Blended particle filters for large-dimensional chaotic dynamical systems

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1073/pnas.1405675111">http://dx.doi.org/10.1073/pnas.1405675111</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>National Academy of Sciences (U.S.)</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sat Jun 24 03:32:49 EDT 2017</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/91955">http://hdl.handle.net/1721.1/91955</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Blended particle filters for large-dimensional chaotic dynamical systems

Andrew J. Majda\textsuperscript{a,1}, Di Qi\textsuperscript{a}, and Themistoklis P. Sapsis\textsuperscript{b,1}

\textsuperscript{a}Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; and \textsuperscript{b}Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Andrew J. Majda, April 4, 2014 (sent for review February 7, 2014)

A major challenge in contemporary data science is the development of statistically accurate particle filters to capture non-Gaussian features in large-dimensional chaotic dynamical systems. Blended particle filters that capture non-Gaussian features in an adaptively evolving low-dimensional subspace through particles interacting with evolving Gaussian statistics on the remaining portion of phase space are introduced here. These blended particle filters are constructed in this paper through a mathematical formalism involving conditional Gaussian mixtures combined with statistically nonlinear forecast models compatible with this structure developed recently with high skill for uncertainty quantification. Stringent test cases for filtering involving the 40-dimensional Lorenz 96 model with a 5-dimensional adaptive subspace for nonlinear blending filtered in various turbulent regimes with at least nine positive Lyapunov exponents are used here. These cases demonstrate the high skill of the blended particle filter algorithms in capturing both highly non-Gaussian dynamical features as well as crucial nonlinear statistics for accurate filtering in extreme filtering regimes with sparse infrequent high-quality observations. The formalism developed here is also useful for multiscale filtering of turbulent systems and a simple application is sketched below.

Some authors have suggested that particle filtering strategies (6–8) based on the Kalman filter (9) have been developed with some success in these extremely complex high-dimensional nonlinear systems. However, these accurate nonlinear particle filtering strategies are not feasible for large-dimensional chaotic dynamical systems because sampling a high-dimensional space is extremely difficult issues when a particle filter with its various up-to-date resampling strategies (2) provide better estimates than the Kalman filter in the presence of strong nonlinearity and highly non-Gaussian distributions. However, these accurate nonlinear particle filtering strategies are not feasible for large-dimensional chaotic dynamical systems because sampling a high-dimensional variable is computationally impossible for the foreseeable future. Recent mathematical theory strongly supports this curse of dimensionality for particle filters (3, 4). In the second direction, Bayesian hierarchical modeling (5) and reduced-order filtering strategies (6–8) based on the Kalman filter (9) have been developed with some success in these extremely complex high-dimensional nonlinear systems. There is an inherently difficult practical issue of small ensemble size in filtering statistical solutions of these complex problems due to the large computational overload in generating individual ensemble members through the forward dynamical operator. Numerous ensemble-based Kalman filters (10–14) show promising results in addressing this issue for synoptic-scale midlatitude weather dynamics by imposing suitable spatial localization on the covariance updates; however, all these methods are very sensitive to model resolution, observation frequency, and the nature of the turbulent signals when a practical limited ensemble size (typically less than 100) is used.

Recent attempts without a major breakthrough to use particle filters with small ensemble size directly on large-dimensional chaotic dynamical systems are surveyed in chap. 15 of ref. 15. The goal here is to develop blended particle filters for large-dimensional chaotic dynamical systems. For the blended particle filters developed below for a state vector $u \in \mathbb{R}^N$, there are two subspaces that typically evolve adaptively in time, where $u = (u_1, u_2)$, $u_1 \in \mathbb{R}^{N_1}$, and $N_1 + N_2 = N$, with the property that $N_1$ is low dimensional enough so that the non-Gaussian statistics of $u_1$ can be calculated from a particle filter whereas the evolving statistics of $u_2$ are conditionally Gaussian given $u_1$. Statistically nonlinear forecast models with this structure with high skill for uncertainty quantification have been developed recently by Sapsis and Majda (16–19) and are used below in the blended filters.

