Multiscale Data Assimilation and Prediction using Clustered Particle Filters

Yoonsang Lee[a], Andrew J. Majda[a]

[a] *Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA*

Abstract

Multiscale data assimilation uses a coarse-resolution forecast model to increase the number of samples in the estimation of large-scale and long time behavior of high-dimensional complex systems along with noisy incomplete observations. A new class of multiscale particle filters, the multiscale clustered particle filter, is developed here as an effective multiscale data assimilation method for capturing non-Gaussian distributions and extreme events of high-dimensional turbulent systems using relatively few particles. The multiscale clustered particle combines the single-scale clustered particle filter with a general multiscale data assimilation framework that can handle mixed observations of both the resolved and unresolved scale components. To test the multiscale data assimilation method, we use a two-layer Lorenz system having 440 modes with important features of turbulent systems such as non-Gaussian statistics including fat-tails and intermittent extreme events. The effect of the observation model error is investigated and it is shown that the multiscale clustered particle filter captures non-Gaussian distributions using a small number of samples while an ensemble-based method fails to capture the correct distribution.

Keywords: data assimilation, filtering, Monte-Carlo, multiscale, clustered particle filter, non-Gaussian

1. Introduction

Data assimilation or filtering of turbulent systems is an important problem in many contemporary applications in science and engineering including real-time prediction of weather and climate as well as the spread of hazardous plumes of pollutants [1]. Data assimilation provides the best statistical estimate of the true signal by combining a numerical forecast model and noisy partial observations of the true signal. Although data assimilation is a well-developed discipline for low-dimensional dynamical systems [2], its application to turbulent systems is challenging due to the characteristics of turbulent systems. Turbulent systems are well-known for a high-dimensional phase space and a large dimensional space of instability with positive Lyapunov exponents [3]. Also turbulent systems show extreme events and non-Gaussian features such as skewed or fat-tailed distributions [4, 5] as observed in nature [6, 7].

Email addresses: ylee@cims.nyu.edu (Yoonsang Lee), jonjon@cims.nyu.edu (Andrew J. Majda)

Turbulent systems have a wide range of spatiotemporal scales in a high-dimensional space and thus resolving all the active scales in a high-dimensional space is computationally prohibitive. Especially for ensemble-based data assimilation methods [8, 9], it is important to use a sufficient number of ensemble to approximate the probability distribution of the system. However, due to the high computational costs to run a forecast model resolving all the active scales of the system, the practical ensemble number is limited and insufficient due to the high computational costs to run each forecast model, which is called "curse of dimensionality" [10] or "curse of small ensemble size" [1]. Therefore, it is indispensable to use low-resolution or coarse-resolution forecast models in data assimilation of turbulent systems to alleviate the curse of small ensemble size. In [11], a cheap and robust coarse-resolution forecast model called stochastic superparameterization [12], which is 200 times cheaper than the full-resolution forecast model, has been successfully applied for a two-layer quasigeostrophic baroclinic turbulent flows with inhomogeneous statistics and zonal jets.

Another important issue in data assimilation of high-dimensional systems is catastrophic filter divergence [13, 14], which drives the filter forecast to machine infinity although the system remains in a bounded set (see [15] for a rigorous mathematical analysis of catastrophic filter divergence). The catastrophic filter divergence can occur when observations are sparse, infrequent and of high-quality, which are typical in many geophysical systems due to the vast area of the geophysical systems and expensive costs to increase the number of observation points. In a recent study [16], it is shown that the coarse-resolution forecast model, stochastic superparameterization, plays an important role in preventing catastrophic filter divergence.

In the use of coarse-resolution forecast models for data assimilation of high-dimensional systems, the imperfect coarse-resolution models lead to several model errors. The first error is the forecast model error related to the numerical truncation error in modeling the resolved large-scale dynamics and the error from unresolved sub-grid scale interactions (see [17] for a study of the information barrier from the sub-grid scales). The error due to imperfect models and insufficient ensemble size often yields underestimation of the uncertainty in the forecast and thus the filter puts more confidence on the forecast than the information given by observations, which is the standard filter divergence. Covariance inflation [18, 19], which adds uncertainty in the forecast by inflating the prior covariance, and localization [20], which calibrates the overestimated correlations between observed and unobserved variables, are essential tools to remedy the filter divergence. In a recent study [11], the effect of covariance inflation and stochastic parameterization of the unresolved scales are investigated to remedy the standard filter divergence and imperfect model errors.

The incorporation of a coarse-resolution forecast model for data assimilation of high-dimensional systems has another model error, an observation model error. The coarse-resolution forecast model provides predictions for only the resolved coarse scales. However, the observation has mixed contributions from both the resolved and unresolved scales and thus there is an observation model error related to the contribution of the unresolved or sub-grid scales to the observation. This error has been known as "representation error" or "representative error" in the data assimilation community and several approaches have been developed to analyze the representation error [21].

