

# Blended particle filters for large-dimensional chaotic dynamical systems

Andrew J. Majda<sup>a,1</sup>, Di Qi<sup>a</sup>, and Themistoklis P. Sapsis<sup>b,1</sup>

<sup>a</sup>Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; and <sup>b</sup>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 blended filtering 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.**

curse of dimensionality | hybrid methods

**M**any contemporary problems in science ranging from protein folding in molecular dynamics to scaling up of small-scale effects in nanotechnology to making accurate predictions of the coupled atmosphere–ocean system involve partial observations of extremely complicated large-dimensional chaotic dynamical systems. Filtering is the process of obtaining the best statistical estimate of a natural system from partial observations of the true signal from nature. In many contemporary applications in science and engineering, real-time filtering or data assimilation of a turbulent signal from nature involving many degrees of freedom is needed to make accurate predictions of the future state.

Particle filtering of low-dimensional dynamical systems is an established discipline (1). When the system is low dimensional, Monte Carlo approaches such as the 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  $\mathbf{u} \in \mathbb{R}^N$ , there are two subspaces that typically evolve adaptively in time, where  $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ ,  $\mathbf{u}_j \in \mathbb{R}^{N_j}$ , and  $N_1 + N_2 = N$ , with the property that  $N_1$  is low dimensional enough so that the non-Gaussian statistics of  $\mathbf{u}_1$  can be calculated from a particle filter whereas the evolving statistics of  $\mathbf{u}_2$  are conditionally Gaussian given  $\mathbf{u}_1$ . 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

## Significance

**Combining large uncertain computational models with big noisy datasets is a formidable problem throughout science and engineering. These are especially difficult issues when real-time state estimation and prediction are needed such as, for example, in weather forecasting. Thus, 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. New blended particle filters are developed in this paper. These algorithms exploit the physical structure of turbulent dynamical systems and capture non-Gaussian features in an adaptively evolving low-dimensional subspace through particles interacting with evolving Gaussian statistics on the remaining portion of the phase space.**

Author contributions: A.J.M. designed research; A.J.M., D.Q., and T.P.S. performed research; and A.J.M. and D.Q. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence may be addressed. E-mail: jonjon@cims.nyu.edu or sapsis@mit.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1405675111/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1405675111/-DCSupplemental).

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  $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ ,  $\mathbf{u}_j \in \mathbb{R}^{N_j}$ , 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

$$p_-(\mathbf{u}) = p_-(\mathbf{u}_1)p_-^G(\mathbf{u}_2|\mathbf{u}_1), \quad [1]$$

where  $p_-^G(\mathbf{u}_2|\mathbf{u}_1)$  is a Gaussian distribution determined by the conditional mean and covariance,  $p_-^G(\mathbf{u}_2|\mathbf{u}_1) = \mathcal{N}(\bar{\mathbf{u}}_2^-(\mathbf{u}_1), R_2^-(\mathbf{u}_1))$ . We assume that the marginal distribution,  $p_-(\mathbf{u}_1)$ , is approximated by  $Q$  particles,

$$p_-(\mathbf{u}_1) = \sum_{j=1}^Q p_{j,-} \delta(\mathbf{u}_1 - \mathbf{u}_{1j}), \quad [2]$$

with nonnegative particle weights,  $p_{j,-}$  with  $\sum_j p_{j,-} = 1$ . 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

$$\mathbf{v} = G(\mathbf{u}) + \boldsymbol{\sigma}_0 = G_0(\mathbf{u}_1) + G_1(\mathbf{u}_1)\mathbf{u}_2 + \boldsymbol{\sigma}_0, \quad [3]$$

where  $G_1(\mathbf{u}_1)$  has rank  $M$  and the observational noise,  $\boldsymbol{\sigma}_0$ , is Gaussian  $\boldsymbol{\sigma}_0 = \mathcal{N}(\mathbf{0}, R_0)$ . 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,  $p_-(\mathbf{u}_1, \mathbf{u}_2)$  given from the forecast distribution

$$p_-(\mathbf{u}) = \sum_{j=1}^Q p_{j,-} \delta(\mathbf{u}_1 - \mathbf{u}_{1j}) \mathcal{N}(\bar{\mathbf{u}}_{2j}^-, R_{2j}^-), \quad [4]$$

and compute the posterior distribution in the analysis step through Bayes' theorem,  $p_+(\mathbf{u}|\mathbf{v}) \sim p(\mathbf{v}|\mathbf{u})p_-(\mathbf{u})$ . 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,  $p_{j,+}$ ,  $1 \leq j \leq Q$ , and conditional mean,  $\bar{\mathbf{u}}_{2j}^+$ , and covariance,  $R_{2j}^+$ , so that

$$p_+(\mathbf{u}) = \sum_{j=1}^Q p_{j,+} \delta(\mathbf{u}_1 - \mathbf{u}_{1j}) \mathcal{N}(\bar{\mathbf{u}}_{2j}^+, R_{2j}^+). \quad [5]$$

In fact, the distributions  $\mathcal{N}(\bar{\mathbf{u}}_{2j}^+, R_{2j}^+)$  are updated by suitable Kalman filter formulas with the mean update for  $\bar{\mathbf{u}}_{2j}^+$  depending nonlinearly on  $\mathbf{u}_{1j}$  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

$$\mathbf{u}_t = L\mathbf{u} + \mathbf{B}(\mathbf{u}, \mathbf{u}) + \mathbf{F}, \quad [6]$$

where  $\mathbf{B}(\mathbf{u}, \mathbf{u})$  involves energy-conserving nonlinear interactions with  $\mathbf{u} \cdot \mathbf{B}(\mathbf{u}, \mathbf{u}) = 0$  whereas  $L$  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 velocity for example, which naturally mix the slow and fast waves at each analysis step (15, 25–27). A small parameter  $\epsilon \ll 1$  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  $\epsilon \rightarrow 0$  with the form

$$\begin{aligned} p(\mathbf{u}_1, \mathbf{u}_2)(t) &= p(\mathbf{u}_1)(t)p(\mathbf{u}_2|\mathbf{u}_1) \\ &= p(\mathbf{u}_1)(t)\mathcal{N}(0, R_2), \end{aligned} \quad [7]$$

where  $p(\mathbf{u}_1)(t)$  satisfies a reduced Fokker–Planck equation for the slow variables alone and  $R_2$  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 time evolves 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

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

where  $\mathbf{e}_j(t)$ ,  $j=1, \dots, s$  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],

$$\frac{d\bar{\mathbf{u}}}{dt} = (L + D)\bar{\mathbf{u}} + \mathbf{B}(\bar{\mathbf{u}}, \bar{\mathbf{u}}) + R_{ij}\mathbf{B}(\mathbf{v}_i, \mathbf{v}_j) + \mathbf{F}. \quad [9]$$

**Equation for the stochastic coefficients and the modes.** Both the stochastic coefficients and the modes evolve according to the DO equations (31). The coefficient equations are obtained by a direct Galerkin projection and the DO condition

$$\begin{aligned} \frac{dY_i}{dt} = & Y_m[(L + D)e_m + \mathbf{B}(\bar{\mathbf{u}}, e_m) + \mathbf{B}(e_m, \bar{\mathbf{u}})] \cdot \mathbf{e}_i \\ & + (Y_m Y_n - C_{mn})\mathbf{B}(e_m, e_n) \cdot \mathbf{e}_i \end{aligned} \quad [10]$$

with  $C_{mn} = \langle Y_m Y_n^* \rangle$ . Moreover, the modes evolve according to the equation obtained by stochastic projection of the original equation to the DO coefficients

$$\frac{\partial \mathbf{e}_i}{\partial t} = \mathbf{M}_i - \mathbf{e}_j(\mathbf{M}_i \cdot \mathbf{e}_j), \quad [11]$$

with

$$\begin{aligned} \mathbf{M}_i = & (L + D)\mathbf{e}_i + \mathbf{B}(\bar{\mathbf{u}}, \mathbf{e}_i) + \mathbf{B}(\mathbf{e}_i, \bar{\mathbf{u}}) \\ & + \mathbf{B}(e_m, e_n)(Y_m Y_n Y_k) C_{ik}^{-1}. \end{aligned}$$

