MULTI-SCALE METHODS FOR DATA ASSIMILATION IN TURBULENT SYSTEMS*

YOONSANG LEE† AND ANDREW J. MAJDA†

Abstract. Data assimilation of turbulent signals is an important challenging problem because of the extremely complicated large dimension of the signals and incomplete partial noisy observations which usually mix the large scale mean flow and small scale fluctuations. Due to the limited computing power in the foreseeable future, it is desirable to use multi-scale forecast models which are cheap and fast to mitigate the curse of dimensionality in turbulent systems; thus model errors from imperfect forecast models are unavoidable in the development of a data assimilation method in turbulence. Here we propose a suite of multi-scale data assimilation methods which use stochastic Superparameterization as the forecast model. Superparameterization is a seamless multi-scale method for parameterizing the effect of small scales by cheap local problems embedded in a coarse grid. The key ingredient of the multi-scale data assimilation methods is the systematic use of conditional Gaussian mixtures which make the methods efficient by filtering a subspace whose dimension is smaller than the full state. The multi-scale data assimilation methods proposed here are tested on a six dimensional conceptual dynamical model for turbulence which mimics interesting features of anisotropic turbulence including two way coupling between the large and small scale parts, intermittencies, and extreme events in the smaller scale fluctuations. Numerical results show that suitable multi-scale data assimilation methods have high skill in estimating the most energetic modes of turbulent signals even with infrequent observation times.

Key words. data assimilation, filtering, multi-scale, Superparameterization, turbulence

AMS subject classifications. 93E11, 62M20, 62L12,65C20, 76F55, 34E13

1. Introduction. Data assimilation is the process of providing the best statistical estimate of a true signal by incorporating observations into a forecast model. Especially the data assimilation of a turbulent system is a formidable task due to the extremely complicated large dimension of the signal and its incomplete partial observations. Turbulent systems are usually chaotic systems with a large number of active degrees of freedom and thus turbulent signals from nature have a wide range of spatio-temporal scales [9, 30, 31]. The behavior of the large scale process depends on nonlinear feedbacks involving the energy from the small scale process and thus it is important to account for the effect of the small scales on the large scale dynamics. But it is computationally impossible to resolve all these small scale processes in the foreseeable future due to the tremendously large number of dimensions of turbulent systems. Therefore, it is necessary to develop filtering or data assimilation methods for turbulent signals using cheap and fast forecast models which are imperfect models with model errors to make accurate predictions of the future state with incomplete observations mixing the large and small scales [22].

As a reduced or cheap forecast model, various multi-scale methods were introduced to mitigate the curse of dimensionality of turbulent systems. Among others, Superparameterization is a multi-scale algorithm that was originally developed for the purpose of parameterizing unresolved cloud process in tropical atmospheric convection [11, 12, 20]. This conventional Superparameterization resolves the large scale mean

*The research of A.J.M. is partially supported by Office of Naval Research (ONR) Grants, ONR-Departmental Research Initiative N0014-10-1-0554, ONR N0014-11-1-0306, and ONR Multidisciplinary University Research Initiative 25-74200-F7112. Y.L. is supported as a postdoctoral research fellow on this last award.

†Department of Mathematics and Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012 (ylee@cims.nyu.edu, jonjon@cims.nyu.edu)

flow on a coarse grid in a physical domain while the fluctuating parts are resolved using a fine grid high resolution simulation on periodic domains embedded in the coarse grid. A much cheaper version of Superparameterization, stochastic Superparameterization [21, 14, 15, 20], replaces the nonlinear eddy terms by quasilinear stochastic processes on formally infinite embedded domains where the stochastic processes are Gaussian conditional to the large scale mean flow. This conditional Gaussian closure approximation results in a seamless algorithm without using the high resolution space grid for the small scales and is much cheaper than the conventional Superparameterization, with significant success in difficult test problems [14, 15].

In this study we will discuss multi-scale data assimilation or filtering methods for turbulent systems, using stochastic Superparameterization as the forecast model in the practically important setting where the observations mix the large and small scales. The key idea of the multi-scale data assimilation method is to use conditional Gaussian mixtures [25, 7] whose distributions are compatible with Superparameterization. The method uses particle filters [3] or ensemble filters on the large scale part whose dimension is small enough so that the non-Gaussian statistics of the large scale part can be calculated from a particle filter whereas the statistics of the small scale part are conditionally Gaussian given the large scale part. This framework is not restricted to Superparameterization as the forecast model and other cheap forecast model can also be employed. See [25] for another multi-scale filter with quasilinear Gaussian dynamically orthogonality method as the forecast method in an adaptively evolving low dimensional subspace without using Superparameterization. We note that data assimilation using Superparameterization has already been discussed in [17] with noisy observations of the large scale part of the signal alone. Here in contrast to [17] we consider multi-scale data assimilation methods with noisy observations with contributions from both the large and small scale parts of the signal, which is a more difficult problem than observing only the large scale because it requires accurate estimation of statistical information of the small scales. Also mixed observations of the large and small scale parts occur typically in real applications. For example, in geophysical fluid applications, the observed quantities such as temperature, moisture, and the velocity field necessarily mix both the large and small scale parts of the signal [6, 22].

Here the multi-scale data assimilation methods are tested for a conceptual dynamical model for turbulence which was proposed and studied by the authors [24]. The conceptual model is the simplest model for anisotropic turbulence and is given by a $K+1$ dimensional stochastic differential equation (SDE) with deterministic energy conserving nonlinear interactions between the large scale mean flow and the smaller scale fluctuating components. The fluctuating parts have statistical equilibrium state for given large scale mean flow but the fluctuating parts develop instability through chaotic behavior of the large mean flow and this instability generates nontrivial nonlinear feedbacks on the large mean flow too. Through this complicated interaction between the large mean flow and the fluctuating turbulent parts, the conceptual model mimics interesting features of anisotropic turbulence. It has a wide range of scales where the large scale mean is usually chaotic but more predictable with sub-Gaussian probability distribution functions (PDFs) than the smaller scale fluctuations which have intermittency and fat tailed non-Gaussian PDFs while the overall PDF is essentially Gaussian. The large scale components contain more energy than the smaller scale components and the large scale fluctuating components decorrelate faster in time than the mean flow while the smaller scale fluctuating components decorrelate faster

in time than the larger scale fluctuating components.

This paper is organized in the following way. The conceptual model is reviewed in Section 2 followed by Superparameterization of the conceptual model using conditional Gaussian closure approximation for small scales. In Section 3, we describe a suite of multi-scale data assimilation algorithms for turbulent systems using Superparameterization as the forecast method. The most complete multi-scale algorithm involves a particle filter on the large scales with simpler variants involving ensemble filters for the large scales [2, 22], and for example, projected filters for the turbulent fluctuations [13]; these multi-scale methods for data assimilation are all derived from the conditional Gaussian framework [25] which is compatible with the Superparameterization forecast models. In Section 4, we show the numerical results of the multi-scale filtering methods applied to the conceptual model. We finish this paper in Section 5 with a brief summary and a discussion of future directions.

2. Conceptual Dynamical Models for Turbulence and Superparameterization. The conceptual dynamical model introduced and studied by the authors in [24] is a simple $K+1$ dimensional SDE mimicking the interesting features of anisotropic turbulence even for a small number K. Thus it is a useful test bed for multi-scale algorithms and strategies for data assimilation. In this section we briefly review the conceptual dynamical model with its interesting features resembling actual turbulence [9, 30, 31]. Also stochastic Superparameterization for the conceptual model using Gaussian closure for the small scales conditional to the large scale variable will be discussed in detail.

2.1. Conceptual Dynamical Model for Turbulence. In [24], a low dimensional stochastic dynamical system is introduced and studied as a conceptual dynamical model for anisotropic turbulence. It is a simple $K+1$ dimensional SDE which captures the key features of vastly more complicated turbulent systems. The model involves a large scale mean flow, $\overline{u}$, and turbulent fluctuations, $\mathbf{u}' = (u_1', u_2', ..., u_K')$, on a wide range of spatial scales with energy-conserving wave-mean flow interactions as well as stochastic forcing in the fluctuations. Although the model is not derived quantitatively from the Navier-Stokes equation, it mimics key statistical features of vastly more complex anisotropic turbulent systems in a qualitative fashion [30, 9, 31]: 1) Large scale mean is usually chaotic but more predictable than the smaller scale fluctuations; 2) The large-scale mean flow and the smaller-scale fluctuations have nontrivial nonlinear interactions which conserve energy; 3) There are wide ranges of scales for the fluctuations. The large scale components contain more energy than the smaller scale components. Also the large scale fluctuating components decorrelate faster in time than the mean flow while the smaller scale fluctuating components decorrelate faster in time than the larger scale components. 4) The overall turbulent field has a nearly Gaussian PDF while the large scale mean flow has a sub-Gaussian PDF. The larger scale components of fluctuations are nearly Gaussian while the smaller scale components are intermittent, and have fat tailed PDFs, i.e., much more extreme events than a Gaussian distribution.

Following the above discussion, the conceptual dynamical model for turbulence introduced in [24] is the following $K+1$ dimensional stochastic differential equation

$$\begin{aligned} \frac{d\overline{u}}{dt} &= -\overline{d}\,\overline{u} + \gamma \sum_{k=1}^{K} (u_k')^2 - \overline{\alpha}\,\overline{u}^3 + \overline{F}, \\ \frac{du_k'}{dt} &= -d_k u_k' - \gamma \overline{u} u_k' + \sigma_k \dot{W}_k, \quad 1 \le k \le K. \end{aligned} \tag{2.1}$$

where $\dot{W}_k$ are independent white noises for each k. Mean scalar variable $\overline{u}$ represents the largest scale and a family of small scale variables, $u'_k, 1 \le k \le K$, represent contributions to the turbulent fluctuations with $u' = \sum_k^K u'_k$ the turbulent fluctuations. The large scale $\overline{u}$ can be regarded as the large scale spatial average of the turbulent dynamics at a single grid point in a more complex system while u'_k is the amplitude of the k-th Fourier cosine mode evaluated at a grid point. Thus it is straightforward to generalize the conceptual model to many large-scale grid points, which yields a coupled system of equations on the large scales [24].

