Blended Particle Filters for Large Dimensional Chaotic Dynamical Systems

Andrew J. Majda [*], **Di Qi** [*], **Themistoklis P. Sapsis** [†]

[*] Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, and [†] Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139

Contributed by Andrew J. Majda

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 which 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 below 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 forty dimensional Lorenz 96 model with a five dimensional adaptive subspace for nonlinear blended filtering in various turbulent regimes with at least nine positive Lyapunov exponents are utilized 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 multi-scale filtering of turbulent systems and a simple application is sketched below.

Data assimilation | Curse of dimensionality | Hybrid filters | Turbulent dynamical systems

Many 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 [9]. When the system is low dimensional, Monte-Carlo approaches such as the particle filter with its various up-to-date resampling strategies [14] 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 since sampling a high dimensional variable is computationally impossible for the foreseeable future. Recent mathematical theory strongly supports this curse of dimensionality for particle filters [8, 6]. In the second direction, Bayesian hierarchical modeling [7] and reduced order filtering strategies [15, 4, 3] based on the Kalman filter [12] 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 [2, 5, 11, 10, 13] 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 Chapter 15 of [1]. 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 $\boldsymbol{u} \in \mathbb{R}^N$, there are two subspaces which typically evolve adaptively in time where $\boldsymbol{u} = (\boldsymbol{u}_1, \boldsymbol{u}_2)$, $\boldsymbol{u}_j \in \mathbb{R}^{N_j}$, $N_1 + N_2 = N$ with the property that N_1 is low dimensional enough so that the non-Gaussian statistics of $\boldsymbol{u}_1$ can be calculated from a particle filter while the evolving statistics of $\boldsymbol{u}_2$ are conditionally Gaussian given $\boldsymbol{u}_1$. Statistically nonlinear forecast models with this structure with high skill for uncertainty quantification have been developed recently by two of the authors [33, 32, 31, 30] and are utilized below in the blended filters.

The mathematical foundation for implementing the analysis step where observations are utilized in the conditional Gaussian mixture framework are 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 forty 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 five dimensional with excellent performance for the blended filters. The mathematical formalism for filtering with conditionally Gaussian mixtures should be useful as a framework for multi-scale 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 which captures the leading variance adaptively [28] while ignoring the other degrees of freedom; simple examples for non-normal linear systems [31] demonstrate the poor skill of such an approach for reduced filtering in general; for filtering linear non-normal systems, the optimal reduced basis instead is defined through balanced truncation [3].

Mathematical Foundations for Blended Particle Filters

Here we consider real time filtering or data assimilation algorithms for a state vector $\boldsymbol{u} \in \mathbb{R}^N$ from a nonlinear turbulent dynamical

Reserved for Publication Footnotes

system where the state $\boldsymbol{u}$ is forecast by an approximate dynamical model between successive observation times, $m\Delta t$, and the state of the system is updated through 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 which typically evolve adaptively in time (although that dependence is suppressed here) where $\boldsymbol{u}=(\boldsymbol{u}_1,\boldsymbol{u}_2)$, $\boldsymbol{u}_j\in\mathbb{R}^{N_j}$, $N_1+N_2=N$ with the property that N_1 is low dimensional enough so that the statistics of $\boldsymbol{u}_1$ can be calculated from a particle filter while the statistics of $\boldsymbol{u}_2$ are conditionally Gaussian given $\boldsymbol{u}_1$. Thus, at any analysis time step $m\Delta t$, we have the prior forecast density given by

$$p_-\left(\boldsymbol{u}\right)=p_-\left(\boldsymbol{u}_1\right)p_-^G\left(\boldsymbol{u}_2\mid\boldsymbol{u}_1\right), \qquad [1]$$

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

$$p_-\left(\boldsymbol{u}_1\right)=\sum_{j=1}^{Q}p_{j,-}\delta\left(\boldsymbol{u}_1-\boldsymbol{u}_{1,j}\right), \qquad [2]$$

with non-negative particle weights, $p_{j,-}$ with $\sum_j p_{j,-}=1$. Below we will 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 $G\left(\boldsymbol{u}\right)$ maps $\mathbb{R}^N$ to $\mathbb{R}^M$, with M observations at the analysis time, $m\Delta t$, and has the form

