THE AMERICAN STATISTICIAN
2016, VOL. 70, NO. 4, 350–357
http://dx.doi.org/10.1080/00031305.2016.1141709

Taylor & Francis
Taylor & Francis Group

Understanding the Ensemble Kalman Filter

Matthias Katzfuss[a], Jonathan R. Stroud[b], and Christopher K. Wikle[c]

[a]Department of Statistics, Texas A&M University, College Station, TX, USA; [b]Department of Statistics, George Washington University, Washington DC, USA; [c]Department of Statistics, University of Missouri, Columbia, MO, USA

ABSTRACT

The ensemble Kalman filter (EnKF) is a computational technique for approximate inference in state-space models. In typical applications, the state vectors are large spatial fields that are observed sequentially over time. The EnKF approximates the Kalman filter by representing the distribution of the state with an ensemble of draws from that distribution. The ensemble members are updated based on newly available data by shifting instead of reweighting, which allows the EnKF to avoid the degeneracy problems of reweighting-based algorithms. Taken together, the ensemble representation and shifting-based updates make the EnKF computationally feasible even for extremely high-dimensional state spaces. The EnKF is successfully used in data-assimilation applications with tens of millions of dimensions. While it implicitly assumes a linear Gaussian state-space model, it has also turned out to be remarkably robust to deviations from these assumptions in many applications. Despite its successes, the EnKF is largely unknown in the statistics community. We aim to change that with the present article, and to entice more statisticians to work on this topic.

ARTICLE HISTORY
Received June 2015
Revised November 2015

KEYWORDS
Bayesian inference; Forecasting; Kalman smoother; Sequential Monte Carlo; State-space models

1. Introduction

Data assimilation involves combining observations with "prior knowledge" (e.g., mathematical representations of mechanistic relationships; numerical models; model output) to obtain an estimate of the true state of a system and the associated uncertainty of that estimate (see, e.g., Nychka and Anderson 2010, for a review). Although data assimilation is required in many fields, its origins as an area of scientific inquiry arose out of the weather forecasting problem in geophysics. With the advent of the digital computer, one of the first significant applications was the integration of the partial differential equations that described the evolution of the atmosphere, for the purposes of short-to-medium range weather forecasting. Such a numerical model requires initial conditions from real-world observations that are physically plausible. Observations of the atmosphere have varying degrees of measurement uncertainty and are fairly sparse in space and time, yet numerical models require initial conditions that match the relatively dense spatial domain of the model. Data assimilation seeks to provide these "interpolated" fields while accounting for the uncertainty of the observations and using the numerical model itself to evolve the atmospheric state variables in a physically plausible manner. Thus, data assimilation considers an equation for measurement error in the observations and an equation for the state evolution, a so-called state-space model (see, e.g., Wikle and Berliner 2007).

In the geophysical problems that motivated the development of data assimilation, the state and observation dimensions are huge and the evolution operators associated with the numerical models are highly nonlinear. From a statistical perspective, obtaining estimates of the true system state and its uncertainty in this environment can be carried out, in principle, by a type of inference called filtering, which attempts to obtain sequentially the posterior distribution of the state at the current time point based on all observations collected so far. The combination of high dimensionality and nonlinearity makes this a very challenging problem.

The ensemble Kalman filter (EnKF) is an approximate filtering method introduced in the geophysics literature by Evensen (1994). In contrast to the standard Kalman filter (Kalman 1960), which works with the entire distribution of the state explicitly, the EnKF stores, propagates, and updates an ensemble of vectors that approximates the state distribution. This ensemble representation is a form of dimension reduction, in that only a small ensemble is propagated instead of the joint distribution including the full covariance matrix. When new observations become available, the ensemble is updated by a linear "shift" based on the assumption of a linear Gaussian state-space model. Hence, additional approximations are introduced when non-Gaussianity or nonlinearity is involved. However, the EnKF has been highly successful in many extremely high-dimensional, nonlinear, and non-Gaussian data-assimilation applications. It is an embodiment of the principle that an approximate solution to the right problem is worth more than a precise solution to the wrong problem (Tukey 1962). For many realistic, highly complex systems, the EnKF is essentially the only way to do (approximate) inference, while alternative exact inference techniques can only be applied to highly simplified versions of the problem.

The key difference between the EnKF and other sequential Monte Carlo algorithms (e.g., particle filters) is the use of a linear updating rule that converts the prior ensemble to a posterior ensemble after each observation. Most other sequential Monte Carlo algorithms use a reweighting or resampling step, but it is well known that the weights degenerate (i.e., all but one weight

CONTACT Matthias Katzfuss *katzfuss@tamu.edu* Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 78843.
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/TAS*.

are essentially zero) in high-dimensional problems (Snyder et al. 2008).

Most of the technical development and application of the EnKF has been in the geophysics literature (e.g., Burgers, van Leeuwen, and Evensen 1998; Houtekamer and Mitchell 1998; Bishop, Etherton, and Majumdar 2001; Tippett et al. 2003; Ott et al. 2004; see also Anderson 2009, for a review) and it has received relatively little attention in the statistics literature. This is at least partially because the jargon and notation in the geophysics literature can be daunting upon first glance. Yet, with the increased interest in approximate computational methods for "big data" in statistics, it is important that statisticians be aware of the power of this relatively simple methodology. We also believe that statisticians have much to contribute to this area of research. Thus, our goal in this article is to provide the elementary background and concepts behind the EnKF in a notation that is more common to the statistical state-space literature. Although the real strength of the EnKF is its application to high-dimensional nonlinear and non-Gaussian problems, for pedagogical purposes, we focus primarily on the case of linear measurement and evolution models with Gaussian errors to better illustrate the approach.

In Section 2, we review the standard Kalman filter for linear Gaussian state-space models. We then show in Section 3 how the basic EnKF can be derived based on the notions of conditional simulation, with an interpretation of shifting samples from the prior to the posterior based on observations. In Section 4, we discuss issues, extensions, and operational variants of the basic EnKF, and Section 5 concludes.

2. State-Space Models and The Kalman Filter

For discrete time points $t = 1, 2, \ldots$, assume a linear Gaussian state-space model,

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{v}_t, \qquad \mathbf{v}_t \sim \mathcal{N}_{m_t}(\mathbf{0}, \mathbf{R}_t), \tag{1}$$

$$\mathbf{x}_t = \mathbf{M}_t \mathbf{x}_{t-1} + \mathbf{w}_t, \qquad \mathbf{w}_t \sim \mathcal{N}_n(\mathbf{0}, \mathbf{Q}_t), \tag{2}$$

where $\mathbf{y}_t$ is the observed m_t-dimensional data vector at time t, $\mathbf{x}_t$ is the n-dimensional unobserved state vector of interest, and the observation and innovation error $\mathbf{v}_t$ and $\mathbf{w}_t$ are mutually and serially independent. We call Equations (1)–(2) the observation model and the evolution model, respectively. The observation matrix $\mathbf{H}_t$ relates the state to the observation, and the evolution (propagator) matrix $\mathbf{M}_t$ determines how the state evolves over time. In weather prediction problems, the state and observation dimensions are often enormous (i.e., $n \geq 10^7$ and $m_t \geq 10^5$). In addition, with the exception of Section 5.3, we assume that $\mathbf{M}_t$, $\mathbf{H}_t$, $\mathbf{R}_t$, and $\mathbf{Q}_t$ are known, which is often the assumption in geophysical applications.

