Conservation of Mass and Preservation of Positivity with Ensemble-Type Kalman Filter Algorithms

TIJANA JANJIĆ AND DENNIS MCLAUGHLIN

Massachusetts Institute of Technology, Cambridge, Massachusetts

STEPHEN E. COHN

NASA Goddard Space Flight Center, Greenbelt, Maryland

MARTIN VERLAAN

Delft Technical University, Delft, Netherlands

(Manuscript received 11 February 2013, in final form 13 September 2013)

ABSTRACT

This paper considers the incorporation of constraints to enforce physically based conservation laws in the ensemble Kalman filter. In particular, constraints are used to ensure that the ensemble members and the ensemble mean conserve mass and remain nonnegative through measurement updates. In certain situations filtering algorithms such as the ensemble Kalman filter (EnKF) and ensemble transform Kalman filter (ETKF) yield updated ensembles that conserve mass but are negative, even though the actual states must be nonnegative. In such situations if negative values are set to zero, or a log transform is introduced, the total mass will not be conserved. In this study, mass and positivity are both preserved by formulating the filter update as a set of quadratic programming problems that incorporate nonnegativity constraints. Simple numerical experiments indicate that this approach can have a significant positive impact on the posterior ensemble distribution, giving results that are more physically plausible both for individual ensemble members and for the ensemble mean. In two examples, an update that includes a nonnegativity constraint is able to properly describe the transport of a sharp feature (e.g., a triangle or cone). A number of implementation questions still need to be addressed, particularly the need to develop a computationally efficient quadratic programming update for large ensemble.

1. Introduction

The importance of respecting physical conservation principles has long been recognized in numerical weather prediction modeling (Arakawa 1972; Arakawa and Lamb 1977; Sadourny 1975; Janjić 1984; Janjić et al. 2011; Janjić and Gall 2012). One of the most basic of these principles is the need to conserve the total mass of air, water in its different phases, and relevant chemical species, in the latter two cases accounting properly for sources and sinks while maintaining the proper sign (positivity or, more strictly, nonnegativity) in every grid volume of a computational model (e.g., Lin and Rood 1996, p. 2047). Smolarkiewicz and Margolin (1998, p. 460) write in a review paper that "the preservation of sign during numerical advection is the essential aspect of the stability and accuracy in modeling water phase-change or chemical processes." Sign preservation should be an implicit requirement in any numerical algorithm that attempts to conserve mass. Although most of the concepts described in this paper apply to the conservation of other quantities, such as angular momentum and energy, we focus on mass conservation to make the discussion more specific and to illustrate ideas with simple examples.

When ensemble Kalman filters are used for data assimilation two distinct mass conservation issues arise. Mass should be conserved in each member (replicate) of

Corresponding author address: Tijana Janjić, Hans Ertel Centre for Weather Research, DWD/LMU, Theresienstr. 37, 80333 Munich, Germany.
E-mail: tijana.janjic-pfander@dwd.de

DOI: 10.1175/MWR-D-13-00056.1

the forecast ensemble produced by propagating states over time with a forecast model, and it should be conserved by the procedure used to update the ensemble members with observations. Conservation of mass during the forecast is dependent on the time- and space-discretized method used to obtain a numerical solution to the governing continuum equations. Although mass is conserved by construction in the original differential equations, it may not be conserved in the numerical solution.

Similarly, conservation of mass during the ensemble filter update depends on the particular numerical method used to generate the replicates of the updated ensemble and to generate the mean of this ensemble (analysis itself). Mass may not be conserved during the update even when the numerical forecast technique is mass conservative. A number of methods have been proposed to deal with this issue. For example, Jacobs and Ngodock (2003) noted that mass in a simplified 1D ocean model can be conserved when the model error in a representer algorithm is expressed in terms of the mass flux due to uncertainty in ocean depth rather than as additive error in the continuity equation. In land surface hydrology, Pan and Wood (2006) showed how to ensure conservation of total water mass by imposing it as a ''perfect observation'' in a two-step Kalman filter approach. In an ocean data assimilation system, Brankart et al. (2003) imposed conservation of total mass, including positive layer thicknesses, through an a posteriori adjustment to the analyzed state. Positivity can also be ensured by introducing a change of state variables, using techniques such as Gaussian anamorphosis (e.g., Simon and Bertino 2009). In atmospheric data assimilation, nonnegativity of the specific humidity has been imposed as a weak constraint in a three-dimensional variational data assimilation (3D-Var) implementation (Liu and Xue 2006; Liu et al. 2007).

It is reasonable to ask if we should conserve mass in an ensemble Kalman filter update if the total mass in the system is uncertain. Measurements introduced during the update may provide useful information about this uncertain mass. If so, this information should be used to adjust and improve mass estimates. But this does not change the fact that a filtering algorithm should be able to preserve a *known* value of total mass through both the forecast and update steps. If the total mass is, in fact, unknown then it should be treated as uncertain and estimated as part of an otherwise mass-conservative filtering procedure. Thus, we distinguish the need to respect conservation laws from the need to properly account for uncertainties in conserved quantities. In the simple examples considered here total mass is a known constant that does not need to be estimated, but this restriction could be relaxed by including uncertain sources/sinks for instance in the state vector so they can be updated when new information becomes available.

Going beyond mass conservation, Cohn (2009) has shown in the context of minimum variance state estimation that conservation of total energy requires including a special term in the evolution equation for the state estimate that couples state and covariance evolution. This requirement was subsequently formulated more generally as the *principle of energetic consistency*, and was used to study certain pathological behavior of ensemble-based data assimilation schemes (Cohn 2010). Cohn (2010) shows that the mild energy dissipation typical of numerical weather prediction models can lead to ensemble collapse through a feedback mechanism introduced by the assimilation of observations, and that this dissipative behavior can in principle be eliminated by including an appropriately scale-selective, anti-dissipative operator in the formulation of the ensemble data assimilation scheme. Using such an operator to maintain ensemble spread is a generalization of the covariance inflation technique now commonly used in ensemble data assimilation schemes (Anderson and Anderson 1999, p. 2747).

One objective of data assimilation is to use observations to correct forecast errors in the vicinity of well-defined natural features such as fronts, filaments of chemical constituents, or plumes from surface emissions of aerosols. But data assimilation schemes have not traditionally been explicitly formulated to preserve such sharp features and, in fact, they often tend to blur and distort sharp interfaces (e.g., Lawson and Hansen 2005). Riishøjgaard (1998) proposed a type of state-dependent, anisotropic covariance modeling as a simple and direct analysis approach in the presence of sharp features; see Liu and Xue (2006) for an implementation. Hoffman et al. (1995) approached feature analysis by defining nontraditional, feature-based measures of spatial forecast error and minimizing them explicitly in variational data assimilation; see Gilleland et al. (2010) for a review of forecast verification methods utilizing such measures of forecast error. Lawson and Hansen (2005) showed that the performance of an ensemble Kalman filter in the presence of a well-defined feature can be improved dramatically by the use of alternative error models to redefine the state estimation problem.

All of these studies highlight the importance of conserving known mass, maintaining nonnegativity, and preserving feature geometry during data assimilation. In an ensemble context it can be argued that individual ensemble members as well as the ensemble mean should meet all of these requirements. Here we examine the issues of mass conservation, nonnegativity, and feature

preservation with two computational experiments that focus on simple features with well-defined shapes. In addition, we propose a new sequential ensemble data assimilation algorithm that enforces mass conservation and maintains nonnegativity by adding constraints to the ensemble Kalman filtering update.

We can view the classical Kalman update at any given time as either an unconstrained regularized minimum variance estimator or, if we take a Bayesian perspective, as a method for generating the maximum a posteriori (MAP) estimate when the a posteriori probability is Gaussian. In either case, the update requires the solution of an unconstrained optimization problem. When the measurement operator is linear the problem objective function depends quadratically on the analysis (which is the decision variable to be determined). In this case the unconstrained optimization problem has a closed form solution given by the classical Kalman filter update equation.

If the Kalman update does not satisfy physically based mass conservation or nonnegativity conditions, it is reasonable to enforce these conditions by adding appropriate constraints to the original unconstrained problem (Simon and Simon 2005). When the constraints are linear equalities and/or inequalities and the objective function Hessian is positive definite the resulting constrained optimization problem is a convex quadratic program with a unique global minimum. It is important to emphasize that this constrained optimization problem is a quadratic program even if the forecast model is nonlinear, so long as the constraints and observation operator are linear.

It is possible to apply similar concepts to ensemble filtering problems, where we can obtain each member of the analysis ensemble by solving a replicate-specific quadratic programming problem, replacing the forecast mean with one of the forecast ensemble members and adding random measurement errors to the actual observations. This procedure is analogous to the ensemble Kalman filter formulation proposed by Burgers et al. (1998).

In our extension of the Burgers et al. (1998) ensemble Kalman update we include nonnegativity constraints when computing the replicates of the analysis ensemble. This ensures that the ensemble members are all physically plausible, making it more likely that sample statistics, such as the covariance of this ensemble, are also physically realistic. From a Bayesian perspective, the nonnegativity constraints provide additional prior information that is not included in the classical formulation of the ensemble filtering problem. Consequently, the analysis replicates are no longer Gaussian but are strictly nonnegative. Our formulation of ensemble Kalman filtering as a sequence of static linear estimation problems makes the incorporation of physically based constraints a natural extension of the classical algorithm.

It is useful to briefly distinguish the ensemble quadratic programming approach proposed here from other approaches that share some of its features. The review article of Simon (2010) gives an overview of various methods for incorporating constraints into the classical Kalman filter, including a version of quadratic programming. These classical methods are able to maintain mass conservation through the filter update but not during the forecast if the system dynamics are nonlinear. Our quadratic programming approach uses an ensemble forecast that is able to conserve mass and nonnegativity through the forecast step for each ensemble member, even for nonlinear problems, if the forecast model is properly formulated.

Another ensemble data assimilation technique known as randomized maximum likelihood (RML) also solves an ensemble of optimization problems (Gu and Oliver 2007; Emerick and Reynolds 2013). In this case each problem minimizes the batch mean-squared measurement misfit computed over all measurement times (perhaps with an additional quadratic regularization term) for a particular replicate. The forecast and observation models are formulated as nonlinear equality constraints and are incorporated into each optimization problem through a set of derived objective function gradients. The problem solution is obtained with an unconstrained nonlinear programming procedure. Our approach uses a different objective function at each measurement time as well as for each ensemble member since it is formulated as a sequence of static updates rather than as a batch algorithm. This makes it possible to formulate the optimization problem as an efficient quadratic programming problem that readily accommodates inequality constraints. It is also compatible with the time-recursive structure of the classical ensemble Kalman filter.

Other options for incorporating ensemble information include a number of different ensemble variational or hybrid data assimilation algorithms (see, e.g., Hamill and Snyder 2000; Lorenc 2003; Buehner 2005; Zupanski 2005; Wang et al. 2007b, 2008; Wang 2010, 2011; Isaksen et al. 2010; Bonavita et al. 2012). These generally use a forecast ensemble to construct the hybrid covariance needed for a variational update. The decision variables in the optimization problem can include the analysis mean (Wang et al. 2008; Wang 2010, 2011) or individual ensemble members (Hamill and Snyder 2000; Isaksen et al. 2010; Bonavita et al. 2012). Additional inequality or equality constraints such as those used in our quadratic programming approach are typically not included in these hybrid methods. Despite these differences, our data

assimilation procedure shares important features with randomized maximum likelihood and ensemble variational methods (Hamill and Snyder 2000; Zupanski 2005) and it is possible to imagine variants that combine aspects of all three approaches.