There are random forces on the fluctuating turbulent modes, $u'_k, 1 \le k \le K$, to mimic the nonlinear interactions between turbulent modes, while the large scale mean flow $\overline{u}$ has only a deterministic constant force $\overline{F}$. But the large scale $\overline{u}$ can have fluctuating, chaotic dynamics in time through interactions with turbulence and its own intrinsic dynamics. The reader easily verifies that the nonlinear interactions in (2.1) also conserve the total energy of the mean and fluctuations

$$E = \frac{1}{2}\left(\overline{u}^2 + \sum_k^K u_k'^2\right). \tag{2.2}$$

The large scale damping, $\overline{d}$, can be positive with $\overline{\alpha} = 0$ or negative with $\overline{\alpha} > 0$ but it is essential to have $d_k > 0$ in order for the turbulence to have a statistical steady state. For a fixed $\gamma > 0$, the large scale can destabilize the smaller scales in the turbulent fluctuations intermittently provided that $-d_k - \gamma\overline{u} > 0$ for the k-th mode and the chaotic fluctuation of $\overline{u}$ creates intermittent instability in $u'_k, 1 \le k \le K$. Thus, the overall system can have a statistical steady state while there is intermittent instability on the small scales creating non-Gaussian intermittent behavior in the system. It can also be shown that (2.1) is geometrically ergodic which means that a unique smooth ergodic invariant measure exists with exponential convergence of suitable statistics from time averages in the long time limit. More details can be found in [24].

Figure 1 shows segments of time series of each mode, $u, \overline{u}$, and $u'_k, 1 \le k \le K = 5$, of a six mode conceptual model using negative large scale damping $\overline{d} = -0.1 < 0$ and their corresponding PDFs in log-scale from a direct numerical simulation with Gaussian fits in dashed lines. The other parameters for this test model are $\alpha = 0.05$, $\gamma = 1.6$ and $\overline{F} = -0.057$. The overall field $u = \overline{u} + \sum_k u_k$ is almost Gaussian while the large scale mean $\overline{u}$ is sub-Gaussian. For the turbulent mode $u_k, 1 \le k \le 5$, the smaller parts of the fluctuating modes have more intermittency than the larger parts which can be seen from the fat tailed PDFs of the smaller parts compared to the slightly sub-Gaussian tails of the larger parts. In this test regime selective small scale damping

$$d_k = 1 + 0.01k^2, \quad 1 \le k \le 5 \tag{2.3}$$

and 5/3 energy spectrum

$$\frac{\sigma_k^2}{2d_k} = E_k = E_0|k|^{-5/3}, 1 \le k \le 5 \tag{2.4}$$

are employed to mimic the Kolmogorov spectrum [9, 22]. The statistical mean, variance, and correlation times of $u, \overline{u}$, and $u'_k, 1 \le k \le 5$ are shown in Table 1. The large mean $\overline{u}$ has more variance and decorrelates slower than the larger parts of fluctuating

parts while the smaller parts have less variance and decorates faster than the larger parts. In this experiment the cross correlation between different modes are less than 5%.

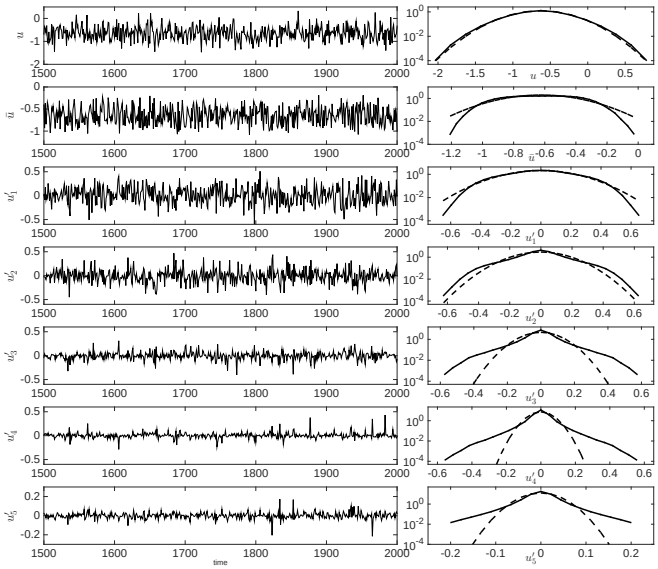

FIG. 1. *Negative large-scale damping. Time series of each mode, $u, \overline{u}$, and $u'_k, 1 \le k \le K$, (left column) and its corresponding PDFs in log-scale (right column). $\overline{d} = -0.1$, $\alpha = 0.05$, $\gamma = 1.6$, and $\overline{F} = -0.057$.*

	mean	variance	correlation time
u	-0.6282	0.1057	25.96
$\overline{u}$	-0.6226	0.0412	38.95
u'_1	-0.0060	0.0355	27.44
u'_2	-0.0003	0.0181	19.21
u'_3	0.0017	0.0071	9.065
u'_4	-0.0005	0.0028	6.036
u'_5	-0.0004	0.0011	4.274

TABLE 1

Negative large-scale damping. Mean, variance, and correlation time of each mode, $u, \overline{u}$, and $u'_k, 1 \le k \le K$. $\overline{d} = -0.1$, $\alpha = 0.05$, $\gamma = 1.6$, and $\overline{F} = -0.057$.

2.2. Superparameterization in the conceptual model. Now we describe the stochastic Superparameterization of the conceptual model (2.1). Stochastic Superparameterization [20, 14, 15] is a seamless multi-scale method parameterizing the small scale eddy terms by quasilinear stochastic processes embedded in a formally infinite domain instead of periodic domains in physical space as in conventional Superparameterization [11, 12]. In stochastic Superparameterization, which we call Superparameterization for the remainder of the paper, the large scale mean flow is resolved on a coarse grid in a physical domain while the fluctuating small scales are

closed by stochastic processes under a Gaussian assumption conditional to the large scale mean flow [21, 20, 14, 15].

In Superparameterization, it is implicitly assumed that there is modest scale separation in time between $\overline{u}$ and $u'_k, 1 \le k \le K$. Due to the scale separation assumption, the nonlinear term containing u'_k in the equation of $\overline{u}$ is replaced by the statistical average, that is, its variance $R'_{k,k}$

$$\frac{d\overline{u}}{dt} = -\overline{d}\overline{u} + \gamma \sum_{k=1}^{K} R'_{k,k} - \alpha\, \overline{u}^3 + \overline{F}. \tag{2.5}$$

For the equation of $u'_k, 1 \le k \le K$, on the other hand, $\overline{u}$ is regarded as a fixed parameter and the equations for the mean and variance of u'_k are closed under the Gaussian assumption

$$\begin{aligned} \frac{d\langle u'_k\rangle}{dt} &= -(d_k + \gamma\overline{u})\langle u'_k\rangle, \quad 1 \le k \le K, \\ \frac{dR'_{k,k}}{dt} &= -2(d_k + \gamma\overline{u})R'_{k,k} + \tilde{\sigma}_k^2 \end{aligned} \tag{2.6}$$

where there is no cross correlation, that is,

$$\frac{dR'_{k,l}}{dt} = 0, \quad k \ne l. \tag{2.7}$$

Due to approximation in Superparameterization, the same noise level $\sigma_k, 1 \le k \le K$ of the true signal does not guarantee the same level of the stationary state variance of Superparameterization and thus Superparameterization uses a tunable noise level $\tilde{\sigma}_k, 1 \le k \le K$ to match the stationary state variance of Superparameterization with the stationary state variance of the true signal. Note that stationary state information of the true signal is usually assumed to be provided in data assimilation such as climatology in atmosphere science and here we also assume that the stationary state information is available for Superparameterization. Instead of tuning each noise level σ_k, Superparameterization mimics the spectrum of the true signal by setting a relation between the frozen $\overline{u} = \langle\overline{u}^\infty\rangle$ stationary state $\frac{\tilde{\sigma}_k^2}{2(d_k+\gamma\langle\overline{u}^\infty\rangle)}$ and the true signal spectrum where $\langle\overline{u}^\infty\rangle$ is the statistical stationary state mean of $\overline{u}$. That is, Superparameterization tunes the following relation coefficient A

$$\frac{\tilde{\sigma}_k^2}{2(d_k + \gamma\langle\overline{u}^\infty\rangle)} = A \times (\text{k-th mode stationary state variance of the true model}) \tag{2.8}$$

so that the actual stationary variance of the numerical solution by Superparameterization matches the true signal variance. Note that A is not necessarily 1 due to variability in $\overline{u}$.

3. Multi-scale Data Assimilation Algorithms. Due to limited computing resources and power, resolving all the small scales of turbulent signals with large statistical ensembles is impossible in the foreseeable future. Thus, it is important to use cheap and fast forecast models which can mitigate the curse of dimensionality of turbulence. In this section multi-scale data assimilation algorithms with Superparameterization as the forecast model are developed. The method uses Gaussian mixtures similar to those in [18, 29] but it is essentially different from those in that

conditional Gaussian distributions are applied in the reduced subspace $\mathbf{u}'$ combined with particle approximations in the lower-dimensional subspace $\overline{\mathbf{u}}$ where $\overline{\mathbf{u}} \in \mathbb{R}^{N_1}$ and $\mathbf{u}' \in \mathbb{R}^{N_2}, N_1 < N_2$, are two components of the turbulent signal $\mathbf{u} \in \mathbb{R}^N = (\overline{\mathbf{u}}, \mathbf{u}')$. Thus the proposed method is more efficient than conventional data assimilation or filtering algorithms which use the whole space $\mathbf{u}$ for the filter.