The mathematical foundation for implementing the analysis step where observations are used in the conditional Gaussian mixture framework is developed in the next section, followed by a summary of the nonlinear forecast models as well as crucial issues for practical implementation. The skill of the blended filters in capturing significant non-Gaussian features as well as crucial nonlinear dynamics is tested below in a 40-dimensional chaotic dynamical system with at least nine positive Lyapunov exponents in various turbulent regimes where the adaptive non-Gaussian subspace with a particle filter is only 5 dimensional with excellent performance for the blended filters. The mathematical formalism for filtering with conditionally Gaussian mixtures should be useful as a framework for multiscale data assimilation of turbulent signals and a simple application is sketched below. An earlier strategy related to the approach developed here is simply to filter the solution on an evolving low-dimensional subspace that captures the leading features in large-dimensional chaotic dynamical systems because sampling a high-dimensional space is extremely difficult issues when a particle filter with its various up-to-date resampling strategies (2) provide better estimates than the Kalman filter in the presence of strong nonlinearity and highly non-Gaussian distributions. However, these accurate nonlinear particle filtering strategies are not feasible for large-dimensional chaotic dynamical systems because sampling a high-dimensional variable is computationally impossible for the foreseeable future. Recent mathematical theory strongly supports this curse of dimensionality for particle filters (3, 4). In the second direction, Bayesian hierarchical modeling (5) and reduced-order filtering strategies (6–8) based on the Kalman filter (9) have been developed with some success in these extremely complex high-dimensional nonlinear systems. There is an inherently difficult practical issue of small ensemble size in filtering statistical solutions of these complex problems due to the large computational overload in generating individual ensemble members through the forward dynamical operator. Numerous ensemble-based Kalman filters (10–14) show promising results in addressing this issue for synoptic-scale midlatitude weather dynamics by imposing suitable spatial localization on the covariance updates; however, all these methods are very sensitive to model resolution, observation frequency, and the nature of the turbulent signals when a practical limited ensemble size (typically less than 100) is used.

Recent attempts without a major breakthrough to use particle filters with small ensemble size directly on large-dimensional chaotic dynamical systems are surveyed in chap. 15 of ref. 15. The goal here is to develop blended particle filters for large-dimensional chaotic dynamical systems. For the blended particle filters developed below for a state vector $u \in \mathbb{R}^N$, there are two subspaces that typically evolve adaptively in time, where $u = (u_1, u_2)$, $u_1 \in \mathbb{R}^{N_1}$, and $N_1 + N_2 = N$, with the property that $N_1$ is low dimensional enough so that the non-Gaussian statistics of $u_1$ can be calculated from a particle filter whereas the evolving statistics of $u_2$ are conditionally Gaussian given $u_1$. Statistically nonlinear forecast models with this structure with high skill for uncertainty quantification have been developed recently by Sapsis and Majda (16–19) and are used below in the blended filters.

The mathematical foundation for implementing the analysis step where observations are used in the conditional Gaussian mixture framework is developed in the next section, followed by a summary of the nonlinear forecast models as well as crucial issues for practical implementation. The skill of the blended filters in capturing significant non-Gaussian features as well as crucial nonlinear dynamics is tested below in a 40-dimensional chaotic dynamical system with at least nine positive Lyapunov exponents in various turbulent regimes where the adaptive non-Gaussian subspace with a particle filter is only 5 dimensional with excellent performance for the blended filters. The mathematical formalism for filtering with conditionally Gaussian mixtures should be useful as a framework for multiscale data assimilation of turbulent signals and a simple application is sketched below. An earlier strategy related to the approach developed here is simply to filter the solution on an evolving low-dimensional subspace that captures the leading
variance adaptively (20) while ignoring the other degrees of freedom; simple examples for nonnormal linear systems (18) demonstrate the poor skill of such an approach for reduced filtering in general; for filtering linear nonnormal systems, the optimal reduced basis instead is defined through balanced truncation (8).

Mathematical Foundations for Blended Particle Filters

Here we consider real-time filtering or data assimilation algorithms for a state vector \( \mathbf{u} \in \mathbb{R}^N \) from a nonlinear turbulent dynamical system where the state \( \mathbf{u} \) is forecast by an approximate dynamical model between successive observation times, \( m\Delta t \), and the state of the system is updated through the use of the observations at the discrete times \( m\Delta t \) in the analysis step. For the blended particle filters developed below, there are two subspaces that typically evolve adaptively in time (although that dependence is suppressed here), where \( \mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2) \), \( \mathbf{u} \in \mathbb{R}^N \), and \( N_1 + N_2 = N \) with the property that \( N_1 \) is low dimensional enough so that the statistics of \( \mathbf{u}_1 \) can be calculated from a particle filter whereas the statistics of \( \mathbf{u}_2 \) are conditionally Gaussian given \( \mathbf{u}_1 \). Thus, at any analysis time step \( m\Delta t \), we have the prior forecast density given by

\[
p_{\sim}(\mathbf{u}) = p_{\sim}(\mathbf{u}_1) p_{Q}(\mathbf{u}_2 | \mathbf{u}_1),
\]