In this paper we compare our constrained Kalman filtering algorithm to an ensemble Kalman filter (EnKF; Evensen 2009; Houtekamer and Mitchell 1998; Burgers et al. 1998). Section 2 provides background and considers mass conservation and sign preservation for ensemble Kalman filters. Section 3 describes two numerical examples that illustrate the consequences of mass balance errors in this class of filters, using results obtained with the EnKF. The first example is implemented both with and without a log transform in order to explore the behavior of an anamorphosis-based approach for maintaining positivity. Section 4 introduces our constrained ensemble quadratic programming algorithm and section 5 shows how this algorithm performs on the numerical experiments introduced in section 3. Section 6 discusses benefits and drawbacks of the proposed algorithm and identifies some open research issues.

2. Problem formulation

Consider a scalar quantity w whose evolution is governed by the continuity equation with no sources or sinks:

$$w_t + \nabla \cdot (\mathbf{v}w) = 0, \tag{1}$$

$$w(\mathbf{x}, t_0) = w_0(\mathbf{x}), \quad \text{for} \quad \mathbf{x} \text{ in } D, \tag{2}$$

where $\mathbf{v}$ is a given velocity field, t_0 is the initial time, and D is the spatial domain, assumed either cyclic or to have no mass flux through the boundaries. Then the total integral of w over D is conserved through time:

$$\int_D w(\mathbf{x}, t)\, d\mathbf{x} = \int_D w_0(\mathbf{x})\, d\mathbf{x}, \tag{3}$$

and if $\langle\ \rangle$ denotes expectation, then

$$\int_D \langle w(\mathbf{x}, t)\rangle\, d\mathbf{x} = \int_D \langle w_0(\mathbf{x})\rangle\, d\mathbf{x}. \tag{4}$$

Here the value of the initial state w_0 at any given location is random but we suppose that the total mass M of the initial state is fixed and deterministic, so that $\int_D \langle w_0(\mathbf{x})\rangle\, d\mathbf{x} = \int_D w_0(\mathbf{x})\, d\mathbf{x} = M$. Note that the spatial distribution of mass at times after the initial time will be random as a result of initial condition uncertainty. We assume that we have access to a numerical model that exactly conserves a discrete version of the total mass integral.

In ensemble data assimilation we typically work with an ensemble of spatially and temporally discretized state n-vectors $\mathbf{w}_k$ that approximate realizations of w at the n grid points of a computational grid that covers the domain D, evaluated at the time t_k. The ensemble members are updated to incorporate information from observations collected at specified measurement times. In sequential assimilation algorithms, such as the ensemble Kalman filter, it is convenient to separate the assimilation process into two steps, carried out at each measurement time: 1) a forecast step that uses a numerical model of Eq. (1) to propagate the analysis ensemble from the previous measurement time forward to the current measurement time and 2) an analysis step that computes a new analysis ensemble from the forecast ensemble and current measurements. The first forecast in this recursion is initialized with a set of specified random initial conditions and all subsequent forecasts are initialized with the most recent analysis ensemble.

The specific operations used to derive the analysis ensemble at each measurement time depend on the particular updating procedure selected. Here we consider the EnKF described in Evensen (2009) and Burgers et al. (1998). In the EnKF update step the analysis ensemble member $\mathbf{w}_k^{a,i}$ is obtained by combining the forecast (prior) ensemble member $\mathbf{w}_k^{f,i}$ with an m_k vector of perturbed measurements $\mathbf{w}_k^{o,i}$, as described by the following update equation:

$$\mathbf{w}_k^{a,i} = \mathbf{w}_k^{f,i} + \mathbf{K}_k(\mathbf{w}_k^{o,i} - \bar{\mathbf{r}}_k^{o} - \mathbf{H}_k\mathbf{w}_k^{f,i}), \tag{5}$$

where $i = 1, \ldots, N_{\text{ens}}$, N_{ens} is the number of ensemble members, and $\bar{\mathbf{r}}_k^{o}$ is a known possibly nonzero measurement error mean. Following usual EnKF practice, each perturbed measurement vector $\mathbf{w}_k^{o,i}$ is a random sample from a specified multivariate normal probability distribution with a mean equal to the m_k vector $\mathbf{w}_k^{o}$ of actual measurements and a covariance given by the specified $m_k \times m_k$ observation error covariance matrix $\mathbf{R}_k$. The gain $\mathbf{K}_k$ is given by

$$\mathbf{K}_k = \mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}}(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k)^{-1}, \tag{6}$$

where $\mathbf{H}_k$ is an $n \times m_k$ observation matrix and $\mathbf{P}_k^f$ is the $n \times n$ forecast error covariance of $\mathbf{w}_k^f$. The mean $\mathbf{w}_k^a = 1/N_{\text{ens}}\sum_{i=1}^{N_{\text{ens}}}\mathbf{w}_k^{a,i}$ of the analysis ensemble, called the analysis, is often selected as an estimate of the uncertain state at t_k. When the new analysis ensemble at t_k is computed one cycle of the filter recursion is completed and the process repeats at t_{k+1}.

In the EnKF the forecast error covariance appearing in Eq. (6) is calculated as follows:

$$\mathbf{P}_k^f = \frac{1}{N_{\text{ens}} - 1} \sum_{i=1}^{N_{\text{ens}}} [\mathbf{w}_k^{f,i} - \mathbf{w}_k^f][\mathbf{w}_k^{f,i} - \mathbf{w}_k^f]^{\mathrm{T}}, \quad (7)$$

where $\mathbf{w}_k^{f,i}$ are the individual forecast ensemble members for $i = 1, \ldots, N_{\text{ens}}$, and $\mathbf{w}_k^f$ is the forecast ensemble mean (i.e., $\mathbf{w}_k^f = 1/N_{\text{ens}} \sum_{i=1}^{N_{\text{ens}}} \mathbf{w}_k^{f,i}$).

Other versions of the ensemble Kalman filter update include square root filters, which generate the analysis ensemble by first calculating the analysis ensemble mean and then adding a random deviation for each replicate. These filters do not require the use of perturbed measurements. An example is the ensemble transform Kalman filter (ETKF; Bishop et al. 2001; Wang et al. 2004, 2007a; Hunt et al. 2007). Here we use the classical perturbed measurement EnKF for comparison with the ensemble quadratic programming algorithm introduced in section 4. This is convenient because the EnKF can be viewed as an important special case of the quadratic programming algorithm. However, the discussion of mass conservation and nonnegativity that follows applies to all common versions of the ensemble Kalman filter, including both the EnKF and the ETKF.

From Eq. (4), we require that the continuous version, $w^a(x, t_k) = \langle w(x, t) \,|\, \mathbf{w}_1^o, \ldots, \mathbf{w}_k^o \rangle$ of an analysis $\mathbf{w}_k^a$ must satisfy

$$\int_D w^a(x, t_k)\, dx = \int_D \langle w_0(x) \rangle\, dx, \quad (8)$$

and that the analysis error covariance function,

$$P^a(x_1, x_2, t_k) \equiv \langle [w(x_1, t_k) - w^a(x_1, t_k)][w(x_2, t_k) - w^a(x_2, t_k)] \rangle, \quad (9)$$

must satisfy

$$\int_D P^a(x_1, x_2, t_k)\, dx_1 = 0, \quad \text{for all } x_2. \quad (10)$$

The latter condition reflects the requirement that every realization of $w^a(x, t_k)$ must conserve total mass. Such an analysis error covariance is called "mass conserving." In the discrete case, the analysis error covariance matrix $\mathbf{P}_k^a$ is mass conserving if

$$\mathbf{P}_k^a \mathbf{e} = 0, \quad (11)$$

where $\mathbf{e} = \mathbf{e}_{n\times 1} = [11\ldots1]^{\mathrm{T}}$. The form of $\mathbf{e}$ given here is chosen for simplicity. The exact form of the definition in Eq. (11) will depend on the grid of our numerical model and the quadrature chosen for Eq. (10).

We can define the total mass of each member in the forecast and analysis ensembles as $M_k^{f,i} = \mathbf{e}^{\mathrm{T}}\mathbf{w}_k^{f,i}$ and $M_k^{a,i} = \mathbf{e}^{\mathrm{T}}\mathbf{w}_k^{a,i}$, respectively. In appendix A we show that both the EnKF and ETKF algorithms produce mass conserving covariances that give the same total mass M for each ensemble member through the update. That is, if $\mathbf{e}^{\mathrm{T}}\mathbf{w}_k^{f,i} = M$ for each forecast ensemble member, then $\mathbf{e}^{\mathrm{T}}\mathbf{w}_k^{f,i} = \mathbf{e}^{\mathrm{T}}\mathbf{w}_k^f = \mathbf{e}^{\mathrm{T}}\mathbf{w}_k^{a,i} = \mathbf{e}^{\mathrm{T}}\mathbf{w}_k^a = M$.

Now suppose that w is a nonnegative scalar quantity such as humidity or the concentration of a chemical constituent. There is no guarantee that the analysis mean or a given analysis replicate produced by an ensemble filter will be nonnegative everywhere, even when the forecast is nonnegative everywhere and total mass is conserved. In fact, ensemble filters often conserve mass by canceling large positive values with negative values. This is easily shown with examples such as those described in the next section. Various methods, such as truncation of negative values to zero or formulating the update step in terms of $\log(w)$ can force nonnegativity but the resulting analysis mean and replicates typically no longer conserve mass. Thus, it is fairly easy to obtain either mass-conservative or nonnegative analyses but much more difficult to obtain analyses that are *both* mass conservative and nonnegative. The problem is illustrated by example in section 3 and a possible solution is presented in section 4.