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

where $G_1\left(\boldsymbol{u}_1\right)$ has rank M and the observational noise, $\boldsymbol{\sigma}_0$, is Gaussian $\boldsymbol{\sigma}_0=\mathcal{N}\left(\boldsymbol{0},R_0\right)$. Note that the form in **[3]** is the leading Taylor expansion around $\boldsymbol{u}_1$ of general nonlinear observation operator. Now, start with the conditional Gaussian particle distribution, $p_-\left(\boldsymbol{u}_1,\boldsymbol{u}_2\right)$ given from the forecast distribution

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

and compute the posterior distribution in the analysis step through Bayes theorem, $p_+\left(\boldsymbol{u}\mid\boldsymbol{v}\right)\sim p\left(\boldsymbol{v}\mid\boldsymbol{u}\right)p_-\left(\boldsymbol{u}\right)$. 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,* $\bar{\boldsymbol{u}}_{2,j}^+$, *and covariance,* $R_{2,j}^+$, *so that*

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

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

The proof of Proposition 1 is a direct calculation similar to those in [27, 26] for other formulations of Gaussian mixtures over the entire state space and the explicit formulas can be found in the supplementary material. However, there is a crucial difference that here conditional Gaussian mixtures are applied in the reduced subspace $\boldsymbol{u}_2$ blended with particle filter approximations only in the lower dimensional subspace, $\boldsymbol{u}_1$, unlike the previous work.

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

$$\boldsymbol{u}_t=L\boldsymbol{u}+\boldsymbol{B}\left(\boldsymbol{u},\boldsymbol{u}\right)+\boldsymbol{F}, \qquad [6]$$

where $\boldsymbol{B}\left(\boldsymbol{u},\boldsymbol{u}\right)$ involves energy conserving nonlinear interactions with $\boldsymbol{u}\cdot\boldsymbol{B}\left(\boldsymbol{u},\boldsymbol{u}\right)=0$ while L is a linear operator including damping as well as anisotropic physical effects; many turbulent dynamical systems in geosciences and engineering have the structure in **[6]** [25, 24].

Application to Multi-scale 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 $\boldsymbol{u}_1$ representing the slow (vortical) waves and a fixed subspace $\boldsymbol{u}_2$ 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 [17, 16, 1]. A small parameter $\epsilon\ll1$ characterizes the ratio of the fast time scale to the slow time scale. A well known formalism for stochastic mode reduction has been developed for such multi-scale systems [18, 19] with a simplified forecast model valid in the limit $\epsilon\to0$ with the form

$$\begin{aligned} p\left(\boldsymbol{u}_1,\boldsymbol{u}_2\right)(t) &= p\left(\boldsymbol{u}_1\right)(t)\,p\left(\boldsymbol{u}_2\mid\boldsymbol{u}_1\right)\\ &= p\left(\boldsymbol{u}_1\right)(t)\,\mathcal{N}\left(\boldsymbol{0},R_2\right), \end{aligned} \qquad [7]$$

where $p\left(\boldsymbol{u}_1\right)(t)$ 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 multi-scale filtering algorithms with this flavor designed to capture unresolved features of turbulence are developed recently in [23].

Blended Statistical Nonlinear Forecast Models. While the above formulation can be applied to hybrid particle filters with conditional Kalman filters on fixed subspaces, defined by $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ as sketched above, a more attractive idea is to utilize statistical forecast models that adaptively change these subspaces as time evolves in response to the uncertainty without a separation of time scales. Recently two of the authors [33, 32, 31, 30] have developed nonlinear statistical forecast models of this type, the quasilinear Gaussian dynamical orthogonality method (QG-DO) and the more sophisticated modified quasilinear Gaussian dynamical orthogonality method (MQG-DO) for turbulent dynamical systems with the structure in **[6]**. It is shown in [33] and [32] 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 [33, 31] for **[6]** or a more statistically accurate modified quasilinear Gaussian (MQG) closure [32, 31]; 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 [32, 31] 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

$$\boldsymbol{u}\left(t\right)=\bar{\boldsymbol{u}}\left(t\right)+\sum_{j=1}^{s}Y_j\left(t;\omega\right)\boldsymbol{e}_j\left(t\right) \qquad [8]$$

where $\boldsymbol{e}_j\left(t\right)$, $j=1,\cdots s$ are time-dependent orthonormal modes and $s \ll N$ is the reduction order. The modes and the stochastic coefficients $Y_j\left(t;\omega\right)$ evolve according to the DO condition [29]. 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]**

$$\frac{d\bar{\boldsymbol{u}}}{dt}=\left(L+D\right)\bar{\boldsymbol{u}}+\boldsymbol{B}\left(\bar{\boldsymbol{u}},\bar{\boldsymbol{u}}\right)+R_{ij}\boldsymbol{B}\left(\boldsymbol{v}_i,\boldsymbol{v}_j\right)+\boldsymbol{F}. \qquad \textbf{[9]}$$