where \( p_{Q}(\mathbf{u}_2 | \mathbf{u}_1) \) is a Gaussian distribution determined by the conditional mean and covariance, \( p_{Q}(\mathbf{u}_2 | \mathbf{u}_1) = \mathcal{N}(\mathbf{u}_2; \mathbf{u}_2^\prime, \mathbf{R}_2^\prime) \). We assume that the marginal distribution, \( p_{\sim}(\mathbf{u}_1) \), is approximated by \( Q \) particles,

\[
p_{\sim}(\mathbf{u}_1) = \sum_{j=1}^{Q} p_{\sim j}(\mathbf{u}_1 - \mathbf{u}_{1j}),
\]

with nonnegative particle weights, \( p_{\sim j} \) with \( \sum p_{\sim j} = 1 \). Below we sometimes abuse notation as in [2] and refer to the continuous distribution in [1] and the particle distribution in [2] interchangeably. Here it is assumed that the nonlinear observation operator \( G(\mathbf{u}) \) maps \( \mathbb{R}^N \) to \( \mathbb{R}^M \), with \( M \) observations at the analysis time, \( m\Delta t \), and has the form

\[
v = G(\mathbf{u}) + \mathbf{v}_0 = G_0(\mathbf{u}_1) + G_2(\mathbf{u}_1) \mathbf{u}_2 + \mathbf{v}_0,
\]

where \( G_1(\mathbf{u}_1) \) has rank \( M \) and the observational noise, \( \mathbf{v}_0 \), is Gaussian \( \mathbf{v}_0 = \mathcal{N}(0, \mathbf{R}_0) \). Note that the form in [3] is the leading Taylor expansion around \( \mathbf{u}_1 \) of the general nonlinear observation operator. Now, start with the conditional Gaussian particle distribution, \( p_{\sim}(\mathbf{u}_1, \mathbf{u}_2) \) given from the forecast distribution

\[
p_{\sim}(\mathbf{u}) = \sum_{j=1}^{Q} p_{\sim j}(\mathbf{u}_1 - \mathbf{u}_{1j}) \mathcal{N}(\mathbf{u}_2; \mathbf{u}_2^\prime, \mathbf{R}_2^\prime),
\]

and compute the posterior distribution in the analysis step through Bayes’ theorem, \( p_{\sim}(\mathbf{u}|v) \sim p(v|\mathbf{u}) p_{\sim}(\mathbf{u}) \). The key fact is the following.

Proposition 1. Assume the prior distribution from the forecast is the blended particle filter conditional Gaussian distribution in [4] and assume the observations have the structure in [3]; then the posterior distribution in the analysis step taking into account the observations in [3] is also a blended particle filter conditional Gaussian distribution; i.e., there are explicit formulas for the updated weights, \( p_{\sim j} \), \( 1 \leq j \leq Q \), and conditional mean, \( \mathbf{u}_2^\prime \), and covariance, \( \mathbf{R}_2^\prime \), so that

\[
p_{\sim}(\mathbf{u}) = \sum_{j=1}^{Q} p_{\sim j}(\mathbf{u}_1 - \mathbf{u}_{1j}) \mathcal{N}(\mathbf{u}_2; \mathbf{u}_2^\prime, \mathbf{R}_2^\prime).
\]

In fact, the distributions \( \mathcal{N}(\mathbf{u}_2; \mathbf{u}_2^\prime, \mathbf{R}_2^\prime) \) are updated by suitable Kalman filter formulas with the mean update for \( \mathbf{u}_2 \) depending non-linearly on \( \mathbf{u}_1 \) in general.

The proof of Proposition 1 is a direct calculation similar to those in refs. 21 and 22 for other formulations of Gaussian mixtures over the entire state space and the explicit formulas can be found in SI Text. However, there is a crucial difference that here conditional Gaussian mixtures are applied in the reduced subspace \( \mathbf{u}_2 \) blended with particle filter approximations only in the lower-dimensional subspace, \( \mathbf{u}_1 \), unlike the previous work.

Here we consider algorithms for filtering turbulent dynamical systems with the form

\[
u_1 = \mathbf{L} \mathbf{u} + B(\mathbf{u}, \mathbf{u}) + \mathbf{F},
\]

where \( B(\mathbf{u}, \mathbf{u}) \) involves energy-conserving nonlinear interactions with \( u \mathbf{B}(\mathbf{u}, \mathbf{u}) = 0 \) whereas \( \mathbf{L} \) is a linear operator including damping as well as anisotropic physical effects; many turbulent dynamical systems in geosciences and engineering have the structure shown in [6] (23, 24).