Another important issue is that the observations mix the large and small scale parts, $\overline{\mathbf{u}}$ and $\mathbf{u}'$ respectively. That is, observations $\mathbf{v} \in \mathbb{R}^M$ mix $\overline{\mathbf{u}}$ and $\mathbf{u}'$ through a nonlinear observation operator $G(\overline{\mathbf{u}}, \mathbf{u}') : \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \to \mathbb{R}^M$

$$\mathbf{v} = G(\overline{\mathbf{u}}, \mathbf{u}') + \sigma_\theta \tag{3.1}$$

where σ_θ is the observation noise error with a probability distribution $p_\theta(\mathbf{v} - G(\overline{\mathbf{u}}, \mathbf{u}'))$ for the observational noise. One approach to deal with mixed observations is to treat the contribution from the small scales as a component of observation error (which is known as 'representation error' or 'representativeness error' [19, 5]) and use the method of [17] which is a multi-scale method with observation of only the large scale variables. But this approach has a limitation in that it only provides the prediction for the large scales. The multi-scale method proposed here, on the other hand, can provide predictions for the energetic small scale modes in addition to the large scales.

We first develop a multi-scale data assimilation method using particle filters on the large scale $\overline{\mathbf{u}}$ and Superparameterization as the forecast method followed by an alternative using ensemble filters (Chapter 9 of [22]) on the large scales which can be used for the case when N_1 is large so that the use of a particle filter is prohibitive. Then we develop an ensemble filter with projected turbulent scales following [13]. We end this section with full-scale filtering methods using Gaussian closure, which provides the statistics of $\overline{\mathbf{u}}$ in addition to $\mathbf{u}'$ without using particles or ensembles, to compare with the multi-scale filtering methods. The proposed multi-scale data assimilation framework is not restricted to the use of Superparameterization as the forecast model and other cheap and fast forecast models can be employed with some modifications. See [25] which uses the same multi-scale framework with for example quasilinear Gaussian closure [28] as the forecast model.

3.1. Particle Filter with Superparameterization (PF-SP). Superparameterization retains the large scale variables by resolving them on a coarse grid while the effect of the small scales on the large scales is parameterized by approximating the small scales on local or reduced spaces. Stochastic Superparameterization discussed in the previous section uses Gaussian closure for the small scales conditional to the large scale variable $\overline{\mathbf{u}}$ [21, 20, 14, 15]. Thus we consider a multi-scale filtering algorithm with forecast prior distributions given by the conditional distribution

$$p^f(\mathbf{u}) = p^f(\overline{\mathbf{u}}, \mathbf{u}') = p^f(\overline{\mathbf{u}}) p^f_G(\mathbf{u}'|\overline{\mathbf{u}}) \tag{3.2}$$

where $p^f_G(\mathbf{u}'|\overline{\mathbf{u}})$ is a Gaussian distribution conditional to $\overline{\mathbf{u}}$

$$p^f_G(\mathbf{u}'|\overline{\mathbf{u}}) = \mathcal{N}(\mathbf{u}'(\overline{\mathbf{u}}), \mathbf{R}'(\overline{\mathbf{u}})). \tag{3.3}$$

Here we assume that N_1 is sufficiently small enough that particle filters can be applied to the large scales. For a low dimensional space $\overline{\mathbf{u}}$, the marginal distribution of $\overline{\mathbf{u}}$ can be approximated by Q particles

$$p^f(\overline{\mathbf{u}}) = \sum_{j=1}^{Q} p^f_j \delta(\overline{\mathbf{u}} - \overline{\mathbf{u}}_j) \tag{3.4}$$

where $p_j^f \geq 0$ are particle weights such that $\sum_j p_j^f = 1$. After the forecast step where Superparameterization is applied to each particle member, we have the following general form for the prior distribution $p^f(\mathbf{u})$

$$\begin{aligned} p^f(\mathbf{u}) = p^f(\overline{\mathbf{u}}, \mathbf{u}') &= \sum_{j=1}^{Q} p_j^f \delta(\overline{\mathbf{u}} - \overline{\mathbf{u}}_j) p_G^f(\mathbf{u}'^f | \overline{\mathbf{u}}_j) \\ &= \sum_{j=1}^{Q} p_j^f \delta(\overline{\mathbf{u}} - \overline{\mathbf{u}}_j) \mathcal{N}(\mathbf{u}'(\overline{\mathbf{u}}_j)^f, \mathbf{R}'(\overline{\mathbf{u}}_j)^f), \end{aligned} \tag{3.5}$$

which is a conditional Gaussian mixture distribution where each summand is a Gaussian distribution conditional to $\overline{\mathbf{u}}_j$. The Gaussian mixture has already been used in data assimilation [29, 18, 1] but the multi-scale method developed here is different in that conditional Gaussian distributions are applied in the reduced subspace $\mathbf{u}'$ with particle approximations only in the lower dimensional subspace $\overline{\mathbf{u}}$. Thus the proposed multi-scale data assimilation method can be highly efficient and fast in comparison with conventional data assimilation methods which use the whole space for the filter.

For a general nonlinear observation operator G of (3.1) and observational noise distribution $p_\theta(\mathbf{v} - G(\overline{\mathbf{u}}_j, \mathbf{u}'))$, the posterior distribution is not necessarily in the same form as the prior distribution. If we restrict the multi-scale observation operator mildly and assume that the observational noise error $p_\theta(\mathbf{v} - G(\overline{\mathbf{u}}_j, \mathbf{u}')) = \mathcal{N}(\mathbf{0}, \mathbf{r}_\theta)$, is Gaussian, the posterior distributions has the same form as the prior distribution, (3.5).

PROPOSITION 3.1. *Assume that the prior distribution from the forecast is in the from* (3.5) *and that the observations have the following structure*

$$\mathbf{v} = G(\overline{\mathbf{u}}, \mathbf{u}') + \sigma_\theta = \overline{\mathbf{G}}\overline{\mathbf{u}} + \mathbf{G}'(\overline{\mathbf{u}})\mathbf{u}' + \sigma_\theta. \tag{3.6}$$

where $\mathbf{G}'(\overline{\mathbf{u}}_j)$ *has rank* M*. Then the posterior distribution in the analysis step taking into account the observations* (3.6) *is in the form of* (3.5)

$$p^a(\mathbf{u}) = p^a(\overline{\mathbf{u}}, \mathbf{u}') = \sum_{j=1}^{Q} p_j^a \delta(\overline{\mathbf{u}} - \overline{\mathbf{u}}_j) \mathcal{N}(\mathbf{u}'(\overline{\mathbf{u}}_j)^a, \mathbf{R}'(\overline{\mathbf{u}}_j)^a), \tag{3.7}$$

The new mixture weights are

$$p_j^a = \frac{p_j^f I_j}{\sum_{k=1}^{Q} p_k^f I_k} \tag{3.8}$$

where $I_j = \int p(\mathbf{v}|\overline{\mathbf{u}}_j, \mathbf{u}') p(\mathbf{u}'|\overline{\mathbf{u}}_j) d\mathbf{u}'$ *and for each particle* $\overline{\mathbf{u}}_j$*, the posterior mean and variance of* $\mathbf{u}'$*,* $\mathbf{u}'(\overline{\mathbf{u}_\mathbf{j}})^a$ *and* $\mathbf{R}'(\overline{\mathbf{u}_\mathbf{j}})^a$ *respectively, are*

$$\begin{aligned} \mathbf{u}'(\overline{\mathbf{u}}_j)^a &= \mathbf{u}'^f + \mathbf{K}'(\mathbf{v} - \overline{\mathbf{G}}\overline{\mathbf{u}}_j^f - \mathbf{G}'(\overline{\mathbf{u}}_j^f)\mathbf{u}') \\ \mathbf{R}'(\overline{\mathbf{u}}_j)^a &= (\mathbf{I} - \mathbf{K}'\mathbf{G}'(\overline{\mathbf{u}}_j^f))\mathbf{R}'(\overline{\mathbf{u}})^f \end{aligned} \tag{3.9}$$

where the Kalman gain matrix $\mathbf{K}'$ *is given by*

$$\mathbf{K}' = \mathbf{R}'^f \mathbf{G}'(\overline{\mathbf{u}}_j^f)^T (\mathbf{G}'(\overline{\mathbf{u}}_j^f)\mathbf{R}'^f \mathbf{G}'(\overline{\mathbf{u}}_j^f)^T + \mathbf{r}_\theta)^{-1}. \tag{3.10}$$

For the proof of Proposition 3.1, see the supplementary material of [25].

Note that the special form of nonlinear observation operator (3.6) is a Taylor expansion around $\overline{\mathbf{u}}$ of a general nonlinear observation operator G. For simplicity of exposition, we consider a linear observation operator which has contributions from both the large and small scales

$$\mathbf{v} = G(\overline{\mathbf{u}}, \mathbf{u}') + \sigma_\theta = \overline{\mathbf{G}}\overline{\mathbf{u}} + \mathbf{G}'\mathbf{u}' + \sigma_\theta. \tag{3.11}$$

In Superparameterization, each mode of the small scale part is approximated by a quasilinear stochastic process which is independent of each other mode. But the posterior covariance $\mathbf{R}'^a$ includes cross correlations among different components of $\mathbf{u}'$ even for a diagonal prior covariance matrix due to the mixed observation of the small scales. Thus, we need to approximate the posterior covariance matrix to be compatible with Superparameterization which does not have cross correlations between different small scale modes

$$\mathbf{R}'^a \approx \mathrm{diag}(r_1, r_2, ..., r_{N_2}) \tag{3.12}$$

The best approximation is given by the minimizer of the information loss according to [23] and in our case, the minimizer is simply the diagonal components of the posterior covariance matrix. That is, in (3.12) we have

$$r_i = \mathbf{R}'^a_{ii}, \quad 1 \le i \le N_2.$$

Another compatibility issue regarding Superparameterization with the multi-scale data assimilation method is the non-zero posterior means for the small scales. In the simplest stochastic Superparameterization algorithms [14, 15], the linearity of the small scale process initialized by zero means guarantee that the small scales have zero mean unless they are reinitialized by non-zero means and thus the means of the small scales are not explicitly solved. Thus, if the means of the small scales are not explicitly solved, we need to approximate the posterior statistics of the small scales of nonzero means with small scale statistics with zero means. For this compatibility issue, we use two different approaches here. The first approach keeps the means of the small scales and evolves them as in the Superparameterization algorithm discussed in Section 2. The other approach ignores the nonzero means of the small scales and instead approximates them with zero means as in general stochastic Superparameterization. To distinguish these two approaches, we denote the first approach which keeps the means by SP while the cheaper version which does not keeps the mean by SP'. That is, SP solves the first and second equations of (2.6) in addition to (2.5) while SP' solves only the second equation of (2.6) with $\langle u'_k \rangle = 0$ and (2.5). Note that the small scales in SP' are always zero and thus the small scale parts cannot be updated to exactly match the full posterior distribution. We want to mention that there is another more elaborate approach which rescales the small scale variables in Superparameterization so that they match a zero-mean relative entropy minimizing estimate of the posterior distribution but we do not use this here.