- *Equation for the stochastic coefficients and the modes*
  Both the stochastic coefficients and the modes evolve according to the DO equations [29]. The coefficient equations are obtained by a direct Galerkin projection and the DO condition

$$\begin{aligned}\frac{dY_i}{dt} &= Y_m\left[\left(L+D\right)\boldsymbol{e}_m+\boldsymbol{B}\left(\bar{\boldsymbol{u}},\boldsymbol{e}_m\right)+\boldsymbol{B}\left(\boldsymbol{e}_m,\bar{\boldsymbol{u}}\right)\right]\cdot\boldsymbol{e}_i \\ &\quad +\left(Y_mY_n-C_{mn}\right)\boldsymbol{B}\left(\boldsymbol{e}_m,\boldsymbol{e}_n\right)\cdot\boldsymbol{e}_i \qquad \textbf{[10]}\end{aligned}$$

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

$$\frac{\partial\boldsymbol{e}_i}{\partial t}=\boldsymbol{M}_i-\boldsymbol{e}_j\left(\boldsymbol{M}_i\cdot\boldsymbol{e}_j\right), \qquad \textbf{[11]}$$

  with

$$\begin{aligned}\boldsymbol{M}_i &= \left(L+D\right)\boldsymbol{e}_i+\boldsymbol{B}\left(\bar{\boldsymbol{u}},\boldsymbol{e}_i\right)+\boldsymbol{B}\left(\boldsymbol{e}_i,\bar{\boldsymbol{u}}\right) \\ &\quad +\boldsymbol{B}\left(\boldsymbol{e}_m,\boldsymbol{e}_n\right)\langle Y_mY_nY_k\rangle C_{ik}^{-1}.\end{aligned}$$

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

$$\frac{dR}{dt}=L_vR+RL_v^*+Q_{F,s} \qquad \textbf{[12]}$$

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

$$\begin{aligned}Q_{F,s} &= \langle Y_mY_nY_k\rangle\left(\boldsymbol{B}\left(\boldsymbol{e}_m,\boldsymbol{e}_n\right)\cdot\boldsymbol{v}_i\right)\left(\boldsymbol{v}_j\cdot\boldsymbol{e}_k\right) \\ &\quad +\langle Y_mY_nY_k\rangle\left(\boldsymbol{B}\left(\boldsymbol{e}_m,\boldsymbol{e}_n\right)\cdot\boldsymbol{v}_j\right)\left(\boldsymbol{v}_i\cdot\boldsymbol{e}_k\right). \textbf{[13]}\end{aligned}$$

  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 [33, 32, 31].

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 $Y_j\left(t\right)$, $1\leq j\leq s$ 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 $E\left(t\right)=\left\{\boldsymbol{e}_1\left(t\right),\cdots,\boldsymbol{e}_s\left(t\right)\right\}$ denote the s-dimensional stochastic subspace in the forecast; at any analysis time, $t=m\Delta t$, $\boldsymbol{u}_1$ denote the projection of $\boldsymbol{u}\in\mathbb{R}^N$ to $E\left(t\right)$. Complete the dynamical basis E with an orthonormal basis $E^\perp$ and define $\boldsymbol{u}_2$ 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:

A) A particle approximation for the marginal distribution

$$p_-\left(\boldsymbol{u}_1\right)=\sum_{j=1}^{Q}p_{j,-}\delta\left(\boldsymbol{u}_1-\boldsymbol{u}_{1,j}\right); \qquad \textbf{[14]}$$

B) The mean $\bar{\boldsymbol{u}}_2^-$ and the covariance matrix

$$R=\begin{pmatrix} R_1 & R_{12} \\ R_{12}^{\mathrm{T}} & R_2\end{pmatrix} \qquad \textbf{[15]}$$

  where R is the covariance matrix in the basis $\left\{E,E^\perp\right\}$.