3. Ensemble Kalman filter performance for two examples

We now consider two examples chosen to illustrate how mass conservation and/or nonnegativity problems can arise during the filter update step. These examples focus on feature-oriented problems where ensemble Kalman filters may have difficulty generating results that conserve mass and/or maintain the proper sign.

a. One-dimensional static analysis with non-Gaussian background and observation errors

In our first example the true feature is a static one-dimensional hat (isosceles triangle) function (cf. with the smooth function used in Lawson and Hansen 2005) of unit height and five grid points wide on a 1D periodic domain. The state vector that describes this feature consists of values defined at 40 uniformly spaced grid points (see Fig. 1). The peak of the true feature is located at grid point 20. In a virtual experiment we generate noisy synthetic observations of the true state at a single time by summing the true (nonnegative) values at specified measurement locations and additive

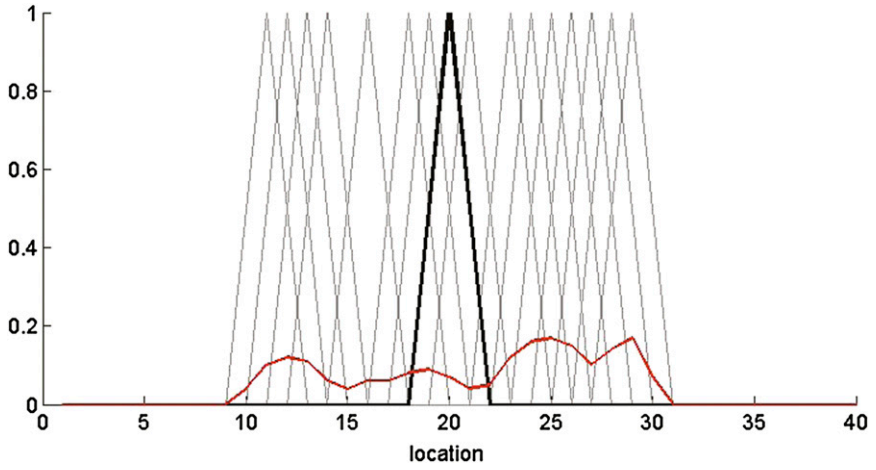

FIG. 1. 1D static forecast/prior. The true state (black), forecast/prior mean (red), and forecast/prior ensemble (gray) are shown. Mass of true, forecast mean, and each forecast ensemble member is equal to 2.

lognormal random measurement errors (also nonnegative), as follows:

$$\mathbf{w}_k^o = \mathbf{H}\mathbf{w}_k + \mathbf{r}_k^o, \tag{12}$$

where $\mathbf{H} = \mathbf{H}_k$ is a time-invariant measurement matrix that consists of appropriately located zeros and ones. The lognormal measurement error $\mathbf{r}_k^o$ has a specified mean $\bar{\mathbf{r}}_k^o$ (equal to 0.02 for each element) in this experiment and a diagonal covariance $\mathbf{R}_k$ (variance values are equal to 0.01). The measurement error mean is included in the filter update expression, as indicated in Eq. (5). The spatial configuration of the measurements that determines $\mathbf{H}$ is discussed below.

We assume that the exact position of the true triangle peak is unknown. This uncertainty is reflected through differences in the state vectors used to define the 50 forecast (or prior) ensemble members, each being identical to the true state except that the peak is located randomly, accordingly to a discrete uniform distribution, over the grid points between locations 10 and 30. Figure 1 shows that the ensemble mean (red line) obtained by averaging over the forecast replicates (light gray lines) is nonnegative but does not preserve the triangular shape of the true feature (black line).

Since neither the forecast nor measurement error distributions are normally distributed in this example the Kalman filter analysis is not the exact mean of the a posteriori probability distribution and the analysis replicates are not samples from this distribution. Consequently, from a Bayesian perspective, the ensemble Kalman filter is suboptimal for this problem (Simon and Bertino 2009; Bocquet et al. 2010). Ensemble estimators are frequently suboptimal in applications, especially those involving nonnegative features with sharp boundaries. Our primary interest here is in the filter's ability to conserve mass, maintain nonnegativity, and capture the shape of the true triangular feature.

With the problem setup described above we first derive the analysis replicates using the traditional EnKF described in Eq. (5). Figure 2 shows the true feature (black line) and analysis mean (red line) and ensemble (light gray lines) generated by the EnKF when measurements (green circles) are taken at every other grid point in the interval between locations 10 and 30. The analysis and the ensemble replicates exhibit spurious positive and negative lobes away from the true peak, although the total mass is conserved by both replicates and mean. The analysis ensemble replicates are generally not positive isosceles triangles. The relatively large anomalies shown in Fig. 2 appear to result from poor interpolation of values to unmeasured locations.

This effect is revealed in a different form in Fig. 3, where measurements are taken at every location over the more limited range from 15 to 25. In this clustered measurement case poor extrapolation yields large anomalies outside the measured region. Figure 4 compares the standard deviation of the EnKF analysis ensemble to the root-mean-square error (RMSE) between the analysis replicates and the true state, for the case with measurement gaps. This comparison indicates that the analysis ensemble generally captures the true degree of variability, with some underestimation at unmeasured locations.

The problem of maintaining nonnegativity in a Kalman filter update lies in the underlying Gaussian assumptions in Eqs. (5) and (A3) (Simon and Bertino 2009; Bocquet et al. 2010). These assumptions are not

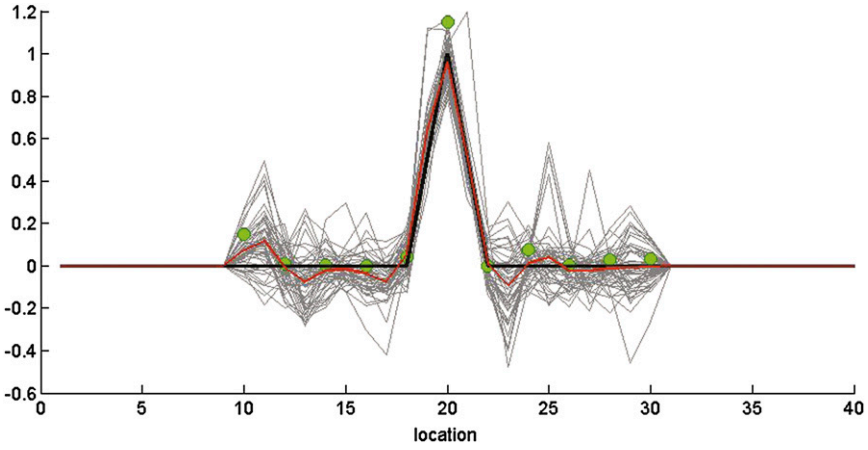

FIG. 2. 1D static analysis results for the EnKF with measurement gaps. The true state (black), observations (green), analysis ensemble (gray), and ensemble mean (red) are shown. Mass is conserved by all analysis ensemble members and analysis mean, but there are significant negative anomalies.

applicable for estimation of a state that has discontinuities, or for a state that is nonnegative, or for problems characterized by an error in the location of a disturbance (Chen and Snyder 2007). There are several techniques for modifying filters to enforce nonnegativity (Cohn 1997; Lauvernet et al. 2009; Bocquet et al. 2010; Schneider 1984). One alternative is to use Gaussian anamorphosis (Simon and Bertino 2009), which introduces a nonlinear change of state variables, such as a log transformation, in order to make the analysis step more consistent with Gaussian assumptions.

We now consider a version of the EnKF, formulated in terms of the transformed state $\tilde{\mathbf{w}}_{kj} = \log(\mathbf{w}_{kj} + \epsilon)$, where ϵ is a small positive number that ensures a finite $\tilde{\mathbf{w}}_{kj}$ value when $\mathbf{w}_{kj} = 0$ and the subscript j refers to the jth scalar component of $\tilde{\mathbf{w}}_k$ or $\mathbf{w}_k$. This transformation is an example of the anamorphosis approach taken by Simon and Bertino (2009). In our experiments we take $\epsilon = 10^{-3}$. The log transformed EnKF works with a vector $\tilde{\mathbf{w}}_k^o$ of transformed synthetic measurements with components $\tilde{\mathbf{w}}_{kj}^o = \log(\mathbf{w}_{kj}^o + \epsilon)$. The measurement equation of the log transformed EnKF assumes that $\tilde{\mathbf{w}}_k^o$ is related to the log transformed state $\tilde{\mathbf{w}}_k$ according to the following additive measurement equation:

$$\tilde{\mathbf{w}}_k^o = \mathbf{H}\tilde{\mathbf{w}}_k + \tilde{\mathbf{r}}_k^o, \tag{13}$$

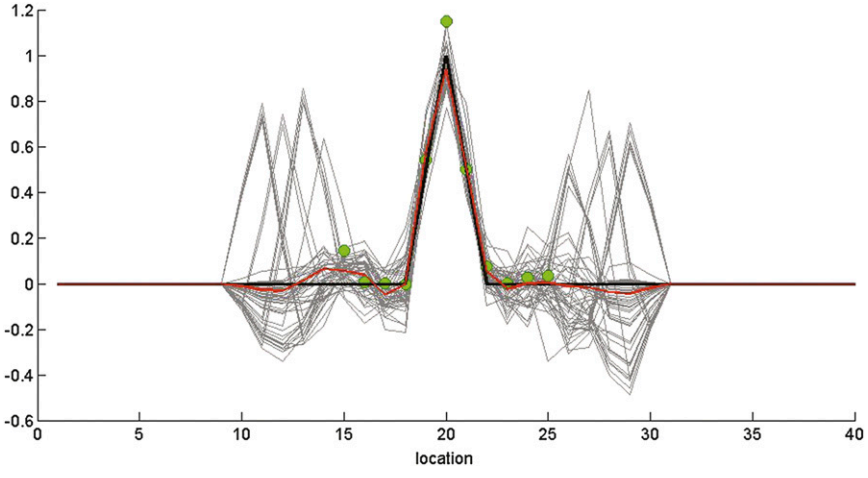

FIG. 3. 1D static analysis results for the EnKF with clustered measurements. The true state (black), observations (green), analysis ensemble (gray), and ensemble mean (red) are shown. Mass is conserved by all analysis ensemble members and the analysis mean, but there are significant negative anomalies.

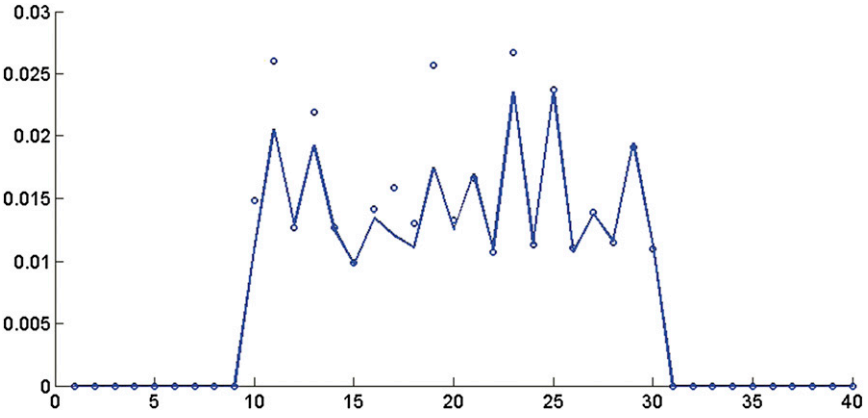

FIG. 4. 1D static analysis results for the EnKF with measurement gaps. Comparison of ensemble standard deviation (solid) and RMSE between analysis ensemble members and true (circles). The EnKF variances are generally comparable to the RMSE, with some underestimation at scattered locations.

where $\tilde{\mathbf{r}}_k^o$ is a vector of additive measurement errors in the observed components of the log transformed variable $\tilde{\mathbf{w}}_k$. This additive error equation for log transformed variables is not equivalent to the additive error equation for untransformed variables given in Eq. (12) but is required by the additive error assumption of the EnKF if the states and measurements are expressed as log transformed variables. Consequently, Eq. (13) should be viewed as an alternative to Eq. (12). This alternative is similar to the measurement error model described by Simon and Bertino (2009).

The individual elements of $\tilde{\mathbf{r}}_k^o$ are assumed to be mutually independent with a uniform mean of zero and a uniform variance that can be adjusted to capture the aggregate effects of measurement error on the log transformed measurements (Simon and Bertino 2009). The need to adjust these measurement error statistics can be viewed as a limitation of the log transform approach since it is difficult to know in advance how they should be selected.

