Ensemble Kalman Filter

APPLICATION TO METEOROLOGICAL DATA ASSIMILATION

S. LAKSHMIVARAHAN and DAVID J. STENSRUD

Meteorological models are used to predict the weather, study atmospheric processes, and provide input to decision makers on the consequences of increased greenhouse gas emissions to the Earth's climate. Predictions of future atmospheric states are accomplished as an initial value or marching problem, where the initial atmospheric state is specified and the variables are advanced in time using numerical techniques. The challenge of data assimilation in meteorology is to estimate, based on a set of limited observations of varying types, the complete three-dimensional atmospheric state at a given time to provide an initial value for a meteorological model.

The observations used to help estimate the atmospheric state are obtained from a variety of sensors and platforms. Both rawinsondes, often called radiosondes, and surface stations provide routine in situ observations of pressure, temperature, relative humidity, and wind speed and direction. A rawinsonde is a balloon-borne

PHOTO COMPOSITE COURTESY OF DAVID STENSRUD

measurement system that provides a vertical profile of the variables several times a day, while surface observations are taken only at the surface but provide observations at hourly intervals. Ships and buoys also provide routine observations near the water surface at hourly intervals. Commercial aircraft in some countries often provide in situ observations along their flight legs, with vertical profiles provided at their departure and arrival locations. Unfortunately, these in situ observations are unevenly distributed over the Earth, being much more dense over land areas and in the northern hemisphere.

Remotely sensed observations from satellites and radars provide additional data sources for estimating an initial atmospheric state. Satellite observations of radiances cover large swaths of several hundred kilometers in width as the satellite circles the globe from pole to pole. These data provide greatly improved coverage over the

Digital Object Identifier 10.1109/MCS.2009.932225

1066-033X/09/$25.00©2009IEEE

oceans but represent vertically averaged quantities that are weakly related to a large number of variables in meteorological models. Since it takes time for the satellites to circle the globe, the observations at various locations are not obtained at the same time. Similarly, radar observations provide data that can be related to the liquid and ice content of precipitating clouds and the motion of the precipitation relative to the radar location. Radars, however, are deployed over land and cover relatively small regions of the globe since they are mainly located in developed countries. Additional sensors exist that provide either in situ or remotely sensed observations, yielding a total of 10^8–10^9 observations within any given three-hour interval, which is roughly the same order of magnitude as the number of model state variables.

Early approaches to data assimilation use simple two-dimensional univariate and multivariate interpolation schemes [1]–[3]. These schemes estimate the value of the variables at each location based on the set of all observations within a specified distance, or radius of influence, from the chosen location. The quality of the estimates is related to the density and distribution of the observations. Imagine a situation in which five observations surround a given location, four of which cluster close to each other to the north with one lone observation located to the south. If all the observations are the same distance from the given location, the estimate is biased toward the four northern observations even though they provide similar information content and are not independent. Similarly, the nonuniform distribution and quality of globally available observations, with much greater information content over land than over water, leads to poorer estimates of the atmospheric state over the oceans. This problem is alleviated by adopting a Bayesian framework in which a first guess, or prior, estimate of the atmospheric state is provided by a model forecast started at an earlier time [2]. A posterior estimate is obtained by cleverly combining the two pieces of information, namely, the old information in the prior, or background, estimate and the new information in the observations. Under the standard Gaussian assumption, this approach leads naturally to the least squares technique, which is known as the three-dimensional variational (3DVAR) method [1]–[3].

While the use of prior estimates greatly improves the three-dimensional analyses of the atmospheric state, problems associated with the nonuniform distribution and quality of observations persist. To reduce the influence of clustered observations, and make better use of nonstandard data sources such as provided from satellites, statistical interpolation schemes are used to account for the covariances between observations. The covariances used in the schemes are typically calculated as the mean over a large sample of differences between two short-range forecasts valid at the same time and properly scaled. These approaches, which are known as optimal interpolation methods in meteorology, are essentially versions of the classical Wiener filtering technique. Spatial and temporal variability in the background error covariances to account for changes in the local atmospheric flow conditions are not included in these schemes [1].

A natural extension to 3DVAR schemes is to add time as a fourth dimension to the formulation. While the model dynamics influence the analysis in 3DVAR schemes through the prior estimate, four-dimensional variational analysis (4DVAR) schemes, which are rooted in the adjoint methodology, allow both observations and the model dynamics to influence the analysis throughout a specified time period or window. The model state at the analysis time is the one that best fits the available observations (in the least squares sense) over a chosen time window that extends backwards from the analysis time. Thus, many more observations are able to influence the analysis than are available at a single time, while the resulting analysis, which is the model state at the end of the assimilation time window must, by design, satisfy the model equations. Most operational meteorological forecast centers use either 3DVAR or 4DVAR schemes to create the atmospheric states needed to start the forecast models. The use of variational analysis schemes in operational centers has led to improvements in model forecast skill.

If the deterministic model dynamics used in a 4DVAR scheme exhibit sensitivity to initial conditions, then the error in the prediction grows rapidly and leads to a loss of prediction skill. Nonlinear filtering techniques, based on extensions of the Kalman filter, offer an improved approach to estimating the state of an uncertain dynamical system based on incomplete and imperfect observations distributed in both time and space. The solution to this state estimation problem, which consists of describing the time evolution of the probability density function of the system state, is given by the Kushner-Zakai (K-Z) equation [4]–[14]. The K-Z equation is a stochastic (parabolic) partial differential equation (PDE), which represents an extension of Kolmogorov's forward equation. Unfortunately, the K-Z equation is very difficult to solve except in special cases [15]–[18], requiring the use of alternative approaches to state estimation problem.

While nonlinear filtering methods are an ideal tool for assimilating noisy data into a nonlinear stochastic model, the dimensionality of the meteorological model space, which is of order 10^8–10^9 makes the implementation of these techniques infeasible. Hence one often seeks reduced-rank approximations to the nonlinear filter. One class of approximations that are widely used in meteorology is ensemble Kalman filtering. Various nonlinear filtering methods currently in use are shown in Figure 1.

An early attempt to use Monte Carlo methods in the context of nonlinear filtering is described in [19]. Monte Carlo filters have been independently developed in several disciplines, including economics [20], geophysical sciences [21]–[22], and statistics [23]. A method based on deterministic ensembles is the unscented Kalman filters [24]. An emerging class of methods that are deeply rooted in the

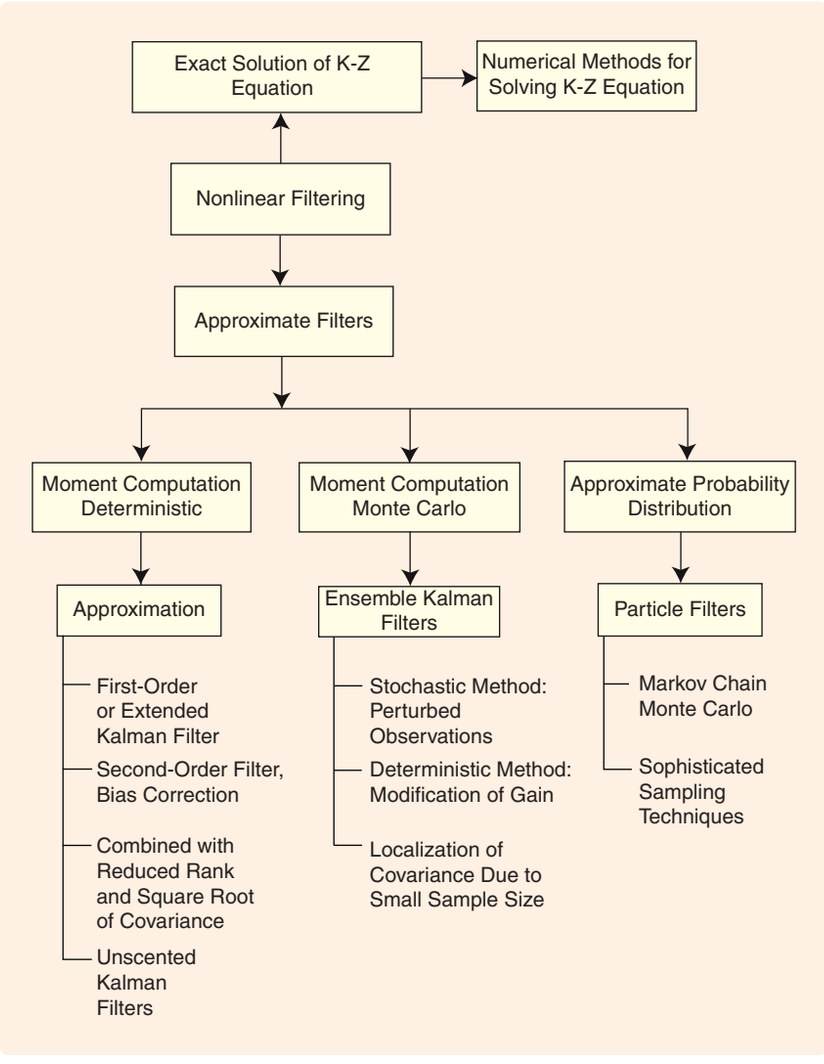

FIGURE 1 A classification of nonlinear filtering methods. Ensemble Kalman filter methods fall into the second of three families of approximate filtering methods, which include approximate deterministic, approximate Monte Carlo, and approximation methods based on the idea of particle filters rooted in advanced sampling techniques. Within ensemble Kalman filters, differences in application arise along three dimensions, the use of perturbed observations, the change of gain without perturbed observations, and localization and stabilization of the covariance calculations resulting from small sample size.