An important form of inference for state-space models is *filtering*. In this framework, the goal at every time point $t = 1, 2, \ldots$, is to obtain the filtering distribution of the state; in Bayesian terms, this is equivalent to the posterior distribution of $\mathbf{x}_t$ given $\mathbf{y}_{1:t} := \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t\}$, the data up to time t. This distribution is the conditional distribution of $\mathbf{x}_t$ given $\mathbf{y}_{1:t}$, which we denote by $\mathbf{x}_t|\mathbf{y}_{1:t}$. Filtering for a linear Gaussian model as in (1)–(2) can be carried out using the Kalman filter, which consists of two steps at every time point: a *forecast step* and an *update step*.

Assuming that the filtering distribution at the previous time $t-1$ is given by

$$\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1} \sim \mathcal{N}_n(\widehat{\boldsymbol{\mu}}_{t-1}, \widehat{\boldsymbol{\Sigma}}_{t-1}), \tag{3}$$

the forecast step computes the forecast distribution for time t based on (2) as

$$\begin{aligned} \mathbf{x}_t|\mathbf{y}_{1:t-1} &\sim \mathcal{N}_n(\widetilde{\boldsymbol{\mu}}_t, \widetilde{\boldsymbol{\Sigma}}_t), \\ \widetilde{\boldsymbol{\mu}}_t &:= \mathbf{M}_t \widehat{\boldsymbol{\mu}}_{t-1}, \\ \widetilde{\boldsymbol{\Sigma}}_t &:= \mathbf{M}_t \widehat{\boldsymbol{\Sigma}}_{t-1} \mathbf{M}_t' + \mathbf{Q}_t. \end{aligned} \tag{4}$$

The update step modifies the forecast distribution using the new data $\mathbf{y}_t$. The update formula can be easily derived by considering the joint distribution of $\mathbf{x}_t$ and $\mathbf{y}_t$ conditional on the past data $\mathbf{y}_{1:t-1}$, which is given by a multivariate normal distribution,

$$\begin{pmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{pmatrix} \Big| \mathbf{y}_{1:t-1} \sim \mathcal{N}_{n+m_t} \left(\begin{pmatrix} \widetilde{\boldsymbol{\mu}}_t \\ \mathbf{H}_t \widetilde{\boldsymbol{\mu}}_t \end{pmatrix}, \begin{pmatrix} \widetilde{\boldsymbol{\Sigma}}_t & \widetilde{\boldsymbol{\Sigma}}_t \mathbf{H}_t' \\ \mathbf{H}_t \widetilde{\boldsymbol{\Sigma}}_t & \mathbf{H}_t \widetilde{\boldsymbol{\Sigma}}_t \mathbf{H}_t' + \mathbf{R}_t \end{pmatrix} \right). \tag{5}$$

Using well-known properties of the multivariate normal distribution, it follows that $\mathbf{x}_t|\mathbf{y}_{1:t} \sim \mathcal{N}_n(\widehat{\boldsymbol{\mu}}_t, \widehat{\boldsymbol{\Sigma}}_t)$, where the update equations are given by

$$\widehat{\boldsymbol{\mu}}_t := \widetilde{\boldsymbol{\mu}}_t + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t \widetilde{\boldsymbol{\mu}}_t) \tag{6}$$

$$\widehat{\boldsymbol{\Sigma}}_t := (\mathbf{I}_n - \mathbf{K}_t \mathbf{H}_t) \widetilde{\boldsymbol{\Sigma}}_t, \tag{7}$$

and $\mathbf{K}_t := \widetilde{\boldsymbol{\Sigma}}_t \mathbf{H}_t' (\mathbf{H}_t \widetilde{\boldsymbol{\Sigma}}_t \mathbf{H}_t' + \mathbf{R}_t)^{-1}$ is the so-called Kalman gain matrix of size $n \times m_t$.

An alternative expression for the update equations is given by

$$\widehat{\boldsymbol{\mu}}_t = \widehat{\boldsymbol{\Sigma}}_t (\widetilde{\boldsymbol{\Sigma}}_t^{-1} \widetilde{\boldsymbol{\mu}}_t + \mathbf{H}_t' \mathbf{R}_t^{-1} \mathbf{y}_t) \tag{8}$$

$$\widehat{\boldsymbol{\Sigma}}_t^{-1} = \widetilde{\boldsymbol{\Sigma}}_t^{-1} + \mathbf{H}_t' \mathbf{R}_t^{-1} \mathbf{H}_t, \tag{9}$$

where the second equation is obtained from (7) using the Sherman–Morrison–Woodbury formula (Sherman and Morrison 1950; Woodbury 1950). These equations allow a nice interpretation of the Kalman filter update. The filtered mean in (8) is a weighted average of the prior mean $\widetilde{\boldsymbol{\mu}}_t$ and the observation vector $\mathbf{y}_t$, where the weights are proportional to the prior precision $\widetilde{\boldsymbol{\Sigma}}_t^{-1}$, and $\mathbf{H}_t' \mathbf{R}_t^{-1}$, a combination of the observation matrix and the data precision matrix. The posterior precision in (9) is the sum of the prior precision and the data precision (projected onto the state space).

In summary, the Kalman filter provides the exact filtering distribution for linear Gaussian state-space models as in (1)–(2). However, if n or m_t are large, calculating and storing the $n \times n$ matrices $\widetilde{\boldsymbol{\Sigma}}_t$ and $\widehat{\boldsymbol{\Sigma}}_t$ and calculating the inverse of the $m_t \times m_t$ matrix in the Kalman gain (below (7)) are extremely expensive, and approximations become necessary.

3. The Ensemble Kalman Filter (EnKF)

The ensemble Kalman filter (EnKF) can be viewed as an approximate version of the Kalman filter, in which the state distribution is represented by a sample or "ensemble" from the distribution. This ensemble is then propagated forward through time and updated when new data become available. The ensemble representation is a form of dimension reduction, which leads

to computational feasibility even for very high-dimensional systems.

Specifically, assume that the ensemble $\widehat{\mathbf{x}}_{t-1}^{(1)}, \ldots, \widehat{\mathbf{x}}_{t-1}^{(N)}$ is a sample from the filtering distribution at time $t-1$ in (3): $\widehat{\mathbf{x}}_{t-1}^{(i)} \sim \mathcal{N}_n(\widehat{\boldsymbol{\mu}}_{t-1}, \widehat{\boldsymbol{\Sigma}}_{t-1})$. Similar to the Kalman filter, the EnKF consists of a forecast step and an update step at every time point t.