Results for the log transformed EnKF with measurement gaps are shown in Figs. 5 and 6. The log transformed filter is very sensitive to the variance specified for its assumed additive measurement error $\tilde{\mathbf{r}}_k^o$. In Fig. 5 we have set the variance of $\tilde{r}_k^o$ equal to the variance of the log of $\mathbf{r}_k^o$. This value is 1.81 for the experiments

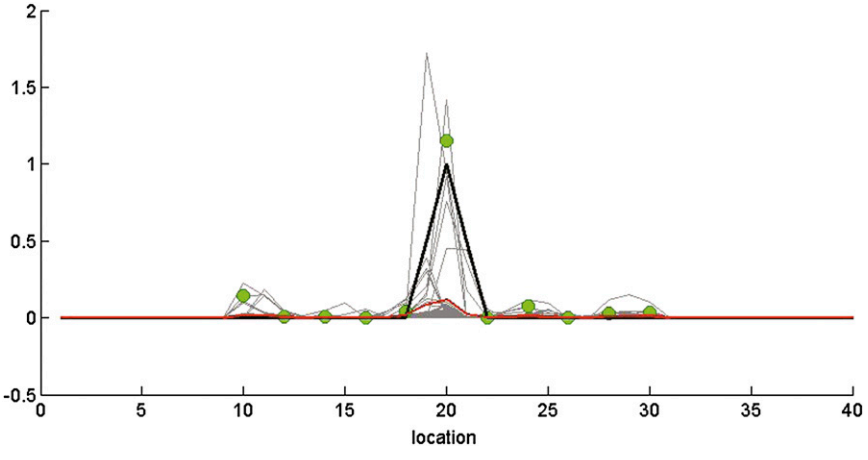

FIG. 5. 1D static analysis results for the log transformed EnKF with measurement gaps and additive log transformed measurement error variance set at the log of the additive untransformed synthetic measurement error. The true state (black), observations (green), analysis ensemble (gray), and ensemble mean (red) are shown. The analysis mean mass is $0.348 < 2$. The mass is not conserved, but the analysis mean and all analysis replicates are nonnegative. The analysis does not capture the triangular feature.

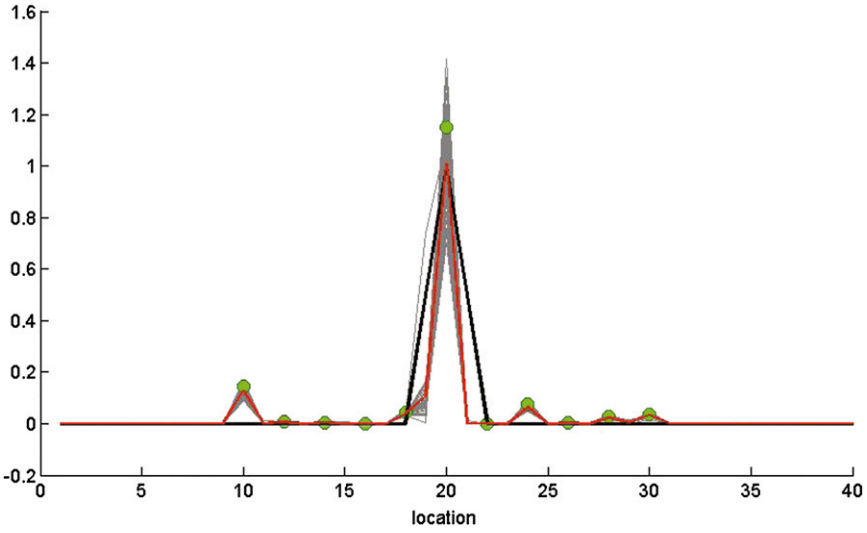

FIG. 6. 1D static analysis results for the log transformed EnKF with measurement gaps and measurement error covariance diagonal set at 0.1 variance of the log measurement error. The true state (black), observations (green), analysis ensemble (gray), and ensemble mean (red) are shown. The analysis mean mass is equal to $1.42 < 2$ is not conserved, but the analysis mean and all analysis ensemble members are nonnegative. The ensemble is greatly constrained by the measurements and ensemble variability is small.

described here. The $\tilde{r}_k^o$ variance used to generate Fig. 6 is 0.181, one-tenth of the value used in Fig. 5. The analysis values in the higher measurement error variance case of Fig. 5 are small and the measurements have relatively little impact. The mass of the analysis is 0.348, which is much lower than the true mass of 2. The lower measurement error variance case shown in Fig. 6 is highly constrained by the measurements, with minimal ensemble spread and an analysis mass of 1.42. This value is still significantly lower than the true mass.

In the limit of very low measurement error variance the log transformed EnKF analysis ignores the forecast, giving a triangle with a smaller base and smaller total mass than the true. The mass conservation results are poorest for intermediate values of the measurement error variance, when the analysis deviates from the forecast but fails to capture the shape of the true feature. Similar results are obtained for the clustered measurement configuration, which is not shown here. When the measurement error random seeds are varied some of the analysis ensemble members can take on positive values significantly higher than 1.0, even for small measurement error variances. Generally speaking, the log transform EnKF appears to be marginally stable for this problem, requiring fine adjustments of the measurement error variance to give reasonable analysis results. In any case, although analysis ensemble members and the analysis are nonnegative everywhere, mass is generally not conserved.

The results for this simple one-dimensional example show the dilemma encountered with classical ensemble Kalman filtering for problems where mass conservation and sign are both important. The classical EnKF conserves mass if the forecasts are mass conservative but it can produce unrealistic analyses (mean and ensemble) that are negative. The log transformed filter gives nonnegative analyses but does not generally conserve mass.

b. *Two-dimensional dynamic analysis*

The previous example shows the behavior of the EnKF solution for only one measurement update. To estimate the effect of analysis errors through time and on forecasts, we perform a second virtual experiment that considers two-dimensional solid-body rotation (Tremback et al. 1987; Janjić et al. 2011). In this experiment a moving cone completes a full clockwise rotation of 2π about the origin (domain center) every 48 h. Synthetic observations are assimilated every quarter revolution (4320 time steps) until seven forecast/analysis cycles are completed (time step 30 241). The experiment is carried out on a uniform numerical grid of 101 by 101 square cells, each of size 8 km by 8 km.

The initial ensemble is specified to be a set of cones, each with a radius of 100 km at the base and a height of 100 units. The true initial cone is centered at the grid point with indices $i = 33$ and $j = 33$. The central location of each cone in the 50-member ensemble is described by an angle and radius defined relative to the domain

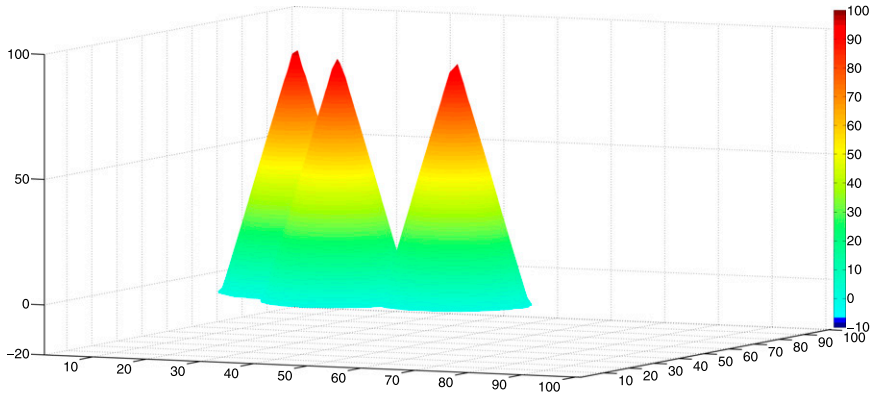

FIG. 7. Three members of the initial 50-member ensemble generated by perturbing the location of the center of the cone.

origin. The angle is perturbed around the true cone angle by a random number normally distributed with mean 0 and standard deviation of 0.5 radians. The distance of the cone from the center of rotation is perturbed as well, with normally distributed random error with mean 0 and standard deviation of 0.5 km. The random angle and the distance define the center of each cone in the ensemble. The ensemble of cones is transported with pure advection by simply changing the angle of their centers relative to the domain origin at the appropriate rate. This ensures that the solution to the governing equation given in Eq. (1) is exact, with no numerical dispersion or other numerical errors. Three of the ensemble members are shown at the initial time in Fig. 7.

The synthetic observations are obtained by perturbing the true solution with a small amount of log normally distributed noise with mean 0.5 and variance 1. The minimum observation value is 0.002. In this experiment the observation operator $\mathbf{H}_k$ varies in time. It uses every 20th synthetic observation at the locations where true cone would have values greater than zero. This corresponds to measurements at 25 grid points per analysis time. The restriction of noisy measurements to the region of the true cone reflects the fact that measurements outside this region will be small compared to the cone amplitude and would normally be removed by truncating measurements below an appropriate threshold.

The forecast mean at the first update time is computed as an average over the ensemble members, producing a much attenuated field with a maximum value of 46.7 and minimum value of zero (Fig. 8). Note that, although every ensemble member generated at the initial time is a perfect cone with the desired properties, the structure of the cone is not preserved in the ensemble average. Since the model is exact, each copy of the cone rotates without losing its structure or its minimum and maximum values. However, the shape of the cone is not preserved through the EnKF analysis step. Although the total mass is conserved, unrealistic negative mass values are obtained after the analysis. Figure 9 shows the analysis and the three analysis ensemble members at the end of the experiment (time step 30 241) from bird's-eye perspective. The minimum, maximum, and RMS error values of the analysis are −11.2, 98.4, and 1.1, respectively. The analysis and each of the ensemble members show spurious positive and unphysical negative values away from the cone structure. Negative values (depicted with dark blue contour lines in Fig. 9) reach −16.3 in the ensemble. The ensemble members that have the lowest (−16.3) and the highest (−7.3) minimum values are shown in Fig. 9, together with an ensemble member with minimum value of −10.5. The difference between the chosen ensemble members is the largest in the area away from

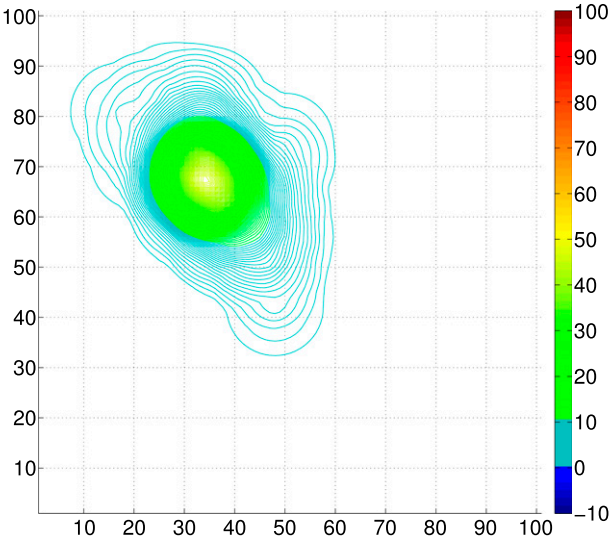

FIG. 8. The forecast ensemble mean at the first measurement time from bird's-eye perspective.