Bayesian framework include the sequential Monte Carlo methods, including particle filters [25], which exploit importance sampling techniques and are becoming an indispensable tool for data assimilation and estimation [26]–[28].

The goal of this article is to survey the growing use of ensemble Kalman filter methods in meteorological data assimilation. We begin with a summary of meteorological models and ensembles and then discuss how observation operators are defined in meteorological data assimilation methods. This discussion is followed by a statement of the data assimilation problem, an overview of the use of traditional Kalman filter methods in meteorology, an exploration of the present state of ensemble Kalman filtering methods, and future opportunities.

MODELS

Meteorological models used to predict future atmospheric states are derived from the primitive equations, which comprise a system of coupled nonlinear PDEs for fluid flow in three-dimensional space. The primitive equations include three momentum equations, a mass balance equation, an equation of state, and the basic thermodynamic relation [29]. These equations describe the time evolution of the basic atmospheric variables, namely, pressure, temperature, wind velocities, and water content, in a three-dimensional layer covering the Earth from the surface to a height of tens of kilometers. Numerical methods for these equations include finite difference, spectral, and finite element methods [1]. Finite-difference methods rely on spatial and time discretizations that conserve properties of the fluid, such as mass and energy. The system of PDEs is converted to a set of algebraic expressions that can be integrated forward in time. In spectral methods, the spatial variations of the physical quantities are captured in a Fourier series with amplitudes that vary in time. Using this orthogonal expansion, the system of PDEs is converted into a system of coupled nonlinear ordinary differential equations (ODEs) that can be integrated forward in time. In finite element methods, a variational framework is used to convert the PDEs into a system of algebraic equations that are integrated numerically.

Consider a three-dimensional grid with n_x, n_y, and n_z points in the x, y, and z directions, respectively, giving rise to $n_g = n_x n_y n_z$ grid points. A fairly small grid of $n_x = 200$, $n_y = 200$, and $n_z = 50$, gives rise to $n_g = 2\times10^6$ grid points. Moreover, at each point, L physical variables, such as pressure and temperature, must be represented. Thus, $n = n_g L$ total unknown variables must be defined. Many models used in meteorology have over 10^8 degrees of freedom. Time in meteorological models is also discretized in units of Δt. Basic consistency and stability conditions dictate the relation between the spatial grid spacings Δx, Δy, and Δz, and the time increment Δt. Regardless of which solution method is selected to convert the set of PDEs into a system of nonlinear algebraic, recurrence, or iterative schemes, the integration process can be represented by

$$x_{k+1} = M(x_k), \tag{1}$$

where $x_k \in R^n$ denotes the real n vector called the model state at time $k = 0, 1, 2, \ldots$, and the n-dimensional Euclidean space R^n is called the model space. The mapping M: $R^n \rightarrow R^n$ denoted by $M(x_k) = (M_1(x_k), M_2(x_k) \ldots M_n(x_k))^T$ is a vector-valued nonlinear function of the vector x_k defining the state transition rule of the model.

In addition to the fact that the primitive equations are approximations to reality, further errors are introduced when these approximate models are converted to (1) resulting from the finite grid length, truncation in the spectral expansion, and finite difference approximations and simplifications. These errors introduce an additional term into the equations modeled as

$$x_{k+1} = M(x_k) + w_{k+1}. \tag{2}$$

When the errors have high frequency, it is useful to consider $w_k \in R^n$ as a sequence of random vectors, in which case (2) defines a stochastic nonlinear or linear model. It is generally assumed that $w_k \in R^n$ has mean $E(w_k) = 0$ and covariance $\mathrm{cov}(w_k) = Q_k$. It is also commonly assumed that w_k is normally distributed, that is, $w_k \sim N(0, Q_k)$.

Ensembles

Small errors in the initial state given to a forecast model can grow rapidly and degrade forecast skill [30]. Since error growth occurs most quickly on the smallest scales, forecast skill is lost most rapidly for small-scale features, such as thunderstorms and rain bands. Forecast skill for large-scale features, such as low pressure systems and jet streams, may extend out to two weeks. Skill beyond two weeks, which typically comes from forcing external to the atmosphere, such as ocean temperature structures associated with El Niño, is often limited to forecasts of above- or below-normal conditions based on past statistical relationships.

Owing to the lack of sufficient observations to perfectly define the model state at each given point in time, Monte Carlo-type ensembles are often used to provide explicit information on forecast uncertainty. An ensemble, which is a group of model forecasts valid over the same time period, is typically created by adding perturbations that sample the probability density function of the initial model state [1]. The statistics of the ensemble forecasts are assumed to be a good estimate of the properties of the future atmospheric probability density function. Since meteorological models are imperfect, multimodel and varied subgrid-scale physics ensembles may be used to help account for model error, and results from these ensembles can yield better forecasts [31]–[32]. The use of ensembles correctly shifts the forecast problem from a deterministic perspective to a probabilistic perspective, which makes the resulting forecasts more amenable to cost-benefit analyses and better decision making [33].

The quality of the ensemble analysis and prediction critically depends on the choice of the initial ensemble. Given this reality, [34]–[36] examine the impact of various choices for the initial ensemble, such as singular vectors, normal modes, perturbed integrations, and random perturbations. It is not clear which approaches are best suited to providing reliable and skillful forecasts for the wide variety of forecast problems that need to be addressed to meet the public's demand for weather information.

OBSERVATION OPERATORS

A key component of atmospheric state estimation is to relate the available observations to the model state variables. While some of the model variables, such as pressure, temperature, relative humidity, and horizontal wind velocities, may be directly observed, other variables are not. These unobserved variables include vertical velocity, liquid water content, ice content, and precipitation particle size and distribution. Indirect information on the model variables, such as provided by satellite radiances, is also available, and is especially useful over the oceans where in situ measurements are impractical or too costly. Forward operators, which are used to relate these observations to the model state variables, often involve complicated physical or empirical relations. For example, reflectivity measured by radars is an empirically derived function of the size and amount of raindrops [37]. On the other hand, the radiated energy measured by satellites in various wavelength bands is related to the skin temperature of the Earth and the composition of the atmosphere in various layers. This energy is related to temperature through Planck's law [38].

Developing and understanding the complex relations between observations and model variables is an important first step in meteorological data assimilation. As an example, let $z \in R^m$ be an m-vector of observations, and let h:$R^n \rightarrow R^m$ be a nonlinear operator that relates the state variable x to the expected value of the observations

$$z_k = h(x_k). \tag{3}$$

When the mapping (3) is linear, the right-hand side of (3) is given by Hx, where $H \in R^{m \times n}$ is an $m \times n$ matrix. Unfortunately, observations are always associated with additive measurement noise, and thus we have

$$z_k = h(x_k) + v_k, \tag{4}$$

where $v_k \in R^m$ is an unbiased random noise vector with $E(v_k) = 0$ and $cov(v_k) = \mathbf{R}_k$. If the observations are biased, then the bias can be removed as part of the analysis procedures. The diagonal entries of the $m \times m$ matrix $\mathbf{R}_k$ are the variances of the observations, which are largely a function of the measuring instruments, whereas the off-diagonal entries are related to the correlations between observations. If

the observational error distributions are uncorrelated, then $\mathbf{R}_k$ is a diagonal matrix.

When the model is unable to represent the conditions sampled by observations, we have to deal with a new form of error called the representative error. An example of this type of error is seen by assuming that a rawinsonde rises into a thunderstorm whose horizontal size is 10 km by 10 km. If the model horizontal grid spacing is greater than 10 km, then the thunderstorm is not represented on the model grid since the storm is sampled by at most one grid point and may occur between grid points. The thunderstorm in this case is a subgrid-scale feature, which cannot be represented by the numerical model. However, even for a model grid spacing of 5 km, the thunderstorm is not represented accurately since the minimum resolvable $2\Delta x$ wavelength is insufficient to accurately describe the complex structures within the thunderstorm in the model integration [1]. Experimental results suggest that ten grid points often are needed to reasonably represent the true amplitude and first horizontal derivatives of a given feature [1]. Thus, a grid spacing of 1 km is required to obtain a reasonable representation of a thunderstorm, although even smaller grid spacing may be required to accurately represent the physical processes within the thunderstorm.