The general multiscale data assimilation framework in [22] addresses the issues related to the use of coarse-resolution forecast models for data assimilation of high-dimensional systems. The multiscale data assimilation framework provides the best statistical estimate of the resolved coarse-scale dynamics using coarse-resolution forecast models and mixed contributions from both the resolved and unresolved scales. The general framework uses particle filtering for the low-dimensional resolved scales while the unresolved scales are filtered using the standard Kalman filter formula and thus it is also called multiscale particle filter (see [23] for multiscale data assimilation using the modified quasi-Gaussian closure model as a forecast model). From the general multiscale data assimilation framework, a simpler version of multiscale data assimilation method, an ensemble multiscale data assimilation method [24] , can be derived under the Gaussian assumption for the forecast and linear observations. The ensemble multiscale data assimilation method treats the contribution of the unresolved scales to the observations as representation errors. The ensemble method has been successfully applied for several difficult problems including one-dimensional wave turbulence with breaking solitons and shallow energy spectrum [24] and turbulence tracers advected by baroclinic turbulent flows with inhomogeneous meridional structures [25]. Another data assimilation method incorporating a coarse-resolution forecast model has been studied and investigated in [26]. However, the observations in [26] depend only on the resolved coarse scales while the general multiscale data assimilation framework can handle mixed contributions from both the resolved and unresolved scales.

Despite the successful application of the multiscale particle filter [22] for the conceptual dynamical models for turbulence [27], which has energy-conserving nonlinear interactions and mimics the interesting features of turbulent flows including non-Gaussian statistics and extreme events, the application of the multiscale particle filter is limited to low-dimensional resolved spaces. The problem is not from the multiscale data assimilation algorithm but from the well-known inapplicability of the standard particle filter for high-dimensional systems (in [28, 10], it is shown that the number of particles increases exponentially with the dimension of the system). The multiscale ensemble data assimilation method is a good workaround with successful results for several difficult test problems mentioned above. However, the method has a difficulty in capturing non-Gaussian features, which are typical in turbulent systems [6, 7], using relatively few samples as it assumes Gaussian prior and observation error statistics.

Recently a new class of particle filter, the clustered particle filter (CPF), has been developed, which can be applied for high-dimensional systems effectively [29]. CPF captures the non-Gaussian features of high-dimensional systems using relatively few particles compared with the standard particle filter and is robust for sparse and high-quality observations. The key features of CPF are coarse-grained localization through clustering of state variables depending on the observation network and particle adjustment that translates forecast particles to prevent particle collapse. In this paper, we combine the multiscale particle filter with CPF (which we call multiscale clustered particle filter

(MsCPF)) to apply the multiscale data assimilation framework for high-dimensional resolved spaces.

A preliminary result of the multiscale clustered particle filter applied for an one-dimensional wave turbulence model with Gaussian large-scale statistics is reported in [29]. To investigate several aspects of the multiscale data assimilation algorithm, including the effect of the observation model error (or representation error), we introduce an advective two-layer Lorenz-96 model as a test model, which contains both large- and small-scale advection to small-scale components. This model is a prototype model for slow-fast systems, which is typical, for example, in atmosphere where a slow advective vortical Rossby wave is coupled with fast inertia-gravity waves [30, 31]. The model has non-Gaussian statistics and extreme events represented by fat-tails and thus serves as a good test model for the multiscale data assimilation method.

The structure of this paper is as follows. In section 2, we briefly review the standard and clustered particle filters and describe the main algorithm, the multiscale clustered particle filter. In section 3, we propose a new test model with two different scales, advective two-layer Lorenz-96 model and discuss test regimes with non-Gaussian statistics and instability and provides linear stability analysis of the model as a guideline. In section 4, we show the data assimilation prediction experiments with a superior performance of MsCPF in capturing non-Gaussian statistics of the true signal, followed by discussions and conclusions in section 5.

2. Multiscale Clustered Particle Filter

In this section, we explain a mathematical setup and introduce notation to describe the main algorithm, the multiscale clustered particle filter. After introducing the basic setup, we briefly review the standard particle filter [2] and the clustered particle filter [29], which are important to derive and understand the multiscale clustered particle filter algorithm.