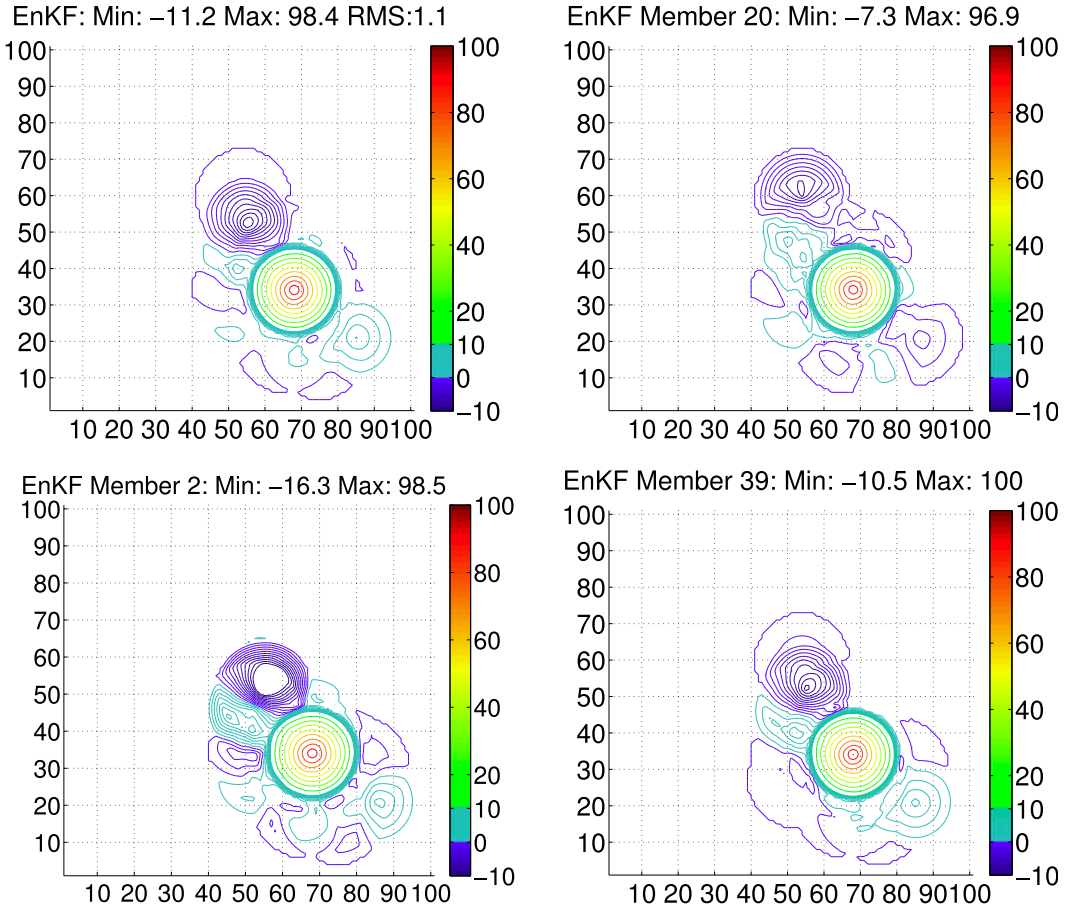

FIG. 9. (top left) The analysis at the end of the solid-body rotation experiment (time step 30 241), obtained with the EnKF algorithm and 50 ensemble members. Ensemble members are shown as examples of replicates with the (top right) highest and (bottom left) lowest minimum values. (bottom right) An ensemble member with a minimum value between the two is depicted. Contour lines in the range from −10 to 10 are shown in steps of 1, and above 10 in steps of 10.

the cone, where spurious error structures appear (see Fig. 9). The maximum value of the true cone (100) is slightly underestimated in the ensemble since the maximum of the ensemble members varies between 96.9 and 100.5. Although all ensemble members show the cone structure in the right location, unphysical negative values away from the cone are large. Both positive and negative spurious values affect a large area of the domain.

4. Problem solution

The update step of the ensemble Kalman filter described in Burgers et al. (1998) can be posed as the solution to a set of regularized least squares optimization problems, one problem for each member in the ensemble. The problem associated with members $i = 1, \ldots, N_{\text{ens}}$ at time t_k can be expressed using the same notation as in Eqs. (5) and (6):

$$\mathbf{w}_k^{a,i} = \mathbf{w}_k^{f,i} + \arg\min_{\delta w^i} \frac{1}{2}[\delta \mathbf{w}^{i^{\mathrm{T}}} (\mathbf{P}^f)^{-1} \delta \mathbf{w}^i + \mathbf{f}^{i^{\mathrm{T}}} \mathbf{R}^{-1} \mathbf{f}^i], \tag{14}$$

where $\delta \mathbf{w}^i = \mathbf{w}_k^{a,i} - \mathbf{w}_k^{f,i}$ is the analysis increment, which serves as the decision variable, and $\mathbf{f}^i = \mathbf{w}_k^{o,i} - \mathbf{H}_k \mathbf{w}_k^{a,i} - \bar{\mathbf{r}}_k^o = \mathbf{w}_k^{o,i} - \mathbf{H}_k \mathbf{w}_k^{f,i} - \mathbf{H}_k \delta \mathbf{w}^i - \bar{\mathbf{r}}_k^o$. Note that $\mathbf{P}^f$ is obtained from the forecast ensemble, as indicated in Eq. (7). Also, the measurements appearing in $\mathbf{f}^i$ are perturbed as in Eq. (5). The solutions to the N_{ens} optimization problems defined in Eq. (14) form the analysis ensemble. When there are no constraints in Eqs. (5) and (6) give a closed form solution to the ensemble optimization problem in Eq. (14) (Zupanski 2005; Wang et al. 2007a).

Our extension of this optimization formulation of the ensemble Kalman filter adds linear inequality constraints in order to enforce nonnegativity in the update.

The resulting constrained analysis step solves the following quadratic programming problem for each member $i = 1, \ldots, N_{\text{ens}}$:

$$\mathbf{w}_k^{a,i} = \mathbf{w}_k^{f,i} + \arg\min_{\delta w^i} \frac{1}{2}[\delta\mathbf{w}^{i\mathrm{T}}(\mathbf{P}^f)^{-1}\delta\mathbf{w}^i + \mathbf{f}^{i\mathrm{T}}\mathbf{R}^{-1}\mathbf{f}^i] \tag{15}$$

subject to the following nonnegativity constraint:

$$\delta\mathbf{w}^i \geq -\mathbf{w}_k^{f,i}. \tag{16}$$

Note that the inequality in Eq. (16) is equivalent to $\mathbf{w}_k^{a,i} \geq 0$. In general the constrained solution to Eq. (15) cannot be expressed in closed form but must be derived numerically.

The quadratic programming objective given above depends on the inverse of the sample covariance $\mathbf{P}^f$, which is singular when $N_{\text{ens}} - 1 < n$ and/or the forecast covariance is mass conserving. The low rank of $\mathbf{P}^f$ allows us to transform the optimization problem by reducing the number of decision variables to $\rho = \text{Rank}(\mathbf{P}^f)$, which is no larger than $N_{\text{ens}} - 1$. Appendix B shows that the quadratic programming solution conserves mass if the forecast covariance is mass conservative and the decision variable $\delta\mathbf{w}^i$ is chosen to lie in the ρ dimensional subspace spanned by the forecast ensemble. For the unconstrained case this solution duplicates the closed form EnKF solution given in Eq. (5).

If $\delta\mathbf{w}^i$ lies in the forecast ensemble subspace it can be expressed in terms of a ρ-dimensional transformed decision variable η^i as follows:

$$\delta\mathbf{w}^i = \mathbf{L}\eta^i, \tag{17}$$

where the columns of the $n \times \rho$ dimensional matrix $\mathbf{L}$ form a basis for the forecast ensemble subspace and the elements of η^i can be viewed as the weights in a linear combination of the basis vectors. Then $\mathbf{P}^f$ can be written in terms of the $\rho \times \rho$ dimensional covariance $\mathbf{Q}$ of η:

$$\mathbf{P}^f = \mathbf{L}\mathbf{Q}\mathbf{L}^{\mathrm{T}}. \tag{18}$$

Since there is flexibility in defining $\mathbf{L}$ we select it to satisfy the requirement that $\mathbf{P}^f = \mathbf{L}\mathbf{L}^{\mathrm{T}}$, so that $\mathbf{Q}$ is the ρ-dimensional identity matrix and $\mathbf{L}$ is the matrix square root of $\mathbf{P}^f$. This square root may be computed using a singular value decomposition of the matrix whose columns are the differences between the vectors of the forecast ensemble members and the forecast ensemble mean.

We use the change of variables given in Eq. (17) to rewrite the constrained quadratic programming problem in terms of η^i:

$$\eta^i = \arg\min_{\eta^i} \frac{1}{2}[\eta^{i\mathrm{T}}\eta^i + \mathbf{f}^{i\mathrm{T}}\mathbf{R}^{-1}\mathbf{f}^i], \tag{19}$$

where $\mathbf{f}^i = \mathbf{w}_k^{o,i} - \mathbf{H}_k\mathbf{w}_k^{f,i} - \mathbf{H}_k\mathbf{L}\eta^i - \bar{\mathbf{r}}_k^o$ subject to the following nonnegativity constraint:

$$-\mathbf{L}\eta^i \leq \mathbf{w}_k^{f,i}. \tag{20}$$

The analysis ensemble can be derived from the solution to Eq. (19):

$$\mathbf{w}_k^{a,i} = \mathbf{w}_k^{f,i} + \delta\mathbf{w}^i = \mathbf{w}_k^{f,i} + \mathbf{L}\eta^i. \tag{21}$$

The analysis ensemble and its mean all lie in the forecast ensemble subspace since $\mathbf{w}_k^{f,i}$ and $\delta\mathbf{w}^i$ lie in this space by construction (see appendix B).

Once the analysis ensemble is calculated using Eq. (21), it is propagated with the forecast model to obtain the new forecast ensemble, as in the classical unconstrained ensemble Kalman filter. For easy reference, we call this ensemble data assimilation method QPEns. The structure of the QPEns algorithm insures that the analysis and each member of the analysis ensemble will have the desired mass conservation and positivity properties if the forecast ensemble is mass conservative.

There are a number of QP algorithms available to compute the optimum for Eqs. (19)–(20). For example, the active-set method can be viewed as an extension of the traditional EnKF analysis. The first iteration starts with an unconstrained optimization. If this solution satisfies all constraints, then the optimum is found. If not, more iterations will follow, where in each iteration some inequality constraints are converted to equality constraints (i.e., made active) or removed as equality constraints (i.e., made inactive). The number of iterations depends on the problem at hand. For the experiments in section 5a, we found that 10 iterations are sufficient on average. For many applications, obtaining the ensemble forecast with the forecast model will dominate the computations, so the additional effort required by the QPEns optimization will likely be affordable. There is much potential in future research for developing more efficient optimization procedures that exploit the special structure of the QPEns problem.