STATEMENT OF THE ASSIMILATION PROBLEM

Consider the model given by (2) with the observations given by (4). Two versions of the data assimilation problem are of interest. In the offline or batch mode, using previously collected data, it is assumed that the dynamical model is deterministic with $w_k \equiv 0$ in (2). Given a set $F_N = \{z_{k_1}, z_{k_2}, z_{k_3}, \ldots, z_{k_N}\}$ of N observations taken at times $0 \le k_1 < k_2 < k_3 < \cdots < k_N$, the goal is to find the initial condition x_0 such that the model solution best describes the observations. This goal is recast as the constrained least squares problem of minimizing

$$f(x_0) = \frac{1}{N}\sum_{i=1}^{N} [z_{k_i} - h(x_{k_i})]^T \mathbf{R}_{k_i}^{-1} [z_{k_i} - h(x_{k_i})],$$

where $\mathbf{R}_{k_i}$ is the covariance of the noise v_{k_i} for $i = 1, 2, \ldots, N$ and x_k evolves according to (2). This problem is solved by optimization methods [3].

For the online or sequential mode, using data as they become available, the model (2) is stochastic, and a set of observations is provided by F_N. The goal is to obtain an unbiased, least squares (minimum variance) estimate

$$\hat{x}_{k_N} = \mathrm{E}[x_{k_N} \mid F_N],$$

and its covariance. When the model and observation operators are linear, and the noise is Gaussian, the solution is given by the Kalman filter equations [39]–[41]. However, in the more realistic situation where the model or observations are nonlinear, the exact solution to this problem involves computing the evolution of the probability density function of the state x_k given the observations. While this problem is conceptually simple and the equation defining the exact solution is given by the K-Z equation [13]–[14], it is infeasible to compute probability distributions in R^n, where $n \approx 10^8$. In view of this difficulty, interest shifts to computing the dynamics of the first two moments, that is, the mean and covariance, of the probability distribution. Even this problem is riddled with moment closure problems, which forces us to settle for only approximate computation of the evolution of the first two moments. The extended Kalman filter is an example of this type of approximation method [3]. Additional approximations and applications of the Kalman filter are found in [4]–[12].

Discussion Points

The 4DVAR method is intimately related to the sequential method based on Kalman filtering. When the model is linear and has no noise, and the observations are a linear function of the state, a sequential version of the 4DVAR method is similar to the Kalman filter equations [3].

The data assimilation problem in meteorology faces challenges coming from two very different directions, namely the inherent nonlinearity of the model or observations and the dimensionality of the state space. While the underlying nonlinearity of the models cannot be altered, a compromise can be made in the selection of the appropriate dimension. This compromise is dictated by the available computing power (measured in megaflops) and the time horizon T over which forecast decisions have to be made. Indeed, we can easily find the largest dimension of the state space for which the data assimilation and subsequent forecast using the best known algorithms can be performed within T hours on a given computer platform that delivers K megaflops.

Two options are available for changing the dimension of the state space. First, we can adjust the size of the three-dimensional grid used in the finite difference approximation of the PDEs that define the model. Second, we can reduce the PDEs to a system of ODEs, called a low-order model, using a Galerkin-type projection method [42]. The order of the reduced-order model can be controlled by a suitable choice of the subspace onto which the projection is obtained. In classical spectral methods, the projection is defined by using the standard Fourier basis functions with trigonometric or spherical harmonics. Model reduction based on empirical orthogonal functions, proper orthogonal decomposition [43], principal interaction patterns [44], and optimal persistent patterns [45] are also possible. Reduced-order models obtained from these bases are not currently exploited in data assimilation.

The fundamental difficulty associated with the nonlinear filtering problem can be illustrated using the following example. Let $x \in R^n$ be a random vector with distribution given by $p_x(x)$, whose mean is m_x and covariance is P_x. Let $M: R^n \to R^n$ and $Y = M(x)$. Knowing M, $p_x(x)$, m_x, and

P_x, we wish to compute the distribution $p_Y(y)$, m_y, and P_y. A basic result from probability theory states that $P_Y(y)$ is given by [46]

$$P_Y(y) = P_x(M^{-1}(y))|D_{M^{-1}}(y)|, \tag{5}$$

where M^{-1} is the inverse of M and $|D_{M^{-1}}(y)|$ is the absolute value of the Jacobian determinant of M^{-1} evaluated at y. Since the computation of $M^{-1}(y)$ is not feasible, $P_Y(y)$ is unknown. Given this reality, we consider estimates of m_y and P_Y. However, even this calculation is difficult since

$$m_y = \mathrm{E}(y) = \mathrm{E}(M(x)) \neq M(\mathrm{E}(x)) = M(m_x),$$

due to the fact that M is nonlinear. We therefore seek an approximation obtained by expanding $M(x)$ in a small neighborhood of m_x using, for example, a first-order Taylor expansion. Accordingly, from

$$M(x) \approx M(m_x) + D_m(m_x)(x - m_x)$$

we obtain

$$m_y = \mathrm{E}(y) = \mathrm{E}(M(x)) \approx M(m_x), \tag{6}$$

and

$$P_y = \mathrm{cov}(y) = \mathrm{E}[(y - m_y)(y - m_y)^T] \approx D_m(m_x)P_xD_m^T(m_x). \tag{7}$$

The first-order filter, that is, the extended Kalman filter, computes the approximate forecast and its covariance using (6) and (7). By using second-order terms in the Taylor expansion, we can obtain the second-order filter, and so on [3], [4], [9].

While (7) is obtained using only a first-order approximation, evaluating the right-hand side of (7) requires $O(n^3)$ operations, where n is the number of degrees of freedom. Using a conservative estimate of $n = 10^6$ for meteorological models, this step requires $O(10^{18})$ operations. Thus, on a computer that can deliver 10^{12} flops (1 teraflop), completing the calculations on the right-hand side of (7) requires $O(10^6)$ seconds or approximately 12 days. In addition, over 4 TB of memory is needed to store the covariance matrices. Thus, the computation of the approximate but full rank covariance takes considerable time and large amounts of computer memory for meteorological problems. To circumvent this difficulty, reduced-rank approximations are often made to the forecast covariance. Several ways to realize the rank reduction are shown in [3].

Besides computation time, yet another hidden challenge in computing P_y using (7) stems from the "bad conditioning" of the matrix P_x. It is well known [3] that if the condition number of P_x is large, then numerical error growth can occur due to round off. By way of controlling this undesirable instability, the matrix update in (7) is often recast in terms of square root matrices. Thus, given the square root factorizations $P_x = S_xS_x^T$ and $P_y = S_yS_y^T$, (7) is equivalent to

$$S_y \approx D_m(m_x)S_x. \tag{8}$$

Since n is of the order 10^8–10^9, the computation of S_y using (8) is infeasible. Hence, a "reduced rank" implementation of (8) is sought by approximating the $n \times n$ matrix S_x by an $n \times q$ matrix with $q << n$. This approximation results in an $n \times q$ approximation to S_y. Square root versions of reduced-rank implementations enjoy the benefits of increased numerical stability while keeping the computation time under control [47].

The need for the Jacobian $D_M(x)$ in (7) for the Kalman filter, or the adjoint operator in the 4DVAR approach, can become problematic or extremely complex to code, however, when the forecast model operator contains parameterized discontinuities resulting from phase changes such as from water to ice [48]–[49]. This computational complexity is avoided by using the ensemble Kalman filter, illustrating another reason why the use of the ensemble Kalman filter in meteorology is growing steadily.

While numerous ideas are available for implementing extended Kalman filters that involve combinations of the first-order or second-order approximations, square root form, and reduced-rank approximations, in this article our goal is to provide a comprehensive review of the state of the art in the particular form of reduced-rank filters resulting from the use of Monte Carlo type ensembles. Before embarking on a discussion of the ensemble filters, we undertake a succinct review of the traditional Kalman filter in meteorology.

PRE-ENSEMBLE ERA, TRADITIONAL KALMAN FILTER APPLICATIONS

An early application of online or sequential statistical estimation based on Kalman filtering in meteorology is considered in [50] where the estimation problem is analyzed related to the one-dimensional linearized shallow water model. This analysis is then extended to the linearized shallow water model in two dimensions in [51].