Application to Multiscale Filtering of a Slow–Fast System. A typical direct use of the above formalism is briefly sketched. In many applications in the geosciences, there is a fixed subspace \( \mathbf{u}_1 \) representing the slow (vortical) waves and a fixed subspace \( \mathbf{u}_2 \) representing the fast (gravity) waves with observations of pressure and velo-city for example, which naturally mix the slow and fast waves at each analysis step (15, 25–27). A small parameter \( \varepsilon \approx 1 \) characterizes the ratio of the fast timescale to the slow timescale. A well-known formalism for stochastic mode reduction has been developed for such multiscale systems (28, 29) with a simplified forecast model valid in the limit \( \varepsilon \to 0 \) with the form

\[
p(\mathbf{u}_1, \mathbf{u}_2)(t) = p(\mathbf{u}_1(t)|\mathbf{p}(\mathbf{u}_2|\mathbf{u}_1)) = p(\mathbf{u}_1(t)) \mathcal{N}(0, \mathbf{R}_2),
\]

where \( p(\mathbf{u}_1(t)) \) satisfies a reduced Fokker–Planck equation for the slow variables alone and \( \mathbf{R}_2 \) is a suitable background covariance matrix for the fast variables. A trivial application of Proposition 1 guarantees that there is a simplified algorithm consisting of a particle filter for the slow variables alone updated at each analysis step through Proposition 1, which mixes the slow and fast components through the observations. More sophisticated multiscale filtering algorithms with this flavor designed to capture unresolved features of turbulence have been developed recently in ref. 30.

Blended Statistical Nonlinear Forecast Models. Whereas the above formulation can be applied to hybrid particle filters with conditional Kalman filters on fixed subspaces, defined by \( \mathbf{u}_1 \) and \( \mathbf{u}_2 \) as sketched above, a more attractive idea is to use statistical forecast models that adaptively change these subspaces as times evolve in response to the uncertainty without a separation of timescales. Recently Sapsis and Majda (16–19) developed nonlinear statistical forecast models of this type, the quasilinear Gaussian dynamical orthogonality (QG-DO) method and the more sophisticated modified quasilinear Gaussian dynamical orthogonality (MQG-DO) method for turbulent dynamical systems with the structure shown in [6]. It is shown in refs. 16 and 17, respectively, that both QG-DO and MQG-DO have significant skill for uncertainty quantification for turbulent dynamical systems with MQG-DO superior to QG-DO although more calibration is needed for MQG-DO in the statistical steady state.

The starting point for these nonlinear forecast models is a quasilinear Gaussian (QG) statistical closure (16, 18) for [6] or a more statistically accurate modified quasilinear Gaussian (MQG) closure (17, 18); the QG and MQG forecast models incorporate only Gaussian features of the dynamics given by mean and covariance. The more sophisticated QG-DO and MQG-DO methods have an adaptively evolving lower-dimensional subspace where non-Gaussian features are tracked accurately and allow for the exchange of statistical information between the evolving subspace with non-Gaussian statistics and the evolving Gaussian statistical background. We illustrate the simpler QG-DO scheme below and refer to refs. 17 and 18 for the details of the more sophisticated
MQG-DO scheme. The QG-DO statistical forecast model is the following algorithm.

The subspace is represented as

$$u(t) = \mathbf{\bar{u}}(t) + \sum_{j=1}^{s} Y_j(t; \omega)e_j(t), \quad [8]$$

where $e_j(t)$, $j = 1, \cdots, s$ are time-dependent orthonormal modes and $s \ll N$ is the reduction order. The modes and the stochastic coefficients $Y_j(t; \omega)$ evolve according to the dynamical orthogonality (DO) condition (31). In particular, the equations for the QG-DO scheme are as follows.

**Equation for the mean.** The equation for the mean is obtained by averaging the original system equation in [6],

$$\frac{d\mathbf{\bar{u}}}{dt} = (L + D)\mathbf{\bar{u}} + B(\mathbf{\bar{u}}, \mathbf{\bar{u}}) + R_s\mathbf{\bar{u}}(v_i, v_j) + F. \quad [9]$$

**Equation for the stochastic coefficients and the modes.** Both the stochastic coefficients and the modes evolve according to the direct Galerkin projection and the DO condition

$$\frac{dY_j}{dt} = Y_m[(L + D)e_m + B(e_m, e_m) + B(e_m, \mathbf{\bar{u}}) + B(e_m, e_m)(Y_mY_m)^{-1}] \cdot e_j + (Y_mY_n - C_{mn})B(e_m, e_n) \cdot e_i, \quad [10]$$

with $C_{mn} = (Y_mY_n)$. Moreover, the modes evolve according to the equation obtained by stochastic projection of the original equation to the DO coefficients

$$\frac{\partial e_j}{\partial t} = M_j - e_j\mathbf{\bar{u}}(M_j \cdot e_j), \quad [11]$$

with $M_j = (L + D)e_j + B(e_j, e_j) + B(e_j, \mathbf{\bar{u}}) + B(e_j, e_j)(Y_mY_n)^{-1}$.\hspace{1em}