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

$$\frac{dR}{dt} = L_v R + R L_v^* + Q_{F,s}, \quad [12]$$

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

$$\begin{aligned} Q_{F,s} = & \langle Y_m Y_n Y_k \rangle (\mathbf{B}(e_m, e_n) \cdot \mathbf{v}_i) (\mathbf{v}_j \cdot e_k) \\ & + \langle Y_m Y_n Y_k \rangle (\mathbf{B}(e_m, e_n) \cdot \mathbf{v}_j) (\mathbf{v}_i \cdot e_k). \end{aligned} \quad [13]$$

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  $Q_{F,s} \equiv 0$  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  $E(t) = \{\mathbf{e}_1(t), \dots, \mathbf{e}_s(t)\}$  denote the  $s$ -dimensional

stochastic subspace in the forecast; at any analysis time,  $t = m\Delta t$ ,  $\mathbf{u}_1$  denotes the projection of  $\mathbf{u} \in \mathbb{R}^N$  to  $E(t)$ . Complete the dynamical basis  $E$  with an orthonormal basis  $E^\perp$  and define  $\mathbf{u}_2$  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

$$p_-(\mathbf{u}_1) = \sum_{j=1}^Q p_{j,-} \delta(\mathbf{u}_1 - \mathbf{u}_{1,j}) \quad [14]$$

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

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

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  $p_-(\mathbf{u})$  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

$$p_-(\mathbf{u}_1, \mathbf{u}_2) = \sum_{j=1}^Q p_{j,-} \delta(\mathbf{u}_1 - \mathbf{u}_{1,j}) \mathcal{N}(\bar{\mathbf{u}}_{2,j}^-, R_2^-) \quad [16]$$

so that [14] is automatically satisfied by [16] whereas  $\bar{\mathbf{u}}_{2,j}^-$  and  $R_2^-$  need to be chosen to satisfy [15]; note that  $R_2^-$  is a constant matrix independent of  $j$ . Let  $\langle g(\mathbf{u}) \rangle$  denote the expected value of  $g$  with respect to  $p_-$  so that, for example,  $\langle \mathbf{u}_1 \rangle = \bar{\mathbf{u}}_1^-$ ,  $\langle \mathbf{u}_2 \rangle = \bar{\mathbf{u}}_2^-$ , and let  $\mathbf{u}'_1 = \mathbf{u}_1 - \langle \mathbf{u}_1 \rangle$ ,  $\mathbf{u}'_2 = \mathbf{u}_2 - \langle \mathbf{u}_2 \rangle$  denote fluctuations about this mean. Part I of the blended filter algorithm consists of two steps:

Solve the following linear system to find the conditional mean  $\bar{\mathbf{u}}_2(\mathbf{u}_{1,j}) = \bar{\mathbf{u}}_{2,j}^-$ :

$$\begin{bmatrix} p_1 u'_{1,1} & \cdots & p_Q u'_{1,1} \\ \vdots & \ddots & \vdots \\ p_1 u'_{1,N_1} & \cdots & p_Q u'_{1,N_1} \\ p_1 & \cdots & p_Q \end{bmatrix} \begin{bmatrix} u'_{2,1} & \cdots & u'_{2,N_2} \\ u'_{2,1} & \cdots & u'_{2,N_2} \\ \vdots & \ddots & \vdots \\ u'_{2,1} & \cdots & u'_{2,N_2} \end{bmatrix} = \begin{bmatrix} R_{12} \\ 0 \end{bmatrix}. \quad [17]$$

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

Calculate the covariance  $R_2^-$  in the  $\mathbf{u}_2$  subspace by requiring from [15] and [16]

$$\begin{aligned} R_2^- = & R_2 + \langle \mathbf{u}_2 \rangle \otimes \langle \mathbf{u}_2 \rangle - \int \bar{\mathbf{u}}_2(\mathbf{u}_1) \otimes \bar{\mathbf{u}}_2(\mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_1 \\ = & R_2 - \int \bar{\mathbf{u}}_2'(\mathbf{u}_1) \otimes \bar{\mathbf{u}}_2'(\mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_1 \\ = & R_2 - \sum_j \bar{\mathbf{u}}_{2,j}^- \otimes \bar{\mathbf{u}}_{2,j}^- p_{j,-}. \end{aligned} \quad [18]$$

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  $\mathbf{u}_2$  subspace to get posterior mean and covariance  $\bar{\mathbf{u}}_{2,j}^+$ ,  $\bar{R}_2^+$ .

Update the particle weights in the  $\mathbf{u}_1$  subspace by  $p_{j,+} \propto p_{j,-} I_j$ , with

$$I_j = \exp \left[ \frac{1}{2} \left( \bar{\mathbf{u}}_{2,j}^{+\text{T}} (\bar{R}_2^+)^{-1} \bar{\mathbf{u}}_{2,j}^+ - \bar{\mathbf{u}}_{2,j}^{-\text{T}} (\bar{R}_2^-)^{-1} \bar{\mathbf{u}}_{2,j}^- - (\mathbf{v} - G E \mathbf{u}_{1,j})^{\text{T}} R_0^{-1} (\mathbf{v} - G E \mathbf{u}_{1,j}) \right) \right]. \quad [19]$$

Normalize the weights  $p_{j,+} = p_{j,-} I_j / \sum_j p_{j,-} I_j$  and resample.

Get the posterior mean and covariance matrix from the posterior particle statistics  $\{\mathbf{u}_{1,j}, \bar{\mathbf{u}}_{2,j}^+, p_{j,+}\}$ .

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  $\mathcal{L}U_2 = \mathcal{F}$ , 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 that the linear system  $\mathcal{L}U_2 = \mathcal{F}$  is a strongly underdetermined linear system. The empirical approach that we use here is to seek the least-squares solution of  $\mathcal{L}U_2 = \mathcal{F}$  that minimizes the weighted  $L^2$  norm,

$$\sum_{j=1}^Q |\bar{\mathbf{u}}_{2,j}^+|^2 p_j. \quad [20]$$

This solution is given as the standard least-squares solution through the pseudoinverse for the auxiliary variable  $\bar{\mathbf{v}}_{2,j}^+ = p_j^{1/2} \bar{\mathbf{u}}_{2,j}^+$ . Such a least-squares solution guarantees that the trace of  $R_2^-$ , defined in [18], is maximized; however, this criterion still does not guarantee that  $R_2^- = R_2 - \sum_j \bar{\mathbf{u}}_{2,j}^+ \bar{\mathbf{u}}_{2,j}^{+\text{T}} p_j$  is realizable; to help guarantee this, we add extra inflation terms  $\alpha_j \in [0, 1]$  such that

$$R_2^- = R_2 - \sum_j \alpha_j p_j \bar{\mathbf{u}}_{2,j}^+ \bar{\mathbf{u}}_{2,j}^{+\text{T}}.$$

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

$$\alpha_j = 1 \quad \text{if} \quad \bar{\mathbf{u}}_{2,j}^{+\text{T}} \left( R_2 - \sum_k \bar{\mathbf{u}}_{2,k}^+ \bar{\mathbf{u}}_{2,k}^{+\text{T}} p_k \right) \bar{\mathbf{u}}_{2,j}^+ > \epsilon_0,$$