It is well known that particle filters suffer from degeneracy where only a few of the particles have a significant weight, and all the other particles have very small weights [27]. Thus, in the multi-scale particle filtering with Superparameterization, various resampling strategies [27] can be employed for the large scale part of the multi-scale particle filter. In addition to the degeneracy problem, a particle filter for a deterministic system with internal instability such as the large scale variable $\overline{u}$ of the conceptual model (2.1) can have duplicate particles after analysis. Especially

when it is solved by Superparameterization which replaces the source of the internal instability, $u_k'^2$, by the variance; thus $\overline{u}$ can suffer from duplicate particles due to the statistical average of the effect from the small scales which has instability. To remedy this problem, we add additional noise by perturbing the large scale variables after the resampling [25]. The perturbation is added as a white noise with a variance, σ^2, proportional to the variance of $\overline{u}$

$$\sigma^2 = \beta \sum_j p_j |\mathbf{u}_j - \langle \mathbf{u} \rangle|^2, \quad \langle \overline{\mathbf{u}} \rangle = \sum_j p_j \overline{\mathbf{u}}_j. \tag{3.13}$$

where p_j is the particle weight of $\overline{u}_j$. Here β can be regarded as an inflation factor similar to the covariance inflation factor in ensemble filters [2, 22].

Now we summarize the multi-scale particle filter with Superparameterization (PF-SP)

PF-SP

At the m-th assimilation step,

1. Evolve the system using Superparameterization, (2.5) and (2.6), for each particle member to provide the prior distributions for the next assimilation step
2. Assimilate the prior distribution with the observations using (3.9)
3. Resample
4. Add additional noise (3.13)
5. Repeat from 1 for the next assimilation cycle

PF-SP'

Similar to PF-SP except that the small scales have zero mean.

3.2. Ensemble Filter with Superparameterization (EF-SP). The use of particle filters is computationally prohibitive for a large dimensional space because a large number of particles are necessary to represent the statistical properties. Thus for the case when N_1 is not sufficiently small for the use of a particle filter, an alternative way such as an ensemble filter is necessary. Here we modify the multi-scale data assimilation framework with Superparameterization in the previous section so that an ensemble filter can be incorporated instead of a particle filter.

To use an ensemble filter on the large scale variables, we assume that $\overline{\mathbf{u}}^f$ is Gaussian with mean and covariance for the overall density

$$\mathbf{u}^f = \begin{pmatrix} \overline{\mathbf{u}}^f \\ \mathbf{u}'^f \end{pmatrix} \quad \text{and} \quad \mathbf{R}^f = \begin{pmatrix} \overline{\mathbf{R}}^f & \mathbf{0} \\ \mathbf{0} & \mathbf{R}'^f \end{pmatrix} \tag{3.14}$$

respectively where we assume that there is no cross correlation between the large and small scale variables. This assumption is somewhat restrictive for non-zero $\mathbf{u}_k'$ because the small scales are directly driven by the large scales. In [13], it is verified for zero mean $\mathbf{u}_k'$ that the dependence of the small scales on the large scales is fully compatible with the no cross correlation assumption between the large and small scales by showing that the covariance matrix of the least-biased Gaussian approximation has no cross correlation. See Appendix A of [13] for more details.

For each ensemble member, Superparameterization gives a prior distribution which is conditionally Gaussian to the large scale variable (3.5). Thus for an ensemble

representation with Q ensemble members, the prior distribution by Superparameterization is given by

$$p^f(\mathbf{u}) = p^f(\overline{\mathbf{u}}, \mathbf{u}') = \frac{1}{Q}\sum_{j=1}^{Q} \delta(\overline{\mathbf{u}} - \overline{\mathbf{u}}_j)\mathcal{N}(\mathbf{u}'(\overline{\mathbf{u}}_j)^f, R'(\overline{\mathbf{u}}_j)^f) \tag{3.15}$$

which can be seen as (3.5) with $p_j^f = \frac{1}{Q}$. From this conditional distribution, the statistics of the small scales, $\mathbf{u}'^f$ and $\mathbf{R}'^f$, of (3.14) are given by the marginalization of the Gaussian mixture with respect to the large scale $\overline{\mathbf{u}}_j$

$$\mathbf{u}'^f = \frac{1}{Q}\sum_{j}^{Q} \mathbf{u}'_j(\overline{\mathbf{u}}_j) \tag{3.16}$$

and

$$\mathbf{R}'^f = \frac{1}{Q}\sum_{j}^{Q} \mathbf{R}'(\overline{\mathbf{u}}_j). \tag{3.17}$$

As in the previous section, we assume that the observation operator is linear (3.11). In this configuration applying the standard Kalman formula to the full state (3.14) with Gaussian observational noise $p_\theta(\mathbf{v} - G(\overline{\mathbf{u}}, \mathbf{u}')) = \mathcal{N}(\mathbf{0}, \mathbf{r}_\theta)$, we can separate the posterior update formulae for the large and small scales

$$\begin{aligned} \overline{\mathbf{u}}^a &= \overline{\mathbf{u}}^f + \overline{\mathbf{K}}(\mathbf{v} - \overline{\mathbf{G}}\overline{\mathbf{u}}^f - \mathbf{G}'\mathbf{u}'^f) \\ \overline{\mathbf{K}} &= \overline{\mathbf{R}}^f\overline{\mathbf{G}}^T(\overline{\mathbf{G}}\,\overline{\mathbf{R}}^f\overline{\mathbf{G}}^T + \mathbf{G}'\mathbf{R}'^f\mathbf{G}'^T + \mathbf{r}_\theta)^{-1} \\ \overline{\mathbf{R}}^a &= (\mathbf{I} - \overline{\mathbf{K}}\,\overline{\mathbf{G}})\overline{\mathbf{R}}^f \end{aligned} \tag{3.18}$$

and

$$\begin{aligned} \mathbf{u}'^a &= \mathbf{u}'^f + \mathbf{K}'(\mathbf{v} - \overline{\mathbf{G}}\overline{\mathbf{u}}^f - \mathbf{G}'\mathbf{u}'^f) \\ \mathbf{K}' &= \mathbf{R}'^f\mathbf{G}'^T(\overline{\mathbf{G}}\,\overline{\mathbf{R}}^f\overline{\mathbf{G}}^T + \mathbf{G}'\mathbf{R}'\mathbf{G}'^T + \mathbf{r}_\theta)^{-1} \\ \mathbf{R}'^a &= (\mathbf{I} - \mathbf{K}'\mathbf{G}')\mathbf{R}'^f \end{aligned} \tag{3.19}$$

respectively. The update of $\overline{u}$ (3.18) is the standard Kalman update formula for $\overline{u}$ except an additional observation error $\mathbf{G}'\mathbf{R}'^f\mathbf{G}'^T$ in addition to the raw observation error $\mathbf{r}_\theta$ (the additional observation error variance induced by the small scale fluctuations is known as 'representation error' or 'representativeness error' [19, 5]). Thus in EF-SP, a standard ensemble filter is used for the update of the large scale $\overline{\mathbf{u}}$ with the prior covariance $\overline{\mathbf{R}}^f$ and an observation error $\mathbf{G}'\mathbf{R}'^f\mathbf{G}'^T + \mathbf{r}_\theta$. With this additional observation noise for the large scales, the large scales trust the dynamics more than without the representation error and the large scale is updated using the ensemble filter while the small scale is updated directly using the Kalman update formula (3.19).

For the small scales, only the marginalized posterior statistics are available and thus the small scales of the ensemble members are updated to have the same marginalized mean and variance as the posterior mean and variance. For the mean, the difference between the prior and posterior marginalized mean is added to each ensemble member,

$$\mathbf{u}'(\overline{\mathbf{u}}_j)^a = \mathbf{u}'(\overline{\mathbf{u}}_j)^f + \mathbf{u}'^a - \mathbf{u}'^f, \quad 1 \le j \le N_1. \tag{3.20}$$

For the variance of each ensemble member, this additive update can make negative variance which is not physical and thus rescaling with the ratio between the prior and posterior marginalized variances is applied to match the posterior variance

$$\mathbf{R}'(\overline{\mathbf{u}}_j)^a = \frac{\mathbf{R}'^a}{\mathbf{R}'^f}\mathbf{R}'(\overline{\mathbf{u}}_j)^f, \quad 1 \le j \le N_1. \tag{3.21}$$

with $\mathbf{R}'^f$ given in (3.17).

Note that as in the multi-scale particle filter in the previous section, the multi-scale ensemble filter has the same compatibility issues - non-zero means for $\mathbf{u}'$ and cross correlations between small scales. For these issues we use the same approaches - keeping the non-zero means for $\mathbf{u}'$ or ignoring them (which are SP and SP' respectively) and approximation of the cross correlation by the diagonal components of the posterior covariance matrix as the minimizer of the information loss.

Now we summarize the multi-scale ensemble filter with Superparameterization (EF-SP).