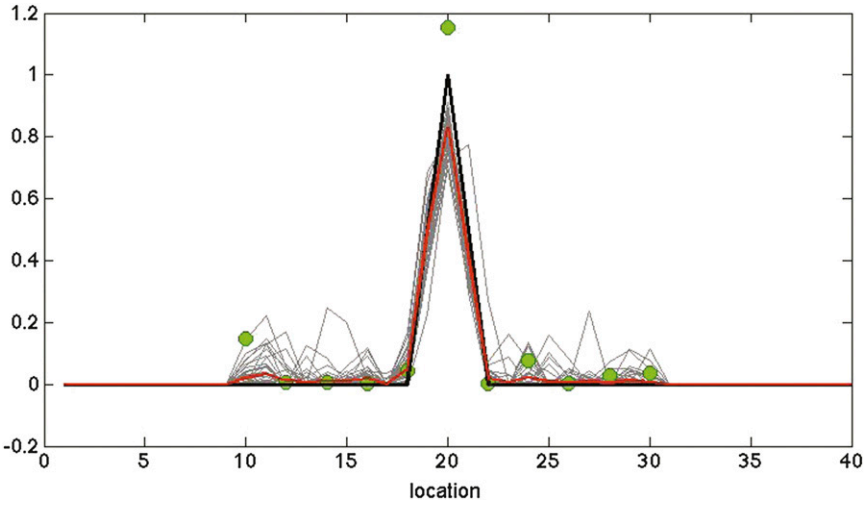

FIG. 10. 1D static analysis results for QP with positivity constraint and nominal assumed measurement error variance 0.01. True state (black), observations (green), analysis ensemble (gray), and analysis ensemble mean (red). The analysis ensemble members and their mean conserve mass and are always nonnegative. The analysis mean underestimates true at the peak in order to conserve total mass since the ensemble members and mean both include some mass in the region where the true state is zero.

5. Quadratic programming performance for two examples

a. One-dimensional static analysis with non-Gaussian background and observation errors

This section considers the static one-dimensional problem discussed in section 3 when the analysis ensemble is derived with the ensemble quadratic programming algorithm (QPEns). All results were obtained with an active set Matlab quadratic programming routine. Figure 10 shows the true state (black) and the analysis mean (red line) and ensemble (light gray lines) generated by QPEns for the measurements already considered in Fig. 2 (green circles). In this case the measurement error variance used in the QPEns is set equal to the variance of the additive lognormal synthetic measurements, without any adjustment. For our computational experiment the mean and variance of the lognormal measurement errors are 0.02 and 0.01, respectively. These are the same values used to generate the untransformed EnKF results presented in section 3. The resulting QPEns analysis mean and ensemble all conserve mass and are nonnegative everywhere. Some of ensemble members take on nonzero values in regions where the true field is zero, leading to nonzero analysis values in these regions. As a result, the analysis mass near the true peak needs to decrease in order to maintain the correct total analysis mass (i.e., the areas under the red and black curves need to be the same).

The QPEns results are sensitive to the specified measurement error covariance $\mathbf{R}$. Figure 11 shows the results obtained for the same problem considered in Fig. 10 with the measurement error variance used in the QPEns algorithm reduced to one-quarter of the actual value (i.e., from 0.01 to 0.0025). In this case, the measurements have more influence and the analysis mean and all the ensemble members are very close to the true values. This behavior reflects the fact that, for this particular problem, the low measurement error covariance quadratic programming objectives are minimized when the values at the peak and the two nearest measurement locations are as close as possible to the corresponding measurements. Mass conservation and nonnegativity can only be satisfied if the values at points farther from the peak are all close to zero. It is interesting that the EnKF still gives significant negative values (not shown here) for the low measurement error variance case. In this example the nonnegativity constraint that distinguishes the QPEns algorithm has an important impact on overall accuracy as well as sign.

Figure 12 compares the QPEns standard deviation of the analysis ensemble to the RMSE between the analysis ensemble and the true state, for the nominal specified measurement error variance value. The analysis ensemble variances are generally comparable to the RMSE, with some underestimation at a few locations near the true peak. Note that the RMSE values for QPEns are significantly lower than the EnKF values plotted in Fig. 4, indicating a better match to the true feature.

The results shown above indicate that the constrained QPEns algorithm provides a major improvement in

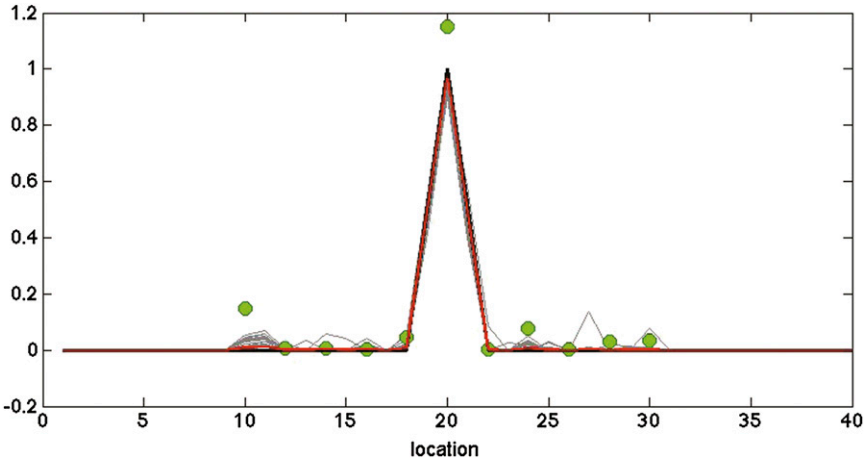

FIG. 11. 1D static analysis results for QPEns with positivity constraint and lowered assumed measurement error variance (0.0025). True state (black), observations (green), analysis ensemble (gray), and analysis ensemble mean (red). The analysis ensemble and their mean conserve mass and are always nonnegative. The analysis mean and ensemble members are very close to the true values.

accuracy and physical realism over either an untransformed or log transformed EnKF, at least for the problem we have considered. The QPEns conserves mass, gives nonnegative values, and provides an accurate description of the true feature of interest.

b. *Two-dimensional dynamic analysis*

In this section we use the QPEns algorithm to solve the two-dimensional dynamic data assimilation problem introduced in section 3b. For implementation of the QPEns algorithm in this example, we constructed the perturbed observations as $\mathbf{w}_k^o + \mathbf{r}_k^{o,i}$, where $\mathbf{r}_k^{o,i}$ is a vector normally distributed $\mathcal{N}(0, \mathbf{R}_k)$ and $\mathbf{R}_k$ is a diagonal matrix with 1^2 on the diagonal. The results of this experiment are summarized in Fig. 13. The QPEns algorithm is able to recover the cone structure, with a maximum value of 94.9 and RMSE of 0.4 at the end of the experiment (cf. Fig. 9). As in Fig. 9 we show an example of three ensemble members. Since the result of the QPEns cannot be negative, we show the ensemble members with lowest and highest maximum values, and one with the maximum value in between. Maximum values of ensemble members vary in the range between 56.6 and 98.5. One of the depicted ensemble member (Fig. 13, bottom-left panel) almost perfectly represents the true cone structure with the maximum value of 98.5. The ensemble member with the lowest maximum value differs from the true cone, with the errors primarily

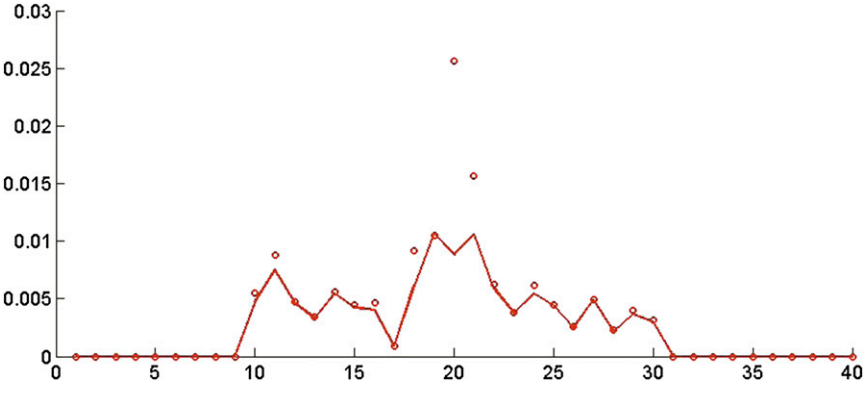

FIG. 12. 1D static analysis results for the QPEns with positivity constraint and nominal measurement error variance. Comparison of ensemble standard deviation (solid) and RMSE between analysis ensemble members and true (circles). The QPEns variances are generally comparable to the RMSE, with some underestimation at a few locations near the true peak. RMSE values for the QPEns are significantly lower than for the EnKF (cf. Fig. 4).

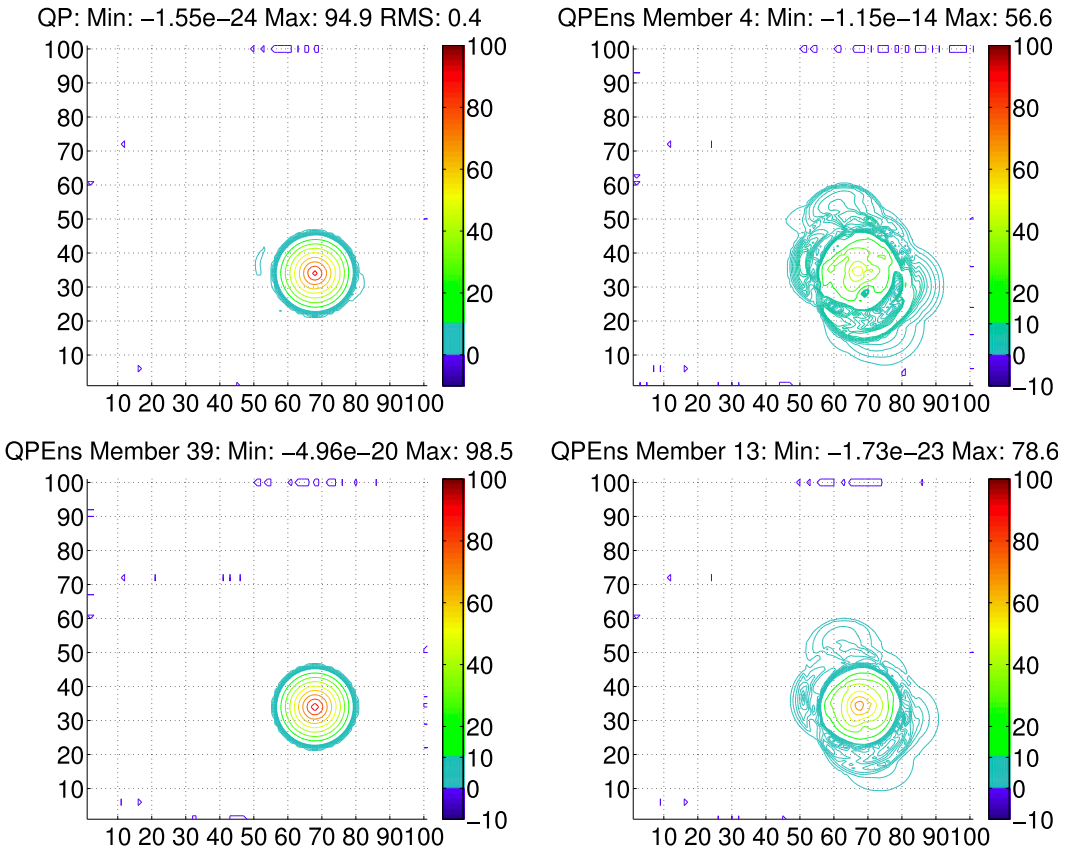

FIG. 13. (top left) The analysis at the end of the solid-body rotation experiment (time step 30 241), obtained with the QPEns algorithm. Ensemble members are shown as examples of replicates with the (top right) lowest and (bottom left) highest maximum values. (bottom right) An ensemble member with maximum value between the two is depicted. Contour lines in the range from −10 to 10 are shown in steps of 1, and above 10 in steps of 10.

localized around the cone in the range from 0 to 10. The third ensemble member shown in Fig. 13 has a maximum value in between the lowest and highest. This member has a clearer cone structure than the one with the lowest maximal value but positive errors are still present around the cone. In all cases, the analysis ensemble members have the same mass as the true cone and are nonnegative, resulting in a more accurate analysis mean than the traditional EnKF.