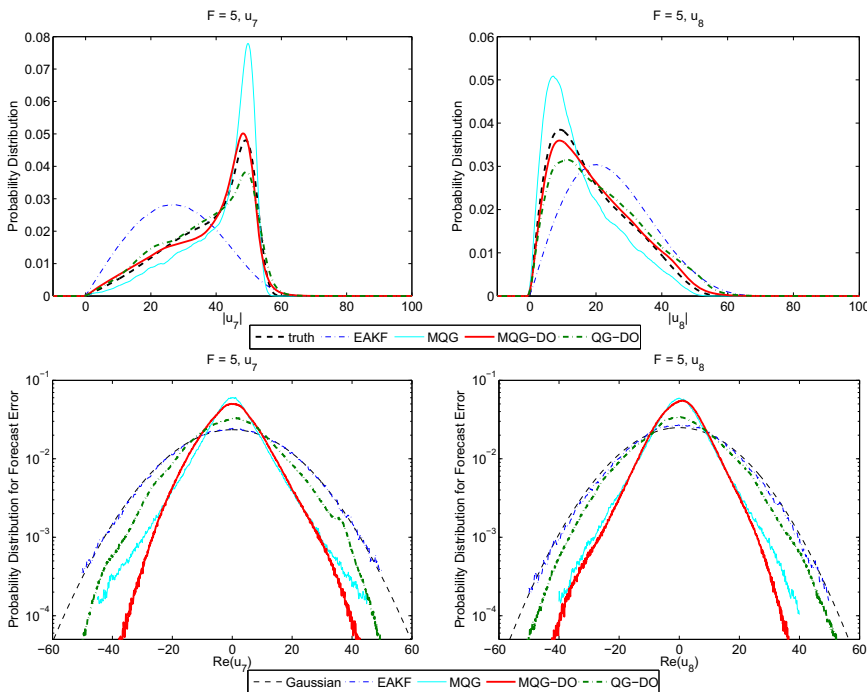
$$\alpha_j = 1 - \frac{\epsilon_0 - \bar{\mathbf{u}}_{2,j}^{+\text{T}} \left( R_2 - \sum_k \bar{\mathbf{u}}_{2,k}^+ \bar{\mathbf{u}}_{2,k}^{+\text{T}} p_k \right) \bar{\mathbf{u}}_{2,j}^+}{p_j |\bar{\mathbf{u}}_{2,j}^+|^4}, \quad \text{otherwise,}$$

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  $\mathcal{L}U_2 = \mathcal{F}$  minimizing [20] comes from the maximum entropy principle (24); the least biased probability density satisfying  $\mathcal{L}U_2 = \mathcal{F}$  in [17] formally maximizes the entropy of  $R_2^-$ ; i.e.,

$$\begin{aligned} \bar{\mathbf{u}}_{2,j}^+ &= \text{argmax} \log \det \left( R_2 - \sum_j p_j \bar{\mathbf{u}}_{2,j}^+ \bar{\mathbf{u}}_{2,j}^{+\text{T}} \right) \\ &= \text{argmax} \log \det \left( I - \sum_j \left( p_j^{1/2} R_2^{-1/2} \bar{\mathbf{u}}_{2,j}^+ \right) \otimes \left( p_j^{1/2} R_2^{-1/2} \bar{\mathbf{u}}_{2,j}^+ \right) \right). \end{aligned} \quad [21]$$

The high-dimensional nonlinear optimization problem in [21] is too expensive to solve directly but the small amplitude expansion,  $\det(I - \epsilon R) = -\epsilon \text{tr} R + O(\epsilon^2)$  of [21] becomes a weighted least-squares optimization problem for the new variable,  $\bar{\mathbf{v}}_{2,j}^+ = p_j^{1/2} R_2^{-1/2} \bar{\mathbf{u}}_{2,j}^+$  constrained by  $\mathcal{L}U_2 = \mathcal{F}$ . If we choose  $R_2 = I$ , the least-squares solution from [20] is recovered. Whereas the algorithm using  $\bar{\mathbf{v}}_{2,j}^+$  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  $\hat{u}_7, \hat{u}_8$  (Upper) and pdfs of the forecast error  $u_i - u_{i,\text{truth}}$  (Lower, only real parts are shown) captured by different filtering methods.

$R_2$ , 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 turbulent 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

$$\frac{du_i}{dt} = u_{i-1}(u_{i+1} - u_{i-2}) - u_i + F, \quad i = 0, \dots, J-1, \quad [22]$$

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

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,  $F = 5$ , with nearly Gaussian statistics for  $F = 16$ , 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  $\hat{u}_7$  and  $\hat{u}_8$  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  $|\hat{u}_7|$ ,  $|\hat{u}_8|$  are plotted, there is significant non-Gaussian behavior in these modes because the pdfs for  $|\hat{u}_7|$ ,  $|\hat{u}_8|$  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  $r_0 = 2$  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  $|\hat{u}_7|$ ,  $|\hat{u}_8|$  as well as forecast error pdfs for the real part of  $\hat{u}_7$ ,  $\hat{u}_8$  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  $|\hat{u}_7|$ ,  $|\hat{u}_8|$  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.

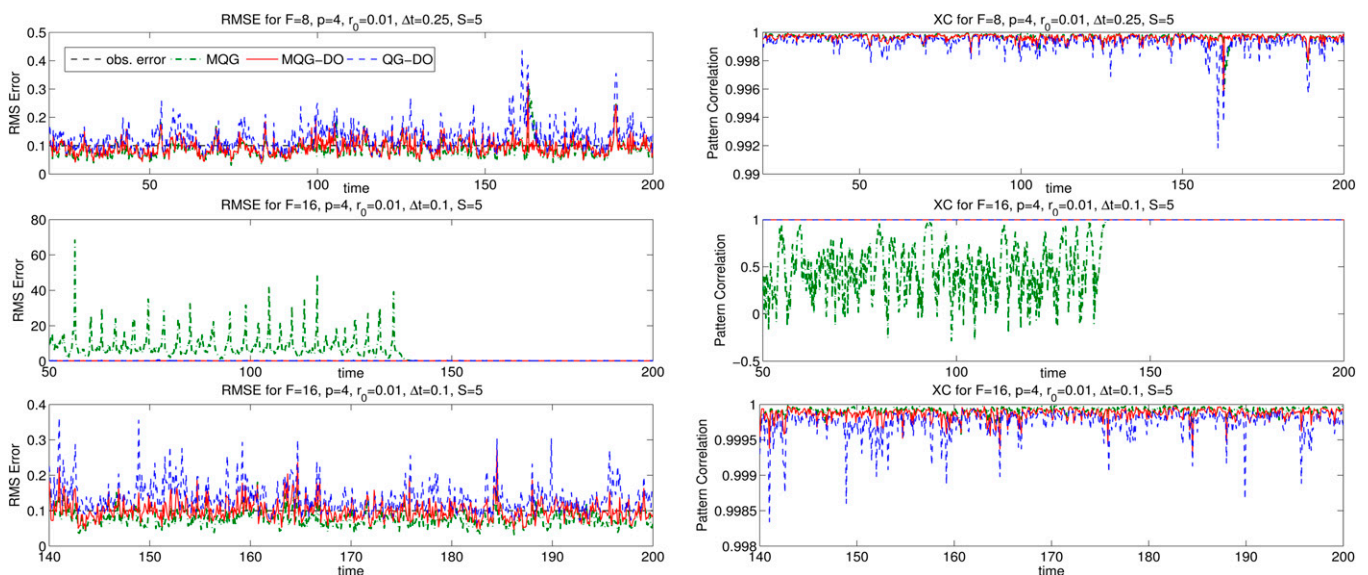


Fig. 2. Comparison of rms errors (Left) and pattern correlations (Right) between different filtering methods in regimes  $F = 8$  (Top) and  $F = 16$  (Middle and Bottom) with sparse infrequent high-quality observations.

### 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 MQG-DO and QG-DO filters as well as the MQG 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 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 MQG-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 QG-DO filtering algorithm, the performance of the blended algorithm with crude variance inflation,  $R_2^+ = R_2$ , significantly outperforms the basic QG-DO blended algorithm due to incorrect energy transfers in the QG forecast models for the long forecast times used here (SI Text). Fig. 2 reports the filtering performance of the MQG-DO and QG-DO blended filters and the MQG 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 MQG-DO and QG-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 MQG filter for  $F=16$  exhibits a long initial regime of filter divergence but eventually settles down to comparable filter per-

formance to that of the blended filters. The blended MQG-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.