EF-SP At the m-th assimilation step,

- Evolve the system using Superparameterization, (2.5) and (2.6), for each ensemble member which yields a prior distribution in the form of (3.15); inflate the large scale variance $\overline{\mathbf{R}}^f$. From this prior distribution, calculate the marginalized prior mean and variance (3.16) and (3.17)
- Update the marginalized mean and variance of the large and small scales separately using (3.18) and (3.19); for the large scale update, use an ensemble filter while for the small scale update, use the direct update
- Update the mean and variance of the small scales for each ensemble member using (3.20) and (3.21)

EF-SP'

Similar to EF-SP except that the small scales have zero mean.

3.2.1. Ensemble filter with projected parameterization (EF-PP). Here we discuss a simpler version of EF-SP, which we will call ensemble filter with projected parameterization (EF-PP) which is first studied in [13]. EF-PP has a stronger assumption than EF-SP that the dimension of the large and small scales are the same by compressing all the modes of $\mathbf{u}'$ into a subspace of dimension N_1, $\mathbf{u}'_{comp}$. In the conceptual model, this corresponds to compressing $u'_k, 1 \le k \le K$, into a turbulent field $u' = \sum_k u'_k$ which has the same dimension of $\overline{u}$. Under this configuration, EF-SP' is used for the reduced space $\mathbf{u}_{comp} = \begin{pmatrix} \overline{\mathbf{u}} \\ \mathbf{u}'_{comp} \end{pmatrix} \in \mathbb{R}^{2N_1}$ which provides the updated posterior means and variances of $\overline{\mathbf{u}}$ and $\mathbf{u}'_{comp}$. Because only the compressed information is available for the small scale part in EF-PP, the Superparameterization cannot be reinitialized using the posterior mean and variance for the small scales after each analysis step as in EF-SP. Instead of using the posterior mean and variance for the small scale part, the method in [13] uses the statistical steady state information for the small scales which is available in advance. Note that the small scale mean is always zero in EF-PP.

3.3. Gaussian closure (GC). In the following, we consider single scale data assimilation methods using the mean and variance for the entire state space evolved explicitly under the Gaussian assumption. This filtering method which we call Gaussian Closure (GC) filter provides the statistical information of the large scales without

use of particles or ensembles and thus it is computationally efficient in making prior distributions compared to the previous methods using Superparameterization. But GC filter involves the inversion of covariance matrix of the whole state vector which can be computationally expensive when the dimension of the whole space is large, which is typical for turbulent signals. Thus we present this filter as a reference filter to compare with the multi-scale filters in Section 4.

We first derive the exact equations for the mean and variance and then close them using the Gaussian assumption for each mode. Using the notation

$$\mathbf{u} = \langle \mathbf{u} \rangle + \tilde{\mathbf{u}} \tag{3.22}$$

for ensemble mean and fluctuations, the exact equation for the mean is obtained by substituting (3.22) into (2.1) and then taking ensemble average

$$\begin{aligned} \frac{d\langle \overline{u} \rangle}{dt} &= -\overline{d}\langle \overline{u} \rangle + \gamma \sum_{k=1}^{K} (\langle u_k' \rangle)^2 + \gamma \sum_{k=1}^{K} R_{k,k} - \overline{\alpha} \left(\langle \overline{u} \rangle^3 + 3\langle \overline{u} \rangle R_{00} + \langle \tilde{\overline{u}}^3 \rangle \right) + F_0 \\ \frac{d\langle \mathbf{u}' \rangle}{dt} &= -\mathbf{D}_K \langle \mathbf{u}' \rangle - \gamma \langle \overline{u} \rangle \langle \mathbf{u}' \rangle - \gamma \mathbf{R}_{01} \end{aligned} \tag{3.23}$$

with a linear damping matrix $\mathbf{D}_K$

$$\mathbf{D}_K = \mathrm{diag}(d_1, d_2, ..., d_K) \tag{3.24}$$

and R_{ij} the ij-th component of the covariance matrix $\mathbf{R}$

$$\mathbf{R} = \begin{pmatrix} R_{00} & \mathbf{R}_{01}^T \\ \mathbf{R}_{01} & \mathbf{R}_{11} \end{pmatrix}, \quad 1 \le i, j \le K. \tag{3.25}$$

On the other hand, the exact equation for the covariance is obtained by taking the ensemble average of the second moment term of the ensemble fluctuating parts [28]

$$\frac{d\mathbf{R}}{dt} = L(\langle \mathbf{u} \rangle)\mathbf{R} + \mathbf{R}L^T(\langle \mathbf{u} \rangle) + \mathbf{Q}_F + \mathbf{Q}_\sigma \tag{3.26}$$

where

- the quasi-linear operator $L(\langle \mathbf{u} \rangle)$ is given by

$$L(\langle \mathbf{u} \rangle)\mathbf{v} = \begin{pmatrix} -\overline{d} & \mathbf{0} \\ \mathbf{0} & -\mathbf{D}_K \end{pmatrix} \mathbf{v} + \begin{pmatrix} -3\overline{\alpha}\langle \overline{u} \rangle^2 & 2\gamma \langle \mathbf{u}' \rangle^T \\ -\gamma \langle \mathbf{u}' \rangle & -\gamma \langle \overline{u} \rangle I_K \end{pmatrix} \mathbf{v} \tag{3.27}$$

  and I_K is the $K \times K$ identity matrix
- $\mathbf{Q}_F$ contains the contributions from third and forth order moments

$$\begin{aligned} \mathbf{Q}_F = &\begin{pmatrix} 2\gamma \langle \tilde{\overline{u}} \sum_k \tilde{u'}_k^2 \rangle & \gamma \langle (-\tilde{\overline{u}}^2 + |\tilde{\mathbf{u}}'|^2) \tilde{\mathbf{u}'}^T \rangle \\ \gamma \langle (-\tilde{\overline{u}}^2 + |\tilde{\mathbf{u}}'|^2) \tilde{\mathbf{u}}' \rangle & -2\gamma \langle \tilde{\overline{u}} \tilde{\mathbf{u}}' \tilde{\mathbf{u}'}^T \rangle \end{pmatrix} \\ &- \overline{\alpha} \begin{pmatrix} 6\langle \overline{u} \rangle \langle \tilde{\overline{u}}^3 \rangle + 2\langle \tilde{\overline{u}}^4 \rangle & 3\langle \overline{u} \rangle \langle \tilde{\overline{u}}^2 \tilde{\mathbf{u}'}^T \rangle + \langle \tilde{\overline{u}}^3 \tilde{\mathbf{u}'}^T \rangle \\ 3\langle \overline{u} \rangle \langle \tilde{\overline{u}}^2 \tilde{\mathbf{u}}' \rangle + \langle \tilde{\overline{u}}^3 \tilde{\mathbf{u}}' \rangle & \mathbf{0} \end{pmatrix} \end{aligned} \tag{3.28}$$

- the positive definite operator due to external stochastic forcing $\mathbf{Q}_\sigma$ is defined as

$$\mathbf{Q}_\sigma = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{0} & \sigma_k^2 \delta_{k,i} \end{pmatrix}. \tag{3.29}$$

The above exact statistical equations (3.23) and (3.26) which include the contributions from the third and fourth moments are not closed. Because the equations for the third or fourth moments include higher moments it is impossible to close the exact equations for the mean and variance in general setting. A commonly used approach to close the exact equations is to assume that all modes are Gaussian [26, 8, 28] so that the third moment is zero while the normalized fourth moment is 3. From these properties of Gaussian distributions, we have the following closed equations for the mean and variance of the conceptual model

$$
\begin{aligned}
\frac{d\langle\overline{u}\rangle}{dt} &= -\overline{d}\langle\overline{u}\rangle + \gamma\sum_{k=1}^{K}(\langle u_k'\rangle)^2 + \gamma\sum_{k=1}^{K}R_{k,k} - \overline{\alpha}\left(\langle\overline{u}\rangle^3 + 3\langle\overline{u}\rangle R_{00}\right) + F_0 \\
\frac{d\langle\mathbf{u}'\rangle}{dt} &= -\mathbf{D}_K\langle\mathbf{u}'\rangle - \gamma\langle\overline{u}\rangle\langle\mathbf{u}'\rangle \\
\frac{dR}{dt} &= L(\langle\mathbf{u}\rangle)R + RL^T(\langle\mathbf{u}\rangle) - \overline{\alpha}\begin{pmatrix} 6R_{00}^2 & 3R_{0k}R_{00} \\ 3R_{k0}R_{00} & \sigma_k^2 \end{pmatrix}
\end{aligned}
\tag{3.30}
$$

Now (3.30) provides the statistical informations for all modes and thus we apply the standard Kalman filter [4] for the whole space yielding the following posterior update formula

$$
\begin{aligned}
\mathbf{u}^a &= \mathbf{u}^f + \mathbf{K}(\mathbf{v} - \mathbf{G}\mathbf{u}^f) \\
\mathbf{K} &= \mathbf{R}^f\mathbf{G}^T(\mathbf{G}\mathbf{R}^f\mathbf{G}^T + \mathbf{r}_\theta)^{-1} \\
\mathbf{R}^a &= (\mathbf{I} - \mathbf{K}\mathbf{G})\mathbf{R}^f
\end{aligned}
\tag{3.31}
$$

which is the GC filter mentioned above.