**Equation for the covariance.** The equation for the covariance starts with the exact equation involving third-order moments with approximated nonlinear fluxes

$$\frac{dR}{dt} = LsR + RL_s, + Q_{Fs}, \quad [12]$$

where the nonlinear fluxes are computed using reduced-order information from the DO subspace

$$Q_{Fs} = (Y_mY_n)(B(e_m, e_n) \cdot v_i)(v_i, v_j) + (Y_mY_n)(B(e_m, e_n) \cdot v_j)(v_i, v_k). \quad [13]$$

The last expression is obtained by computing the nonlinear fluxes inside the subspace and projecting those back to the full $N$-dimensional space.

The QG statistical forecast model is the special case with $s = 0$ so that [9] and [12] with $Q_{Fs} \equiv 0$ are approximate statistical dynamical equations for the mean and covariance alone. The MQG-DO algorithm is a more sophisticated variant based on MQG with significantly improved statistical accuracy (16-18).

**Blended Particle Filter Algorithms**

The QG-DO and MQG-DO statistical forecast models are solved by a particle filter or Monte Carlo simulation of the stochastic coefficients $Y_j(t)$, $1 \leq j \leq s$ from [8] through the equations in [10] coupled to the deterministic equations in [9] and [12] for the statistical mean and covariance and the DO basis equations in [11]. Let $E(t) = \{e_1(t), \cdots, e_i(t)\}$ denote the $s$-dimensional stochastic subspace in the forecast; at any analysis time, $t = m\Delta t$, $u_1$ denotes the projection of $u \in \mathbb{R}_s^N$ to $E(t)$. Complete the dynamical basis $E$ with an orthonormal basis $E^\perp$ and define $u_2$ at any analysis time as the projection on $E^\perp$ (the flexibility in choosing $E^\perp$ can be exploited eventually). Thus, forecasts by QG-DO or MQG-DO lead to the following data from the forecast statistics at each analysis time: (i) a particle approximation for the marginal distribution

$$p_-(u_1) = \sum_{j=1}^{Q} p_j\delta(u_1 - u_{1,j}). \quad [14]$$

and (ii) the mean $\mathbf{\bar{u}}_2$ and the covariance matrix

$$R = \begin{pmatrix} R_1 & R_{12} \\ R_{12}^T & R_2 \end{pmatrix}. \quad [15]$$

where $R$ is the covariance matrix in the basis $\{E, E^\perp\}$. To apply Proposition 1 in the analysis step, we need to find a probability density $p_-(u)$ with the form in [1] recovering the statistics in [14] and [15]. Below, for simplicity in exposition, we assume linear observations in [3]. We seek this probability density in the form

$$p_-(u_1, u_2) = \sum_{j=1}^{Q} p_j\delta(u_1 - u_{1,j})\mathcal{N}(\mathbf{\bar{u}}_2, R^2). \quad [16]$$

so that [14] is automatically satisfied by [16] whereas $\mathbf{\bar{u}}_2$, and $R^2$ need to be chosen to satisfy [15]; note that $R_2$ is a constant matrix independent of $j$. Let $g(u)$ denote the expected value of $g$ with respect to $p_-$ so that, for example, $u_1 = \mathbf{\bar{u}}_1$, $u_2 = \mathbf{\bar{u}}_2$, and let $u_1^\perp = u_1 - \mathbf{\bar{u}}_1$, $u_{2,1} = u_2 - \mathbf{\bar{u}}_2$ denote fluctuations about this mean.

Part I of the blended filter algorithm consists of two steps:

1. Solve the following linear system to find the conditional mean $\mathbf{\bar{u}}_2(u_1) = \mathbf{\bar{u}}_2$:

$$\begin{pmatrix} p_1u_1^{11} & \cdots & pQ_1u_1^{1Q} \\
\vdots & \ddots & \vdots \\
p_1u_{N_1}^{11} & \cdots & pQ_{N_1}^{1Q} \end{pmatrix} \begin{pmatrix} u_1^{11} \\
\vdots \\
u_{N_1}^{11} \end{pmatrix} = \begin{pmatrix} R_{12} \\
0 \end{pmatrix}. \quad [17]$$

Note that this is an underdetermined system for a sufficiently large number of particles, $Q$, and “$\sim$” notation is suppressed here.