- Bain A, Crisan D (2009) *Fundamentals of Stochastic Filtering*, Stochastic Modeling and Applied Probability (Springer, New York), Vol 60.
- Moral PD, Jacod J (2001) Interacting particle filtering with discrete observations. *Sequential Monte Carlo Methods in Practice*, eds Doucet A, de Freitas N, Gordon N (Springer, New York), pp 43–75.
- Bengtsson T, Bickel P, Li B (2008) Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. *IMS Collections in Probability and Statistics: Essays in Honor of David A. Freedman*, eds Nolan D, Speed T (Inst Math Stat, Beachwood, OH), Vol 2, 316–334.
- Bickel P, Li B, Bengtsson T (2008) Sharp failure rates for the bootstrap particle filter in high dimensions. *IMS Collections: Pushing the Limits of Contemporary Statistics: Contributions in Honor of J. K. Ghosh*, eds Clarke B, Ghosal S (Inst Math Stat, Beachwood, OH), Vol 3, pp 318–329.
- Berliner LM, Milliff RF, Wikle CK (2003) Bayesian hierarchical modeling of air-sea interaction. *J Geophys Res Oceans* (1978–2012) 108:3104–3120.
- Ghil M, Malanotte-Rizzoli P (1991) Data assimilation in meteorology and oceanography. *Adv Geophys* 33:141–266.
- Chorin AJ, Krause P (2004) Dimensional reduction for a Bayesian filter. *Proc Natl Acad Sci USA* 101(42):15013–15017.
- Farrell BF, Ioannou PJ (2001) State estimation using a reduced-order Kalman filter. *J Atmos Sci* 58:3666–3680.
- Anderson B, Moore J (1979) *Optimal Filtering* (Prentice-Hall, Englewood Cliffs, NJ).
- Evensen G (2003) The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn* 53:343–367.
- Bishop CH, Etherton BJ, Majumdar SJ (2001) Adaptive sampling with the ensemble transform Kalman filter, part I: Theoretical aspects. *Mon Weather Rev* 129:420–436.
- Anderson JL (2001) An ensemble adjustment Kalman filter for data assimilation. *Mon Weather Rev* 129:2884–2903.
- Anderson JL (2003) A local least squares framework for ensemble filtering. *Mon Weather Rev* 131:634–642.
- Szunyogh I, et al. (2005) Assessing a local ensemble Kalman filter: Perfect model experiments with the national centers for environmental prediction global model. *Tellus Ser A Dyn Meteorol Oceanogr* 57:528–545.
- Majda AJ, Harlim J (2012) *Filtering Complex Turbulent Systems* (Cambridge Univ Press, Cambridge, UK).
- Sapsis TP, Majda AJ (2013) Blended reduced subspace algorithms for uncertainty quantification of quadratic systems with a stable mean state. *Physica D* 258:61–76.
- Sapsis TP, Majda AJ (2013) Blending modified Gaussian closure and non-Gaussian reduced subspace methods for turbulent dynamical systems. *J Nonlin Sci* 23(6):1039–1071.
- Sapsis TP, Majda AJ (2013) A statistically accurate modified quasilinear Gaussian closure for uncertainty quantification in turbulent dynamical systems. *Physica D* 252: 34–45.
- Sapsis TP, Majda AJ (2013) Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems. *Proc Natl Acad Sci USA* 110(34): 13705–13710.
- Lermusiaux PF (2006) Uncertainty estimation and prediction for interdisciplinary ocean dynamics. *J Comput Phys* 217(1):176–199.
- Sorenson HW, Alspach DL (1971) Recursive Bayesian estimation using Gaussian sums. *Automatica* 7:465–479.
- Hoteit I, Luo X, Pham DT (2012) Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters. *Mon Weather Rev* 140:528–542.
- Salmon R (1998) *Lectures on Geophysical Fluid Dynamics* (Oxford Univ Press, Oxford, UK).
- Majda AJ, Wang X (2006) *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows* (Cambridge Univ Press, Cambridge, UK).
- Majda AJ (2003) *Introduction to PDEs and Waves for the Atmosphere and Ocean*, Courant Lecture Notes in Mathematics (Am Math Soc, Providence, RI), Vol 9.
- Gershgorin B, Majda AJ (2008) A nonlinear test model for filtering slow-fast systems. *Commun Math Sci* 6:611–649.
- Subramanian AC, Hoteit I, Cornuelle B, Miller AJ, Song H (2012) Linear versus nonlinear filtering with scale-selective corrections for balanced dynamics in a simple atmospheric model. *J Atmos Sci* 69(11):3405–3419.
- Majda AJ, Timofeyev I, Vanden-Eijnden E (2001) A mathematical framework for stochastic climate models. *Commun Pure Appl Math* 54:891–974.
- Majda AJ, Franzke C, Khouider B (2008) An applied mathematics perspective on stochastic modelling for climate. *Philos Trans R Soc A Math Phys Eng Sci* 366: 2427–2453.
- Grooms I, Lee Y, Majda AJ (2014) Ensemble Kalman filters for dynamical systems with unresolved turbulence. *J Comput Phys*, in press.
- Sapsis TP, Lermusiaux PF (2009) Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Physica D* 238:2347–2360.
- Lorenz EN (1996) Predictability: A problem partly solved. *Proc Semin Predict* 1(1):1–18.
- Lorenz EN, Emanuel KA (1998) Optimal sites for supplementary weather observations: Simulation with a small model. *J Atmos Sci* 55:399–414.
- Abramov RV, Majda AJ (2007) Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems. *Nonlinearity* 20(12):2793–2821.
- Doucet A, Godsill S, Andrieu C (2000) On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat Comput* 10(3):197–208.

# Supporting Information

Majda et al. 10.1073/pnas.1405675111

## SI Text

### 1. Proof of Proposition 1

**Proposition 1.** Assume the prior distribution from the forecast is the blended particle filter conditional Gaussian distribution

$$p_-(\mathbf{u}) = \sum_{j=1}^Q p_{j,-} \delta(\mathbf{u}_1 - \mathbf{u}_{1,j}) \mathcal{N}(\bar{\mathbf{u}}_{2,j}^-, R_{2,j}^-) \quad [\text{S1}]$$

and assume the observations have the structure

$$\mathbf{v} = G(\mathbf{u}) + \boldsymbol{\sigma}_0 = G_0(\mathbf{u}_1) + G_1(\mathbf{u}_1)\mathbf{u}_2 + \boldsymbol{\sigma}_0; \quad [\text{S2}]$$

then the posterior distribution in the analysis step taking into account the observations in [S2] is also a blended particle filter conditional Gaussian distribution; i.e., there are explicit formulas for the updated weights,  $p_{j,+}$ ,  $1 \leq j \leq Q$ , and conditional mean,  $\bar{\mathbf{u}}_{2,j}^+$ , and covariance,  $R_{2,j}^+$ , so that

$$p_+(\mathbf{u}) = \sum_{j=1}^Q p_{j,+} \delta(\mathbf{u}_1 - \mathbf{u}_{1,j}) \mathcal{N}(\bar{\mathbf{u}}_{2,j}^+, R_{2,j}^+). \quad [\text{S3}]$$

In fact, the distributions  $\mathcal{N}(\bar{\mathbf{u}}_{2,j}^+, R_{2,j}^+)$  are updated by suitable Kalman filter formulas with the mean update for  $\bar{\mathbf{u}}_{2,j}$  depending nonlinearly on  $\mathbf{u}_{1,j}$  in general.