Throughout this paper, we consider the data assimilation of the true signal $\mathbf{u} \in \mathbb{R}^N$ at a discrete time (or observation time) $n\Delta T, n \in \mathbb{N}$, where ΔT is the observation interval, whose dynamics is given by a nonlinear map ψ

$$\mathbf{u}^{n+1} = \psi(\mathbf{u}^n). \tag{1}$$

As we are concerned with high-dimensional systems with turbulent behavior, the dimension of the system, N, is assumed to be large $N \gg 1$, and ψ has chaotic characteristics such as a large dimensional space of instability with positive Lyapunov exponents. As the system is difficult to estimate and predict due to the chaotic behavior, we use observations $\mathbf{v} = \{v_1, ..., v_{N_o}\} \in \mathbb{R}^{N_o}$, $N_o \leq N$, which are available at each observation time. We assume that the observation operator, $\mathbf{H} : \mathbb{R}^N \to \mathbb{R}^{N_o}$ is local, that is, each observation variable y_j, depends on only the corresponding state variable at the same location

$$\mathbf{v} = \mathbf{H}(\mathbf{u}) + \xi = (h(x_{i_1} + \xi_1, h(x_{i_2}) + \xi_2, ..., h(x_{i_{N_o}}) + \xi_{N_o}) \tag{2}$$

where ξ_j is I.I.D. Gaussian with mean zero and variance r_o. In real applications, a full recovery of the true state from observations is impossible due to incomplete observations; the observations are noisy and sparse, i.e., the number of observation N_o is smaller than the dimension of the full state N for high-dimensional systems $N \gg 1$, along with the nonlinear dependence of the observation on the true signal. Thus the goal of data assimilation is to provide the best statistical estimate combining the forecast PDF from a numerical prediction model and incomplete partial observations.

The standard particle filter [2] is a well-developed discipline for filtering low-dimensional non-Gaussian systems using different weights for different samples (or particles) to effectively represent the PDF of the system. Using K particles and scalar particle weights $\{w_k \geq 0, k = 1, 2, ..., K\}$, the standard particle filter approximates a probability density using the following form of PDF

$$p(\mathbf{u}) = \sum_{k}^{K} w_k \delta(\mathbf{u} - \mathbf{u}_k), \tag{3}$$

where δ is the Dirac delta function. In comparison with the standard Monte-Carlo or ensemble-based method, which uses the same weight $\frac{1}{K}$ for each sample, the standard particle filter can represent non-Gaussian distributions more efficiently using non-constant particle weights for each sample. The standard particle filer shows robust performance in many applications in science and engineering [2]. However, its applications are limited to low-dimensional systems as the number of particles increases exponentially with the dimension of the system [28, 10]; in the application of the standard particle filter for high-dimensional systems, the standard particle filter suffers from particle collapse where only a small fraction of particles have the most weights while the rest of the particles have nearly zero weights.

2.1. Clustered particle filter

There are several attempts to overcome the limitation of the standard particle filter in the application for high-dimensional systems including the method that solves an optimal transport problem for the transition before the posteior to avoid the random sampling aspects of the standard particle filter [32], hybrid ensemble transform particle filter [33], and the localized particle filter [34]. Recently a new class of particle filter, clustered particle filter (CPF), has been proposed and it shows robust filtering performance with successful application for difficult test regimes, sparse and high-quality observation networks, in [29]. CPF also does not need ad-hoc tuning parameters.

Coarse-grained localization

The main features of the clustered particle filter are coarse-grained localization and particle adjustment, which enable the method to use relatively few particles to capture non-Gaussian statistics of high-dimensional systems even with sparse and infrequent observations. In the formulation of CPF, we assume that the observations are so sparse that each observation at different locations is uncorrelated with each other.

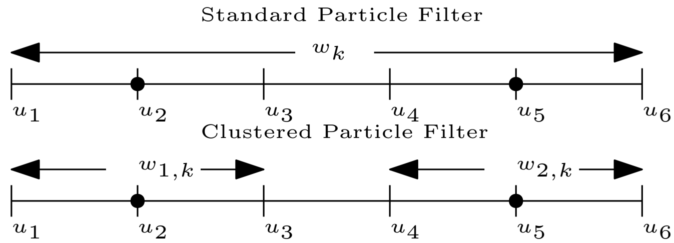


Figure 1: Schematics of particle weight for the k-th particle. Total dimension is 6 and there are two observations at u_2 and u_5, which yields two clusters in CPF. The standard particle filter uses the same particle weight at different locations whereas the clustered particle filter uses different weights in different clusters but the weights are the same in the same cluster.

Thus, if there are N_o observation points, CPF partitions the state vectors into N_o non-overlapping clustered $\{C_l, l = 1, 2, ..., N_o\}$ according to the observation location. Each cluster, C_l, is centered at the observation point and the cluster boundary is chosen as the middle point of the two adjacent observation locations, which can be applied to irregularly spaced observation networks. For the subspace state vector of each cluster, $\mathbf{u}_{C_l} = \{u_i | u_i \in C_l\}$ after clustering of the state variable, each cluster uses its own cluster particle weights $\{w_{l,k}\}$ to represent the marginalized probability distribution of each cluster (see Figure 1 which compares the schematics of the particle weights of the standard and the clustered particle filters for a 6 dimensional system with two observation points).