2. Calculate the covariance $R_2$ in the $u_2$ subspace by requiring from [15] and [16]

$$R_2 = R_2 + \langle u_2 \rangle \otimes \langle u_2 \rangle - \int \mathbf{\bar{u}}_2(u_1) \otimes \mathbf{\bar{u}}_2(u_1)p(u_1)du_1. \quad [18]$$

Any solution of [17] and [18] with $R_2 \geq 0$ automatically guarantees that [14] and [15] are satisfied by the probability density in [16]. Note that $R_2$ is a constant matrix that does not depend on the individual particle weights. This crucial fact makes part II inexpensive.

Part II of the analysis step for the blended particle filter algorithm is an application of Proposition 1 to [16] (details in SI Text).

Use Kalman filter updates in the $u_2$ subspace to get posterior mean and covariance $\mathbf{\bar{u}}_2^\perp$, $R_2^\perp$.\hspace{1em}

Majda et al.
Update the particle weights in the $u_1$ subspace by $p_{j,+} \propto p_j I_j$, with

$$I_j = \exp \left[ \frac{1}{2} \left( n_{2,j}^T (R_j 2) - 1 n_{2,j} - n_{2,j}^T (R_j 2) - 1 n_{2,j} \right) - (v - GEu_{1,j})^T R_0^{-1} (v - GEu_{1,j}) \right].$$

[19]

Normalize the weights $p_{j,+} / \sum p_j I_j$ and resample.

Get the posterior mean and covariance matrix from the posterior particle statistics $\{ u_{1,j}, u_{2,j}, p_{j,+} \}$. This completes the description of the blended particle filter algorithms.

**Realizability in the Blended Filter Algorithms.** A subtle issue in implementing the blended filter algorithms occurs in [17] and [18] from part I of the analysis step: a particular solution of the linear system in [17], denoted here as $LU_2 = F$, may yield a candidate covariance matrix, $R_2$, defined in [18] that is not positive definite; i.e., realizability is violated. Here we exploit the fact that the subspace for particle filtering is low dimensional so that in general, the number of particles satisfies $Q \geq N_1 + 1$ so the linear system $LU_2 = F$ is a strongly underdetermined linear system. The empirical approach that we use here is to seek the least-squares solution of $LU_2 = F$ that minimizes the weighted $L^2$ norm,

$$\sum_{j=1}^Q |\tilde{R}_{2,j}|^2 / p_j.$$

[20]

This solution is given as the standard least-squares solution through the pseudoinverse for the auxiliary variable $\tilde{R}_{2,j} = p_j^{-1/2} u_{2,j}$. Such a least-squares solution guarantees that the trace of $\tilde{R}_2$, defined in [18], is maximized; however, this criterion still does not guarantee that $R_2 = R_2 - \sum \tilde{R}_{2,j} \otimes \tilde{R}_{2,j} / p_j$ is realizable; to help guarantee this, we add extra inflation terms $a_j \in [0,1]$ such that

$$R_2 = R_2 - \sum_j \alpha_j \tilde{R}_{2,j} \otimes \tilde{R}_{2,j}.$$

Here the inflation coefficients $\alpha_j$ can be chosen according to

$$\alpha_j = 1 \text{ if } \tilde{R}_{2,j} \left( R_2 - \sum_k \tilde{R}_{2,k} \otimes \tilde{R}_{2,k} / p_k \right) \tilde{R}_{2,j} > \epsilon_0,$$

$$\alpha_j = 1 - \frac{\epsilon_0 - \tilde{R}_{2,j} \left( R_2 - \sum_k \tilde{R}_{2,k} \otimes \tilde{R}_{2,k} / p_k \right) \tilde{R}_{2,j}}{p_j \tilde{R}_{2,j}}$$

otherwise,

where $\epsilon_0 \ll 1$ is a small number chosen to avoid numerical errors, and $1 \leq j \leq Q$. We find that this empirical approach works very well in practice as shown in subsequent sections. An even simpler but cruder variance inflation algorithm is to set $R_2 = R_2 \geq 0$. Further motivation for the constrained least-squares solution of $LU_2 = F$ minimizing [20] comes from the maximum entropy principle (24); the least biased probability density satisfying $LU_2 = F$ in [17] formally maximizes the entropy of $R_2$; i.e.,

$$\tilde{u}_{2,j} = \arg\max \log \det \left( R_2 - \sum_j p_j \tilde{R}_{2,j} \otimes \tilde{R}_{2,j} \right)$$

$$= \arg\max \log \det \left( I - \sum_j (p_j^{1/2} R_2^{-1/2} \tilde{R}_{2,j}) \otimes (p_j^{1/2} R_2^{-1/2} \tilde{R}_{2,j}) \right).$$

[21]