**Proof.** Let  $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)$ ,  $\mathbf{u}_j \in \mathbb{R}^{N_j}$ ,  $N_1 + N_2 = N$ , and  $\mathbf{u}_1 = \{\mathbf{u}_{1,j}\}_{j=1}^Q$  be sampled in  $Q$  discrete states such that [S1] is satisfied. Using discrete representation of  $\mathbf{u}_1$ , the posterior distribution can be written as

$$p_+(\mathbf{u}|\mathbf{v}) = \sum_{j=1}^Q p_+(\mathbf{u}, \mathbf{u}_{1,j}|\mathbf{v}) = \sum_{j=1}^Q p_+(\mathbf{u}|\mathbf{v}, \mathbf{u}_{1,j}) p_{j,+}, \quad [\text{S4}]$$

where  $p_{j,+} = p_+(\mathbf{u}_{1,j}|\mathbf{v})$  is the posterior probability of the  $j$ th state.

i) First, consider the general likelihood function  $p(\mathbf{v}|\mathbf{u})$  (not possibly Gaussian). A simple application of Bayes' theorem gives

$$\begin{aligned} p_+(\mathbf{u}|\mathbf{v}, \mathbf{u}_{1,j}) &= \frac{p(\mathbf{v}|\mathbf{u}, \mathbf{u}_{1,j}) p_-(\mathbf{u}, \mathbf{u}_{1,j})}{p(\mathbf{v}, \mathbf{u}_{1,j})} \\ &= \frac{p(\mathbf{v}|\mathbf{u}, \mathbf{u}_{1,j}) p_-(\mathbf{u}|\mathbf{u}_{1,j})}{\int p(\mathbf{v}|\mathbf{u}_{1,j}, \mathbf{u}_2) p(\mathbf{u}_2|\mathbf{u}_{1,j}) d\mathbf{u}_2} \\ &= \frac{p(\mathbf{v}|\mathbf{u}_{1,j}, \mathbf{u}_2) \Pi_{j,-}}{\tilde{I}_j}, \end{aligned} \quad [\text{S5}]$$

with  $\tilde{I}_j = \int p(\mathbf{v}|\mathbf{u}_{1,j}, \mathbf{u}_2) \mathcal{N}(\bar{\mathbf{u}}_{2,j}^-, R_{2,j}^-) d\mathbf{u}_2$ ,  $1 \leq j \leq Q$ , and  $\Pi_{j,-} = p_-(\mathbf{u}|\mathbf{u}_{1,j}) = \delta(\mathbf{u}_1 - \mathbf{u}_{1,j}) \mathcal{N}(\bar{\mathbf{u}}_{2,j}^-, R_{2,j}^-)$  by the conditional Gaussian assumption.

ii) Then, to determine the posterior weights  $p_{j,+}$ , the formula of conditional probability for the posterior gives

$$\begin{aligned} p(\mathbf{u}, \mathbf{u}_{1,j}, \mathbf{v}) &= p(\mathbf{u}|\mathbf{v}, \mathbf{u}_{1,j}) p_+(\mathbf{u}_{1,j}|\mathbf{v}) p(\mathbf{v}) \\ &= p(\mathbf{u}|\mathbf{v}, \mathbf{u}_{1,j}) p_{j,+} p(\mathbf{v}), \end{aligned} \quad [\text{S6}]$$

and for the prior

$$\begin{aligned} p(\mathbf{u}, \mathbf{u}_{1,j}, \mathbf{v}) &= p(\mathbf{v}|\mathbf{u}, \mathbf{u}_{1,j}) p_-(\mathbf{u}, \mathbf{u}_{1,j}) \\ &= p(\mathbf{v}|\mathbf{u}, \mathbf{u}_{1,j}) p_{j,-} \Pi_{j,-}. \end{aligned} \quad [\text{S7}]$$

Comparing [S6] with [S7], we get the identity

$$p(\mathbf{v}|\mathbf{u}, \mathbf{u}_{1,j}) p_{j,-} \Pi_{j,-} = p(\mathbf{u}|\mathbf{v}, \mathbf{u}_{1,j}) p_{j,+} p(\mathbf{v}). \quad [\text{S8}]$$

Using the fact that  $\sum_{j=1}^Q \int p(\mathbf{u}|\mathbf{v}, \mathbf{u}_{1,j}) p_{j,+} d\mathbf{u}_2 = 1$  and integrating both sides of [S8],

$$p(\mathbf{v}) = \sum_{j=1}^Q \int p(\mathbf{v}|\mathbf{u}_{1,j}, \mathbf{u}_2) p_{j,-} \Pi_{j,-} d\mathbf{u}_2 = \sum_{j=1}^Q p_{j,-} \tilde{I}_j. \quad [\text{S9}]$$

Therefore, substituting [S5] and [S9] into [S8],

$$\begin{aligned} p_{j,+} &= \frac{p_{j,-} \int p(\mathbf{v}|\mathbf{u}_{1,j}, \mathbf{u}_2) p(\mathbf{u}_2|\mathbf{u}_{1,j}) d\mathbf{u}_2}{p(\mathbf{v})} \\ &= \frac{p_{j,-} \tilde{I}_j}{\sum_{k=1}^Q p_{k,-} \tilde{I}_k}. \end{aligned} \quad [\text{S10}]$$

iii) Finally, with observations in the form of [S2] and Gaussian likelihood function  $p(\mathbf{v}|\mathbf{u}_{1,j}, \mathbf{u}_2) = p_G(\mathbf{v} - G_0(\mathbf{u}_{1,j}) - G_1(\mathbf{u}_{1,j})\mathbf{u}_2)$ ,  $p_G \sim \mathcal{N}(0, R_0)$ , the numerator of [S5] becomes the standard process of the Kalman filter. That is,

$$\begin{aligned} p_+(\mathbf{u}|\mathbf{v}, \mathbf{u}_{1,j}) &= \frac{\delta(\mathbf{u}_1 - \mathbf{u}_{1,j}) \mathcal{N}(\bar{\mathbf{u}}_{2,j}^-, R_{2,j}^-) p_G(\mathbf{v} - G_0(\mathbf{u}_{1,j}) - G_1(\mathbf{u}_{1,j})\mathbf{u}_2)}{\tilde{I}_j} \\ &= \delta(\mathbf{u}_1 - \mathbf{u}_{1,j}) \mathcal{N}(\bar{\mathbf{u}}_{2,j}^+, R_{2,j}^+), \end{aligned} \quad [\text{S11}]$$

where  $(\bar{\mathbf{u}}_{2,j}^+, R_{2,j}^+)$ , depending nonlinearly on  $\mathbf{u}_{1,j}$  in general, are the posterior mean and covariance through a suitable Kalman filtering process. Also the weight updating factor  $\tilde{I}_j$  can be calculated explicitly by integrating the multiplication of two Gaussian distributions,

$$\begin{aligned} \tilde{I}_j &= \int \exp\left(-\frac{1}{2}(\mathbf{v} - G(\mathbf{u}))^T R_0^{-1} (\mathbf{v} - G(\mathbf{u}))\right) \exp\left(-\frac{1}{2}(\mathbf{u}_2 - \bar{\mathbf{u}}_{2,j}^-)^T R_{2,j}^- (\mathbf{u}_2 - \bar{\mathbf{u}}_{2,j}^-)\right) d\mathbf{u}_2 \\ &= \det(R_{2,j}^+)^{-1/2} \exp\left[\frac{1}{2} \left( \bar{\mathbf{u}}_{2,j}^{+T} R_{2,j}^{+1} \bar{\mathbf{u}}_{2,j}^+ - \bar{\mathbf{u}}_{2,j}^{-T} R_{2,j}^{-1} \bar{\mathbf{u}}_{2,j}^- - (\mathbf{v} - G E \mathbf{u}_{1,j}^-)^T R_0^{-1} (\mathbf{v} - G E \mathbf{u}_{1,j}^-) \right)\right]. \end{aligned} \quad [\text{S12}]$$

Therefore, the expression for the posterior distribution [S4] is derived explicitly in the form [S10], [S11], and [S12].