In order to apply Proposition 1 in the analysis step, we need to find a probability density $p_-\left(\boldsymbol{u}\right)$ 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

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

so that **[14]** is automatically satisfied by **[16]** while $\bar{\boldsymbol{u}}_{2,j}^-$ and R_2^- need to be chosen to satisfy **[15]**; note that R_2^- is a constant matrix independent of j. Let $\langle g\left(\boldsymbol{u}\right)\rangle$ denote the expected value of g with respect to p_- so that for example, $\langle\boldsymbol{u}_1\rangle=\bar{\boldsymbol{u}}_1^-$, $\langle\boldsymbol{u}_2\rangle=\bar{\boldsymbol{u}}_2^-$ and let $\boldsymbol{u}_1'=\boldsymbol{u}_1-\langle\boldsymbol{u}_1\rangle$, $\boldsymbol{u}_2'=\boldsymbol{u}_2-\langle\boldsymbol{u}_2\rangle$ 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 $\bar{\boldsymbol{u}}_2\left(\boldsymbol{u}_{1,j}\right)=\bar{\boldsymbol{u}}_{2,j}^-$

$$\begin{bmatrix} p_1u_{1,1}'^1 & \cdots & p_Qu_{1,1}'^Q \\ \vdots & \ddots & \vdots \\ p_1u_{1,N_1}'^1 & \cdots & p_Qu_{1,N_1}'^Q \\ p_1 & \cdots & p_Q\end{bmatrix}\begin{bmatrix} u_{2,1}'^1 & \cdots & u_{2,N_2}'^1 \\ u_{2,1}'^2 & \cdots & u_{2,N_2}'^2 \\ \vdots & \ddots & \vdots \\ u_{2,1}'^Q & \cdots & u_{2,N_2}'^Q\end{bmatrix}=\begin{bmatrix} R_{12} \\ 0\end{bmatrix}. \qquad \textbf{[17]}$$

  Note that this is an underdetermined system for a sufficiently large number of particles, Q, and '–' notation is suppressed here.
- Calculate the covariance R_2^- in the $\boldsymbol{u}_2$ subspace by requiring from **[15]** and **[16]**

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

  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]**.

Part II of the analysis step for the blended particle filter algorithm is an application of Proposition 1 to **[16]**.

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

$$\bar{\boldsymbol{u}}_{2,j}^+=\bar{\boldsymbol{u}}_{2,j}^-+K\left(\boldsymbol{v}-GE\boldsymbol{u}_{1,j}-GE^\perp\bar{\boldsymbol{u}}_{2,j}^-\right), \qquad \textbf{[19a]}$$

$$\tilde{R}_2^+=\left(I-KGE^\perp\right)R_2^-, \qquad \textbf{[19b]}$$

$$K=R_2^-\left(GE^\perp\right)^{\mathrm{T}}\left(GE^\perp R_2^-\left(GE^\perp\right)^{\mathrm{T}}+R_0\right)^{-1}, \qquad \textbf{[19c]}$$

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

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

$$I_j = \exp\left[\frac{1}{2}\left(\bar{\boldsymbol{u}}_{2,j}^{+\mathrm{T}}\left(\tilde{R}_2^+\right)^{-1}\bar{\boldsymbol{u}}_{2,j}^+ - \bar{\boldsymbol{u}}_{2,j}^{-\mathrm{T}}\left(R_2^-\right)^{-1}\bar{\boldsymbol{u}}_{2,j}^-\right) - \left(\boldsymbol{v} - GE\boldsymbol{u}_{1,j}\right)^{\mathrm{T}} R_0^{-1}\left(\boldsymbol{v} - GE\boldsymbol{u}_{1,j}\right)\right]. \quad [20]$$

- Normalize the weights $p_{j,+} = \frac{p_{j,-} I_j}{\sum_j p_{j,-} I_j}$ and use residual resampling (see the supplementary material for the details).
- Get the posterior mean and covariance matrix from the posterior particle presentation

$$\bar{\boldsymbol{u}}_1^+ = \sum_j \boldsymbol{u}_{1,j} p_{j,+},\ \bar{\boldsymbol{u}}_2^+ = \sum_j \bar{\boldsymbol{u}}_{2,j}^+ p_{j,+}, \quad [21a]$$

and

$$R_{1,ij}^+ = \sum_k Y_{i,k} Y_{j,k}^* p_{k,+},\ 1 \le i, j \le s, \quad [21b]$$

$$R_{12,ij}^+ = \sum_k Y_{i,k} \bar{u}_{j,k}^{\prime+*} p_{k,+},\ 1 \le i \le s,\ s+1 \le j \le N, \quad [21c]$$

$$R_2^+ = \tilde{R}_2^+ + \sum_j \bar{\boldsymbol{u}}_{2,j}^{\prime+} \otimes \bar{\boldsymbol{u}}_{2,j}^{\prime+} p_{j,+}. \quad [21d]$$