The high-dimensional nonlinear optimization problem in [21] is too expensive to solve directly but the small amplitude expansion, $det(I - eR) = -etr R + O(e^2)$ of [21] becomes a weighted least-squares optimization problem for the new variable, $\tilde{u}_{2,j} = p_j^{1/2} R_2^{-1/2} \tilde{R}_{2,j}$ constrained by $LU_2 = F$. If we choose $R_2 = I$, the least-squares solution from [20] is recovered. Whereas the algorithm using $\tilde{u}_{2,j}$ has a nice theoretical basis, it requires the singular value decomposition of the large covariance matrix,

![Fig. 1. Comparison of pdfs of the absolute values of the first two leading Fourier modes $\tilde{u}_1, \tilde{u}_2$ (Upper) and pdfs of the forecast error $\tilde{u} - u_{\text{truth}}$. (Lower, only real parts are shown) captured by different filtering methods.](image-url)
and the solution in [20] avoids this expensive procedure. Incidentally, the max-entropy principle alone does not guarantee realizability.

**Numerical Tests of the Blended Particle Filters**

Major challenges for particle filters for large-dimensional turbulent dynamical systems involve capturing substantial non-Gaussian features of the partially observed dynamical system as well as skillful filtering for spatially sparse infrequent high-quality observations of the turbulent signal (chap. 15 of ref. 15). In this last setting, the best ensemble filters require extensive tuning to avoid catastrophic filter divergence and quite often cheap filters based on linear stochastic forecast models are more skillful (15). Here the performance of the blended particle filter is assessed for two stringent test regimes, elucidating the above challenges, for the Lorenz 96 (L-96) model (32, 33); the L-96 model is a 40-dimensional turbulent dynamical system that is a popular test model for filter performance for turbulent dynamical systems (15). The L-96 model is a discrete periodic model given by

\[
\frac{du_i}{dt} = u_{i-1}(u_{i+1} - u_{i-2}) - u_i + F, \quad i = 0, \cdots, J - 1,
\]  

with \( J = 40 \) and \( F \) the deterministic forcing parameter. The model is designed to mimic baroclinic turbulence in the midlatitude atmosphere with the effects of energy conserving nonlinear advection and dissipation represented by the first two terms in [22]. For sufficiently strong constant forcing values such as \( F = 5, 8, \) or 16, the L-96 model is a prototype turbulent dynamical system that exhibits features of weakly chaotic turbulence (\( F = 5 \)), strongly chaotic turbulence (\( F = 8 \)), and strong turbulence (\( F = 16 \)) (15, 24, 34). Because the L-96 model is translation invariant, 20 discrete Fourier modes can be used to study its statistical properties. In all filtering experiments described below with the blended particle filters, we use \( s = 5 \) with 10,000 particles. Thus, non-Gaussian effects are captured in a 5-dimensional subspace through a particle filter interacting with a low-order Gaussian statistical forecast model in the remaining 35 dimensions. The numbers of positive Lyapunov exponents on the attractor for the forcing values \( F = 5, 8, 16 \) considered here are 9, 13, and 16, respectively (34), so the 5-dimensional adaptive subspace with

\[
\text{particle filtering can contain at most half of the unstable directions on the attractor; also, non-Gaussian statistics are most prominent in the weakly turbulent regime, } F = 5, \text{ with nearly Gaussian statistics for } F = 16, \text{ the strongly turbulent regime, and intermediate statistical behavior for } F = 8.
\]

**Capturing Non-Gaussian Statistics Through Blended Particle Filters.**

As mentioned above, the L-96 model in the weakly turbulent regime with \( F = 5 \) has nine positive Lyapunov exponents on the attractor whereas Fourier modes \( \bar{u}_7 \) and \( \bar{u}_8 \) are the two leading empirical orthogonal functions (EOFs) that contain most of the energy. As shown in Fig. 1, where the probability density functions (pdfs) of \( [\bar{u}_7], [\bar{u}_8] \) are plotted, there is significant non-Gaussian behavior in these modes because the pdfs for \( [\bar{u}_7], [\bar{u}_8] \) are far from a Rayleigh distribution; see SI Text for the scatter plot of their joint distribution exhibiting strongly non-Gaussian behavior. For the filtering experiments below, sparse spatial observations are used with every fourth grid point observed with moderate observational noise variance \( n_0 = 2 \) and moderate observation frequency \( \Delta t = 1 \) compared with the decorrelation time 4.4 for \( F = 5 \). We tested the QG-DO and MQG-DO blended filters as well as the Gaussian MQG filter and the ensemble adjustment Kalman filter (EAKF) with optimal tuned inflation and localization with 10,000 ensemble members. All four filters were run for many assimilation steps and forecast pdfs for \( [\bar{u}_7], [\bar{u}_8] \) as well as forecast error pdfs for the real part of \( \bar{u}_7, \bar{u}_8 \) are plotted in Fig. 1. The blended MQG-DO filter accurately captures the non-Gaussian features and has the tightest forecast error pdf, the Gaussian MQG filter outperforms the blended QG-DO filter with a tighter forecast error distribution, and EAKF yields incorrect Gaussian distributions with the largest forecast error spread. The rms error and pattern correlation plots reported in SI Text confirm the above behavior as well as the scatter plots of the joint pdf for \( [\bar{u}_7], [\bar{u}_8] \) reported there. This example illustrates that a Gaussian filter like MQG with an accurate statistical forecast operator can have a sufficiently tight forecast error pdf and be a very good filter yet can fail to capture significant non-Gaussian features accurately. On the other hand, for the QG-DO blended algorithm in this example, the larger forecast errors of the QG dynamics compared with MQG swamp the effect of the blended particle filter. However, all three methods significantly improve upon the performance of EAKF with many ensemble members.