6. Discussion and conclusions

In this paper we propose using constraints to enforce mass conservation, nonnegativity, and other physical requirements in an ensemble Kalman filter update. Forecast means and covariances convey useful but sometimes incomplete information about the physical requirements imposed by conservation laws. As a result, ensemble Kalman filter updates that rely only on these moments and scattered noisy measurements can produce unphysical analyses and analysis ensemble members. It is possible to deal with some physical requirements, such as mass conservation, through proper construction of the forecast error covariance and the subsequent update. But this does not generally insure that analysis results are nonnegative. Conversely, it is possible to impose nonnegativity by using transform methods such as anamorphosis, but these methods do not necessarily conserve mass. If the classical unconstrained ensemble Kalman filter update is appropriately constrained it is possible to conserve mass and also to maintain the correct sign.

When measurements are linearly related to the state, the ensemble Kalman filter update can be posed as a set of unconstrained quadratic programming problems, one for each replicate. The solutions to these unconstrained problems can be expressed in closed form. The quadratic programming structure of the problem is maintained if linear equality and inequality constraints are added, but the solutions must generally be obtained from a numerical optimization procedure rather than from a closed form expression. Fortunately, many important

physical constraints, including mass conservation and nonnegativity, are linear and fit into a quadratic programming framework. This makes it convenient to add constraints to existing sequential data assimilation algorithms based on ensemble Kalman filters. The quadratic programming formulation has a number of important advantages, including the availability of very efficient solution algorithms and the guarantee that the problem has a unique minimum for linear observation operators when the Hessian of the quadratic objective function is positive definite.

The benefits of including constraints to enforce nonnegativity are apparent in the results obtained in our two synthetic experiments. In both cases, ensemble quadratic programming captures the shape and mass of a distinctive spatial feature through a filter update with noisy measurements. The quadratic programming approach works much better than a classical ensemble Kalman filter, which gives negative estimates even though all measurements and forecasts are nonnegative. The classical EnKF conserves total mass by generating inflated positive masses in some locations in order to cancel negative masses generated in other locations. The log transformed EnKF is able to maintain nonnegativity but does not generally conserve mass and requires adjustment of its measurement error covariance in order to obtain reasonable ensemble members and to capture the approximate shape of the true feature.

The quadratic programming approach described here has some limitations that are important to note. It is appropriate when the observation operator is linear and when all constraints included in the update are linear. The forecast model can be nonlinear, as in other ensemble filtering methods that compute forecast statistics by propagating ensemble with nonlinear dynamic models. The basic concept of constraining the Kalman filter update can be extended to accommodate nonlinear measurement operators and constraints. However, the optimization must then be performed with a more expensive nonlinear programming algorithm and there is no longer a guarantee of a unique minimum. It is possible that the quadratic programming formulation adopted here could be retained for nonlinear measurement operators if the objective function to be minimized is expressed in terms of the analysis error covariance rather than the forecast and observation error covariances (Zupanski 2005). Linear equality and inequality constraints could then be included as described in section 4.

As mentioned above, the ensemble quadratic programming approach requires numerical solution of a different quadratic programming problem at every analysis time, for every ensemble member. This is not a significant limitation for our simple examples but it could require substantially greater computational effort than the standard closed form ensemble Kalman filter update, especially for spatially distributed problems with many degrees of freedom. However, the relative increase in overall computational effort may not be significant because in many high-dimensional ensemble filtering problems computational effort is dominated by the forecast rather than the analysis step. Larger problems will have to be investigated before we can assess the overall impact of solving a quadratic programming problem for every analysis ensemble member. It may be possible to reduce computational effort by taking advantage of the fact that quadratic programming solutions for the different ensemble members tend to be clustered around a common mean. Also, the different quadratic programming solutions required in the analysis step can be computed in parallel.

Our quadratic programming approach for including constraints in an ensemble Kalman filter is related to other ensemble data assimilation methods including hybrid variational methods and randomized maximum likelihood. These various approaches are complementary and it is likely that they could be combined in various ways. The distinctive aspects of our approach are an emphasis on the need to include physical constraints during the update and a formulation that takes advantage of the computational benefits of quadratic programming. Together, these provide a practical and effective way to ensure that data assimilation results satisfy fundamental physical requirements.

Acknowledgments. Tijana Janjić is grateful to the Max Kade foundation for providing partial support for this study and to Hans-Ertel Centre for Weather Research. This research network of universities, research institutes, and the Deutscher Wetterdienst is funded by the BMVBS (Federal Ministry of Transport, Building and Urban Development). Stephen E. Cohn gratefully acknowledges the support of the NASA Modeling and Analysis Program, provided through the Global Modeling and Assimilation Office core funding.

APPENDIX A

Mass Conservation in Ensemble Kalman Filters

In ensemble Kalman filters the sample forecast error covariance matrix in Eq. (7) is derived from an ensemble of states produced by integration of a numerical model. Many numerical integration schemes can conserve the total (global) mass of tracers (e.g., Schneider 1984; Lin and Rood 1997). Each replicate in a forecast ensemble

computed with a conservative scheme conserves total mass and the sample covariance derived from the forecast ensemble is mass conserving. That is, $\mathbf{e}^{\mathrm{T}}\mathbf{w}_k^{f,i} = M$ for each ensemble member and the mean $\mathbf{w}_k^f$ over any number of ensemble members has the same mass M, so $\mathbf{e}^{\mathrm{T}}(\mathbf{w}_k^{f,i} - \mathbf{w}_k^f) = 0$ for all i and Eq. (7) gives $\mathbf{e}^{\mathrm{T}}\mathbf{P}_k^f = 0$.

Since $\mathbf{e}^{\mathrm{T}}\mathbf{P}_k^f = 0$, Eq. (6) gives $\mathbf{e}^{\mathrm{T}}\mathbf{K}_k = 0$ and it follows from Eq. (5) that $\mathbf{e}^{\mathrm{T}}\mathbf{w}_k^{a,i} = \mathbf{e}^{\mathrm{T}}\mathbf{w}_k^{f,i} = M$ for the classical EnKF. Therefore, the analysis ensemble and analysis mean of the EnKF all conserve mass if the forecast ensemble members conserve mass or, equivalently, if the forecast covariance is mass conserving. In this case the EnKF analysis covariance is also mass conserving. This covariance can be expressed, for any $\mathbf{K}_k$, as

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^{\mathrm{T}}. \tag{A1}$$

Since $\mathbf{P}_k^f$ is mass conserving and the gain $\mathbf{K}_k$ satisfies $\mathbf{e}^{\mathrm{T}}\mathbf{K}_k = 0$, then from Eq. (A1) we have that $\mathbf{e}^{\mathrm{T}}\mathbf{P}_k^a = 0$, so $\mathbf{P}_k^a$ is mass conserving.

Similar reasoning applies to the ETKF. In this case the analysis mean is computed directly from

$$\mathbf{w}_k^a = \mathbf{w}_k^f + \mathbf{K}_k(\mathbf{w}_k^o - \overline{\mathbf{r}}_k^o - \mathbf{H}_k\mathbf{w}_k^f). \tag{A2}$$

Consequently, it follows that $\mathbf{e}^{\mathrm{T}}\mathbf{w}_k^a = \mathbf{e}^{\mathrm{T}}\mathbf{w}_k^f$. This indicates that the ETKF analysis is mass conserving if the forecast model conserves mass.

The ETKF generates its analysis ensemble as follows:

$$\mathbf{w}_k^{a,i} = \mathbf{w}_k^a + \sqrt{N_{\mathrm{ens}} - 1}[\mathbf{W}_k^f\mathbf{T}_k]_i, \tag{A3}$$

where $\mathbf{W}_k^f = 1/\sqrt{N_{\mathrm{ens}} - 1}[\mathbf{w}_k^{f,1} - \mathbf{w}_k^f, \ldots, \mathbf{w}_k^{f,N_{\mathrm{ens}}} - \mathbf{w}_k^f]$ is the $n \times N_{\mathrm{ens}}$ matrix of deviations of the forecast ensemble members from their mean. We take the $N_{\mathrm{ens}} \times N_{\mathrm{ens}}$ transformation matrix $\mathbf{T}_k$ as in Wang et al. (2004, 2007a) and Hunt et al. (2007). Namely, if $\mathbf{C}_k$ and $\mathbf{D}_k$ are the matrices of eigenvectors and corresponding eigenvalues of the matrix $(\mathbf{W}_k^f)^{\mathrm{T}}\mathbf{H}_k^{\mathrm{T}}\mathbf{R}_k^{-1}\mathbf{H}_k\mathbf{W}_k^f = \mathbf{C}_k\mathbf{D}_k(\mathbf{C}_k)^{\mathrm{T}}$, respectively, then $\mathbf{T}_k = \mathbf{C}_k(\mathbf{I}_{N_{\mathrm{ens}}} + \mathbf{D}_k)^{-1/2}\mathbf{C}_k^{\mathrm{T}}$ is the matrix that transforms deviations of forecast ensemble members from the forecast mean into deviations of analysis ensemble members from the analysis mean. Here, $\mathbf{I}_{N_{\mathrm{ens}}}$ denotes the $N_{\mathrm{ens}} \times N_{\mathrm{ens}}$ identity matrix.

The ETKF analysis error covariance can be shown to be $\mathbf{P}_k^a = \mathbf{W}_k^f\mathbf{T}_k\mathbf{T}_k^{\mathrm{T}}\mathbf{W}_k^{f^{\mathrm{T}}}$ (Bishop et al. 2001; Wang et al. 2007a). It follows that the ETKF analysis covariance is mass conserving since $\mathbf{e}^{\mathrm{T}}\mathbf{P}_k^a = \mathbf{e}^{\mathrm{T}}\mathbf{W}_k^f\mathbf{T}_k\mathbf{T}_k^{\mathrm{T}}\mathbf{W}_k^{f^{\mathrm{T}}}$ and $\mathbf{e}^{\mathrm{T}}\mathbf{W}_k^f = 0$. Also, from Eq. (A3) we have $\mathbf{e}^{\mathrm{T}}\mathbf{w}_k^{a,i} = \mathbf{e}^{\mathrm{T}}\mathbf{w}_k^a = M$.

The above discussion indicates that the analysis ensemble and analysis mean of both the EnKF and the ETKF conserve mass and their analysis covariances are mass conserving if the forecast ensemble members conserve mass or, equivalently, if the forecast covariance is mass conserving. For this derivation we made no assumptions on linearity of the model dynamics. The dynamics can be nonlinear as long as the numerical discretization scheme conserves mass.