Excessive computational requirements present major impediments to using Kalman filtering methods for large-scale operational forecasting problems. This situation provides the impetus for developing computationally feasible suboptimal filters, a trend that continues to this day. These ventures can be categorized as covariance modeling, simplification of model dynamics, local approximations, and the use of steady-state approximations, where possible. In the area of covariance modeling, one idea is to use only the advective part of the model dynamics $M(\cdot)$ in computing $D_m(\cdot)$, in (7) while another technique is to use a two-level

system of grids, where the forecast is derived using the finer grid but the covariance is updated using the coarser grid and interpolated to the finer grid. Yet another idea is to assume zero correlation between each pair of grid points separated by a distance larger than a prespecified threshold. When the steady state can be computed, the limiting values of the covariance and the filter gain are used to sequentially estimate the state. In this case, the estimator is equivalent to classical Wiener filtering. Details on these and related issues are given in [51]–[54], while an overview of estimation theory and data assimilation for meteorology are given in [55] and [56]. The spectral properties of Kalman filters are explored in [57], and application of Kalman filtering to Burgers' nonlinear advection equation is found in [57] and [58].

ENSEMBLE FILTERING FRAMEWORK

The general principles that underlie many algorithms for reduced-rank ensemble filtering are discussed in [3], [22], and [59]. An overview of these principles begins by assuming that, at time $k = 0$, $\hat{x}_0$ is the initial estimate of the unknown atmospheric state and $\hat{S}_0 = \hat{S}_0(1{:}q) \in R^{n\times q}$ is the rank q square root of the covariance $\hat{P}_0$ of the estimate $\hat{x}_0$. The matrix $\hat{S}_0$ is called a full-rank or reduced-rank square root depending on whether $q = n$ or $q < n$.

We now describe the underpinnings of a framework for ensemble filtering in the following steps.

Step 0: Create an Initial Ensemble

Given $(\hat{x}_0, \hat{S}_0)$, compute the initial (one-sided) ensemble of size N, where $1 < N < n$. The ith member of the initial ensemble is given by

$$\hat{\xi}_0(i) = \hat{x}_0 + \hat{S}_0 y_0(i), \tag{9}$$

where $y_0(i) \in R^q$, $y_0(i) \sim N(0, I)$ for $1 \le i \le N$, and $q < N$. Note that $\hat{x}_0$ is the mean of this initial ensemble $\{\hat{\xi}_0(i){:}1 \le i \le N\}$. Let

$$\hat{a}_0(i) = \hat{\xi}_0(i) - \hat{x}_0 = \hat{S}_0 y(i) \tag{10}$$

for $1 \le i \le N$ denote the anomaly, which is the difference between the ith ensemble member $\hat{\xi}_0(i)$ and the ensemble mean. Let

$$\hat{A}_0 = [\hat{a}_0(1), \ldots, \hat{a}_0(N)] \in R^{n\times N},$$

denote the $n\times N$ matrix of anomaly vectors. In ensemble filtering, information about the state and its covariance is extracted from the pair $(\hat{x}_0, \hat{A}_0)$. Since $\hat{x}_0$ is the ensemble mean, it can be verified that $\hat{A}_0 \mathbf{1} = 0$, where $\mathbf{1} = [1, \ldots, 1]^T \in R^N$, and

$$\frac{1}{N-1}\hat{A}_0\hat{A}_0^T = \hat{S}_0(\hat{S}_0)^T.$$

That is, $\hat{S}_0 = (N-1)^{-1/2}\hat{A}_0$, which is the scaled anomaly matrix. This relation shows a natural relation between ensemble methods and reduced-rank filtering.

Instead of the one-sided ensemble generated using (9), it is easy to generate a symmetric ensemble of size $2N$ using

$$\hat{\xi}_0(\pm i) = \hat{x}_0 \pm \hat{S}_0 y(i),\ 1 \le i \le N.$$

For simplicity we consider only a one-sided ensemble, although the discussion readily carries over to symmetric ensembles as well. Furthermore, the relation $\hat{A}_0\mathbf{1} = 0$ indicates that the initial ensemble is centered at its mean value. If $\hat{A}_0\mathbf{1} \ne 0$, then the ensemble center differs from the mean, and the ensemble is uncentered. A discussion of the impact of centered versus uncentered ensembles on data assimilation is given in [60]–[61].

Step 1: Create a Forecast Ensemble

At time k, assume inductively that we are given the pair $(\hat{x}_k, \hat{A}_k)$, where $\hat{A}_k\mathbf{1} = 0$. For $1 \le i \le N$, compute the deterministic forecast ensemble

$$\xi^f_{k+1}(i) = M(\hat{\xi}_k(i)) = M(\hat{x}_k + \hat{a}_k(i)), \tag{11}$$

using the nonlinear model $M(\cdot)$, where $\hat{a}_k(i)$ is the ith column of the anomaly matrix $\hat{A}_k$. Expanding (11) in a first-order Taylor series yields

$$\xi^f_{k+1}(i) \approx M(\hat{x}_k) + D_M(\hat{x}_k)\hat{a}_k(i),$$

where $D_M(x)$ is the Jacobian of $M(x)$ with respect to x. Combining this expression with the relation $\hat{A}_k\mathbf{1} = 0$ indicates that the sample average of this forecast ensemble is given by

$$x^f_{k+1} = M(\hat{x}_k).$$

Define A^f_{k+1} as the forecast anomaly matrix whose ith column $a^f_{k+1}(i)$ is given by

$$a^f_{k+1}(i) = \xi^f_{k+1}(i) - x^f_{k+1}(i), 1 \le i \le N.$$

Then, the pair (x^f_{k+1}, A^f_{k+1}) constitutes the forecast ensemble at time $k+1$. It can be verified that $\hat{A}^f_{k+1}\mathbf{1} = 0$ and

$$(N-1)^{-1/2}A^f_{k+1} = S^f_{k+1},$$

which is the rank q square root of P^f_{k+1}.

The deterministic ensemble generated by (11) does not account for potential model errors, which can be large in meteorological models. One way to account for model errors is to define the one-sided forecast ensemble

$$\xi^f_{k+1}(i) = M(\hat{\xi}_k(i)) + w_{k+1}, \tag{12}$$

where $w_{k+1} \sim N(0, Q_{k+1})$ and Q_{k+1} is the covariance of the model errors. For realistic applications, the model error covariance is not known and is instead estimated. We can also consider a symmetric version of (12).

Step 2: Data Assimilation to Create a New Analysis Ensemble

Given the ensemble forecast (x^f_{k+1}, A^f_{k+1}) and the new observation $z_{k+1} \in R^m$, where it is tacitly assumed that

$$z_k = H_k x_k + v_k,$$

with $H \in R^{m \times n}$ and $v_k \sim N(0, R_k)$, we compute the new analysis ensemble $(\hat{x}_{k+1}, \hat{A}_{k+1})$. This cycle between steps 1 and 2 is repeated for as long as new observations can be assimilated. The available algorithms for ensemble filtering, which essentially differ in the details of this data assimilation step, can be classified into two groups, stochastic and deterministic methods.

Stochastic Methods

In this type of method the ith member of the analysis ensemble at time $k+1$ is computed as

$$\hat{\xi}_{k+1}(i) = \xi^f_{k+1}(i) + K_{k+1}[z_{k+1}(i) - H_{k+1}\xi^f_{k+1}(i)], \quad (13)$$

where K is the Kalman gain in the square root form [3], [22], [59] and $z_{k+1}(i)$ is the ith perturbed observation given by

$$z_{k+1}(i) = z_{k+1} + v_{k+1}(i),$$

where $v_{k+1}(i) \sim N(0, R_{k+1})$.

In the early application of ensemble filtering in geophysical problems [21], the same observation z_{k+1} is assimilated in each of the forecast ensemble members. This approach results in an ensemble whose covariance and hence spread is less than the theoretical values dictated by the Kalman filter. As the filter evolves in time, this reduction in covariance leads to the collapse of the ensemble, and the spread shrinks rapidly. As noted in [62] and [63], the introduction of perturbed observations compensates for this deficiency. Indeed, it can be verified [3], [62], [63] that the covariance of the analysis ensemble generated using (13) matches the theoretical value given by the Kalman filter as $N \to \infty$.

While the use of perturbed observations improves the performance of the ensemble Kalman filter, side effects can occur due to sampling errors, especially when N is small. Since ensemble sizes used in meteorological applications are typically small, alternative strategies are needed to compensate for underestimation of the analysis covariance. Another stochastic method for implementing ensemble filtering is the singular evolutive interpolated Kalman (SEIK) filter as described in [64] and applied in [65].