The EnKF forecast step obtains a sample from the forecast distribution (4) by simply applying the evolution equation in (2) to each ensemble member:

$$\widetilde{\mathbf{x}}_t^{(i)} = \mathbf{M}_t \widehat{\mathbf{x}}_{t-1}^{(i)} + \mathbf{w}_t^{(i)}, \quad \mathbf{w}_t^{(i)} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{Q}_t), \qquad i = 1, \ldots, N. \tag{10}$$

It is easy to verify that $\widetilde{\mathbf{x}}_t^{(i)} \sim \mathcal{N}_n(\widetilde{\boldsymbol{\mu}}_t, \widetilde{\boldsymbol{\Sigma}}_t)$ holds exactly.

Then, this forecast ensemble $\widetilde{\mathbf{x}}_t^{(1)}, \ldots, \widetilde{\mathbf{x}}_t^{(N)}$ must be updated based on $\mathbf{y}_t$, the new data at time t. This update step can be carried out stochastically or deterministically.

3.1. Stochastic Updates

We focus first on stochastic updates, as these are more natural to statisticians, and can be easily motivated as an approximate form of conditional simulation (e.g., Journel 1974).

Specifically, to conditionally simulate from the state filtering distribution, we use the state forecast ensemble in (10), together with a set of simulated observations $\widetilde{\mathbf{y}}_t^{(1)}, \ldots, \widetilde{\mathbf{y}}_t^{(N)}$ from the observation forecast distribution. Setting $\mathbf{v}_t^{(i)} \sim \mathcal{N}_{m_t}(\mathbf{0}, \mathbf{R}_t)$, it can be easily verified that $\widetilde{\mathbf{x}}_t^{(i)}$ and $\widetilde{\mathbf{y}}_t^{(i)} = \mathbf{H}_t \widetilde{\mathbf{x}}_t^{(i)} - \mathbf{v}_t^{(i)}$ follow the correct joint distribution in (5). Conditional simulation then shifts the forecast ensemble based on the difference between the simulated and actual observations:

$$\widehat{\mathbf{x}}_t^{(i)} = \widetilde{\mathbf{x}}_t^{(i)} + \mathbf{K}_t(\mathbf{y}_t - \widetilde{\mathbf{y}}_t^{(i)}), \qquad i = 1, \ldots, N.$$

It is straightforward to show that $\widehat{\mathbf{x}}^{(i)} \sim \mathcal{N}_n(\widehat{\boldsymbol{\mu}}_t, \widehat{\boldsymbol{\Sigma}}_t)$ by considering the first two moments: Clearly, $\mathrm{E}(\widehat{\mathbf{x}}_t^{(i)}) = \widehat{\boldsymbol{\mu}}_t$, and the covariance matrix is given by

$$\begin{aligned}\mathrm{var}(\widehat{\mathbf{x}}_t^{(i)}) &= \mathrm{var}(\widetilde{\mathbf{x}}_t^{(i)}) + \mathrm{var}(\mathbf{K}_t\widetilde{\mathbf{y}}_t^{(i)}) - 2\mathrm{cov}(\widetilde{\mathbf{x}}_t^{(i)}, \mathbf{K}_t\widetilde{\mathbf{y}}_t^{(i)}) \\ &= \widetilde{\boldsymbol{\Sigma}}_t + \mathbf{K}_t\mathbf{H}\widetilde{\boldsymbol{\Sigma}}_t - 2\mathbf{K}_t\mathbf{H}\widetilde{\boldsymbol{\Sigma}}_t = \widehat{\boldsymbol{\Sigma}}_t.\end{aligned}$$

Hence, conditional simulation provides a simple way to update the forecast ensemble to obtain a filtering ensemble that is an exact sample from the filtering distribution. This requires computation of the Kalman gain, $\mathbf{K}_t$, which in turn requires computation and storage of the $n \times n$ forecast covariance matrix, $\widetilde{\boldsymbol{\Sigma}}_t$.

To avoid calculating this potentially huge matrix, the update step of the EnKF is an approximate version of conditional simulation, for which the Kalman gain $\mathbf{K}_t$ is replaced by an estimate $\widehat{\mathbf{K}}_t$ based on the forecast ensemble. Often, the estimated Kalman gain has the form

$$\widehat{\mathbf{K}}_t := \mathbf{C}_t\mathbf{H}_t'(\mathbf{H}_t\mathbf{C}_t\mathbf{H}_t' + \mathbf{R}_t)^{-1}, \tag{11}$$

where $\mathbf{C}_t$ is an estimate of the state forecast covariance matrix $\widetilde{\boldsymbol{\Sigma}}_t$. The simplest example is $\mathbf{C}_t = \widetilde{\mathbf{S}}_t$, where $\widetilde{\mathbf{S}}_t$ is the sample covariance matrix of $\widetilde{\mathbf{x}}_t^{(1)}, \ldots, \widetilde{\mathbf{x}}_t^{(N)}$. For more details, see Section 4.1.

Then, given an initial ensemble $\widehat{\mathbf{x}}_0^{(1)}, \ldots, \widehat{\mathbf{x}}_0^{(N)}$, that can be taken as draws from the posterior distribution at time $t = 0$, or, in the case of complex nonlinear models, from various "spin-up" algorithms (e.g., Hoteit et al. 2015), one can implement the EnKF in Algorithm 1.

Algorithm 1: Stochastic EnKF

Start with an initial ensemble $\widehat{\mathbf{x}}_0^{(1)}, \ldots, \widehat{\mathbf{x}}_0^{(N)}$. Then, at each time $t = 1, 2, \ldots$, given an ensemble $\widehat{\mathbf{x}}_{t-1}^{(1)}, \ldots, \widehat{\mathbf{x}}_{t-1}^{(N)}$ of draws from the filtering distribution at time $t-1$, the stochastic EnKF carries out the following two steps for $i = 1, \ldots, N$:

1. Forecast Step: Draw $\mathbf{w}_t^{(i)} \sim \mathcal{N}_n(\mathbf{0}, \mathbf{Q}_t)$ and calculate $\widetilde{\mathbf{x}}_t^{(i)} = \mathbf{M}_t\widehat{\mathbf{x}}_{t-1}^{(i)} + \mathbf{w}_t^{(i)}$.

2. Update Step: Draw $\mathbf{v}_t^{(i)} \sim \mathcal{N}_{m_t}(\mathbf{0}, \mathbf{R}_t)$ and calculate $\widehat{\mathbf{x}}_t^{(i)} = \widetilde{\mathbf{x}}_t^{(i)} + \widehat{\mathbf{K}}_t(\mathbf{y}_t + \mathbf{v}_t^{(i)} - \mathbf{H}_t\widetilde{\mathbf{x}}_t^{(i)})$, where $\widehat{\mathbf{K}}_t$ is given in (11).