## 2. Details of Blended Filter Algorithms

Here we describe the details about the blended filter algorithms. After the forecast step, we get the predicted values for the mean states  $\bar{\mathbf{u}}$  and covariance matrix  $R$ , together with the particle presentation under dynamical orthogonality (DO) basis  $E(t) = \{\mathbf{e}_1(t), \dots, \mathbf{e}_s(t)\}$ ,

$$\mathbf{u}(t) = \bar{\mathbf{u}}(t) + \sum_{i=1}^s Y_i(t; \omega) \mathbf{e}_i(t),$$

where  $\{\mathbf{e}_i(t)\}_{i=1}^s$  is the subdimensional dynamical basis, and  $\{Y_i\}$  are the corresponding stochastic coefficients achieved through Monte Carlo simulations. The particle statistics presented by  $Y_i$  must be consistent with the covariance matrix  $R$ , and  $Y_i, Y_j$  are normalized to be independent with each other; that is,

$$\langle Y_i Y_j^* \rangle = \mathbf{e}_i \cdot R \mathbf{e}_j \delta_{ij}, \quad 1 \leq i, j \leq s.$$

For the analysis step, filtering is implemented in two separate subspaces. The projection operator is calculated by completing the dynamical basis  $P = [E, E^\perp]$ . The orthogonal complements  $E^\perp$  are chosen freely here. Initially each particle is uniformly weighted as  $(Y_i, p_i) = (Y_i, 1/Q)$ . Following is the analysis step algorithm for the blended methods.

**Algorithm:** Blended filtering (analysis step):

Project the mean and covariance matrix to the two subspaces

$$\begin{pmatrix} \bar{\mathbf{u}}_1 \\ \bar{\mathbf{u}}_2 \end{pmatrix} = P^T \bar{\mathbf{u}}, \quad R_t = P^T R P = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}. \quad [\text{S13}]$$

Solve the following linear system to find the conditional mean  $\bar{\mathbf{u}}_2(\mathbf{u}_1) = \bar{\mathbf{u}}_{2,j}^+$  [denote  $\mathbf{u}'_1 = \mathbf{u}_1 - \langle \mathbf{u}_1 \rangle$ ,  $\bar{\mathbf{u}}'_2(\mathbf{u}_1) = \bar{\mathbf{u}}_2(\mathbf{u}_1) - \langle \mathbf{u}_2 \rangle$  as the fluctuations about the mean]:

$$\begin{bmatrix} p_1 u_{1,1}^1 & \cdots & p_Q u_{1,1}^Q \\ \vdots & \ddots & \vdots \\ p_1 u_{1,N_1}^1 & \cdots & p_Q u_{1,N_1}^Q \\ p_1 & \cdots & p_Q \end{bmatrix}_{(N_1+1) \times Q} \begin{bmatrix} u_{2,1}^1 & \cdots & u_{2,N_2}^1 \\ u_{2,1}^2 & \cdots & u_{2,N_2}^2 \\ \vdots & \ddots & \vdots \\ u_{2,1}^Q & \cdots & u_{2,N_2}^Q \end{bmatrix}_{Q \times N_2} = \begin{bmatrix} R_{12} \\ 0 \end{bmatrix}. \quad [\text{S14}]$$

Note that this is an underdetermined system for a sufficiently large number of particles,  $Q \gg N_1 + 1$ , and “-“ notation is suppressed here.

Calculate the conditional covariance  $R_2^-$  in  $\mathbf{u}_2$  subspace by

$$\begin{aligned} R_2^- &= R_2 + \langle \mathbf{u}_2 \rangle \otimes \langle \mathbf{u}_2 \rangle - \int \bar{\mathbf{u}}_2(\mathbf{u}_1) \otimes \bar{\mathbf{u}}_2(\mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_1 \\ &= R_2 - \int \bar{\mathbf{u}}'_2(\mathbf{u}_1) \otimes \bar{\mathbf{u}}'_2(\mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_1 \\ &= R_2 - \sum_j \bar{\mathbf{u}}'_{2,j} \otimes \bar{\mathbf{u}}'_{2,j} p_{j,-}. \end{aligned} \quad [\text{S15}]$$

Use Kalman filter updates in the  $\mathbf{u}_2$  subspace

$$\bar{\mathbf{u}}_{2,j}^+ = \bar{\mathbf{u}}_{2,j}^- + K \left( \mathbf{v} - G E \mathbf{u}_{1,j}^- - G E^\perp \bar{\mathbf{u}}_{2,j}^- \right), \quad [\text{S16a}]$$

$$\tilde{R}_2^+ = (I - K G E^\perp) R_2^-, \quad [\text{S16b}]$$

$$K = R_2^- (G E^\perp)^\top \left( G E^\perp R_2^- (G E^\perp)^\top + R_0 \right)^{-1}, \quad [\text{S16c}]$$

with the (linear) observation operator  $G(\mathbf{u}_1, \mathbf{u}_2) = G E \mathbf{u}_1 + G E^\perp \mathbf{u}_2$ , where  $R_0$  is the covariance matrix for the observation noise.

Update the particle weights in the  $\mathbf{u}_1$  subspace by  $p_{j,+} \propto p_{j,-} I_j$ , with

$$I_j = \exp \left[ \frac{1}{2} \left( \bar{\mathbf{u}}_{2,j}^{+\top} (\tilde{R}_2^+)^{-1} \bar{\mathbf{u}}_{2,j}^+ - \bar{\mathbf{u}}_{2,j}^{-\top} (R_2^-)^{-1} \bar{\mathbf{u}}_{2,j}^- - (\mathbf{v} - G E \mathbf{u}_{1,j})^\top R_0^{-1} (\mathbf{v} - G E \mathbf{u}_{1,j}) \right) \right]. \quad [\text{S17}]$$

Normalize the weights  $p_{j,+} = p_{j,-} I_j / \sum_k p_{k,-} I_k$  and use residual resampling.

Get the posterior mean and covariance matrix from the posterior particle presentation

$$\bar{\mathbf{u}}_1^+ = \sum_j \mathbf{u}_{1,j} p_{j,+}, \quad \bar{\mathbf{u}}_2^+ = \sum_j \bar{\mathbf{u}}_{2,j}^+ p_{j,+}, \quad [\text{S18}]$$

and

$$R_{1,ij}^+ = \sum_k Y_{i,k} Y_{j,k}^* p_{k,+}, \quad 1 \leq i, j \leq s,$$

$$R_{12,ij}^+ = \sum_k Y_{i,k} \bar{u}_{j,k}^{+\ast} p_{k,+}, \quad 1 \leq i \leq s, \quad s+1 \leq j \leq N,$$

$$R_2^+ = \tilde{R}_2^+ + \sum_j \bar{\mathbf{u}}'_{2,j} \otimes \bar{\mathbf{u}}'_{2,j} p_{j,+}.$$

Rotate the stochastic coefficients and basis to principal directions in the  $s$ -dimensional dynamical subspace.

**Remark 1:** As derived in [S12] of Proposition 1, the weight updating factor  $I_j$  has an additional component  $\det(R_{2,j}^+)^{-(1/2)}$ , which is expensive to compute given the high dimensionality of the orthogonal subspace  $\mathbf{u}_2$ . For the blended algorithms in the main text,  $R_2^-$  is independent of the choice of particles  $j$ , so  $R_2^+$  is also independent of  $j$  by [S16b]. Therefore, the expensive determinant term can be neglected in [S17] as a normalization factor for the updated weights  $p_{j,+}$ .

**Remark 2:** By the blended forecast model, the dynamical basis  $\mathbf{e}_j(t)$  keeps tracking the principal directions of the system [that is, the direction with the largest eigenvalues of the covariance matrix  $R(t)$ ]. However, after the analysis step, the principal directions are changed according to the observation data. Therefore, it is necessary to rotate the basis according to the posterior covariance after every analysis step as described in the last step above. Although numerical tests show reasonable results without the rotation, corrections at every step can make sure that the scheme is robust and more accurate.

### 2.1. The Conditional Covariance $R_2(\mathbf{u}_1)$ Given Particles in $\mathbf{u}_1$ Subspace.

The forecast model gives the particle representation  $\mathbf{u}_{1,j}$  in the low-dimensional subspace. And the conditional Gaussian mixture  $(\bar{\mathbf{u}}_{2,j}, R_{2,j})$  in the orthogonal subspace is achieved by solving the least-squares solution of the linear system [S14] with minimum weighted  $L^2$  norm,  $\sum_{j=1}^Q |\bar{\mathbf{u}}'_{2,j}|^2 p_j$ . Such an idea comes from the maximum entropy principle. Another advantage of this max-entropy solution is that it concludes that the conditional



covariance  $R_2(\mathbf{u}_1) = R_2^-$  is independent of the particles in  $\mathbf{u}_1$  subspace.