Deterministic Methods

Given $A^f_{k+1} \in R^{n \times N}$, the goal of this method is to find a transformation T such that

$$\hat{A}_{k+1} = T(A^f_{k+1}), \quad (14)$$

where $\hat{A}_{k+1}$ satisfies $\hat{A}_{k+1}\mathbf{1} = 0$ and

$$\frac{1}{N-1}\hat{A}_{k+1}(\hat{A}_{k+1})^T \to P^f_{k+1},$$

as N increases. Algorithms of this type realize $T(\cdot)$ as a linear transformation. Accordingly, $\hat{A}_{k+1}$ is obtained from A^f_{k+1} by either a left multiplication by a matrix $T_L \in R^{n \times n}$ or by a right multiplication by $T_R \in R^{N \times N}$. Thus, the ensemble transform Kalman filter (ETKF) as described by [66] computes

$$\hat{A}_{k+1} = A^f_{k+1}T_R, \quad (15)$$

whereas the ensemble adjustment Kalman filter (EAKF) as described by [67] computes

$$\hat{A}_{k+1} = T_L A^f_{k+1}. \quad (16)$$

Since $T_R \in R^{N \times N}$ and $T_L \in R^{n \times n}$, ETKF requires fewer computations compared to EAKF unless EAKF is implemented in the sequential least squares framework, in which case the computational requirements are the same. However, $\hat{A}_{k+1}$ generated by EAKF automatically satisfies $\hat{A}_{k+1}\mathbf{1} = 0$. It is thus necessary to impose additional conditions on T_L to ensure that $\hat{A}_{k+1}\mathbf{1} = 0$.

These deterministic methods and related issues are discussed in [60], [61], and [68]. One unifying theme among deterministic methods is that they all exploit combinations of ideas involving the square root of covariance matrices [68]–[69] along with the reduced-rank approximations resulting from the small ensemble size. Ideas relating to matrix factorizations that are crucial in this approach are discussed in [70] and [71]. Another class of deterministic ensemble filters, called singular evolutive extended Kalman (SEEK) filters, is described in [72]. In this approach, the N columns of the anomaly matrices $\hat{A}_k$ and A^f_k span the N-dimensional subspace of the full n-dimensional error space corresponding to the full covariance matrix. The quality of the ensemble filters with respect to the decomposition of the complete error space is discussed in [73] and [74]. Several variations of the reduced rank filter implementations are described in [75] and [76].

Localization and Stabilization

Owing to the computational burden associated with running large, nonlinear meteorological models, ensemble sizes in the range of 50–100 are typical in meteorological applications. Estimation of the $n \times n$ (with $n \approx 10^7$) forecast and analysis covariance matrices using only $N(\approx 100)$

The goal of this article is to survey the growing use of ensemble Kalman filter methods in meteorological data assimilation.

samples can lead to errors in the estimates of the off-diagonal entries of these matrices. The source of these errors can be seen by considering two correlated normal random variables with a correlation ρ, and recognizing that an estimate $\hat{\rho}$ of ρ using N ensemble members has the variance

$$\overline{(\rho - \hat{\rho})^2} = \frac{1}{N}(1 - \rho^2)^2. \tag{17}$$

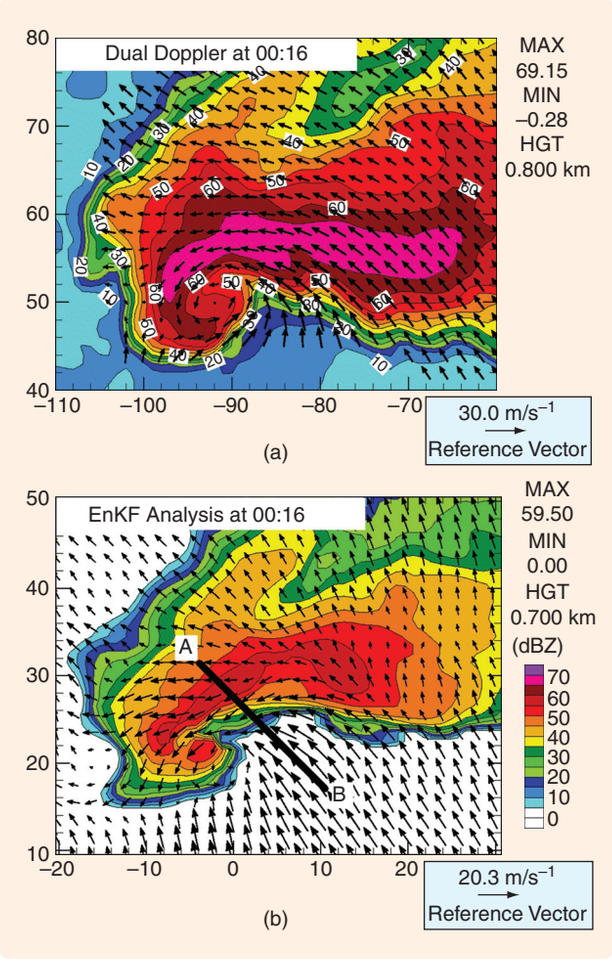

FIGURE 2 Comparison of (a) observed dual-Doppler analysis and (b) ensemble Kalman filter analysis of reflectivity and wind data from a supercell thunderstorm at 0016 UTC 30 May 2004 over central Oklahoma. The plots are from a horizontal slice at about 750 m above ground level of radar reflectivity (dBZ) and horizontal winds (vectors). The good agreement between the dual-Doppler and ensemble Kalman filter analysis indicates that the filter is successful at inserting these thunderstorms into numerical models. The ensemble Kalman filter assimilates only the observed reflectivity and horizontal wind velocity measurements. (Courtesy of Kristin Kuhlman, Louis Wicker, and David Dowell.)

The identity (17) indicates that accurate estimation of smaller correlations requires hundreds and even thousands of ensemble members when using large, nonlinear models [39], [63]. Since such large ensembles are computationally impractical, localization is used in which only observations within a selected distance r from the location of interest are used [77]. This class of ideas is predicated on the fact that grid points that are separated by a larger distance are much less correlated than those that are in a small neighborhood. The Kalman gain is multiplied by additional terms that decrease smoothly to zero as the distance between the observation and the grid point approach r [78]–[79]. The value of r chosen for a given observation needs to be carefully determined through systematic experimentation since its value clearly influences the quality of the resulting assimilation [80]. Efficient localization approaches are also considered in spectral space [81] and wavelet space [82].

Another effect of small sample size, perhaps combined with the effects of imperfect initial ensemble perturbation methods and model error, is that the ensemble spread tends to underestimate the true forecast error. This underestimation can lead to filter divergence, in which the observation is so many standard deviations away from the ensemble mean forecast that it is effectively ignored in the ensemble analysis step. To address this problem, stabilization methods are used to artificially inflate the ensemble spread [83]–[84] or adaptively correct the inflation [85] to stabilize the estimates of the mean of the analysis ensemble.

METEOROLOGICAL APPLICATIONS OF THE ENSEMBLE KALMAN FILTER

The inherent constraints involved in making numerical forecasts of the weather, including limited and irregular observations and the computational expense of running meteorological models, has led to the development and application of several unique applications of the ensemble Kalman filter. A few of these applications are summarized below.

Adaptive Observations

It is well known that the atmosphere is a chaotic system that is very sensitive to small changes in initial conditions [30]. These initially small errors may grow rapidly and eventually dominate the numerical forecast, leading to the loss of forecast skill. Unfortunately, observations are expensive and therefore limited in number and coverage. One of several approaches to making the maximum use of any additional observations is the ETKF, which uses the forecast ensemble to generate approximations to the

Parallel Ensemble Kalman Filtering

Ensemble Kalman filtering readily lends itself to parallel implementation. One straightforward approach is to assign the task of evolving each ensemble member to one processor in a cluster of distributed memory processors [101]–[102]. This approach is natural when the model is implemented in a serial code suitable for running on vector supercomputers and the number N of ensemble members is less than the number of processors. While this approach speeds up the nonlinear forecast phase of the ensemble, the data assimilation phase requires output from all the members of the ensemble to be available at a specific forecast time. This requires each processor to send the forecast it just computed to a master processor so that the latter can assimilate the observation. In addition to incurring excessive communication cost, since no single processor has enough memory to store all the members of the ensemble, in such an implementation, we will be forced to store the forecast ensemble members in a secondary storage device (say, a large disk) attached to the mater processor. But then the time required to repeatedly transfer the data between the main and the secondary memory would further exacerbate overall computation time.

The sequential nature of ensemble Kalman filtering algorithms originally led to concern that this approach may not be useful for large meteorological problems run on massively parallel computers. However, parallelism in both the forecast and the data assimilation phase can be obtained using a domain decomposition approach [103]. Assume that we have P processors. In this approach the spatial domain over which the model operates is naturally divided into P subdomains, and each subdomain is allocated to one processor. The state vector is also naturally partitioned into P contiguous parts, and each processor computes only the portion allocated to it. The same model code resides in each processor but each processor computes the same portion of the state vector allocated to it on all of the N ensemble members. While this technique would require communications between processors sharing domain boundaries, this burden is often considerably smaller compared to sending the entire state vector to the master processor. Since each processor has a portion of all the ensemble members, the data assimilation phase is also implemented in parallel.