It is well known that filtering with direct models, such as (3.30) and (3.31), can generate spurious correlations and some type of localization is introduced to overcome this [2, 22]. This effect is exaggerated here due to the non-Gaussian features of the conceptual model. One strategy for localization discussed in Chapter 7 from [22] is to ignore all ill-conditioned cross correlation. In the conceptual dynamical model for turbulence discussed in Section 2, the reference simulation shows that the cross correlations between different modes are at most 5%. Thus it is useful in this test regime to ignore the cross correlations between different modes in (3.30) so that the forecast model has the following form

$$
\begin{aligned}
\frac{d\langle\overline{u}\rangle}{dt} &= -\overline{d}\langle\overline{u}\rangle + \gamma\sum_{k=1}^{K}(\langle u_k'\rangle)^2 + \gamma\sum_{k=1}^{K}R_{kk} - \overline{\alpha}\left(\langle\overline{u}\rangle^3 + 3\langle\overline{u}\rangle R_{00}\right) + F_0 \\
\frac{d\langle u_k'\rangle}{dt} &= -d_k\langle u_k'\rangle - \gamma\langle\overline{u}\rangle\langle u_k'\rangle, \quad 1 \le k \le K \\
\frac{dR_{00}}{dt} &= -2(\overline{d} + 3\overline{\alpha}\langle\overline{u}\rangle^2)R_{00} - 6\overline{\alpha}R_{00}^2 \\
\frac{dR_{kk}}{dt} &= -2(d_k + \gamma\langle\overline{u}\rangle)R_{kk} + \sigma_k^2, \quad 1 \le k \le K.
\end{aligned}
\tag{3.32}
$$

We call this new single scale filtering method Diagonal Gaussian Closure (DGC) filter which replaces (3.30) of the GC filter by (3.32). Note that the multi-scale ensemble filter EF-SP can be seen as a finite ensemble approximation of the DGC filter. In Section 4 it is numerically verified that the ensemble filter shows comparable results to DGC filter.

GC filter At the m-th assimilation step,

- Evolve the system using (3.30) which provides the statistics of both the large and small scales
- Update the mean and variance using (3.31) which is the standard Kalman formula applied to the full state vector

DGC filter

Similar to GC filter except the system is evolved by (3.32) instead of (3.30).

3.4. Reduced mode filters. In filtering of turbulent signals, we are often interested in estimating the most energetic modes which correspond to the modes with large variances. For example, in the test case of the conceptual model discussed in Section 2, $\overline{u}$ and u_1' contain the most energies accounting for more than 75% of the total variance. Thus for problems with a steep energy spectrum, a more efficient filtering method might be achieved by actively filtering only the most energetic modes without filtering all the small scale modes. In [16] this type of reduced mode filter is introduced for theoretical purposes in a different context and it shows that observability is improved which affects the filter performance in addition to the apparent computational advantage.

In the assimilation step, the reduced mode filter updates only the most energetic small scale modes in addition to the large scale modes, while it trusts the dynamics for the other small scale modes. The reduced mode multi-scale filter is much less expensive compared to the full mode multi-scale filters especially when the dimension of the small scale, N_2, is large because it will filter only the most energetic modes out of N_2 modes. In Section 4.2, it will be shown that the reduced PF-SP and EF-SP have comparable results to the full mode multi-scale filters. Also, the reduced ensemble filter shows more stable filter performance than the full mode ensemble filter.

3.5. List of methods. As a summary of this section, we list the multi-scale data assimilation methods discussed so far in Table 2. For multi-scale data assimilation, there are two different methods 1) the multi-scale particle filter (PF) in Section 3.1 and 2) multi-scale ensemble filter (EF) in Section 3.2. These two different methods can employ the general Superparameterization which keeps the non-zero mean of $\mathbf{u}_k'$ from the analysis (SP) or the simplified Superparameterization which ignores the non-zero mean of $\mathbf{u}_k'$ by setting them to zero (SP'). Thus there are four main methods PF-SP, PF-SP', EF-SP, and EF-SP'. In addition to these methods, a simplified version of EP-SP' by considering a collective small scale variable rather than keeping all different modes (projected parameterization) was also considered. As single-scale filters to compare with the multi-scale data assimilation methods, we also considered Gaussian closure (GC) and its simplified version diagonal Gaussian closure (DGC) methods. Note that the reduced mode filtering strategy is applicable to all the methods.

Multiscale filtering	Particle filter (PF)	Ensemble filter (EF)
Superparameterization (SP)	PF-SP	EF-SP
SP with zero $\mathbf{u}'$ (SP')	PF-SP'	EF-SP'
Projected parameterization (PP)	N/A	EF-PP

TABLE 2

List of multi-scale data assimilation methods. Particle or ensemble filters for the analysis update and SP, SP', and PP for the forecast model.

4. Data assimilation results on the test model. In this section we test the multi-scale filtering methods, PF-SP, PF-SP', EF-SP, and EF-SP' along with the Gausisan closure filters discussed in Section 3. As a test bed, the conceptual model with negative large scale damping and -5/3 spectrum of Section 2 with $K = 5$ is used for all numerical tests (see Table 1 and Figure 1 for its stationary state information; mean, variance, and PDFs). In our numerical tests below, we use linear observations

$$v = \overline{u} + \sum_k u_k' + \sigma_\theta \tag{4.1}$$

where the observation noise variance r_θ is 5% of the total variance. The true signal is generated by a realization of (2.1) by solving with a fine time step $\delta = 5 \times 10^{-3}$ using Euler-Maruyama method.

The ensemble filter is not restricted to any specific ensemble method and we use the ensemble adjust Kalman filter (EAKF) by Anderson [2] with covariance inflation on the large scales. Empirically obtained 10% of covariance inflation value is used for all ensemble filters - EF-SP, EF-SP', and EF-PP while particle filters need only marginal additional noise level, that is, only 1% additional noise with $\beta = 0.01$ in (3.13), along with residual resampling [27]. Although this additional noise level for particle filters is small, it is essential because the particle filters have poor filter performance with no additional noise $\beta = 0$ for the reasons discussed earlier in the paragraph above (3.13). In all tests, both particle filters and ensemble filters use 100 particles and 100 ensemble members to compare performance with a practical moderate ensemble size.

To measure the filter performance, we use the root mean square error (RMSE) and pattern correlation (PC) between the posterior mean state $\mathbf{x^a} = (x^a_{m_0}, x^a_{m_0+1}, ..., x^a_M)$ and the true signal $\mathbf{x} = (x_{m_0+1}, x_{m_0+2}, ..., x_M)$, where $x^a_m = x^a(t_m)$ and $x_m = x(t_m)$, by ignoring the first 400 assimilation steps out of 2000 assimilation steps, that is, $m_0 = 400$ and $M = 1000$. The formulas for RMSE and PC are defined by

$$\text{RMSE} = \sqrt{\frac{1}{M - m_0} \sum_{m=m_0+1}^{M} (x^a_m - x_m)^2} \tag{4.2}$$

and

$$\text{PC} = \frac{\sum_{m=m_0+1}^{M} (x^a_m - \langle x^a \rangle)(x_m - \langle x \rangle)}{\sqrt{\sum_{m=m_0+1}^{M} (x^a_m - \langle x^a \rangle)^2 \sum_{m=m_0+1}^{M} (x_m - \langle x \rangle)^2}} \tag{4.3}$$

respectively where $\langle x^a \rangle = \dfrac{1}{M - m_0} \displaystyle\sum_{m=m_0+1}^{M} x^a_m$ and $\langle x \rangle = \dfrac{1}{M - m_0} \displaystyle\sum_{m=m_0+1}^{M} x_m$ are time averages of the posterior and true means.

After discussing filter performance in Section 4.1 using the full mode filtering which was discussed in Section 3, we will also show the filter performance of the reduced filter of Section 3.4 where only the large scale, $\overline{u}$, and the first small scale mode, u_1' , are actively filtered.

4.1. Filter performance using all modes. The RMS errors and pattern correlations in estimating $\overline{u}$ and the first two small scale modes, u_1' and u_2', by PF-SP, EF-SP', EF-PP, and DGC are shown in Figure 2 as functions of observation times ranging from 1 to 8. The observation times from 1 to 4 are frequent observations which are shorter than the shortest correlation time of u_5' while the observation times from 5 to 8 are infrequent which are comparable to the correlation times of the intermediate small scale modes. The dashed line is the climatological error which is the square root of the stationary state variance of the true model. The climatological error is the expected error when we estimate the variable by the stationary state mean.

For all observation times, PF-SP has the best filter performance while EF-SP' and DGC are the second best comparable methods in the estimation of $\overline{u}$ and u_1'. In the estimation of u_2', no method shows any significant filter performance better than the other methods. Note that EF-SP' has catastrophic filter divergence at the observation 6 which can occur for ensemble filters (see [22, 10] for the study of catastrophic filter divergence for ensemble filters). On the other hand, EF-PP exhibits poor performance with this regime of filter parameters due to standard filter divergence with little variability in $\overline{u}$; see Figure 6 below. As illustrated earlier [17, 22], EF-PP does not have practical controllability with the present set of parameters. In Section 4.3 we improve the filter performance of EF-PP by increasing the inflation factor significantly which restores practical controllability.

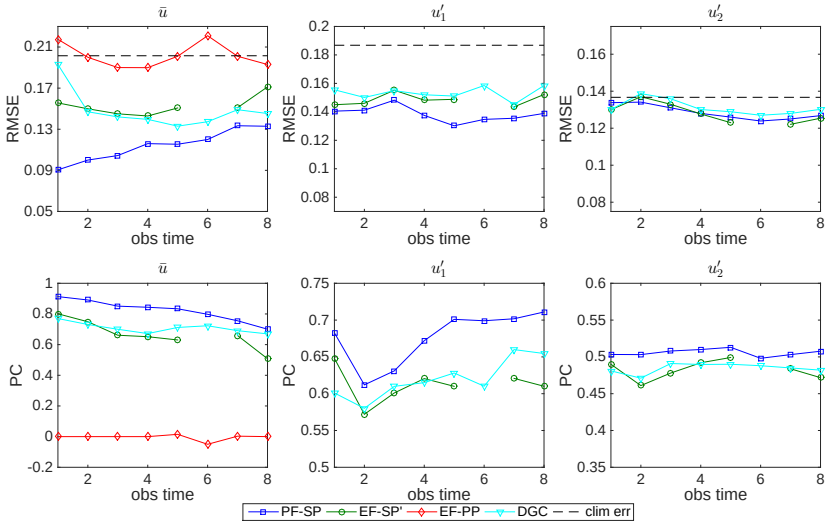

FIG. 2. *Full mode filters. RMS errors (top) and pattern correlation (bottom) in the estimation of* $\overline{u}, u_1', u_2'$ *as functions of observation time.*

Among the other methods which are not shown here, PF-SP' has results comparable to EF-SP' measured by RMS errors and PCs. EF-SP which captures the nonzero mean of the small scales has filter performance comparable to PF-SP which also captures the nonzero mean for short observation times but the performance degrades for longer observation times. The last method whose results are not shown here, Gaussian closure, has the worst filter performance of all methods with RMS errors comparable to or larger than the climatological error. Thus the updated posterior mean of GC is worse than just using the stationary mean of the estimate. The

poor performance of GC compared to DGC can be explained by the falsely amplified cross correlations between different modes [17, 22]. In the conceptual model, different small scale modes have small correlations of less than 5% and thus can be regarded as having no correlation.