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

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]** which 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 $Q \ge N_1 + 1$ so the linear system $\mathcal{L}U_2 = \mathcal{F}$ is a strongly underdetermined linear system. The empirical approach which we utilize here is to seek the least squares solution of $\mathcal{L}U_2 = \mathcal{F}$ which minimizes the weighted L^2-norm,

$$\sum_{j=1}^{Q} \left|\bar{\boldsymbol{u}}_{2,j}'\right|^2 p_j. \quad [22]$$

This solution is given as the standard least squares solution through the pseudo-inverse for the auxiliary variable $\bar{\boldsymbol{v}}_{2,j}' = p_j^{\frac{1}{2}} \bar{\boldsymbol{u}}_{2,j}'$. 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 $R_2^- = R_2 - \sum_j \bar{\boldsymbol{u}}_{2,j}' \otimes \bar{\boldsymbol{u}}_{2,j}' p_j$ is realizable; to help guarantee this, we add extra inflation terms $\alpha_j \in [0,1]$ such that

$$R_2^- = R_2 - \sum_j \alpha_j p_j \bar{\boldsymbol{u}}_{2,j}' \otimes \bar{\boldsymbol{u}}_{2,j}'.$$

Here the inflation coefficients α_j can be chosen according to

- $\alpha_j = 1$, if $\bar{\boldsymbol{u}}_{2,j}^{\prime\mathrm{T}}\left(R_2 - \sum_k \bar{\boldsymbol{u}}_{2,k}' \otimes \bar{\boldsymbol{u}}_{2,k}' p_k\right)\bar{\boldsymbol{u}}_{2,j}' > \epsilon_0$;
- $\alpha_j = 1 - \frac{\epsilon_0 - \bar{\boldsymbol{u}}_{2,j}^{\prime\mathrm{T}}\left(R_2 - \sum_k \bar{\boldsymbol{u}}_{2,k}' \otimes \bar{\boldsymbol{u}}_{2,k}' p_k\right)\bar{\boldsymbol{u}}_{2,j}'}{p_j \left|\bar{\boldsymbol{u}}_{2,j}'\right|^4}$, otherwise;

where $\epsilon_0 \ll 1$ is a small number chosen to avoid numerical errors, and $1 \le j \le Q$.

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 $R_2^- = R_2 \ge 0$. Further motivation for the constrained least squares solution of $\mathcal{L}U_2 = \mathcal{F}$ minimizing **[22]** 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.

$$\begin{aligned}\bar{\boldsymbol{u}}_{2,j}' &= \operatorname{argmax} \log\det\left(R_2 - \sum_j p_j \bar{\boldsymbol{u}}_{2,j}' \otimes \bar{\boldsymbol{u}}_{2,j}'\right) \\ &= \operatorname{argmax} \log\det\left(I - \sum_j \left(p_j^{\frac{1}{2}} R_2^{-\frac{1}{2}} \bar{\boldsymbol{u}}_{2,j}'\right) \otimes \left(p_j^{\frac{1}{2}} R_2^{-\frac{1}{2}} \bar{\boldsymbol{u}}_{2,j}'\right)\right). \quad [23]\end{aligned}$$

The high dimensional nonlinear optimization problem in **[23]** is too expensive to solve directly but the small amplitude expansion, $\det\left(I - \epsilon R\right) = -\epsilon \mathrm{tr} R + O\left(\epsilon^2\right)$ of **[23]** becomes a weighted least squares optimization problem for the new variable, $\bar{\boldsymbol{v}}_{2,j}' = p_j^{\frac{1}{2}} R_2^{-\frac{1}{2}} \bar{\boldsymbol{u}}_{2,j}'$ constrained by $\mathcal{L}U_2 = \mathcal{F}$. If we choose $R_2 = I$, the least squares solution from **[22]** is recovered. While the algorithm utilizing $\bar{\boldsymbol{v}}_{2,j}'$ has a nice theoretical basis, it requires the singular value decomposition of the large covariance matrix, R_2, and the solution in **[22]** 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 (see Chapter 15 of [1]). 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 [1]. 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 [22, 21]; the L-96 model is a forty dimensional turbulent dynamical system which is a popular test model for filter performance for turbulent dynamical systems [1]. The L-96 model is a discrete periodic model given by

$$\frac{du_i}{dt} = u_{i-1}\left(u_{i+1} - u_{i-2}\right) - u_i + F, \quad i = 0, \cdots, J-1, \quad [24]$$