A demonstration of this idea using a two-layer spectral shallow water model is given in [97]. In this study, synthetic data are assimilated using a parallel implementation of the ensemble Kalman filter. In particular, the implementation uses the method of domain decomposition described above using the standard message-passing interface (MPI) [103]. Similar parallel formulations of ensemble filters over local regions are developed in [104] and [105], while another approach divides the observations into batches that are assimilated sequentially [106]. Parallel ensemble Kalman filter algorithms have been shown to yield results equivalent to the sequential algorithm when the forward observation operator $h(x)$ is linear and qualitatively similar results when the forward observation operator is nonlinear [107].

prediction error covariance matrices associated with a variety of observational deployment strategies [86]. This approach is used in [86] to develop adaptive observation strategies in which additional observations are taken in selected regions to minimize the subsequent forecast error. Owing to the time required to reach the desired location and take the observations, this approach focuses on atmospheric features that are crossing the Northern Hemisphere oceans in order to improve one–three-day forecasts over either North America or Europe. Results from actual targeting campaigns, in which airplanes fly over the oceans and deploy dropsondes to take observations as they fall from the airplane to the ocean surface, indicate that forecasts are improved on 70% of the days in which targeted observations are made and with an average forecast improvement of 10% when evaluated using root-mean-square errors and anomaly correlations [87].

Multimodel Ensembles

One of the advantages of the ensemble Kalman filter over 4DVAR schemes is the relatively easy inclusion of multimodel ensembles in the ensemble Kalman filter. Model error is a persistent problem in numerical weather prediction, since many physical processes operate at scales below those that the model can represent and hence need to be parameterized. Ensemble forecasts are found to be more skillful when several different numerical models or different physical process parameterizations are used as compared to ensembles that use only a single model [31]–[32]. This result leads naturally to the use of multimodel ensembles within data assimilation methods, such as the ensemble Kalman filter. As found with model forecasts, the results from the ensemble Kalman filter are improved when multimodels or different physical process parameterization schemes are used [88]–[90].

Recent Applications

The assimilation of Doppler radar radial-velocity and reflectivity data into cloud-scale numerical weather prediction models is an active area of research that may lead to improved thunderstorm forecasts and increased lead times for severe weather warnings. Longer warning lead times are expected to help save lives, reduce injuries, positively impact air traffic and surface transportation routing, and provide improved local flood warnings. The current national Doppler radar network provides volumetric scans of individual thunderstorms every five minutes. Using an ensemble Kalman filter, the observed reflectivity and radial velocity can be assimilated into a numerical model (Figure 2). Results indicate that the filter can successfully retrieve the

The meteorological community has made great strides in adapting filtering methods to their problems.

unobserved variables, such as temperature and the full three-dimensional wind field [80]. In contrast, model simulations without data assimilation must create the thunderstorm in an ad hoc manner using a warm bubble, are not able to quickly develop the observed storm characteristics, and the model thunderstorms often diverge from observations. Thus, data assimilation is a key element in producing reasonable predictions of observed thunderstorms.

While the use of ensemble Kalman filters to assimilate Doppler radar data into numerical models to make short-term predictions of thunderstorms is an exciting application, numerous challenges exist. The computational demands of running 50–100 ensemble members using models with 1-km grid spacing or less are huge, making real-time testing and evaluation impossible without access to large supercomputers. Rapid scanning radars, such as phased array radars [91], are capable of producing volumetric scans every 30 s, yielding observation counts over a 15-min time window of $\sim 10^9$ from just three radars. For these data the number of m observations likely will exceed the dimensionality n of the numerical model, suggesting that new filter designs will be needed for computational efficiency. Methods to create an initial ensemble for thunderstorm-scale filtering also need to be more fully developed. Other recent applications of the ensemble Kalman filter involve the use of standard surface observations [92], accounting for model error in the ensemble Kalman filter [93], incorporating constraints into the ensemble Kalman filter for water budget estimation [94], the estimation of snow water equivalent [95], and land data assimilation [96].

CONCLUSIONS

The application of nonlinear filtering techniques to meteorology has helped to unify several subdisciplines including ensemble methods, probabilistic forecasting methods, data assimilation, and adaptive observations. This unification in turn has helped to develop new approximate methods for filtering in large-scale problems such as ensemble filtering. In this article, we have provided an overview of the ensemble filtering approach designed to tackle the high-dimensional problems of interest to the meteorological community. Unfortunately, the current ensemble methods fail to track the evolution of the system states that exhibit multiple modes with transitions between these modes [98]–[99]. It is shown in [100] that in such cases sampling from the posterior distribution to create the analysis ensemble yields good results. This idea belongs to the class of particle filters that lies at the heart of the general sequential Monte Carlo methods. Competing ideas involving unscented filters [24] are gaining increasing attention, and only recently have made entry into the meteorology community. Further, ensemble filtering is a natural candidate for parallel implementation; see "Parallel Ensemble Kalman Filtering" for details. We expect future research to involve comparative analyses of the performance of these new approaches.

The meteorological community has made great strides in adapting filtering methods to their problems. Interesting applications, such as targeting for adaptive observations and the use of multiple models within ensembles, highlight the unique contributions this community is making to the filtering literature. The large dimension of the meteorological data assimilation and filtering problem will continue to be challenging and likely lead to more novel applications and improvements to filtering approaches.

ACKNOWLEDGMENTS

The authors greatly appreciate the helpful and constructive comments provided by three anonymous reviewers and colleagues John Lewis, Qin Xu, and Jeffrey Anderson that led to an improved manuscript.

AUTHOR INFORMATION

S. Lakshmivarahan (varahan@ou.edu) received the B.S. in 1964 in physics from the MDT Hindu College, Madras University and the B.E., M.E., and Ph.D. from the Department of Electrical Engineering, Indian Institute of Science, Bangalore, in 1967, 1969, and 1973, respectively. He joined the School of Electrical Engineering and Computer Science at the University of Oklahoma, Norman, in 1978 where he is currently a George Lynn Cross Research Professor at the School of Computer Science. From 1973 to 1978 he held a faculty position at the Department of Computer Science, Indian Institute of Technology, Madras, and postdoctoral visiting positions at the Department of Applied Mathematics at Brown University and Yale University. He is the author of *Learning Algorithms: Theory and Applications* (Springer Verlag, 1982) and a coauthor of *Analysis and Design of Parallel Algorithms* (McGraw-Hill, 1990), *Parallel Computing using the Prefix Problem*, (Oxford University Press, 1994), and *Dynamic Data Assimilation* (Cambridge University Press, 2006). He is a Fellow of the IEEE and the ACM. His interests include parallel processing, computational sciences, and computational finance. He can be contacted at the School of Computer Science, University of Oklahoma, 200 Felgar Street, Room 146, Norman, OK 73019-6151 USA.

David J. Stensrud received the B.S. in 1983 in mathematics and meteorology from the University of Wisconsin-Madison, and the M.S. and the Ph.D. in meteorology from Pennsylvania State University in 1985 and 1992, respectively. He joined the National Severe Storms Laboratory of the National Oceanic and Atmospheric Administration in 1986 as a research meteorologist and presently leads the Forecast Research and Development Division. He is an affiliate professor in the School of Meteorology at the University of Oklahoma and was the editor of the American Meteorological Society's journal *Weather and Forecasting* (1999–2006). He was awarded the Clarence Leroy Meisinger award from the American Meteorological Society in 1998. He is author of *Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models* (Cambridge University Press, 2007). His interests include short-range ensemble forecasting, data assimilation, and the prediction of severe weather events.

REFERENCES

[1] E. Kalnay, *Atmospheric Modeling, Data Assimilation, and Predictability*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[2] R. Daley, *Atmospheric Data Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

[3] J. M. Lewis, S. Lakshmivarahan, and S. Dhall, *Dynamic Data Assimilation: A Least Squares Approach*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[4] R. S. Bucy and P. D. Joseph, *Filtering for Stochastic Processes with Applications to Guidance*. Hoboken, NJ: Interscience, 1961.

[5] A. H. Jazwinskii, *Stochastic Processes and Filtering Theory*. New York: Academic, 1970.

[6] A. Gelb, *Applied Optimal Estimation*. Boston, MA: MIT Press, 1974.

[7] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice Hall, 1979.

[8] P. S. Maybeck, *Stochastic Models: Estimation and Control*. New York: Academic, vol. 1, 1979 and vol. 2, 1982.

[9] R. S. Bucy, *Lectures on Discrete Time Filtering*. New York: Springer-Verlag, 1994.

[10] D. E. Catlin, *Estimation Control and Discrete Kalman Filter*. New York: Springer-Verlag, 1989.

[11] V. Krishnan, *Nonlinear Filtering and Smoothing*. Hoboken, NJ: Wiley, 1984.