To use the particle adjustment step explained later in this section, CPF considers only the marginalized probability distribution of each cluster

$$p(\mathbf{u}_{C_l}) = \sum_{k}^{K} w_{l,k} \delta(\mathbf{u}_{C_l} - \mathbf{u}_{X_{C_l}}). \tag{4}$$

When we sequentially assimilate each observation v_j (which is possible as each observation error is spatially uncorrelated), the observation v_j affects the marginalized PDF of the corresponding cluster C_j while the other clusters remain unaffected. From the forecast particle weights $\{w^f_{l,k}\}$ for the cluster C_j, the posterior particle weights $\{w^a_{j,k}\}$ are given by

$$\omega^a_{l,k} = \begin{cases} \frac{\omega^f_{l,k} p(v_j|\mathbf{u}_k)}{\sum_m^K \omega^f_{l,m} p(v_j|\mathbf{u}_m)} & l = j, \\ \omega^f_{l,k} & l \neq j. \end{cases} \tag{5}$$

Therefore the clustering of the state variables plays the role of coarse-grained localization.

Particle adjustment

Another important key ingredient of the clustered particle filter is the particle adjustment step, which translates and shrink the forecast particles instead of reweighing when a special criterion related to the forecast statistics is satisfied. An important observation for the standard particle filter is that the posterior statistics by combining the forecast statistics and observations is given by reweighing the forecast samples, which is a convex combination of the forecast samples. This fact implies that if the posterior mean cannot be represented by a convex combination of the forecast samples, it is not possible to represent the accurate posterior statistics using only the reweighing of the forecast samples. This situation can happen when the observation is of high-quality, i.e., the observation error variance is small and thus the observation is close to the true value. In that case, it is straightforward to check whether the observation can be represented by a convex combination of the forecast samples. Otherwise, another method to represent the accurate posterior statistics is necessary.

The particle adjustment step of the hard threshold version clustered particle filter checks whether each observation v_j is in the convex hull of the forecast samples in the corresponding cluster C_j

$$v_j \in \{\sum_k^K q_k \mathbf{H}(\mathbf{u}^f_{C_j,k})|, \forall q_k \geq 0 \text{ such that } \sum_k q_k = 1\}. \tag{6}$$

If (6) is not satisfied, we trigger the particle adjustment step, which updates the forecast samples $\{\mathbf{u}^f_{C_l,k}\}$ through an adjustment matrix A (see the supporting information of [29] for a way to find the adjustment matrix A)

$$\mathbf{u}^a_{C_j,k} = \mathbf{u}^a_{C_j} + A(\mathbf{u}^f_{C_j,k} - \mathbf{u}^f_{C_j}). \tag{7}$$

to match the Kalman analysis mean $\mathbf{x}^a_{C_j}$ and covariance $R^a_{C_j}$ which are given as

$$\mathbf{u}^a_{C_j} = \mathbf{u}^f_{C_j} + G(y_j - H\mathbf{u}^f_{C_j}) \tag{8}$$

and

$$R^a_{C_j} = (I - GH)R^f_{C_j} \tag{9}$$

respectively, where $G = R^f H^T (HR^f H^T + r_o I)^{-1}$ is the Kalman gain matrix, $\mathbf{u}^f_{C_j} = \sum_k^K \omega_{j,k}\mathbf{u}^f_{C_j,k}$ is the forecast mean and $R^f_{C_j} = \sum_k^K \omega_{j,k}(\mathbf{u}^f_{C_j,k} - \mathbf{u}^f_{C_j})(\mathbf{u}^f_{C_j,k} - \mathbf{u}^f_{C_j})^T$ is the forecast covariance. In the particle adjustment step, the particle weights remain unchanged. There are other criteria to trigger particle adjustment than (6) (such as the soft threshold criterion in [29]). In our study, we use only the hard threshold criterion (6) as it shows robust results in our tests. Now we summarize the hard threshold clustered particle filter

Hard Threshold Clustered Particle Filter Algorithm - one step assimilation.

Given :

1) N_o observations $\{v_1, v_2, ..., v_{N_o}\}$

2) prior K particles $\{\mathbf{u}^f_{C_j,k}, k = 1, 2, ..., K\}$ and weight vectors $\{\omega^f_{l,k}, k = 1, 2, ..., K\}$ for each cluster $C_l, l = 1, 2, .., N_{obs}$

For v_j from $j = 1$ to N_o
If The hard threshold criterion (6) is satisfied
Update the prior particles using (7) to match the Kalman update (8) and (9)
Else Use particle filtering
Update $\{\omega^f_{j,k}\}$ using (5)
If $K_{eff} = \frac{1}{\sum_k (\omega^a_{l,k})^2} < \frac{K}{2}$
Do resampling
Add additional noise to the resampled particles

$$\mathbf{u}_{C_l,Resample(k)} \leftarrow \mathbf{u}_{C_l,Resample(k)} + \epsilon \tag{10}$$

where ϵ is IID Gaussian noise with zero mean and variance r_{noise}
End If
End If
End For

Note that there is a potential issue, dynamic imbalance of CPF through the coarse-grained localization [35, 36]. We emphasize that we consider sparse observations where each observation point is uncorrelated with each other (which is typical in geophysical systems due to the vast area of the system). Thus the effect of dynamic imbalance is marginal. In our tests in section 4, we do not find any issues related to dynamic imbalance.