As we have seen, the EnKF update can be nicely interpreted as an approximate version (because $\mathbf{K}_t$ is estimated by $\widehat{\mathbf{K}}_t$) of conditional simulation. For alternative interpretations, rewrite the update step as

$$\begin{aligned}\widehat{\mathbf{x}}_t^{(i)} &= \widetilde{\mathbf{x}}_t^{(i)} + \widehat{\mathbf{K}}_t(\mathbf{y}_t^{(i)} - \mathbf{H}_t\widetilde{\mathbf{x}}_t^{(i)}) &(12)\\ &= (\mathbf{I}_n - \widehat{\mathbf{K}}_t\mathbf{H}_t)\widetilde{\mathbf{x}}_t^{(i)} + \widehat{\mathbf{K}}_t\mathbf{y}_t^{(i)}, &(13)\end{aligned}$$

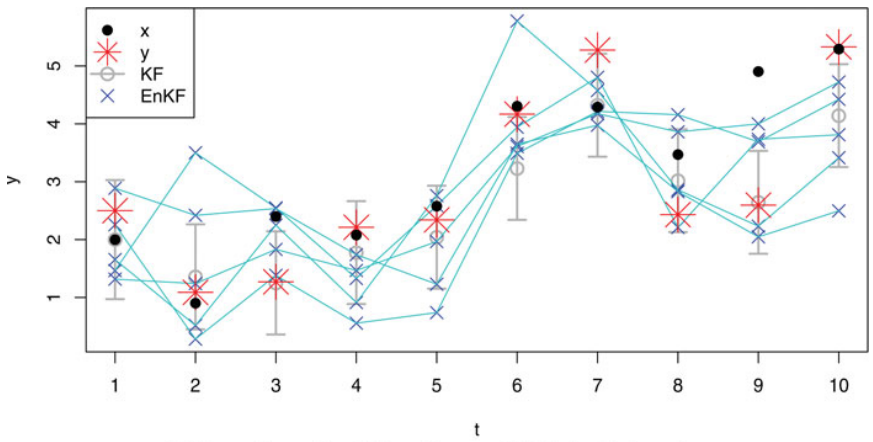
where $\mathbf{y}_t^{(i)} = \mathbf{y}_t + \mathbf{v}_t^{(i)}$ is a "perturbed" observation. From (12), the EnKF update can be viewed as a stochastic version of the Kalman filter mean update (6), where the prior mean and the observation, $(\widetilde{\boldsymbol{\mu}}_t, \mathbf{y}_t)$, are replaced with a prior draw and a perturbed observation, $(\widetilde{\mathbf{x}}_t^{(i)}, \mathbf{y}_t^{(i)})$, respectively. That is, the EnKF combines a draw from the prior or forecast distribution, $\widetilde{\mathbf{x}}_t^{(i)}$, with a draw from the "likelihood," $\mathbf{y}_t^{(i)} \sim \mathcal{N}_{m_t}(\mathbf{y}_t, \mathbf{R}_t)$, to obtain a posterior draw from the filtering distribution as in (12). From (13), we further see that the filtering ensemble is a linear combination or "shift" of the forecast ensemble and the observation.

3.2. Deterministic Updates

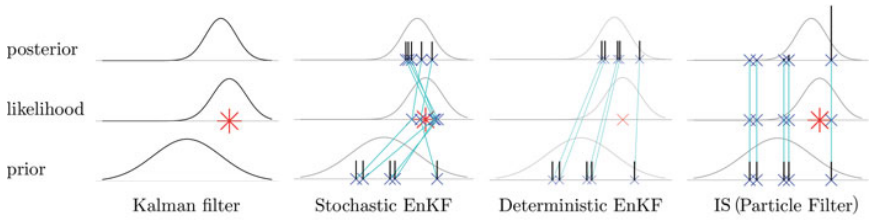
The update step of the stochastic EnKF in Algorithm 1 can be replaced by a deterministic update, leading to the widely used class of *deterministic EnKFs*. These methods obtain an approximate ensemble from the posterior by deterministically shifting the prior ensemble, without relying on simulated or perturbed observations.

The main idea behind the deterministic filter in the univariate case is as follows. Suppose we have prior draws, $\tilde{x}^{(i)} \sim \mathcal{N}(\tilde{\mu}, \tilde{\sigma}^2)$, and we want to convert them to posterior draws, $\hat{x}^{(i)} \sim \mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$. To do this we first standardize the prior draws as $z^{(i)} = (\tilde{x}^{(i)} - \tilde{\mu})/\tilde{\sigma}$, and then "unstandardize" them as $\hat{x}^{(i)} = \hat{\mu} + \hat{\sigma}z^{(i)}$. Combining the two steps gives $\hat{x}^{(i)} = \hat{\mu} + \hat{\sigma}/\tilde{\sigma}\,(\tilde{x}^{(i)} - \tilde{\mu})$. Thus, the deterministic update involves shifting and scaling each prior draw so that the resulting posterior draws are shifted toward the data and have a smaller variance than the prior. Figure 1(b) illustrates the idea in a simple univariate example.

There are many variants of deterministic updates, including the ensemble adjustment Kalman filter (EAKF) and ensemble transform Kalman filter (ETKF). The details of these algorithms differ slightly when $N < n$, but they all belong to the family of square root filters and are based on the same idea (Tippett et al.



(a) State, observation, Kalman filter, and EnKF for 10 time points



(b) Illustration of different updating schemes at time point $t = 1$

Figure 1. Simulation from a one-dimensional state-space model with $n = m_t \equiv 1$, $\mathbf{H}_t \equiv 1$, $\mathbf{M}_t \equiv 0.9$, $\mathbf{R}_t \equiv 1$, and $\mathbf{Q}_t \equiv 1$. The top panel (a) shows the state, observations, Kalman filter, and the $N = 5$ EnKF ensemble members for the first 10 time points. For the Kalman filter, we show the filtering mean and 75% confidence intervals (which correspond to the average spread from $N = 5$ samples from a normal distribution). For the first time point $t = 1$, the bottom panel (b) compares the Kalman filter and stochastic EnKF updates (with perturbed observations) to a deterministic EnKF (see Section 3.2) and to the importance sampler (IS), which is the basis for the update step in most particle filters (e.g., Gordon, Salmond, and Smith 1993). The bars above the ensemble/particles are proportional to the weights. All approximation methods start with the same, equally weighted $N = 5$ prior samples. Even for this one-dimensional example, importance sampling degenerates and represents the posterior distribution with essentially only one particle with significant weight, while the EnKF methods shift the ensemble members and thus obtain a better representation of the posterior distribution.