[12] G. Kallianpur, *Stochastic Filtering Theory*. New York: Springer-Verlag, 1980.

[13] H. J. Kushner, "On the differential equations satisfied by the conditional probability densities of Markov processes with applications," *SIAM J. Control*, vol. 2, pp. 106–119, 1962.

[14] M. Zakai, "On optimal filtering of diffusion processes," *Zietschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 11, pp. 230–243, 1969.

[15] V. Benes, "Exact finite dimensional filter for certain diffusions with nonlinear drift," *Stochastics*, vol. 5, pp. 65–92, 1981.

[16] F. E. Daum, "New exact nonlinear filters," in *Bayesian Analysis of Time Series and Dynamic Models*, J. C. Spall, Ed. New York: Marcel Dekker, 1988, pp. 199–226.

[17] J. F. Bennaton, "Discrete Galerkin approximations to nonlinear filtering solution," *J. Math. Anal. Appl*., vol. 110, pp. 364–383, 1985.

[18] F. LeGland, "Time discretization of nonlinear filtering equations," in *Proc. 28th Decision and Control Conf*., 1989, pp. 2601–2606.

[19] J. E. Handschin and D. Q. Mayne, "Monte Carlo techniques to estimate the conditional expectation in multi-stage nonlinear filtering," *Int. J. Control*, vol. 9, pp. 547–559, 1969.

[20] H. Tanizaki, *Nonlinear Filters: Estimation and Applications*. New York: Springer-Verlag, 1996.

[21] G. Evensen, "Sequential data assimilation with nonlinear quasi-geostrophic model using Monte Carlo methods for forecast error statistics," *J. Geophys. Res*., vol. 99, pp. 10143–10162, 1994.

[22] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*. New York: Springer-Verlag, 2007.

[23] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *J. Comput. Graph. Stat*., vol. 5, pp. 1–25, 1996.

[24] S. Julier, J. Uhlmann, and H. F. Durrant-Whyte, "A new method for the nonlinear transformation of mean and covariances in filters and estimators," *IEEE Trans. Automat. Contr*., vol. 45, pp. 477–482, 2002.

[25] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Proc. Inst. Elec. Eng., F*, vol. 140, pp. 107–113, 1993.

[26] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond Kalman Filter: Particle Filters and Tracking Applications*. Boston, MA: Artech House, 2004.

[27] S. Haykim, Ed., *Kalman Filtering and Neural Networks*. Hoboken, NJ: Wiley, 2001.

[28] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.

[29] J. R. Holton, *An Introduction to Dynamic Meteorology*. New York: Academic, 2004.

[30] E. Lorenz, "Deterministic nonperiodic flow," *J. Atmos. Sci*., vol. 20, pp. 130–141, 1963.

[31] F. Atger, "The skill of ensemble prediction systems," *Month. Weather Rev*., vol. 127, pp. 1941–1953, 1999.

[32] D. J. Stensrud, J.-W. Bao, and T. T. Warner, "Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems," *Month. Weather Rev*., vol. 128, pp. 2077–2107, 2000.

[33] A. H. Murphy and R. L. Winkler, "Probabilistic temperature forecasts: The case for an operational program," *Bull. Amer. Meteorol. Soc*., vol. 60, pp. 12–19, 1979.

[34] J. L. Anderson, "Selection of initial conditions for ensemble forecasts in a simple perfect model framework," *J. Atmos. Sci*., vol. 53, pp. 22–36, 1996.

[35] M. Zupanski, S. J. Fletcher, I. M. Navon, B. Uzunoglu, R. P. Heikes, D. A. Randall, T. D. Ringler, and D. Daescu, "Initiation of ensemble data assimilation," *Tellus A*, vol. 58, pp. 159–170, 2006.

[36] N. E. Bowler, "Comparison of error breeding, singular vectors, random perturbations and ensemble Kalman filter perturbation strategies on a simple model," *Tellus A*, vol. 58, pp. 538–548, 2006.

[37] R. J. Doviak and D. S. Zrnic, *Doppler Radar and Weather Observations*. Mineola, NY: Dover, 1993.

[38] L. Garand, D. S. Turner, C. Chouinard, and J. Hallé, "A physical formulation of atmospheric transmittances for the massive assimilation of satellite infrared radiances," *J. Appl. Meteorol*., vol. 38, pp. 541–554, 1999.

[39] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Eng. D*, vol. 82, pp. 35–45, 1960.

[40] R. E. Kalman, and R. S. Bucy, "New results in linear filtering and prediction theory," *Trans. ASME, J. Basic Eng. D*, vol. 83, pp. 95–108, 1961.

[41] T. Kailath, "A view of three decades of linear filtering theory," *IEEE Trans. Inform. Theory*, vol. 20, pp. 146–181, 1974.

[42] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, *Spectral Methods: Fundamentals in Single Domain*. New York: Springer-Verlag, 2006.

[43] P. Holmes, J. L. Lumley, and G. Berkooz, *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*. Cambridge, U.K.: Cambridge Univ. Press, 1996.

[44] K. Hasselmann, "PIPs and POPs: Reduction of complex dynamical systems using principal interaction and oscillation patterns," *J. Geophys. Res*., vol. 93(D9), pp. 11015–11021, 1988.

[45] T. DelSole, "Optimally persistent patterns in time-varying fields," *J. Atmos. Sci*., vol. 58, pp. 1341–1356, 2001.

[46] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York: McGraw Hill, 1984.

[47] G. J. Bierman, *Factorization Methods for Discrete Sequential Estimation*. New York: Academic, 1977.

[48] Q. Xu, "Generalized adjoint for physical processes with parameterized discontinuities—Part I: Basic issues and heuristic examples," *J. Atmos. Sci*., vol. 53, pp. 1123–1142, 1996.

[49] Q. Xu and J. Gao, "Generalized adjoint for physical processes with parameterized discontinuities—Part VI: Minimization problems in multidimensional space," *J. Atmos. Sci*., vol. 56, pp. 994–1002, 1999.

[50] S. E. Cohn, M. Ghil, and E. Isaacson, "Optimal interpolation and Kalman filter," in *Proc. 5th Conf. Numerical Weather Prediction*, 1981, pp. 36–42.

[51] S. E. Cohn and D. F. Parrish, "The behavior of the forecast error covariance for a Kalman filter in two dimensions," *Month. Weather Rev*., vol. 119, pp. 1757–1785, 1991.

[52] S. E. Cohn, A. Da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, "Assessing the effects of data selection with DAO physical-space statistical analysis system," *Month. Weather Rev*., vol. 126, pp. 2913–2926, 1998.