2.2. Multiscale clustered particle filter

The basic idea of the multiscale clustered particle filter is to use the same coarse-grained localization and particle adjustment as in CPF. The only difference is that the particle weights in each cluster are updated using the multiscale particle filer method [22] in each cluster.

For the subspace state vector $\mathbf{u}_{C_l}$ corresponding to the cluster C_l, we assume that there is a decomposition of the full state vector into resolved large-scale component $\mathbf{x}_{C_l}$ and unresolved small-scale component $\mathbf{y}_{C_l}$. Using this decomposition into the resolved and unresolved scales, the marginalized PDF of $\mathbf{u}_{C_l}$ is represented by the following conditional Gaussian mixture distribution (compare (11) with (4))

$$p(\mathbf{u}_{C_l}) = \sum_k^K w_{l,k}\delta(\mathbf{x} - \mathbf{x}_l)\mathcal{N}(\mathbf{y}_l(\mathbf{x}_{l,k}), \mathbf{R}'(\mathbf{x}_{l,k})). \tag{11}$$

where each summand is a Gaussian distribution conditional to the resolved scale $\mathbf{x}_{C_l,k}$. Note that the interactions between the resolved and unresolved scales through the dependence of the unresolved scale PDFs on the resolved scale can make non-trivial behavior including non-Gaussian distributions.

When the observation $\mathbf{v}$ has the following form (which can be regarded as a Taylor expansion of general nonlinear observation operators around the resolved scale)

$$\mathbf{v} = \mathbf{H}(\mathbf{x}, \mathbf{y}) + \xi = \overline{\mathbf{H}}\mathbf{x} + \mathbf{H}'(\mathbf{x})\mathbf{y} + \xi, \tag{12}$$

where $\mathbf{H}'$ has rank N_o, the posterior marginalized distribution of $\mathbf{u}_{C_l}$ taking into account the observation v_j is in the same form as the forecast PDF (see Proposition 3.1 of [22]) and its analysis weight is given by

$$w^a_{l,k} = \begin{cases} \frac{w^f_{l,k} I_k}{\sum_k w^f_{l,k} I_k} & l = j, \\ w^f_{l,k} & l \neq j \end{cases} \tag{13}$$

where $I_k = \int p(v_j|\mathbf{x}_{C_l,k}, \mathbf{y}_{C_l}) p(\mathbf{y}_{C_l}|\mathbf{x}_k) d\mathbf{y}_{C_l}$.

To trigger particle adjustment for the multiscle clustered particle filter, we use the hard threshold version in the observation space

$$v_j \in \{\sum_k^K q_k \mathbf{H}(\mathbf{x}^f_{C_j,k}, \mathbf{y}^f_{C_j,k})|, \forall q_k \geq 0 \text{ such that } \sum_k q_k = 1\}, \tag{14}$$

that is, we check whether each observation is in the convex combination of the full state vector as the observation does not separate the resolved and unresolved scales. When this criterion (14) is satisfied, we trigger particle adjustment, which is the standard particle adjustment step (7) except that the posterior mean and covariance is given by (8) and (9) with an increased observation error [24, 22]

$$G = R^f H^T (H R^f H^T + r_o I + R')^{-1} \tag{15}$$

accounting for the contribution from the unresolved small-scales, i.e., the representation error.

Hard Threshold Multiscale Clustered Particle Filter Algorithm - one step assimilation.

Given :

1) N_o observations $\{v_1, v_2, ..., v_{N_o}\}$

2) prior K particles $\{(\mathbf{x}^f_{C_j,k}, \mathbf{y}^f_{C_j,k}), k = 1, 2, ..., K\}$ and weight vectors $\{\omega^f_{l,k}, k = 1, 2, ..., K\}$ for each cluster $C_l, l = 1, 2, .., N_{obs}$

For v_j from $j = 1$ to N_o

If The hard threshold criterion (14) is satisfied

Update the prior particles using (7) to match the Kalman update (8) and (9) with the Kalman gain G is given by (15)

Else Use particle filtering

Update $\{\omega^f_{j,k}\}$ using (13)

If $K_{eff} = \frac{1}{\sum_k (\omega^a_{l,k})^2} < \frac{K}{2}$

Do resampling

Add additional noise to the resampled particles

$$\mathbf{u}_{C_l,Resample(k)} \leftarrow \mathbf{u}_{C_l,Resample(k)} + \epsilon \tag{16}$$

where ϵ is IID Gaussian noise with zero mean and variance r_{noise}

End If
End If
End For

2.3. Multiscale ensemble filter

As a benchmark method, we use the multiscale ensemble method [22, 24], which uses a Gaussian assumption for the multiscale forecast PDF. Under this assumption, the multiscale ensemble filter becomes the standard ensemble filter except that the update formula uses an increased observation variance, i.e., the representation error, coming from the contribution of the unresolved scales. As we believe that the qualitative behavior of the multiscale ensemble filter is not strongly dependent on the particular choice of ensemble filters, we choose the ensemble adjustment Kalman filter [37] for the multiscale ensemble filter (we call it Multiscale EAKF (MsEAKF) hereafter).