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 **[24]**. 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 ($F = 16$) [20, 24, 1]. Because the L-96 model is translation invariant, twenty discrete Fourier modes can be utilized 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 five dimensional subspace through a particle filter interacting with a low order Gaussian statistical forecast model in the remaining thirty-five 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 [20] so the five 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 while 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 Figure 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 since the pdfs for $|\hat{u}_7|$, $|\hat{u}_8|$ are far from a Rayleigh distribution; see the supplementary material for the scatter plot of their joint distribution exhibiting strongly non-Gaussian behavior. For the filtering experiments below, sparse spatial observations are utilized 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, 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 Figure 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 while EAKF yields incorrect Gaussian distributions with the largest forecast error spread. The RMS error and pattern correlation plots reported in the supplementary material 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.

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, 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 the rank histogram and maximum entropy particle filters has already been assessed for this difficult test problem in Figure 15.14 of [1] with large intervals in time with filter divergence (RMS errors much larger than one) 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 utilized here. In all examples with L-96 tested here, we find that the MQG-DO algorithm with the approximation described in the paragraph below **[18]** 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 utilized here (see the supplementary material). Figure 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 $F = 8, 16$ in contrast to other methods as shown in Figure 15.14 from [1]. The much cheaper MQG filter for $F = 16$ exhibits a long initial regime of filter divergence but eventually settles down to comparable filter performance as 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 the supplementary material.

Concluding Discussion

Blended particle filters which 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 combined with statistically nonlinear forecast models developed recently [33, 32, 31] with high skill for uncertainty quantification which are compatible with this structure. Stringent test cases for filtering involving the forty dimensional L-96 model with a five dimensional adaptive subspace for nonlinear filtering in various regimes of chaotic dynamics with at least nine positive Lyapunov exponents are utilized 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 here is also useful for multi-scale filtering of turbulent systems and a simple application has been sketched here.

ACKNOWLEDGMENTS. This 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 on the first grant while this research of T. P. S. was partially supported through the second grant, when T. P. S. was a postdoctoral fellow at the Courant Institute.

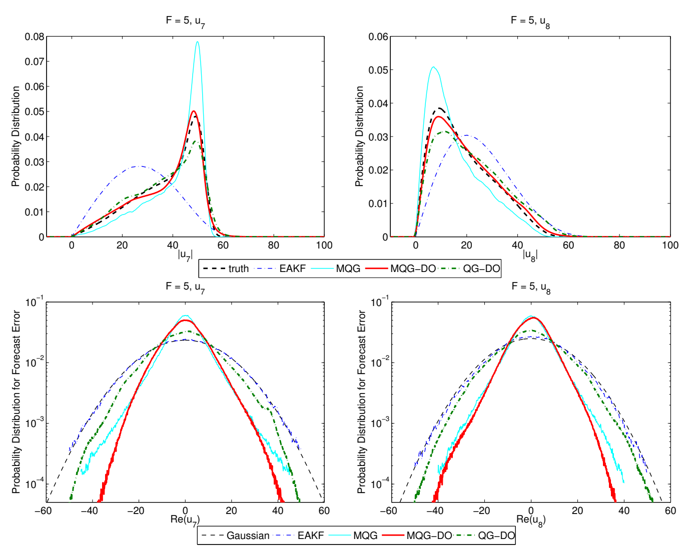

Fig. 1. Comparison of pdfs of the absolute values of the first two leading Fourier modes $\hat{u}_7$, $\hat{u}_8$ (first row), and pdfs of the forecast error $u_j^- - u_{\mathrm{truth}}$ (second row, only real parts are shown) captured by different filtering methods.