2003). Omitting the subscript t for notational simplicity, let $\widetilde{\mathbf{L}}$ be a matrix square root of the forecast (prior) covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ (i.e., $\widetilde{\boldsymbol{\Sigma}} = \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}'$), and define $\mathbf{D} := \mathbf{H}_t\widetilde{\boldsymbol{\Sigma}}_t\mathbf{H}_t' + \mathbf{R}_t$. Then we can write (7) as

$$\widehat{\boldsymbol{\Sigma}} = (\mathbf{I}_n - \widetilde{\mathbf{L}}\widetilde{\mathbf{L}}'\mathbf{H}'\mathbf{D}^{-1}\mathbf{H})\widetilde{\mathbf{L}}\widetilde{\mathbf{L}}' = \widetilde{\mathbf{L}}(\mathbf{I}_n - \widetilde{\mathbf{L}}'\mathbf{H}'\mathbf{D}^{-1}\mathbf{H}\widetilde{\mathbf{L}})\widetilde{\mathbf{L}}' = \widehat{\mathbf{L}}\widehat{\mathbf{L}}',$$

and so $\widehat{\mathbf{L}} = \widetilde{\mathbf{L}}\mathbf{W}$ is a matrix square root of the posterior (filtering) covariance matrix, where $\mathbf{W}\mathbf{W}' = \mathbf{I}_n - \widetilde{\mathbf{L}}'\mathbf{H}'\mathbf{D}^{-1}\mathbf{H}\widetilde{\mathbf{L}}$. That is, the filtering covariance matrix can be obtained by post-multiplying the forecast covariance with the matrix $\mathbf{W}$.

In the deterministic EnKF variants, the $n \times n$ matrix $\widetilde{\boldsymbol{\Sigma}}$ is obtained based on the forecast ensemble with N $\ll$ n members, and the resulting low-rank structure is computationally exploited in different ways by the different variants.

3.3. Summary of The Basic EnKF

In summary, the EnKF only requires storing and operating on N vectors (ensemble members) of length n, and the estimated Kalman gain can often be calculated quickly (see Section 4.1). In theory, as $N \to \infty$, the EnKF converges to the (exact) Kalman filter for linear Gaussian models, but large values of N are typically infeasible in practice. Updating the ensemble by shifting makes the algorithm much less prone to degeneration than alternatives that rely on reweighting of ensemble members (e.g., particle filters). This is illustrated in a simple one-dimensional example in Figure 1.

Deterministic EnKF variants generally have less sampling variability and are more accurate than stochastic filters for very small ensemble sizes (e.g., Furrer and Bengtsson 2007, sec. 3.4). However, if the prior distribution is non-Gaussian (e.g., multimodal or skewed), then stochastic filters can be more accurate than deterministic updates (Lei, Bickel, and Snyder 2010).

For readers interested in applying the EnKF to large real-world problems, we recommend user-friendly software such as the data assimilation research testbed, or DART (Anderson et al. 2009).

4. Operational Variants and Extensions

4.1. Variance Inflation and Localization

For dimension reduction and computational feasibility, the ensemble size, N, typically is much smaller than the state dimension, n. If the estimated Kalman gain $\widehat{\mathbf{K}}_t$ is (11) with the estimated forecast covariance $\mathbf{C}_t$ simply taken to be $\widetilde{\mathbf{S}}_t$ (the sample covariance matrix of the forecast ensemble), then $\widehat{\mathbf{K}}_t$ is often a poor approximation of the true Kalman gain. This has two adverse effects that require adjustments in practice. First, small ensemble sizes lead to downwardly biased estimates of the posterior state covariance matrix (Furrer and Bengtsson 2007). This can be alleviated by *covariance inflation* (e.g., Anderson 2007a),

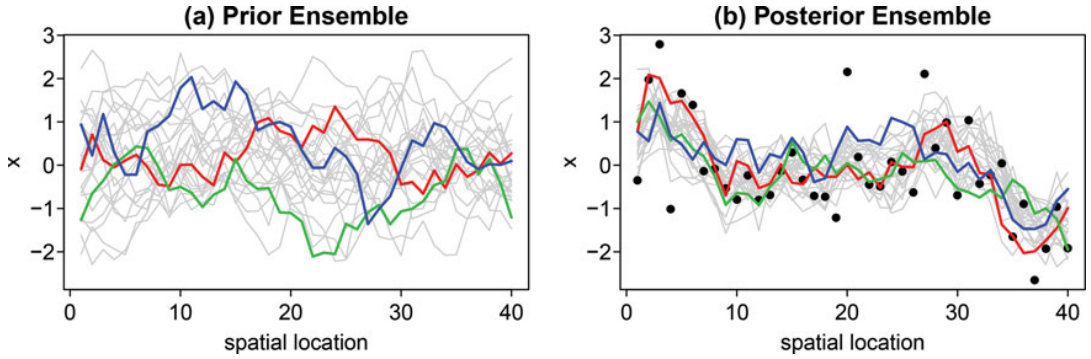


Figure 2. For the spatial example in Section 4.1, (a) $N = 25$ prior ensemble members $\widetilde{\mathbf{x}}_t^{(1)}, \ldots, \widetilde{\mathbf{x}}_t^{(25)}$ and (b) observations $\mathbf{y}$ and posterior ensemble members $\widehat{\mathbf{x}}_t^{(1)}, \ldots, \widehat{\mathbf{x}}_t^{(25)}$ from the stochastic EnKF algorithm with tapering radius $r = 10$. Note that the x-axis represents spatial location, not time. For illustration, three members of the ensemble are shown in different colors. The prior ensemble members have constant mean and variance, while the mean of the posterior ensemble is shifted toward the data and the variance is smaller than for the prior ensemble.

where $\widetilde{\mathbf{S}}_t$ is multiplied by a constant greater than one. Second, small ensemble sizes lead to rank deficiency in $\widetilde{\mathbf{S}}_t$, and often result in spurious correlations appearing between state components that are physically far apart. In practice, this is usually avoided by "localization." Several localization approaches have been proposed, but many of them can be viewed as a form of *tapering* from a statistical perspective (Furrer and Bengtsson 2007). The most common form of localization sets $\mathbf{C}_t = \widetilde{\mathbf{S}}_t \circ \mathcal{T}_t$, where $\circ$ denotes the Hadamard (entrywise) product, and $\mathcal{T}_t$ is a sparse positive definite correlation matrix (e.g., Furrer, Genton, and Nychka 2006; Anderson 2007b).