[53] R. Todling and S. E. Cohn, "Suboptimal schemes for atmospheric data assimilation based on the Kalman filter," *Month. Weather Rev.*, vol. 122, pp. 2530–2557, 1994.
[54] D. P. Dee, "Simplification of the Kalman filter for meteorological data assimilation," *Q. J. R. Meteorol. Soc.*, vol. 117, pp. 365–384, 1991.
[55] S. E. Cohn, "An introduction to estimation theory," *J. Meteorol. Soc. Jpn.*, vol. 75, pp. 257–288, 1997.
[56] M. Ghil and P. Malanotte-Rizzoli, "Data assimilation in meteorology and oceanography," *Adv. Geophys.*, vol. 33, pp. 141–265, 1991.
[57] R. Dayley and R. Menard, "Spectral characteristics of Kalman filter systems for atmospheric data assimilation systems," *Month. Weather Rev.*, vol. 121, pp. 1554–1565, 1993.
[58] M. Verlaan and A. W. Heemink, "Nonlinearity in data assimilation: A practical method for analysis," *Month. Weather Rev.*, vol. 129, pp. 1578–1589, 2001.
[59] A. Segers, *Data Assimilation in Atmospheric Chemistry Models using Kalman Filtering*. Amersterdam, Netherlands: Delft Univ. Press, 2002.
[60] W. Wang, C. H. Bishop, and S. J. Julier, "Which is better, an ensemble of positive-negative pairs or a centered spherical ensemble?," *Month. Weather Rev.*, vol. 132, pp. 1590–1605, 2004.
[61] P. Sakov and P. R. Oke, "Implications of the form of the ensemble transformation in the ensemble square root filters," *Month. Weather Rev.*, vol. 136, pp. 1042–1053, 2008.
[62] G. Burgers, J. van Leeuwen, and G. Evensen, "Analysis scheme in ensemble Kalman filter," *Month. Weather Rev.*, vol. 126, pp. 1719–1724, 1998.
[63] P. L. Houtekamer and H. L. Mitchell, "Data assimilation using an ensemble Kalman filter technique," *Month. Weather Rev.*, vol. 126, pp. 796–811, 1998.
[64] D. T. Pham, "Stochastic methods for sequential data assimilation in strongly nonlinear systems," *Month. Weather Rev.*, vol. 129, pp. 1194–1207, 2001.
[65] I. Hoteit, D. T. Pham, and J. Blum, "A simplified reduced order Kalman filtering and application to altimetric data assimilation in tropical Pacific," *J. Mar. Syst.*, vol. 36, pp. 101–127, 2002.
[66] C. H. Bishop, B. J. Etherton, and S. J. Majumdar, "Adaptive sampling with the ensemble transform Kalman filter—Part I: Theoretical aspects," *Month. Weather Rev.*, vol. 129, pp. 420–436, 2001.
[67] J. L. Anderson, "An ensemble adjustment filter for data assimilation," *Month. Weather Rev.*, vol. 129, pp. 2884–2903, 2001.
[68] M. K. Tippett, J. L. Anderson, T. M. Hamill, and J. S. Whitaker, "Ensemble square root filters," *Month. Weather Rev.*, vol. 131, pp. 1485–1490, 2003.
[69] J. S. Whitaker and T. M. Hamill, "Ensemble data assimilation without perturbed observations," *Month. Weather Rev.*, vol. 130, pp. 1913–1924, 2002.
[70] J. F. Bellatoni and K. W. Dodge, "A square root formulation of Kalman-Schmidt filter," *AIAA J.*, vol. 5, pp. 1309–1314, 1967.
[71] A. Andrews, "A square root formulation of Kalman covariance equations," *AIAA J.*, vol. 6, pp. 1165–1166, 1968.
[72] D. T. Pham, J. Verron, and M. C. Roubau, "A singular evolutive extended Kalman filter for data assimilation in oceanography," *J. Mar. Syst.*, vol. 16, pp. 323–340, 1998.
[73] P. Lermusiaux and A. Robinson, "Data assimilation via error subspace statistical estimation—Part I: Theory and schemes," *Month. Weather Rev.*, vol. 127, pp. 1385–1407, 1999.
[74] P. Lermusiaux and A. Robinson, "Data assimilation via error subspace statistical estimation—Part II: Middle Atlantic bight shelfbreak front simulations and ESSE validation," *Month. Weather Rev.*, vol. 127, pp. 1408–1432, 1999.
[75] A. W. Heemink, M. Verlaan, and A. J. Segers, "Variance reduced ensemble Kalman filtering," *Month. Weather Rev.*, vol. 129, pp. 1718–1728, 2001.
[76] A. C. Lorenc, "The potential of the ensemble Kalman filter for NWP—A comparison with 4D-var," *Q. J. R. Meteorol. Soc.*, vol. 129, pp. 3183–3203, 2003.
[77] P. L. Houtekamer and H. L. Mitchell, "A sequential ensemble Kalman filter for atmospheric data assimilation," *Month. Weather Rev.*, vol. 129, pp. 123–137, 2001.
[78] H. L. Mitchell, P. L. Houtekamer, and G. Pellerin, "Ensemble size, balance, and model-error representation in an ensemble Kalman filter," *Month. Weather Rev.*, vol. 130, pp. 2791–2808, 2002.
[79] D. M. Barker, "Southern high-latitude ensemble data assimilation in the Antarctic mesoscale prediction system," *Month. Weather Rev.*, vol. 133, pp. 3431–3449, 2005.
[80] D. C. Dowell, F. Zhang, L. J. Wicker, C. Snyder, and N. A. Crook, "Wind and temperature retrievals in the 17 May 1981 Arcadia, Oklahoma, supercell: Ensemble Kalman filter experiments," *Month. Weather Rev.*, vol. 132, pp. 1982–2005, 2004.
[81] M. Buehner and M. Charron, "Spectral and spatial localization of background-error correlations for data assimilation," *Q. J. R. Meteorol. Soc.*, vol. 133, pp. 615–630, 2007.
[82] O. Pannekoucke, L. Berre, and G. Desroziers, "Filtering properties of wavelets for local background-error correlations," *Q. J. R. Meteorol. Soc.*, vol. 133, pp. 363–379, 2007.
[83] X. Wang and C. H. Bishop, "Improvement of ensemble reliability with a new dressing kernel," *Q. J. R. Meteorol. Soc.*, vol. 131, pp. 965–986, 2005.
[84] J. L. Anderson and S. L. Anderson, "A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts," *Month. Weather Rev.*, vol. 127, pp. 2741–2758, 1999.
[85] J. L. Anderson, "An adaptive covariance inflation error correction algorithm for ensemble filters," *Tellus A*, vol. 59, pp. 210–224, 2007.
[86] C. Bishop and Z. Toth, "Ensemble transformation and adaptive observations," *J. Atmos. Sci.*, vol. 56, pp. 1748–1765, 1999.
[87] R. H. Langland, "Issues in targeted observing," *Q. J. R. Meteorol. Soc.*, vol. 131, pp. 3409–3425, 2005.
[88] J. A. Hansen, "Accounting for model error in ensemble-based state estimation and forecasting," *Month. Weather Rev.*, vol. 130, pp. 2373–2391, 2002.
[89] Z. Meng and F. Zhang, "Tests of an ensemble Kalman filter for mesoscale and regional-scale data assimilation—Part II: Imperfect model experiments," *Month. Weather Rev.*, vol. 135, pp. 1403–1423, 2007.
[90] T. Fujita, D. J. Stensrud, and D. C. Dowell, "Surface data assimilation using an ensemble Kalman filter approach with initial condition and model physics uncertainties," *Month. Weather Rev.*, vol. 135, pp. 1846–1868, 2007.
[91] D. S. Zrnic, J. F. Kimpel, D. E. Forsyth, A. Shapiro, G. Crain, R. Ferek, J. Heimmer, W. Benner, T. J. McNellis, and R. J. Vogt, "Agile-beam phased array radar for weather observations," *Bull. Amer. Meteorol. Soc.*, vol. 88, pp. 1753–1766, 2007
[92] J. P. Hacker and C. Snyder, "Ensemble Kalman filter assimilation of fixed screen-height observations in a parameterized PBL," *Month. Weather Rev.*, vol. 133, pp. 3260–3275, 2005.
[93] H. L. Mitchell and P. L. Houtekamer, "An adaptive ensemble Kalman filter," *Month. Weather Rev.*, vol. 128, pp. 416–433, 2000.
[94] M. Pan and E. F. Wood, "Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter," *J. Hydrometeorol.*, vol. 7, pp. 534–547, 2006.
[95] A. G. Slater and M. P. Clark, "Snow data assimilation via an ensemble Kalman filter," *J. Hydrometeorol.*, vol. 7, pp. 478–493, 2006.
[96] Y. Zhou, D. McLaughlin, and D. Entekhabi, "Assessing the performance of the ensemble Kalman filter for land surface data assimilation," *Month. Weather Rev.*, vol. 134, pp. 2128–2142, 2006.
[97] C. L. Keppenne, "Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter," *Month. Weather Rev.*, vol. 128, pp. 1971–1981, 2000.
[98] R. N. Miller, M. Ghil, and F. Gauthiez, "Data assimilation into nonlinear stochastic models," *Tellus A*, vol. 51, pp. 167–194, 1999.
[99] G. Evensen and P. J. van Leeuwen, "An ensemble Kalman smoother for nonlinear dynamics," *Month. Weather Rev.*, vol. 128, pp. 1852–1867, 2000.
[100] S. Kim, G. L. Eyink, J. M. Restrepo, F. J. Alexander, and G. Johnson, "Ensemble filtering for nonlinear dynamics," *Month. Weather Rev.*, vol. 131, pp. 2586–2594, 2003.
[101] S. Lakshmivarahan and S. K. Dhall, *Analysis and Design of Parallel Algorithms*. New York: McGraw Hill, 1990.
[102] J. Dongarra, I. Duff, P. Gaffney, and S. McKee, *Vector and Parallel Computing*. New York: Ellis Horwood Ltd., 1989.
[103] W. Gropp, E. Lusk, and A. Skjellum, *Using MPI*. Cambridge, MA: MIT Press, 1994.
[104] C. L. Keppenne and M. M. Rienecker, "Initial testing of a massively parallel ensemble Kalman filter with the Poseidon isopycnal ocean general circulation model," *Month. Weather Rev.*, vol. 130, pp. 2951–2965, 2002.
[105] E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. J. Patil, and J. A. Yorke, "A local ensemble Kalman filter for atmospheric data assimilation," *Tellus A*, vol. 56, pp. 415–428, 2004.
[106] P. L. Houtekamer and H. L. Mitchell, "A sequential ensemble Kalman filter for atmospheric data assimilation," *Month. Weather Rev.*, vol. 129, pp. 123–137, 2001.
[107] J. L. Anderson and N. Collins, "Scalable implementations of ensemble filter algorithms for data assimilation," *J. Atmos. Ocean. Technol.*, vol. 24, pp. 1452–1463, 2007.