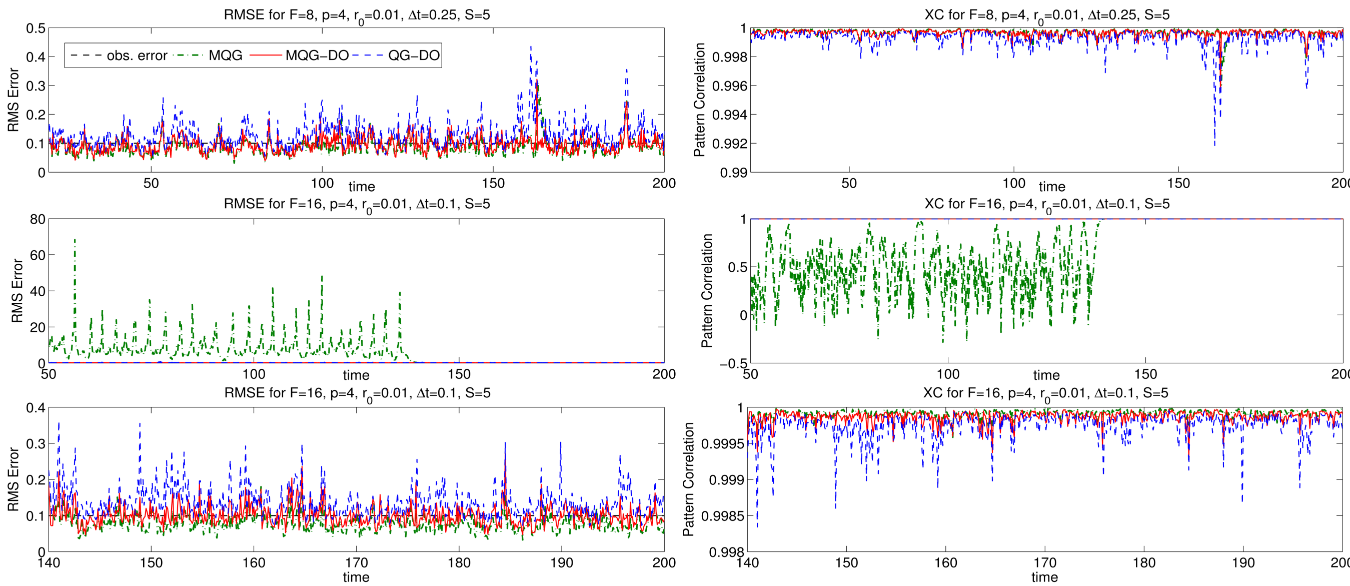

Fig. 2. Comparison of RMS errors (left) and pattern correlations (right) between different filtering methods in regimes $F = 8$ (first row) and $F = 16$ (second and third rows) with sparse infrequent high quality observations.