Figures 2 and 3 illustrate the EnKF update step with tapering localization in a one-dimensional spatial example at a single time step. Omitting time subscripts for notational simplicity, we define the state vector as $\mathbf{x} = (x_1, \ldots, x_{40})'$, which represents the "true spatial field" at $n = 40$ equally spaced locations along a line. Observations are taken at each location, so $m = 40$, and we assume that $\mathbf{H} = \mathbf{I}$ and $\mathbf{R} = \mathbf{I}$. The prior distribution for the state is $\mathbf{x} \sim \mathcal{N}_{40}(\mathbf{0}, \widetilde{\mathbf{\Sigma}})$, with $\widetilde{\mathbf{\Sigma}}_{ij} = \phi^{|i-j|}$, where $\phi = 0.9$. That is, the state follows a stationary spatial process with an exponential correlation function. We implement the EnKF algorithm with localization using $N = 25$ ensemble members, where the tapering matrix $\mathcal{T}$ is based on the compactly supported 5th-order piecewise polynomial correlation function of Gaspari and Cohn (1999), with a radius of $r = 10$. No covariance inflation is used. Figure 2 shows the prior and posterior ensemble members from the EnKF algorithm, along with the observed data. Figure 3 shows the true and ensemble-based prior covariance matrix, $\widetilde{\mathbf{\Sigma}}$ and $\mathbf{C}$, and the Kalman gain, $\mathbf{K}$. Note that the Kalman gain is equivalent to the posterior covariance (i.e., $\mathbf{K} = \widehat{\mathbf{\Sigma}}$) in this example because $\mathbf{H} = \mathbf{R} = \mathbf{I}$.

4.2. Serial Updating

In many high-dimensional data assimilation systems (such as DART), the EnKF is implemented using *serial updating* (e.g., Houtekamer and Mitchell 2001), where at each time t, the ensemble is updated after each scalar observation, y_{ti}, $i = 1, \ldots, m_t$, rather than simultaneously using the entire vector $\mathbf{y}_t$. Serial assimilation requires that the observation errors are uncorrelated (i.e., $\mathbf{R}_t$ is diagonal). If $\mathbf{R}_t$ is nondiagonal, the observation Equation (1) can be premultiplied by $\mathbf{R}_t^{-1/2}$ to obtain uncorrelated observation errors.

For linear Gaussian models, it can be shown that the simultaneous and serial Kalman filter schemes yield identical posteriors after m_t observations. The advantage of serial updating is that it avoids calculation and storage of the $n \times m_t$ Kalman gain matrix $\mathbf{K}_t$ (which requires inverting an $m_t \times m_t$ matrix). Serial methods require computing an $n \times 1$ Kalman gain vector (and hence inverting only a scalar) for each of the m_t scalar observations. Stochastic and deterministic EnKFs can be implemented serially by applying the updating formulas to each row of (1) one at a time. Covariance inflation and localization can also be applied in the serial setting, but they are only applied to one column of the covariance matrix after each scalar update, and the serial updates are then generally not equivalent to joint updates.

4.3. Smoothing

While we have focused on filtering inference so far, there is another form of inference in state-space models called smoothing. Based on data $\mathbf{y}_{1:T}$ collected in a fixed time period $\{1, \ldots, T\}$, smoothing attempts to find the posterior distribution of the state $\mathbf{x}_t$ conditional on $\mathbf{y}_{1:T}$ for any $t \in \{1, \ldots, T\}$. For linear Gaussian state-space models, this can be done exactly using the Kalman smoother, which consists of a Kalman filter, plus subsequent recursive "backward smoothing" for $t = T, T-1, \ldots, 1$ (e.g., Shumway and Stoffer 2006). Again, this becomes computationally challenging when the state or observation dimensions are large. For moderate dimensions, smoothing inference can then be carried out using an ensemble-based approximation of the Kalman smoother, which also starts with an EnKF and then carries out backward recursions using the estimated filtering ensembles (e.g., Stroud et al. 2010).

If the state dimension is very large or the evolution is nonlinear, the most widely used smoother is the ensemble Kalman smoother (EnKS) of Evensen and van Leeuwen (2000). The EnKS is a forward-only (i.e., no backward pass is required) algorithm that relies on the idea of *state augmentation*. At each time point t, the state vector is augmented to include the lagged states

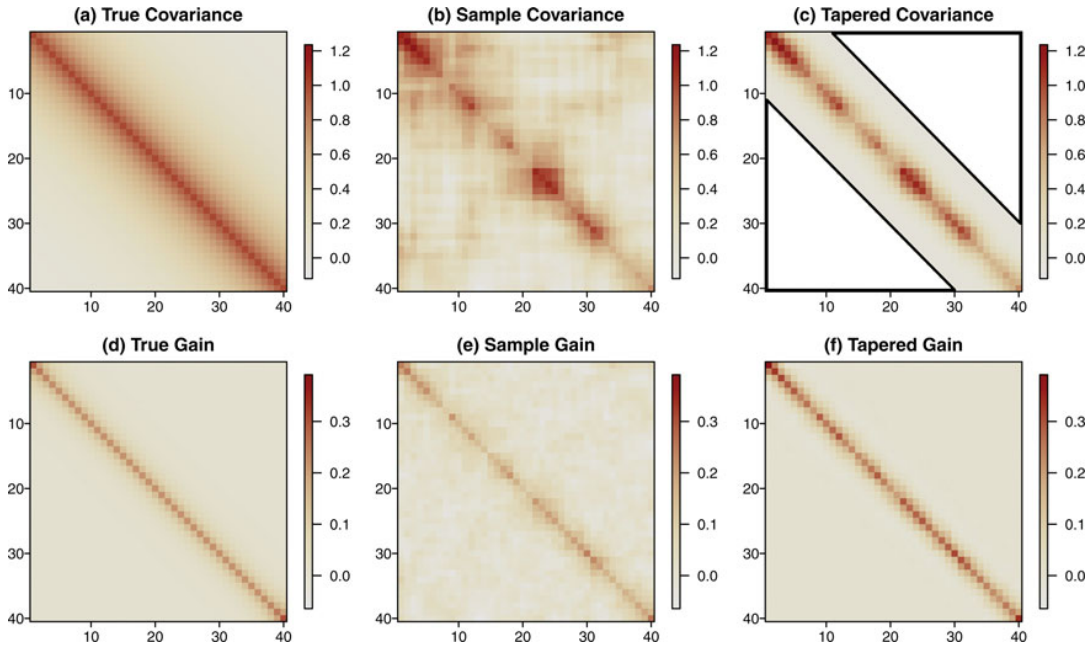


Figure 3. For the spatial example in Section 4.1, comparison of true and ensemble-based prior covariance and Kalman gain matrices. Top row: (a) true covariance $\widetilde{\mathbf{\Sigma}}$, (b) sample covariance $\widetilde{\mathbf{S}}$, (c) tapered sample covariance $\mathbf{C}$ with $r = 10$. Bottom row: Kalman gain matrices $\mathbf{K}$ and $\widehat{\mathbf{K}}$ obtained using the prior covariances in the top row (i.e., plots (d)–(f) correspond to covariances in (a)–(c), respectively). The sample forecast covariance matrix in (b) and the corresponding Kalman gain matrix in (e) exhibit a large amount of sampling variability. The tapered covariance matrix in (c) has less sampling variability, the correlations die out faster, and become identically zero (shown in white) for locations more than $r = 10$ units apart. The corresponding gain matrix in (f) is a more accurate estimate of the true gain than the untapered version in (e), but it is no longer sparse.

back to time 1: $\mathbf{x}_{1:t} = (\mathbf{x}_1', \ldots, \mathbf{x}_t')'$. The update step is analogous to the EnKF, but instead of updating only $\mathbf{x}_t$, it is necessary to update the entire history $\mathbf{x}_{1:t}$. To avoid having to update the entire history, sometimes a moving-window approach is applied, in which the update at each time point only considers the recent history up to a certain time lag.