The proof that the analysis error covariance is mass conserving if the forecast error covariance is mass conserving is applicable to any mass conserving forecast error covariance, not only those derived from an ensemble. Another example of a covariance formulation that would conserve mass is one obtained by eigenvalue decomposition on a sample that has a constant spatial integral, since then $\mathbf{e}$ would be an eigenvector corresponding to a zero eigenvalue of the covariance matrix computed from the sample, and, therefore, would be orthogonal to the other eigenvectors that are part of the low-rank modeled covariance. Since this is a usual technique for initializing ensemble square root Kalman filter algorithms, we assume that we start initially with a mass conserving covariance matrix. Note that if the forecast error covariance is modeled in such a way that all its elements are positive [Gaussian, second order autoregressive (SOAR), third order autoregressive (TOAR)] or nonnegative, then it cannot be mass conserving, since $\mathbf{e}$ cannot be a null vector of a matrix with all positive elements. Similarly, if localization is applied to the ensemble-derived forecast error covariance through a Schur multiplication, then the mass will not be conserved.

APPENDIX B

Mass Conservative Properties of the QPEns Analysis

The mass conservation properties of the QPEns analysis are related to the properties of the ρ-dimensional subspace spanned by the ensemble of N_{ens} forecast ensemble members. This subspace is also spanned by the ensemble of forecast deviations $\mathbf{w}_k^{f,i} - \mathbf{w}_k^f$, since the mean $\mathbf{w}_k^f$ also lies in the same subspace as the forecast ensemble members. The analysis ensemble produced by the QPEns algorithm or, equivalently, the analysis deviations $\mathbf{w}_k^{a,i} - \mathbf{w}_k^a$, are constrained by construction to also lie in the forecast ensemble subspace [see Eq. (17)]. Since the analysis deviations lie in the subspace spanned by the forecast deviations, they may be written as a linear combination of the following form:

$$\mathbf{w}_k^{a,i} - \mathbf{w}_k^a = \sum_{j=1}^{N_{\mathrm{ens}}} \alpha_{i,j}(\mathbf{w}_k^{f,j} - \mathbf{w}_k^f), \tag{B1}$$

where the $\alpha_{i,j}$, $i = 1, \ldots, N_{\mathrm{ens}}$ and $j = 1, \ldots, N_{\mathrm{ens}}$ are the scalar coefficients of the linear combination for replicate

i. This implies that the analysis ensemble is mass conservative if the forecast ensemble is mass conservative since

$$\mathbf{e}^{\mathrm{T}}(\mathbf{w}_k^{a,i} - \mathbf{w}_k^{a}) = \mathbf{e}^{\mathrm{T}} \sum_{j=1}^{N_{\mathrm{ens}}} \alpha_{i,j}(\mathbf{w}_k^{f,j} - \mathbf{w}_k^{f}) = \sum_{j=1}^{N_{\mathrm{ens}}} \alpha_{i,j}\mathbf{e}^{\mathrm{T}}(\mathbf{w}_k^{f,j} - \mathbf{w}_k^{f}) = 0. \quad \text{(B2)}$$

The final equality is a mathematical statement of the assumption that the forecast ensemble is mass conservative. Note that this result applies to any quadratic programming problem with a vector decision variable that is constructed to lie in the forecast ensemble subspace, as is done in Eq. (17).

REFERENCES

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

Arakawa, A., 1972: Design of the UCLA general circulation model. Tech. Rep., Department of Meteorology, University of California, Los Angeles, Tech. Rep. 7, 116 pp.

——, and V. R. Lamb, 1977: Computational design of the basic dynamical processes of the UCLA general circulation model. *Methods Comput. Phys.,* **17,** 173–265.

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

Bocquet, M., C. A. Pires, and L. Wu, 2010: Beyond Gaussian statistical modeling in geophysical data assimilation. *Mon. Wea. Rev.,* **138,** 2997–3023.

Bonavita, M., L. Isaksen, and E. Holm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. *Quart. J. Roy. Meteor. Soc.,* **138** (667), 1540–1559, doi:10.1002/qj.1899.

Brankart, J.-M., C.-E. Testut, P. Brasseur, and J. Verron, 2003: Implementation of a multivariate data assimilation scheme for isopycnic coordinate ocean models: Application to a 1993–1996 hindcast of the North Atlantic Ocean circulation. *J. Geophys. Res.,* **108,** 3074, doi:10.1029/2001JC001198.

Buehner, M., 2005: Ensemble-derived stationary and flow dependent background error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.,* **131,** 1013–1043.

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

Chen, Y., and C. Snyder, 2007: Assimilating vortex position with an ensemble Kalman filter. *Mon. Wea. Rev.,* **135,** 1828–1845.

Cohn, S. E., 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan,* **75,** 257–288.

——, 2009: Energetic consistency and coupling of the mean and covariance dynamics. *Handbook of Numerical Analysis,* R. M. Temam and J. J. Tribbia, Eds., Vol. XIV, *Computational Methods for the Atmosphere and the Oceans,* Elsevier, 443–478.

——, 2010: The principle of energetic consistency in data assimilation. *Data Assimilation—Making Sense of Observations,* W. Lahoz, R. Swinbank, and B. Khattatov, Eds., Springer-Verlag, 145–223.

Emerick, A. A., and A. Reynolds, 2013: Ensemble smoother with multiple data assimilation. *Comput. Geosci.,* **55,** 3–15.

Evensen, G., 2009: *Data Assimilation: The Ensemble Kalman Filter.* Springer, 308 pp.

Gilleland, E., D. A. Ahijevych, B. G. Brown, and E. E. Ebert, 2010: Verifying forecasts spatially. *Bull. Amer. Meteor. Soc.,* **91,** 1365–1373.

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

Hamill, T. M., and C. Snyder, 2000: A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon. Wea. Rev.,* **128,** 2905–2919.

Hoffman, R. N., L. Zheng, J.-F. Louis, and C. Grassoti, 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.,* **123,** 2758–2770.

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

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

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud, 2010: Ensemble of data assimilations at ECMWF. Tech. Rep., ECMWF Tech. Memo. 636, 48 pp.

Jacobs, G. A., and H. E. Ngodock, 2003: The maintenance of conservative physical laws within data assimilation systems. *Mon. Wea. Rev.,* **131,** 2595–2607.

Janjić, Z., 1984: Non-linear advection schemes and energy cascade on semi-staggered grids. *Mon. Wea. Rev.,* **112,** 1234–1245.

——, and R. Gall, 2012: Scientific documentation of the NCEP nonhydrostatic multiscale model on the B grid (NMMB). Part 1: Dynamics. Tech. Rep. NCAR/TN-489+STR, NCAR Tech. Note, 80 pp.

——, T. Janjić, and R. Vasić, 2011: A class of conservative fourth-order advection schemes and impact of enhanced formal accuracy on extended-range forecasts. *Mon. Wea. Rev.,* **139,** 1556–1568.

Lauvernet, C., J.-M. Brankart, F. Castruccio, G. Broquet, P. Brasseur, and J. Verron, 2009: A truncated Gaussian filter for data assimilation with inequality constraints: Application to the hydrostatic stability condition in ocean models. *Ocean Modell.,* **27,** 1–17.

Lawson, W. G., and J. A. Hansen, 2005: Alignment error models and ensemble-based data assimilation. *Mon. Wea. Rev.,* **133,** 1687–1709.

Lin, S.-J., and R. B. Rood, 1996: Multidimensional flux-form semi-Lagrangian transport schemes. *Mon. Wea. Rev.,* **124,** 2046–2070.

——, and ——, 1997: An explicit flux-form semi-Lagrangian shallow-water model on the sphere. *Quart. J. Roy. Meteor. Soc.,* **123,** 2477–2498.

Liu, H., and M. Xue, 2006: Retrieval of moisture from slant-path water vapor observations of a hypothetical GPS network using a three-dimensional variational scheme with anisotropic background error. *Mon. Wea. Rev.,* **134,** 933–949.

——, ——, R. J. Purser, and D. F. Parrish, 2007: Retrieval of moisture from simulated GPS slant-path water vapor observations using 3DVAR with anisotropic recursive filters. *Mon. Wea. Rev.,* **135,** 1506–1521.

Lorenc, A. C., 2003: The potential of the ensemble Kalman filter for NWP—A comparison with 4D-Var. *Quart. J. Roy. Meteor. Soc.,* **129,** 3183–3203.

Pan, M., and E. F. Wood, 2006: Data assimilation for estimating the terrestrial water budget using a constrained ensemble Kalman filter. *J. Hydrometeor.,* **7,** 534–547.

Riishøjgaard, L.-P., 1998: A direct way of specifying flow dependent background error correlations for meteorological analysis systems. *Tellus,* **50A,** 42–57.

Sadourny, R., 1975: The dynamics of finite-difference models of the shallow-water equations. *J. Atmos. Sci.,* **32,** 680–689.

Schneider, H. R., 1984: A numerical transport scheme which avoids negative mixing ratios. *Mon. Wea. Rev.,* **112,** 1206–1217.

Simon, D., 2010: Kalman filtering with state constraints: A survey of linear and nonlinear algorithms. *IET Control Theory Appl. IET,* **4,** 1303–1318, doi:10.1049/iet-cta.2009.0032.

——, and D. L. Simon, 2005: Aircraft turbofan engine health estimation using constrained Kalman filtering. *J. Eng. Gas Turbines Power,* **127,** 323–328.

Simon, E., and L. Bertino, 2009: Application of the Gaussian anamorphosis to assimilation in a 3-D coupled physical-ecosystem model of the North Atlantic with the EnKF: A twin experiment. *Ocean Sci.,* **5,** 495–510, doi:10.5194/os-5-495-2009.

Smolarkiewicz, P. K., and L. G. Margolin, 1998: MPDATA: A finite-difference solver for geophysical flows. *J. Comput. Phys.,* **140,** 459–480.

Tremback, C. J., J. Powell, W. R. Cotton, and R. A. Pielke, 1987: The forward-in-time upstream advection scheme: Extension to higher orders. *Mon. Wea. Rev.,* **115,** 540–555.

Wang, X., 2010: Incorporating ensemble covariance in the Gridpoint Statistical Interpolation (GSI) variational minimization: A mathematical framework. *Mon. Wea. Rev.,* **138,** 2990–2995.

——, 2011: Application of the WRF hybrid ETKF-3DVAR data assimilation system for hurricane track forecasts. *Wea. Forecasting,* **26,** 868–884.

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

——, T. M. Hamill, J. S. Whitaker, and C. H. Bishop, 2007a: A comparison of hybrid ensemble transform Kalman filter–optimum interpolation and ensemble square root filter analysis schemes. *Mon. Wea. Rev.,* **135,** 1055–1076.

——, C. Snyder, and T. M. Hamill, 2007b: On the theoretical equivalence of differently proposed ensemble–3DVAR hybrid analysis schemes. *Mon. Wea. Rev.,* **135,** 222–227.

——, D. Barker, C. Snyder, and T. M. Hamill, 2008: A hybrid ETKF–3DVAR data assimilation scheme for the WRF model. Part I: Observing system simulation experiment. *Mon. Wea. Rev.,* **136,** 5116–5131.

Zupanski, M., 2005: Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.,* **133,** 1710–1726.