1. Majda, A. J., Harlim, J. (2012) *Filtering complex turbulent systems.* (Cambridge University Press, Cambridge, UK).
2. Evensen, G. (2003) *The ensemble Kalman filter: Theoretical formulation and practical implementation.* Ocean dynamics 53, 343–367.
3. Farrell, B. F., Ioannou, P. J. (2001) *State estimation using a reduced-order Kalman filter.* J Atmos Sci 58, 3666–3680.
4. Chorin, A. J., Krause, P. (2004) *Dimensional reduction for a Bayesian filter.* P Natl Acad Sci 101, 15013–15017.
5. Bishop, C. H., Etherton, B. J., Majumdar, S. J. (2001) *Adaptive sampling with the ensemble transform Kalman filter, part I: Theoretical aspects.* Mon Weather Rev 129, 420–436.
6. Bickel, P., Li, B., Bengtsson, T. (2008) *Sharp failure rates for the bootstrap particle filter in high dimensions.* In *IMS Collections: Pushing the Limits of Contemporary Statistics: Contributions in Honor of J. K. Ghosh*, vol. 3, Institute of Mathematical Statistics, 318–329.
7. Berliner, L. M., Milliff, R. F., Wikle, C. K. (2003) *Bayesian hierarchical modeling of air-sea interaction.* J Geophys Res: Oceans (1978–2012) 108, 3104–3120.
8. Bengtsson, T., Bickel, P., Li, B. (2008) *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems.* In *IMS Collections in Probability and Statistics: Essays in honor of David A. Freedman*, vol. 2, Institute of Mathematical Statistics, 316–334.
9. Bain, A., Crisan, D. (2009) *Fundamentals of stochastic filtering, stochastic modeling and applied probability.* (Springer, New York), vol. 60.
10. Anderson, J. L. (2003) *A local least squares framework for ensemble filtering.* Mon Weather Rev 131, 634–642.
11. Anderson, J. L. (2001) *An ensemble adjustment Kalman filter for data assimilation.* Mon Weather Rev 129, 2884–2903.
12. Anderson, B., Moore, J. (1979) *Optimal filtering.* (Prentice-Hall Englewood Cliffs, NJ).
13. Szunyogh, I., Kostelich, E. J., Gyarmati, G., Patil, D., Hunt, B. R., Kalnay, E., Ott, E., Yorke, J. A. (2005) *Assessing a local ensemble Kalman filter: perfect model experiments with the national centers for environmental prediction global model.* Tellus A 57, 528–545.
14. Moral, P. D., Jacod, J. (2001) *Interacting particle filtering with discrete observations.* In *Sequential Monte Carlo methods in practice.* (Springer, New York), 43–75.
15. Ghil, M., Malanotte-Rizzoli, P. (1991) *Data assimilation in meteorology and oceanography.* Adv Geophys 33, 141–266.
16. Gershgorin, B., Majda, A. J. (2008) *A nonlinear test model for filtering slow-fast systems.* Commun Math Sci 6, 611–649.
17. Majda, A. J. (2003) *Introduction to PDEs and Waves for the Atmosphere and Ocean.* Courant Lecture Notes in Mathematics, vol. 9, (American Mathematical Society).
18. Majda, A. J., Timofeyev, I., Vanden-Eijnden, E. (2001) *A mathematical framework for stochastic climate models.* Commun Pur Appl Math 54, 891–974.
19. Majda, A. J., Franzke, C., Khouider, B. (2008) *An applied mathematics perspective on stochastic modelling for climate.* Philos T Roy Soc A: Mathematical, Physical and Engineering Sciences 366, 2427–2453.
20. Abramov, R. V., Majda, A. J. (2007) *Blended response algorithms for linear fluctuation-dissipation for complex nonlinear dynamical systems.* Nonlinearity 20(12), 2793–2821.
21. Lorenz, E. N., Emanuel, K. A. (1998) *Optimal sites for supplementary weather observations: Simulation with a small model.* J Atmos Sci 55, 399–414.
22. Lorenz, E. N. (1996) *Predictability: A problem partly solved.* Proceedings of a Seminar on Predictability (European Centre for Medium-Range Weather Forecasts, Reading, UK), vol. 1, 1–18.
23. Grooms, I., Lee, Y., Majda, A. J. (2014) *Ensemble Kalman filters for dynamical systems with unresolved turbulence.* J Comput Phys, in press.
24. Majda, A. J., Wang, X. (2006) *Nonlinear dynamics and statistical theories for basic geophysical flows.* (Cambridge University Press, Cambridge, UK).
25. Salmon, R. (1998) *Lectures on geophysical fluid dynamics.* (Oxford University Press, Oxford, UK), vol. 378.
26. Hoteit, I., Luo, X., Pham, D.-T. (2012) *Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters.* Mon Weather Rev 140, 528–542.
27. Sorenson, H. W., Alspach, D. L. (1971) *Recursive Bayesian estimation using Gaussian sums.* Automatica 7, 465–479.
28. Lermusiaux, P. F. (2006) *Uncertainty estimation and prediction for interdisciplinary ocean dynamics.* J Comput Phys 217, 176–199.
29. Sapsis, T. P., Lermusiaux, P. F. (2009) *Dynamically orthogonal field equations for continuous stochastic dynamical systems.* Physica D: Nonlinear Phenomena 238, 2347–2360.
30. Sapsis, T. P., Majda, A. J. (2013) *Statistically accurate low-order models for uncertainty quantification in turbulent dynamical systems.* P Natl Acad Sci 110, 13705–13710.
31. Sapsis, T. P., Majda, A. J. (2013) *A statistically accurate modified quasilinear Gaussian closure for uncertainty quantification in turbulent dynamical systems.* Physica D: Nonlinear Phenomena 252, 34–45.
32. Sapsis, T. P., Majda, A. J. (2013) *Blending modified Gaussian closure and non-Gaussian reduced subspace methods for turbulent dynamical systems.* J of Nonlin Sci, 10.1007/s00332-013-9178-1.
33. Sapsis, T. P., Majda, A. J. (2013) *Blended reduced subspace algorithms for uncertainty quantification of quadratic systems with a stable mean state.* Physica D: Nonlinear Phenomena 258, 61–76.