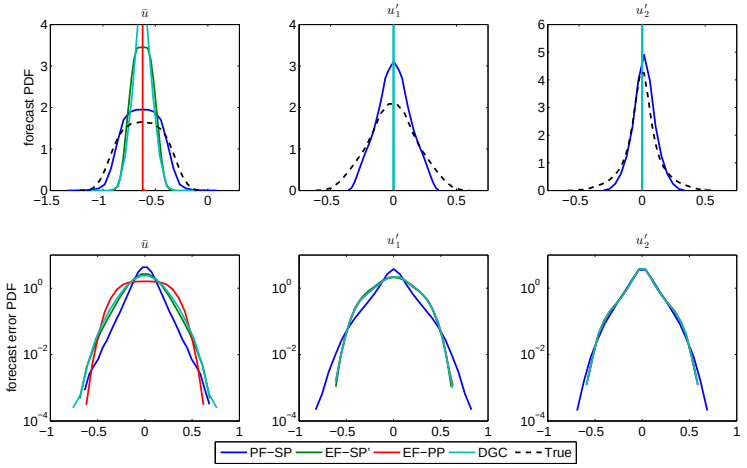

FIG. 3. *Full mode filters. Forecast (top) and forecast error (bottom) PDFs in the estimation of* $\overline{u}, u_1'$, *and* u_2'. *Observation time is 2.*

The forecast, x^f, PDF of a filtering method and the forecast error, $x^f - x$, PDF are more refined measures of filter performance beyond the RMS errors and PC [25]. The forecast PDF measures the capability of the filter to capture non-Gaussian features while methods with tighter forecast error PDFs have higher skill. Figure 3 shows the forecast and forecast error PDFs by PF-SP, EF-SP', EF-PP, and DGC for the observation time 2 in the estimation of $\overline{u}, u_1'$, and u_2'. In the estimation of $\overline{u}$, PF-SP has the best fit to the true signal PDF while EF-SP' and DGC have sharper peaks and narrow tails. Also PF-SP has the tightest forecast error bound. In the estimation of u_1' and u_2', EF-SP' always has zero mean and thus it does not provide any meaningful prior estimate. DGC keeps the nonzero mean of the small scales but the values are on the order of 10^{-4} which can be regarded as zero here. PF-SP, which keeps the nonzero mean for the small scales, does not have perfect fit but it has nontrivial forecast error PDFs with a better fit in the estimation of u_2' than u_1'.

4.2. Reduced filters. In the test problem, the large scale $\overline{u}$ and the first small scale mode u_1' contain the most energies and the full mode filter does not show any apparent filter skill for the smallest scales. Thus, we may expect to have comparable filtering skill (or better skill due to improved observability) using the reduced filters of Section 3.4 which actively filter only $\overline{u}$ and u_1'.

The filter performance of the reduced filters measured by RMS errors and PCs are comparable to the filter skill of the full mode. In Figure 4, RMS errors and PCs by the reduced filters are shown as functions of observation times. PF-SP is the best robust method for all observation times while EF-SP' and DGC are the second best comparable methods. The reduced filter does not improve the performance of EF-PP; EF-PP still has filter divergence for most observation times with no variation in the large scale. However, the reduced EF-SP' does not have filter divergence at the

observation time 6 which can be seen with the full mode filter.

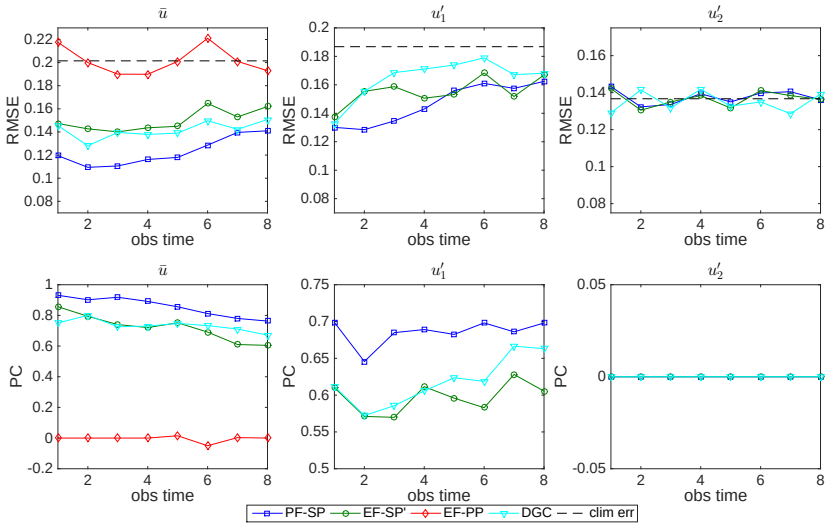

FIG. 4. *Reduced filters. RMS errors (top) and pattern correlation (bottom) in the estimation of* $\overline{u}, u_1'$, *and* u_2' *as functions of observation time.*

In the estimation of u_1', the reduced PF-SP has improved filter performance for short observation times compared to the full mode filter but in the estimation of u_2' no reduced filters show any significant filter skill; the RMS errors are on the order of the climatological error and pattern correlations are almost 0%. Although the full-mode PF-SP has nontrivial forecast PDFs of u_2' by keeping the nonzero mean in Superparameterization, the reduced mode PF-SP has zero forecast means for u_2' because without the update of the man by nonzero values from the posterior estimate, the quasilinear system of u_2' mode has eventually zero mean. Thus, with the reduced filters, no methods using superparametrization give nontrivial estimates for the modes which are not actively filtered.

Regarding the forecast and forecast error PDFs as a more demanding measure of filter performance, the forecast PDFs of $\overline{u}$ by all reduced filters are improved. Especially, the reduced PF-SP has a perfect match to the true signal PDF (see Figure 5 for the forecast and forecast error PDFs by the reduced filters for the observation time 2). But the forecast error PDFs of $\overline{u}$ are slightly wider than the full mode results.

4.3. Improving EF-PP with variance inflation. In [13], EF-PP is applied to a one dimensional tough test problem with a -5/6 shallow spectrum where more than two thirds of the total variance is carried by the small scale part and shows excellent filtering skill in the estimation of the large scales. But in the two previous tests applied to the conceptual dynamical model for turbulence, EF-PP has filter divergence with no variability or little variability in $\overline{u}$. Figure 6 shows segments of time series of the prior, posterior and true signals of $\overline{u}$ by the full mode PF-SP, EF-SP', EF-PP, and DGC for the observation time 7. As we can deduce from the filter performance, PF-SP has a good match with the true signal for both the prior and posterior means while EF-SP' and DGC have slightly less variable means. EF-PP, which has the worst performance in the estimation of $\overline{u}$, has filter divergence with no variability in $\overline{u}$.

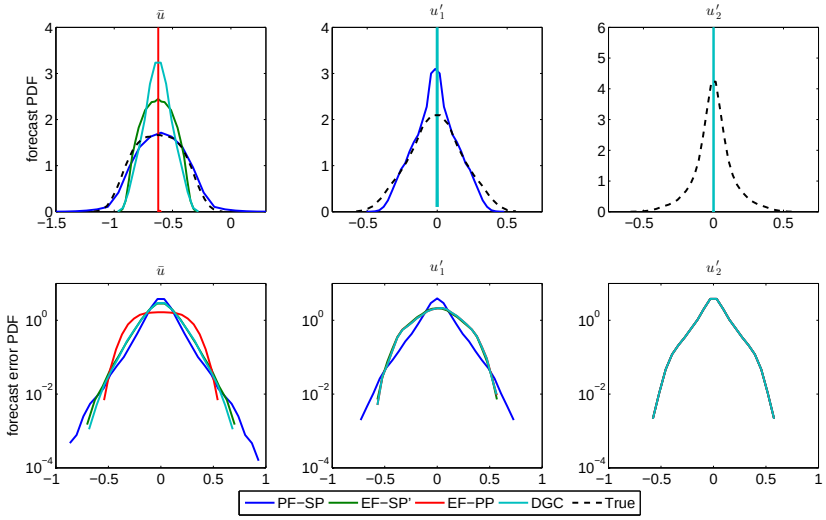

FIG. 5. *Reduced filters. Forecast (top) and forecast error (bottom) PDFs in the estimation of* $\overline{u}, u_1'$, *and* u_2'. *Observation time is 2.*

The difference between the two test models - the test problem from [13] and the conceptual dynamical model in this paper - is that the first test model has significant time scale separation where the correlation time of the large scale is 100 times longer than the correlation times of the small scales but with very energetic small scales due to the shallow energy spectrum while the conceptual dynamical model has moderate scale separation with gradually changing correlation times of different modes and a decaying energy spectrum. EF-PP uses the stationary state variance for the initial value of the small scale parts in superparametrization at each assimilation step. Thus compared to the filtering methods whose initial values for the variances in Superparameterization dynamically change, EF-PP may require a longer time to get enough small scale instability. If there is significant time scale separation between the large and small scales, a moderate observation time comparable to the large scale correlation time is long enough compared to the correlation times of the small scales and thus EF-PP has enough variability for $\overline{u}$. If there is mild scale separation, on the other hand, EF-PP is not expected to be an accurate filter for short observation times.

To overcome the filter divergence of EF-PP in the estimation of $\overline{u}$ and restore practical controllability [17, 22], we increase the inflation factor of EAKF. In fact, the 10% inflation factor was obtained empirically to provides the best results of EF-SP and EF-SP'. Figure 7 shows the RMS errors and pattern correlations with various values of inflation factor ranging from 0% to 50%. By inflating the variance of the large scale with a larger factor, the performance of EF-PP increases for large observation times where the small scale modes can effectively decorrelate. See Chapter 7 of [22] for more intuition and examples with such reduced filters.