4.4. Parameter Estimation

So far, we have assumed that the only unknown quantities in the state-space model (1)–(2) are the state vectors $\mathbf{x}_t$, and that there are no other unknown parameters. In practice, of course, this is often not the case. The matrices $\mathbf{H}_t$, $\mathbf{M}_t$, $\mathbf{R}_t$, and $\mathbf{Q}_t$ often include some unknown parameters $\boldsymbol{\theta}$ (e.g., autoregressive coefficients, variance parameters, spatial range parameters) that also need to be estimated.

There are two main approaches to parameter estimation within the EnKF framework. The first is a very popular approach called state augmentation (Anderson 2001). Here, the parameters are treated as time-varying quantities with small artificial evolution noise. We then combine the states and parameters in an augmented state vector $\boldsymbol{z}_t = (\mathbf{x}_t', \boldsymbol{\theta}_t')'$ and run an EnKF on the augmented state vector to obtain posterior estimates of states and parameters at each time t. This approach works well in many examples; however, it implicitly assumes that the states and parameters *jointly* follow a linear Gaussian state-space model. For some parameters, such as covariance parameters, this assumption is violated, and the method fails completely (Stroud and Bengtsson 2007).

The second approach to parameter estimation is based on approximate likelihood functions constructed using the output from the EnKF. It can in principle be used for any type of parameter. The parameters are estimated either by maximum likelihood (ML) or Bayesian methods. Examples of these approaches include sequential ML (Mitchell and Houtekamer 2000), offline ML (Stroud et al. 2010), and sequential Bayesian methods (Stroud and Bengtsson 2007; Frei and Künsch 2012). In general, these methods have been successful in examples with a relatively small number of parameters, and more work is needed for cases where the parameter and state are both high dimensional.

4.5. Non-Gaussianity and Nonlinearity

As we have shown above in Section 3, the EnKF updates are based on the assumption of a linear Gaussian state-space model of the form (1)–(2). However, the EnKF is surprisingly robust to deviations from these assumptions, and there are many examples of successful applications of such models in the geophysical literature.

In general, nonlinearity of the state-space model means that $\mathbf{H}_t\mathbf{x}_t$ in (1) is replaced by a nonlinear observation operator $\mathcal{H}_t(\mathbf{x}_t)$, or $\mathbf{M}_t\mathbf{x}_{t-1}$ in (2) is replaced by a nonlinear evolution operator $\mathcal{M}_t(\mathbf{x}_{t-1})$. An advantage of the EnKF is that these operators do not have to be available in closed form. Instead, as can be seen in Algorithm 1, these operators simply have to be "applied" to each ensemble member. In addition, it is easy to show that if $\hat{\boldsymbol{x}}_{t-1}$ is a sample from the filtering distribution, then $\tilde{\boldsymbol{x}}_t = \mathcal{M}_t(\hat{\boldsymbol{x}}_{t-1}) + w_t$ is an (exact) sample from the forecast distribution.

The effectiveness of the EnKF and its variants in the non-Gaussian/nonlinear case is a consequence of so-called linear Bayesian estimation (e.g., Hartigan 1969; Goldstein and Wooff 2007). In its original form, linear Bayesian estimation seeks to find linear estimators given only the first and second moments of the prior distribution and likelihood, and in that sense it is "distribution-free." This is a Bayesian justification for kriging methods in spatial statistics (Omre 1987) and non-Gaussian/nonlinear state-space models in time-series analysis (e.g., West, Harrison, and Migon 1985; Fahrmeir 1992). When considering linear Gaussian priors and likelihoods, this approach obviously gives the true Bayesian posterior. In other contexts, it is appealing in that it allows one to update prior information without explicit distributional assumptions. However, as described in O'Hagan (1987), the linear Bayesian approach is potentially problematic in the context of highly non-Gaussian distributions (e.g., skewness, heavy tails, multimodality, etc.).

If the EnKF approximation is not satisfactory in the presence of non-Gaussianity, it is possible to employ normal mixtures (e.g., Alspach and Sorenson 1972; Anderson and Anderson 1999; Bengtsson, Snyder, and Nychka 2003), hybrid particle-filter-EnKFs (Hoteit et al. 2008; Stordal et al. 2011; Frei and Künsch 2012), and other approaches (see, e.g., Bocquet, Pires, and Wu 2010, for a review).

5. Conclusions

The EnKF is a powerful tool for inference in high-dimensional state-space models. The key idea of the EnKF relative to other sequential Monte Carlo methods is the use of shifting instead of reweighting in the update step, which allows the algorithm to remain stable in high-dimensional problems. In practice, the algorithm requires the choice of important tuning parameters (e.g., tapering radius and variance inflation factor).

As noted in the introduction, much of the development of the EnKF has been outside the statistics community. There are many outstanding problems and questions associated with EnKFs (e.g., parameter estimation, highly nonlinear and non-Gaussian models), and statisticians surely have much to contribute.

Funding

Katzfuss' research was partially supported by National Science Foundation (NSF) Grant DMS-1521676 and by NASA's Earth Science Technology Office AIST-14 program. Wikle acknowledges the support of NSF grant DMS-1049093 and Office of Naval Research (ONR) grant ONR-N00014-10-0518.

Acknowledgment

The authors thank Thomas Bengtsson, Jeffrey Anderson, and three anonymous reviewers for helpful comments and suggestions.

References

Alspach, D. L., and Sorenson, H. W. (1972), "Nonlinear Bayesian Estimation Using Gaussian Sum Approximations," *IEEE Transactions on Automatic Control*, 17, 439–448. [356]

Anderson, J., Hoar, T., Raeder, K., Liu, H., Collins, N., Torn, R., and Avellano, A. (2009), "The Data Assimilation Research Testbed: A Community Facility," *Bulletin of the American Meteorological Society*, 90, 1283–1296. [353]

Anderson, J. L. (2001), "An Ensemble Adjustment Kalman Filter for Data Assimilation," *Monthly Weather Review*, 129, 2884–2903. [355]

——— (2007a), "An Adaptive Covariance Inflation Error Correction Algorithm for Ensemble Filters," *Tellus A*, 59, 210–224. [353]

——— (2007b), "Exploring the Need for Localization in Ensemble Data Assimilation Using an Hierarchical Ensemble Filter," *Physica D*, 230, 99–111. [354]

——— (2009), "Ensemble Kalman Filters for Large Geophysical Applications," *IEEE Control Systems Magazine*, 29, 66–82. [351]

Anderson, J. L., and Anderson, S. L. (1999), "A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts," *Monthly Weather Review*, 127, 2741–2758. [356]

