as well as the discrepancy between the true distribution and the Gaussian assumption. This will most likely require an additional increase in Q_k^{filt} above the system noise covariance Q.

We believe that the eventual solution of this difficult problem will involve computational approximation of the Lyapunov exponents of the model as the filter proceeds. Such techniques may require Q_k^{filt} to enter non-additively or even to vary over the state space. Improving filter performance for nonlinear problems may require reinterpretation of the classical meanings of Q and R in order to find the choice that leads to the most efficient shadowing. In combination with the results of (Gonzalez-Tokman and Hunt 2013), this would realize the long-sought nonlinear analogue of the optimal Kalman filter.

REFERENCES

Anderson, J., 2001: An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, **128**, 2884–2903.

Anderson, J., 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, **A59**, 210–224.

Anderson, J., 2008: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **A61**, 72–83.

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

Bélanger, P., 1974: Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*, **10**, 267–275.

Constatinescu, E., A. Sandu, T. Chai, and G. Carmichael, 2007a: Ensemble-based chemical data assimilation. i: General approach. *Quar. J. Roy. Met. Soc.*, **133**, 1229–1243.

Constatinescu, E., A. Sandu, T. Chai, and G. Carmichael, 2007b: Ensemble-based chemical data assimilation. ii: Covariance localization. *Quar. J. Roy. Met. Soc.*, **133**, 1245–1256.

Daley, R., 1992: Estimating model-error covariances for application to atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 1735–1748.

Dee, D., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.

Desroziers, G., L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quar. J. Roy. Met. Soc.*, **131**, 3385–3396.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Geophys. Res.*, **99**, 10 143–10 162.

Evensen, G., 2003: The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, **53**, 343–367.

Evensen, G., 2009: *Data Assimilation: The Ensemble Kalman Filter, 2nd ed.* Springer, Berlin.

Gonzalez-Tokman, D. and B. Hunt, 2013: Ensemble data assimilation for hyperbolic systems. *Physica*, **D243**, 128–142.

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

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

Hunt, B., E. Kostelich, and I. Szunyogh, 2007: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica*, **D230**, 112–126.

Hunt, B., et al., 2004: Four-dimensional ensemble Kalman filtering. *Tellus*, **A56**, 273–277.

Jiang, Z., 2007: A novel adaptive unscented Kalman filter for nonlinear estimation. *46th IEEE Conference on Decision and Control*, 4293–4298.

Julier, S., J. Uhlmann, and H. Durrant-Whyte, 2000: A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Trans. Automat. Control*, **45**, 477–482.

Julier, S., J. Uhlmann, and H. Durrant-Whyte, 2004: Unscented filtering and nonlinear estimation. *Proc. IEEE*, **92**, 401–422.

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

Kalnay, E., 2003: *Atmospheric modeling, data assimilation, and predictability.* Cambridge Univ. Press.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quar. J. Roy. Met. Soc.*, **135**, 523–533.

Lorenz, E., 1996: Predictability: A problem partly solved. *ECMWF Seminar on Predictability, Reading, UK*, **1**, 1–18.

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

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

Luo, X. and I. Hoteit, 2012: Ensemble Kalman filtering with residual nudging. *preprint.*

Mehra, R., 1970: On the identification of variances and adaptive kalman filtering. *IEEE Trans. Auto. Cont.*, **15**, 175–184.

Mehra, R., 1972: Approaches to adaptive filtering. *IEEE Trans. Auto. Cont.*, **17**, 693–698.

Mitchell, H. and P. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128**, 416–433.

Ott, E., et al., 2004: A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, **A56**, 415–428.

Sauer, T., J. Yorke, and M. Casdagli, 1991: Embedology. *J. Stat. Phys.*, **65**, 579–616.

Simon, D., 2006: *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches.* John Wiley and Sons.

Wang, X. and C. Bishop, 2003: A comparison of breeding and ensemble transform Kalman filter ensemble forecast schemes. *J. Atmos. Sci.*, **60**, 1140–1158.