3. Multiscale Dynamical Systems with Non-Gaussianity and Extreme Events : A Paradigm Model

A preliminary result of the multiscale clustered particle filter is reported in [29] with a successful application of the multiscale CPF for an one-dimensional wave turbulence model with breaking solitons and shallow energy spectrum but with a Gaussian distribution. Here we propose a multiscale turbulence model with interesting features of geophysical turbulence flows such as non-Gaussian statistics and extreme events to test the multiscale data assmilation method.

Our test model, which we call advective two-layer Lorenz-96 model, is given by the following two-layer coupled Lorenz-96 system

$$\begin{aligned} \frac{dx_i}{dt} &= x_{i-1}(x_{i+1} - x_{i-2}) + \lambda_1 \sum_{j=1}^{J} y_{i,j} - d_1 x_i + F, \quad i = 1, 2, ..., I \\ \frac{dy_{i,j}}{dt} &= \frac{a_L x_i + a_S y_{i,j+1}}{\epsilon}(y_{i,j-1} - y_{i,j+2}) - \lambda_2 x_i - d_2 y_{i,j}, \quad j = 1, 2, ..., J \end{aligned} \tag{17}$$

where x_i is periodic in i and $y_{i,j}$ is periodic in both i and j. This model is characterized by two sets of variables, slow-climate variable $\mathbf{x} = \{x_i\}$ of size I and fast-weather variable $\mathbf{y} = \{y_{i,j}\}$ of size IJ. Here $\epsilon > 0$ is an explicit time-scale separation parameter, F is an external slow forcing (which is constant in our study), λ_1 and λ_2 (which are not necessarily equal) are coupling parameters, and $d_1 > 0$ and $d_2 > 0$ are damping coefficients to stabilize the system. For the fast variable $\mathbf{y}$, there are large- and small-scale advection corresponding to the terms a_L and a_S respectively, which yields the slow-fast system when $a_L = 0$.

In our study, we fix $I = 40$ and $J = 10$ so that there are 440 variables in total (40 x_i's and 400 $y_{i,j}$'s). Note that when $\lambda_1 = 0$, the equation of x_i is the standard Lorenz-96 model designed to mimic baroclinic turbulence in the midlatitude atmosphere with energy-conserving nonlinear advection and dissipation [38, 3]. As the coupling parameters are set to nonzero values ($\lambda_1 \neq 0, \lambda_2 \neq 0$), this model problem is a good test model for filtering slow variables influenced by fast variables, which is crucial for the problems of medium-range weather prediction that is given by both the slow advective wave and the slowly varying envelope of the fast gravity waves. Note that without damping ($d_1 = d_2 = 0$) and no large-scale advection to the small-scale ($a_L = 0$) along with the same coupling parameters $\lambda_1 = \lambda_2$, this equation becomes the inviscid full Lorenz-96 model designed to study high skill prediction using FDT in [39].

3.1. Linear stability

To find interesting test regimes with extreme events and intermittency, which are represented by non-Gaussian fat-tails, we use the linear stability analysis of the model. First we consider the equation for the stationary homogeneous solution, $x_i = \overline{x}$ and $y_{ij} = \overline{y}$. As this solution has no spatial dependence, the equation of the homogeneous solution becomes

$$\frac{d\overline{x}}{dt} = \lambda_1 J\overline{y} - d_1\overline{x} + F = 0 \tag{18}$$

$$\frac{d\overline{y}}{dt} = -\lambda_2\overline{x} - d_2\overline{y} = 0, \tag{19}$$

which yields

$$\overline{x} = \frac{F}{d1 - \lambda_1\lambda_2 J/d_2}, \quad \overline{y} = \frac{\lambda_2}{d_2}\overline{x}. \tag{20}$$

If we denote the perturbations of x_i and y_{ij} around the steady state by x_i' and y_{ij}' respectively so that

$$x_i = \overline{x} + x_i' \quad \text{and} \quad y_{ij} = \overline{y} + y_{ij}',$$

the equations of x_i' and y_{ij}' are given by

$$\frac{dx_i'}{dt} = (\overline{x} + x_i')(x_{i+1}' - x_{i-2}') + \lambda_1 \sum_j y_{ij}' - d_1 x_i' \tag{21}$$

$$\frac{dy_{ij}'}{dt} = (a_L(\overline{x} + x_i') + a_S(\overline{y} + y_{ij}'))(y_{ij-1}' - y_{ij+2}') - \lambda_2 x_i' - d_2 y_{ij}' \tag{22}$$