One important point to note is that the prior covariance  $R_2$  from [S13] is different from the conditional covariance  $R_2(\mathbf{u}_1) = R_2^-$  given the particle values in the  $\mathbf{u}_1$  subspace. This can be seen clearly from their definitions

$$R_2 = \int \int (\mathbf{u}_2 - \langle \mathbf{u}_2 \rangle) \otimes (\mathbf{u}_2 - \langle \mathbf{u}_2 \rangle) p(\mathbf{u}_2 | \mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_2 d\mathbf{u}_1, \quad [\text{S19}]$$

$$R_2(\mathbf{u}_1) = \int (\mathbf{u}_2 - \bar{\mathbf{u}}_2(\mathbf{u}_1)) \otimes (\mathbf{u}_2 - \bar{\mathbf{u}}_2(\mathbf{u}_1)) p(\mathbf{u}_2 | \mathbf{u}_1) d\mathbf{u}_2, \quad [\text{S20}]$$

with  $\langle \mathbf{u}_2 \rangle = \int \bar{\mathbf{u}}_2(\mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_1$ . By a simple calculation combining [S19] and [S20] we have

$$\begin{aligned} \int R_2(\mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_1 &= R_2 + \langle \mathbf{u}_2 \rangle \otimes \langle \mathbf{u}_2 \rangle - \int \bar{\mathbf{u}}_2(\mathbf{u}_1) \otimes \bar{\mathbf{u}}_2(\mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_1 \\ &= R_2 - \int \bar{\mathbf{u}}_2'(\mathbf{u}_1) \otimes \bar{\mathbf{u}}_2'(\mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_1, \end{aligned}$$

with  $\bar{\mathbf{u}}_2'(\mathbf{u}_1) = \bar{\mathbf{u}}_2(\mathbf{u}_1) - \langle \mathbf{u}_2 \rangle$ . Noting that  $R_2(\mathbf{u}_1) = R_2^-$  is independent of the variable  $\mathbf{u}_1$ , we get the approximation for the conditional covariance

$$\begin{aligned} R_2^- &= R_2 - \int \bar{\mathbf{u}}_2'(\mathbf{u}_1) \otimes \bar{\mathbf{u}}_2'(\mathbf{u}_1) p(\mathbf{u}_1) d\mathbf{u}_1 \\ &= R_2 - \sum_j \bar{\mathbf{u}}_{2,j}' \otimes \bar{\mathbf{u}}_{2,j}' p_j. \end{aligned} \quad [\text{S21}]$$

This conditional covariance formula is theoretically consistent, whereas it will introduce the problem of realizability. Then as discussed in the main text, we may need to introduce a realizability check and correction method for  $R_2^-$  after each analysis step or simply inflate the matrix as  $R_2^- = R_2$  (labeled as ‘‘inflated  $R_2^-$ ’’ in Fig. S4 below). Both strategies work well for the modified quasilinear Gaussian dynamical orthogonality (MQG-DO) method whereas the quasilinear Gaussian dynamical orthogonality (QG-DO) method requires the larger inflation approach in practical implementations as shown in the results below.

**2.2. Resampling Strategies.** Here we add some details about the resampling process. In the resampling step, particles with small weights are abandoned, whereas the other particles are duplicated according to their weights to keep the ensemble size unchanged. One useful and popular resample approach is the residual resampling method. The idea is to replicate the  $j$ th particle to have  $\lfloor p_j Q \rfloor$  copies according to its weight in the first step, and for the rest  $Q - \sum_j \lfloor p_j Q \rfloor$  members particle values are assigned randomly with probability proportional to their residual weights  $p_j Q - \lfloor p_j Q \rfloor$ .

An additional issue arises in the process of duplicating particles. After resampling, we will get several particles with the same value. For stochastic systems, this will not be a problem because the random external forcing term will add uncertainty to each particle and produce different prediction results even though the initial values are the same. However, for deterministic systems with internal instability [for example, the Lorenz 96 (L-96) system], such resampling would be of no real value, because no external randomness is introduced to the particles to generate different outputs in the next step. One possible strategy to avoid this problem is to add small perturbations to the duplicated particles. Due to the internal instability inside the system, small perturbations in the initial time can end up with large deviations

in the forecasts before the next analysis step so that the duplicated particles can get separated. The amplitude of the perturbation is added according to the uncertainties of the variables. This can be approximated by the filtered results before the resampling step. After updating the weights, we get the set of particles together with their new weights,  $\{\mathbf{u}_j', p_j\}$ , and then a perturbation is added to each particle as a white noise with variance  $\sigma^2 = \sum_j \mathbf{u}_j'^2 p_j$ . Note that this perturbation can be viewed as an inflation to the particles. We calculate the posterior covariance matrix according to the resampled particles, and therefore the covariance will also be inflated accordingly. From numerical simulations it is shown that this strategy is effective to avoid filter divergence.

### 3. Blended Filter Performance on the L-96 Model

Here we present the additional results for the blended filter methods compared with the MQG filter as well as the ensemble adjustment Kalman filter (EAKF) with inflation and localization. The L-96 model is used as the test model

$$\frac{du_i}{dt} = u_{i-1}(u_{i+1} - u_{i-2}) - u_i + F, \quad i = 0, \dots, J-1 \quad [\text{S22}]$$

with  $J = 40$  and  $F$  the deterministic forcing parameter. We test the filter performances in both a weakly chaotic regime ( $F = 5$ ) and a strongly chaotic regime ( $F = 8$ ) as described in the main text. We use the root mean-square error (rmse) and pattern correlation (XC) to measure the filter skill. rmse is defined via

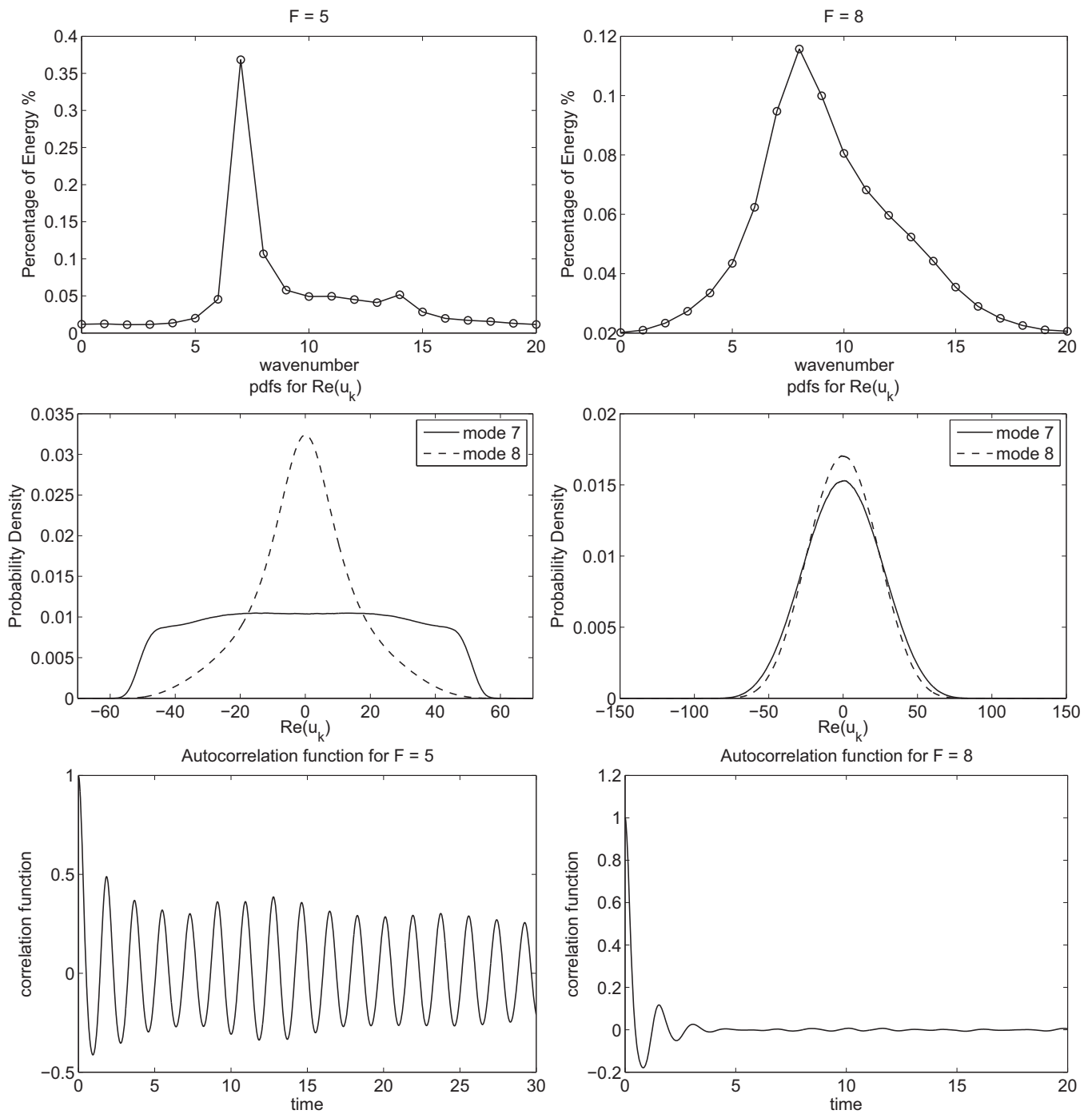
$$\text{rmse}(\mathbf{x} - \mathbf{y}) = \left( \frac{1}{N} \sum_{j=1}^N |x_j - y_j|^2 \right)^{1/2}, \quad [\text{S23}]$$