5. Conclusions. In this study we study multi-scale data assimilation methods with Superparameterization as a cheap and fast imperfect forecast model to mitigate the curse of dimensionality of turbulent signals from nature. The multi-scale method incorporates observations involving contributions from both the large and small scales of the true signal which is general in many real applications. The key idea of the

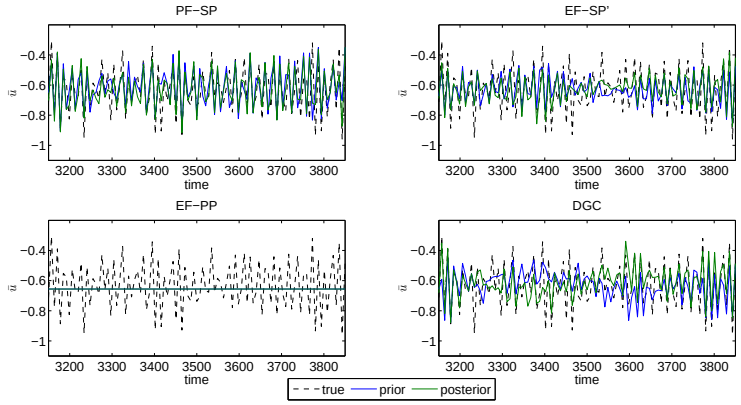

FIG. 6. *Time series the prior, posterior, and true mean of* $\overline{u}$ *by the full mode PF-SP, EF-SP', EF-PP with observation time 7.*

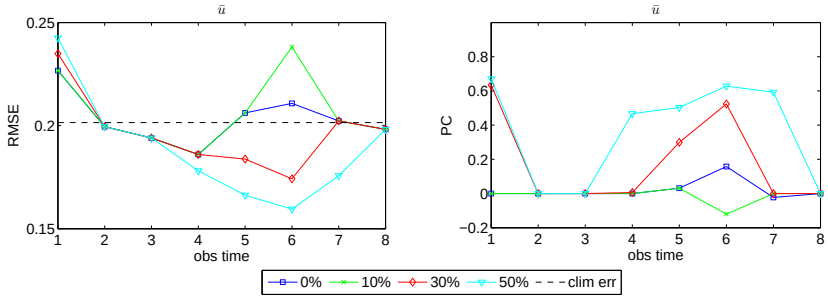

FIG. 7. *RMS errors and pattern correlations by EF-PP using various large-scale inflation factors as functions of observation times.*

method is conditional Gaussian mixtures where the fluctuating parts are approximated by Gaussian distributions conditional to the large scale variables which makes the method efficient by filtering through smaller subspaces instead of the large dimensional full state vector. We emphasize that the multi-scale data assimilation strategy developed here is a general framework and any other imperfect forecast models which provides the statistics of the small scales can be used instead of Superparameterization (see [25] for example).

As a test model for the multi-scale data assimilation methods, a simple conceptual dynamical model for turbulence is used which despite of its simplicity captures the interesting features of anisotropic turbulence. The six dimensional test problem which is used throughout this paper has a large scale mean flow with a sub-Gaussian PDF while the smaller parts of the fluctuations have fat tailed PDFs indicating extreme instabilities. Also, this test problem has mild time scale separation where the correlation time gradually increases for larger parts of the signal. The test problem utilized here was only a typical example; the conceptual dynamical model can be utilized in other parameter regimes as a test model to capture other phenomena in turbulence through filtering.

Various multi-scale filtering methods including particle filters (PF-SP and PF-SP') and ensemble filters (EF-SP, EF-SP', and EF-PP) are tested in addition to

the single scale GC and DGC filters. Among various filtering methods discussed here, PF-SP shows the best robust filter skill for all observation times tested here while EF-SP' and DGC show the second best comparable results; EF-SP' imposes a Gaussian assumption on the large scales and thus EF-SP' can be regarded as a finite number ensemble approximation of DGC. In addition to the full mode multi-scale filters, reduced mode filters are also discussed which are more efficient and stable than the full mode filters yet have comparable filter skill. We also see that EF-PP can have filter divergence due to violation of practical controllability with little variability in the large scales; this filter divergence can be avoided by using a covariance inflation factor larger than the one of EF-SP and EF-SP'.

The multi-scale data assimilation methods developed here can be used with more realistic problems, for example, the quasi-geostrophic (QG) two-layer equations [22, 31]. For the QG equation, there are various imperfect coarse models including stochastic Superparameterization [15, 20] and thus it would be interesting to investigate the effect of various imperfect models on the multi-scale filter performance.

REFERENCES

[1] D.L. Alspach, V.S. Samant, and H.W. Sorenson, *Practical control algorithms for nonlinear stochastic systems and investigations of nonlinear filters*, technical report, DTIC Document, 1980.

[2] J. Anderson, *An ensemble adjustment Kalman filter for data assimilation*, Mon. Weather Rev., 129 (2001), pp. 2884–2903.

[3] A. Bain and D. Crisan, Fundamentals of Stochastic Filtering, Stochastic Modeling and Applied Probability, Springer, New York, 2009.

[4] C. Chui and G. Chen, Kalman Filtering: With Real-Time Applications, Springer, 1999.

[5] S. Cohn, *An introduction to estimation theory*, J. Meteorol. Soc. Jpn., 75 (1997), pp. 257–288.

[6] R. Daley, *Atmospheric Data Analysis*, Cambridge University Press, New York, 1991.

[7] A. Doucet A, S. Godsill, and C. Andrieu, *On sequential Monte Carlo sampling methods for Bayesian filtering*, Stat Comput, 10(3) (2000), pp. 197–208.

[8] E.S. Epstein, *Stochastic dynamic prediction*, Tellus, 21(6) (1969), pp. 739–759.

[9] U. Frisch U, Turbulence, Cambridge Univ Press, New York, 1995.

[10] G.A. Gottwald and A.J. Majda, *A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks*, Nonlinear Processes in Geophysics, 20 (2013), pp. 705–712.

[11] W. Grabowski and P. Smolarkiewicz, *CRCP: a cloud resolving convection parameterization for modeling the tropical convecting atmosphere*, Physica D, 133 (1999), pp. 171–178.

[12] W. Grabowski, *An improved framework for Superparameterization*, J. Atmos. Sci., 61 (2004), pp. 1940–1952.

[13] I. Grooms, Y. Lee and A.J. Majda, *Ensemble Kalman filters for dynamical systems with unresolved turbulence*, J. Comp. Phys., 273(15) (2014), pp. 435–452.

[14] I. Grooms and A.J. Majda, *Stochastic Superparameterization in a one-dimensional model for wave turbulence*, Commun. Math. Sci., 12 (2014). pp. 509–525.

[15] I. Grooms and A.J. Majda, *Stochastic Superparameterization in quasigeostrophic turbulence*, Journal of Computational Physics, 271 (2014), pp. 78–98.

[16] M.J. Grote and A.J. Majda, *Stable time filtering of strongly unstable spatially extended systems*, Proce. Nat. Acad. Sci., 103 (2006), pp. 7548–7553.

[17] J. Harlim and A.J. Majda, *Test models for filtering with Superparameterization*, SIAM J. Multiscale Model. Sim., 11(1) (2013), pp. 282–308.

[18] I. Hoteit, X. Luo, and D.T. Pham, *Particle Kalman filtering: A nonlinear Bayesian framework for ensemble Kalman filters*, Mon Weather Rev, 140 (2012), pp. 528–542.

[19] A. Lorenc, *Analysis-methods for numerical weather prediction*, Q. J. R. Meteorol. Soc., 112 (1986), pp. 1177–1194.

[20] A.J. Majda and I. Grooms, *New perspectives on Superparameterization for geophysical turbulence*, J. Comp. Phys., 271 (2013), pp. 60–77.

[21] A.J. Majda, M.J. Grote, *Mathematical test models for superparametrization in anisotropic turbulence*, Proc. Natl. Acad. Sci. USA, 106 (2009), pp. 5470–5474.

[22] A.J. Majda and J. Harlim, Filtering Complex Turbulent Systems, Cambridge Univ Press, New York, 2012.
[23] A.J. Majda, R. Kleeman, and D. Cai, *A mathematical framework for quantifying predictability through relative entropy*, Methods and Applications of Analysis, 9(3) (2002), pp. 425–444.
[24] A.J. Majda and Y. Lee, *Conceptual dynamical models for turbulence*, Proc. Natl. Acad. Sci. USA, 111(18) (2014), pp. 6548–6553.
[25] A.J. Majda, Di Qi, and T.P. Sapsis, *Blended Particle Filters for Large Dimensional Chaotic Dynamical Systems*, Proc. Natl. Acad. Sci. USA, 111(21) (2014), pp. 7511–7516.
[26] M.D. Millionshchikov, *On the theory of homogeneous isotropic turbulence*, Dokl. Akad. Nauk SSSR, 32 (1941), pp. 615-618.
[27] P.D. Moral and J. Jacod, *Interacting particle filtering with discrete observations*. Sequential Monte Carlo Methods in Practice, eds Doucet A, de Freitas N, Gordon N, Springer, New York, 2001, pp. 43–75.
[28] T.P. Sapsis, and A.J. Majda, *A statistically accurate modified quasi-linear gaussian closure for uncertainty quantification in turbulent dynamical systems*, Physica D: Nonlinear Phenomena, 252(1) (2013), pp. 34–45.
[29] H.W. Sorenson and D.L. Alspach, *Recursive Bayesian estimation using Gaussian sums*, Automatica 7 (1971), pp. 465–479.
[30] A.A. Townsend, The Structure of Turbulent Shear Flow, Cambridge Univ Press, Cambridge, MA, 1976.
[31] G.K. Vallis, Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large- Scale Circulation, Cambridge Univ Press, Cambridge, MA, 2006.