Bengtsson, T., Snyder, C., and Nychka, D. (2003), "Toward a Nonlinear Ensemble Filter for High-Dimensional Systems," *Journal of Geophysical Research*, 108, 8775. [356]

Bishop, C., Etherton, B., and Majumdar, S. (2001), "Adaptive Sampling With the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects," *Monthly Weather Review*, 129, 420–436. [351]

Bocquet, M., Pires, C. A., and Wu, L. (2010), "Beyond Gaussian Statistical Modeling in Geophysical Data Assimilation," *Monthly Weather Review*, 138, 2997–3023. [356]

Burgers, G., van Leeuwen, P. J., and Evensen, G. (1998), "Analysis Scheme in the Ensemble Kalman Filter," *Monthly Weather Review*, 126, 1719–1724. [351]

Evensen, G. (1994), "Sequential Data Assimilation With a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics," *Journal of Geophysical Research*, 99, 10143–10162. [350]

Evensen, G., and van Leeuwen, P. J. (2000), "An Ensemble Kalman Smoother for Nonlinear Dynamics," *Monthly Weather Review*, 128, 1852–1867. [354]

Fahrmeir, L. (1992), "Posterior Mode Estimation by Extended Kalman Filtering for Multivariate Dynamic Generalized Linear Models," *Journal of the American Statistical Association*, 87, 501–509. [356]

Frei, M., and Künsch, H. R. (2012), "Sequential State and Observation Noise Covariance Estimation Using Combined Ensemble Kalman and Particle Filters," *Monthly Weather Review*, 140, 1476–1495. [355,356]

Furrer, R., and Bengtsson, T. (2007), "Estimation of High-Dimensional Prior and Posterior Covariance Matrices in Kalman Filter Variants," *Journal of Multivariate Analysis*, 98, 227–255. [353]

Furrer, R., Genton, M. G., and Nychka, D. (2006), "Covariance Tapering for Interpolation of Large Spatial Datasets," *Journal of Computational and Graphical Statistics*, 15, 502–523. [354]

Gaspari, G., and Cohn, S. E. (1999), "Construction of Correlation Functions in Two and Three Dimensions," *Quarterly Journal of the Royal Meteorological Society*, 125, 723–757. [354]

Goldstein, M., and Wooff, D. (2007), *Bayes Linear Statistics, Theory & Methods* (Vol. 716), New York: Wiley. [356]

Gordon, N., Salmond, D., and Smith, A. (1993), "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEE Proceedings F Radar and Signal Processing,*, 140, 107–113. [353]

Hartigan, J. (1969), "Linear Bayesian Methods," *Journal of the Royal Statistical Society*, Series B, 3, 446–454. [356]

Hoteit, I., Pham, D.-T., Gharamti, M., and Luo, X. (2015), "Mitigating Observation Perturbation Sampling Errors in the Stochastic EnKF," *Monthly Weather Review*. [352]

Hoteit, I., Pham, D. T., Triantafyllou, G., and Korres, G. (2008), "A New Approximate Solution of the Optimal Nonlinear Filter for Data Assimilation in Meteorology and Oceanography," *Monthly Weather Review*, 136, 317–334. [356]

Houtekamer, P. L., and Mitchell, H. L. (1998), "Data Assimilation Using an Ensemble Kalman Filter Technique," *Monthly Weather Review*, 126, 796–811. [351]

——— (2001), "A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation," *Monthly Weather Review*, 129, 123–137. [354]

Journel, A. G. (1974), "Geostatistics for Conditional Simulation of Ore Bodies," *Economic Geology*, 69, 673–687. [352]

Kalman, R. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Journal of Basic Engineering*, 82, 35–45. [350]

Lei, J., Bickel, P., and Snyder, C. (2010), "Comparison of Ensemble Kalman Filters Under Non-Gaussianity," *Monthly Weather Review*, 138, 1293–1306. [353]

Mitchell, H. L., and Houtekamer, P. L. (2000), "An Adaptive Ensemble Kalman Filter," *Monthly Weather Review*, 128, 416–433. [355]

Nychka, D., and Anderson, J. L. (2010), "Data Assimilation," in *Handbook of Spatial Statistics*, eds. A. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes, New York, NY: Chapman & Hall/CRC, pp. 477–492. [350]

O'Hagan, A. (1987), "Bayes Linear Estimators for Randomized Response Models," *Journal of the American Statistical Association*, 82, 580–585. [356]

Omre, H. (1987), "Bayesian Kriging merging Observations and Qualified Guesses in Kriging," *Mathematical Geology*, 19, 25–39. [356]

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., Kalnay, E., Patil, D. J., and Yorke, J. A. (2004), "A Local Ensemble Kalman Filter for Atmospheric Data Assimilation," *Tellus*, 56A, 415–428. [351]

Sherman, J., and Morrison, W. (1950), "Adjustment of an Inverse Matrix Corresponding to a Change in One Element of a Given Matrix," *Annals of Mathematical Statistics*, 21, 124–127. [351]

Shumway, R. H., and Stoffer, D. S. (2006), *Time Series Analysis and its Applications With R Examples* (2nd ed.), New York: Springer. [354]

Snyder, C., Bengtsson, T., Bickel, P., and Anderson, J. L. (2008), "Obstacles to High-dimensional Particle Filtering," *Monthly Weather Review*, 136, 4629–4640. [351]

Stordal, A. S., Karlsen, H. A., Nævdal, G., Skaug, H. J., and Vallès, B. (2011), "Bridging the Ensemble Kalman Filter and Particle Filters: The Adaptive Gaussian Mixture Filter," *Computational Geosciences*, 15, 293–305. [356]

Stroud, J. R., and Bengtsson, T. (2007), "Sequential State and Variance Estimation Within the Ensemble Kalman Filter," *Monthly Weather Review*, 135, 3194–3208. [355]

Stroud, J. R., Stein, M. L., Lesht, B. M., Schwab, D. J., and Beletsky, D. (2010), "An Ensemble Kalman Filter and Smoother for Satellite Data Assimilation," *Journal of the American Statistical Association*, 105, 978–990. [354,355]

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M., and Whitaker, J. S. (2003), "Ensemble Square-Root Filters," *Monthly Weather Review*, 131, 1485–1490. [351,353]

Tukey, J. W. (1962), "The Future of Data Analysis," *The Annals of Mathematical Statistics*, 33, 1–67. [350]

West, M., Harrison, P. J., and Migon, H. S. (1985), "Dynamic Generalized Linear Models and Bayesian Forecasting," *Journal of the American Statistical Association*, 80, 73–83. [356]

Wikle, C. K., and Berliner, L. M. (2007), "A Bayesian Tutorial for Data Assimilation," *Physica D: Nonlinear Phenomena*, 230, 1–16. [350]

Woodbury, M. (1950), "Inverting Modified Matrices," Memorandum Report 42, Statistical Research Group, Princeton University, Princeton, NJ. [351]