where  $\mathbf{x}$  and  $\mathbf{y}$  are the vectors (usually the truth and the filtering result) to be compared and  $N$  is the length of each vector. XC is defined by

$$\text{XC}(\mathbf{x}, \mathbf{y}) = \frac{\sum_{j=1}^N x_j' y_j'}{\left( \sum_{j=1}^N x_j'^2 \sum_{j=1}^N y_j'^2 \right)^{1/2}}, \quad [\text{S24}]$$

with  $x_j' = x_j - \bar{x}$ ,  $y_j' = y_j - \bar{y}$  being the fluctuations about the mean. Therefore, if the two signals are close to each other, the rmse of their difference is then close to zero and their XC is close to one. On the other hand, for two different signals, the rmse diverges from zero and in principle is unbounded and XC approaches zero or negative value.

The statistics for these two cases are shown in Fig. S1. Strong non-Gaussian statistics can be seen from the weakly chaotic regime  $F = 5$  whereas near Gaussian distribution is shown in the strongly chaotic regime  $F = 8$ . shows the rms errors and pattern correlations for various filtering methods in regime  $F = 5$  with moderate observational error and observation frequency  $r_0 = 2$ ,  $\Delta t = 1$ ,  $p = 4$ . And Fig. S3 gives the joint distributions of the first two principal Fourier modes  $\hat{u}_7$  and  $\hat{u}_8$  in this regime by scatter plots to further visualize the non-Gaussian statistics captured through different schemes. Finally, Fig. S4 gives additional examples for the performance of the filters in regimes with sparse infrequent high-quality observations with slightly larger observation noise  $r_0 = 0.25$ . In addition, we compare the performances of the QG-DO method with crude covariance inflation  $R_2^- = R_2$ . Incorrect energy transfers in the QG forecast model for the long forecast time using the original covariance approximation as in [S21] in this case end up with serious filter divergence.



**Fig. S1.** Statistics for the L-96 model. (Top) Energy spectra in Fourier space; (Middle) the probability density functions (pdfs) for the first two principal Fourier modes; (Bottom) the autocorrelation functions of the state variable. Regimes for weakly chaotic ( $F=5$ , Left) and strongly chaotic ( $F=8$ , Right) dynamics are shown.



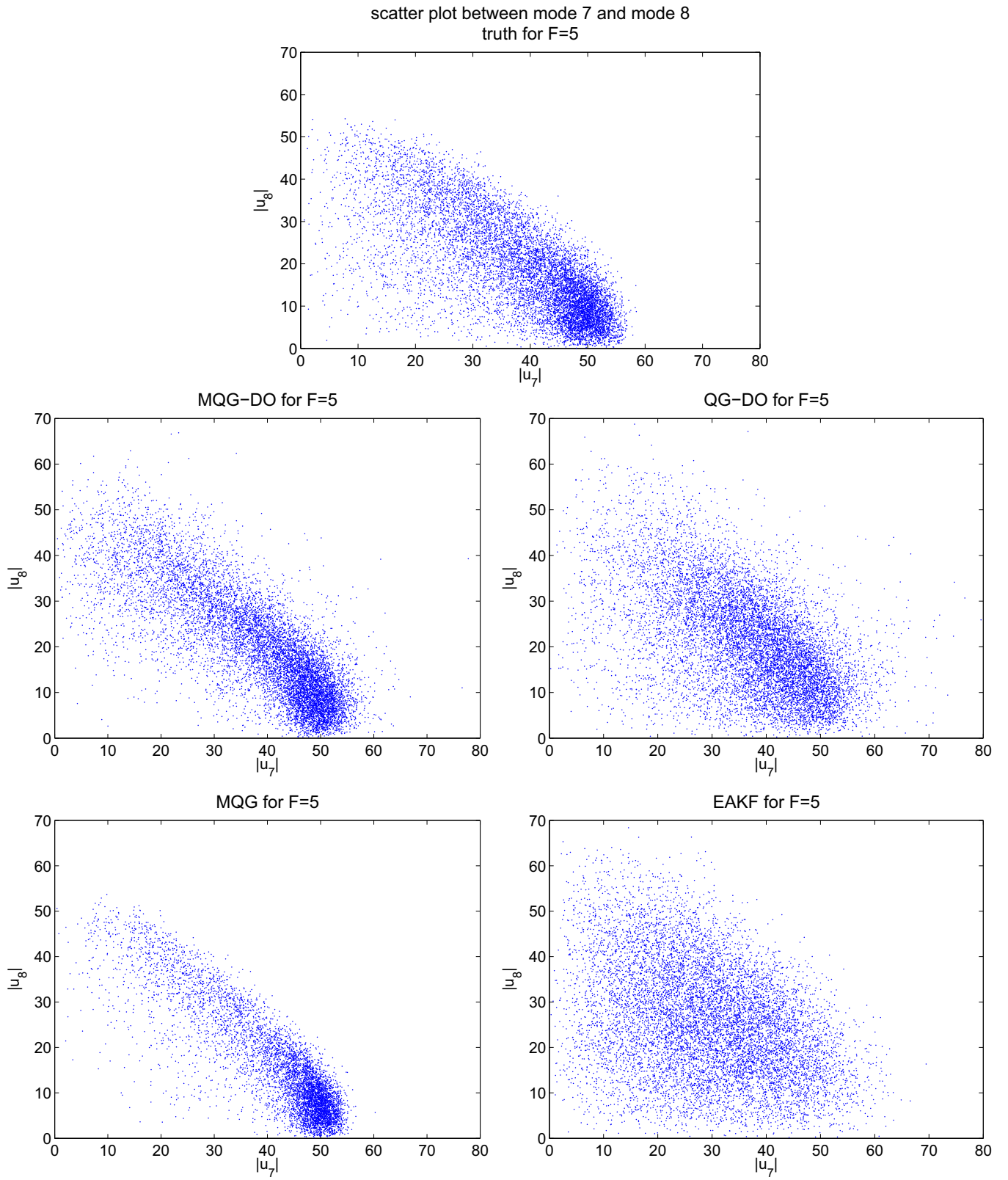


Fig. S3. Joint distributions of the first two principal Fourier modes  $\hat{u}_7$  and  $\hat{u}_8$  shown by scatter plots with different filter schemes in regime  $F=5$ .