To check the linear stability, we linearize (21) and (22) and obtain

$$\begin{aligned} \frac{dx_i'}{dt} &= \overline{x}(x_{i+1}' - x_{i-2}') + \lambda_1 \sum_j y_{ij}' - d_1 x_i' \\ \frac{dy_{ij}'}{dt} &= (a_L\overline{x} + a_S\overline{y})(y_{ij-1}' - y_{ij+2}') - \lambda_2 x_i' - d_2 y_{ij}' \end{aligned} \tag{23}$$

Now we define Y_j as the average of y_{ij}' over j

$$Y_i := \frac{1}{J}\sum_j y_{ij}'.$$

By summing the second equation of (23) over j and divide it by J, we obtain the following system

$$\begin{aligned}\frac{dx'_i}{dt} &= \overline{x}(x'_{i+1} - x'_{i-2}) + \lambda_1 \sum_j y'_{ij} - d_1 x'_i \\ \frac{dY_i}{dt} &= -\lambda_2 x'_i - d_2 Y_i\end{aligned} \tag{24}$$

Using Fourier series expansions of $x'_i = \sum_k \hat{x}'_k \exp(\frac{2\pi i k i}{I})$ and $Y_i = \sum_k \hat{Y}_k \exp(\frac{2\pi i k i}{I})$, plug them in (24), which yields the following equations for the Fourier coefficients

$$\frac{d}{dt}\begin{pmatrix}\hat{x}'_k \\ \hat{Y}_k\end{pmatrix} = A\begin{pmatrix}\hat{x}'_k \\ \hat{Y}_k\end{pmatrix} = \begin{pmatrix}\overline{x}(\exp(\frac{2\pi i k}{I}) - \exp(-\frac{4\pi i k}{I})) - d_1 & \lambda_1 J \\ -\lambda_2 & -d_2\end{pmatrix}\begin{pmatrix}\hat{x}'_k \\ \hat{Y}_k\end{pmatrix} \tag{25}$$

The real and imaginary parts of the matrix A are given by

$$\Re(A) = \begin{pmatrix}\overline{x}(\cos(\frac{2\pi k}{I}) - \cos(\frac{4\pi k}{I})) - d_1 & \lambda_1 J \\ -\lambda_2 & -d_2\end{pmatrix}, \tag{26}$$

and

$$\Im(A) = \begin{pmatrix}\overline{x}(\sin(\frac{2\pi k}{I}) + \sin(\frac{4\pi k}{I})) - d_1 & 0 \\ 0 & 0\end{pmatrix} \tag{27}$$

respectively. Note that the real and imaginary parts commute and thus the linear stability is related to the eigenvalues of the real part matrix (26). For simplicity, we use the following notations for the components of the real part matrix

$$\begin{aligned}a_{11} &= \overline{x}(\cos(\frac{2\pi k}{I}) - \cos(\frac{4\pi k}{I})) - d_1, \\ a_{12} &= \lambda_1 J, \\ a_{21} &= -\lambda_2, \\ a_{22} &= -d_2.\end{aligned} \tag{28}$$

If the discriminant of the characteristic function of the real part matrix

$$D := (a_{11} + a_{22})^2 - 4(a_{11}a_{22} + a_{12}a_{21}) \tag{29}$$

is positive there are two real eigenvalues. In this case, the condition for one positive and one negative eigenvalues is

$$a_{11}a_{22} - a_{12}a_{21} < 0$$

that is,

$$\frac{\lambda_1\lambda_2 J}{d_2} < \overline{x}(\cos(\frac{2\pi k}{I}) - \cos(\frac{4\pi k}{I})) - d_1. \tag{30}$$

On the other hand, the condition for two positive eigenvalues for linear instability is

$$a_{11} + a_{22} > 0 \quad \text{and} \quad a_{11}a_{22} - a_{12}a_{21} > 0$$

that is,

$$\frac{\lambda_1\lambda_2 J}{d_2} > \overline{x}(\cos(\frac{2\pi k}{I}) - \cos(\frac{4\pi k}{I})) - d_1 > d_2 \tag{31}$$