**Figure 2.** Comparison of rms errors (Left) and pattern correlations (Right) between different filtering methods in regimes \( F = 8 \) (Top) and \( F = 16 \) (Middle and Bottom) with sparse infrequent high-quality observations.

Majda et al.
Filter Performance with Sparse Infrequent High-Quality Observations.

Demanding tests for filter performance are the regimes of spatially sparse, infrequent in time, high-quality (low observational noise) observations for a strongly turbulent dynamical system. Here the performances of the blended MOG-DO and DO-DO filters as well as the MOG filter are assessed in this regime. For the strongly chaotic regime, $F = 8$, for the L-96 model, observations are taken every fourth grid point with variance $r_0 = 0.01$ and observation time $\Delta t = 0.25$, which is nearly the decorrelation time, 0.33; the performance of EAKF is as well as that of the rank histogram and maximum entropy particle filters has already been assessed for this difficult test problem in figure 15.14 of ref. 15 with large intervals in time with filter divergence (rms errors much larger than 1) for all three methods. A similar test problem for the strongly turbulent regime, $F = 16$, with spatial observations every fourth grid point with $r_0 = 0.01$ and $\Delta t = 0.1$ compared with the decorrelation time, 0.12, is used here. In all examples with L-96 tested here, we find that the MOG-DO algorithm with the approximation described in the paragraph below (Eq. 20) is always realizable and is the most robust accurate filter; on the other hand, for the DO-DO filtering algorithm, the performance of the blended algorithm with crude variance inflation, $R_1 = R_2$, significantly outperforms the basic DO-DO blended algorithms due to incorrect energy transfers in the OG forecast models for the long forecast times used here (SI Text). Fig. 2 reports the filtering performance of the MOG-DO and DO-DO blended filters and the MOG Gaussian filter in these tough regimes for $F = 8, 16$ through the rms error and pattern correlation. There are no strong filter divergences with the MOG-DO and DO-DO blended filters for both $F = 8$ and $F = 16$ in contrast to other methods as shown in figure 15.14 of ref. 15. The much cheaper MOG filter for $F = 16$ exhibits a long initial regime of filter divergence but eventually settles down to comparable filter performance to that of the blended filters. The blended MOG-DO filter is the most skillful robust filter over all these strongly turbulent regimes, $F = 8, 16$, as the observational noise and observation time are varied; see the examples in SI Text.

Concluding Discussion

Blended particle filters that capture non-Gaussian features in an adaptive evolving low-dimensional subspace through particles interacting with evolving Gaussian statistics on the remaining phase space are introduced here. These blended particle filters have been developed here through a mathematical formalism involving conditional Gaussian mixtures (35) combined with statistically nonlinear forecast models developed recently (16–18) with high skill for uncertainty quantification, which are compatible with this structure. Stringent test cases for filtering involving the 40-dimensional L-96 model with a 5-dimensional adaptive subspace for nonlinear filtering in various regimes of chaotic dynamics with at least nine positive Lyapunov exponents are used here. These test cases demonstrate the high skill of these blended filters in capturing both non-Gaussian dynamical features and crucial nonlinear statistics for accurate filtering in extreme regimes with sparse infrequent high-quality observations.

The formalism developed is also useful for multiscale filtering of turbulent systems and a simple application has been sketched here.

ACKNOWLEDGMENTS. The research of A.J.M. is partially supported by Office of Naval Research (ONR) grants, ONR-Multidisciplinary University Research Initiative 25-74200-F7112, and ONR-Departmental Research Initiative N0014-10-1-0554. D.Q. is supported as a graduate research assistant in the first grant and the research of T.P.S. was partially supported through the second grant, when T.P.S. was a postdoctoral fellow at the Courant Institute.