If D is negative (or zero), the eigenvalues are complex (or repeated real) and thus the condition for a positive real part of the eigenvalues (or positive repeated real), which guarantee linear instability, becomes

$$a_{11} + a_{22} > 0. \tag{32}$$

In addition to the linear stability analysis of x_i and the local average of y_{ij}, Y_i, we check the linear stability analysis of y_{ij} conditional to x_i. If we assume that there is time scale separation between x'_i and y'_{ij}, that is, x'_i can be assumed to be constant compared with y'_{ij}, we can check the linear stability of y'_{ij} directly from the second equation of (23). For fixed x'_i (and i), we use the Fourier series expansion of $y'_{ij} = \frac{-\lambda_2 x'_i}{d_2} + \sum_m \hat{y}'_m \exp(\frac{2\pi imj}{J})$ (where the first term $\frac{-\lambda_2 x'_i}{d_2}$ is the steady state solution to the second equation of (23)) and plug it into the second equation of (23), which yields

$$\frac{d}{dt}\hat{y}'_m = \left((a_L\overline{x} + a_S\overline{y})(\exp(-\frac{2\pi im}{J}) - \exp(\frac{4\pi im}{J})) - d_2\right)\hat{y}'_m. \tag{33}$$

Thus $\hat{y}'_m$ is linearly unstable when

$$\begin{aligned} &\Re\left((a_L\overline{x} + a_S\overline{y})(\exp(-\frac{2\pi im}{J}) - \exp(\frac{4\pi im}{J})) - d_2\right) \\ &= \left((a_L\overline{x} + a_S\overline{y})(\cos(\frac{2\pi m}{J}) - \cos(\frac{4\pi m}{J})) - d_2\right) > 0 \end{aligned} \tag{34}$$

3.2. Three parameter regimes

Depending on the presence of the large-scale and small-scale advection to the small-scale variable, we consider three parameter regimes. For each combination of advection, the other parameters are chosen to make instability in the system of x_i and y_{ij} (23) or the system of x_i and Y_i (24) (see Table 1 for the parameters of each regime). For the slow-fast system case, where $(a_L = 0, a_S = 1)$, λ_1 and λ_2 are equal and thus the interaction terms conserve the energy. This regime is a slow-fast system, which is typical in geophysical systems, for example, in atmosphere where a slow advective vortical Rossby wave is coupled with fast inertia-gravity waves [30, 31]. It is straightforward to check that the discriminant (29) is negative and thus the real part matrix has two complex eigenvalues. Further analysis shows that the real part of these complex numbers are negative and thus the linearized x_i and Y_i system is stable. However, if we assume that there is time-scale separation between x_i and y_{ij}, which is true for this system (see Table 2 for

	Slow-fast system	Strongly chaotic	Weakly chaotic
a_L	0	1	1
a_S	1	1	0
F	1	5	5
λ_1	-3	1/4	1/4
λ_2	-3	-1	-1
d_1	0.01	1	1.5
d_2	0.1	2	2.5
ϵ	0.1	1	1

Table 1: Three parameter regimes of the test model (17). I and J are fixed at 10 and 40 respectively.

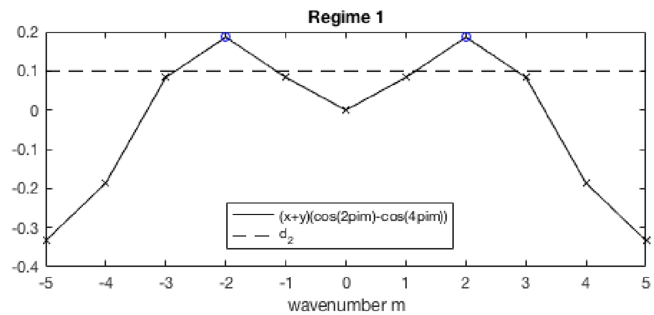
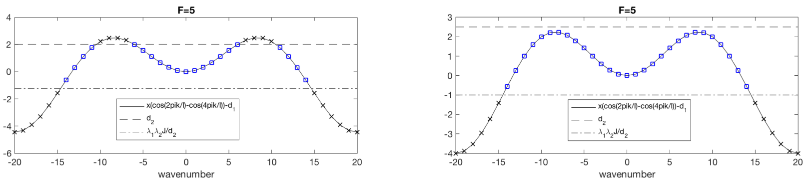


Figure 2: Slow-fast system. Linear stability of y_{ij} by assuming scale-separation between x_i and y_{ij} (34) . Wavenumber 2 is linearly unstable. Solid line : $\overline{y}(\cos(\frac{2\pi m}{J}) - \cos(\frac{4\pi m}{J}))$. Dash line : d_2.

the decorrelation times of x_i and y_{ij}), the linear stability analysis of y_{ij} (34) shows that y_{ij} is unstable (Figure 2 shows linearly unstable modes of y_{ij} for fixed i). Note that in this regime, only a small number of fast waves corresponding to wavenumber 2 are unstable.

When $\lambda_1 > 0$ and $\lambda_2 < 0$ (strongly chaotic and weakly chaotic cases), the discriminant (29) is positive and thus the system is unstable when (30) or (31) are satisfied. Figure 3 shows linearly unstable modes of the x_i and Y_i system (marked with blue circles) of the strongly chaotic and weakly chaotic cases. In the weakly chaotic regime, for low wavenumber k, the system of x_i and Y_i is unstable with one positive eigenvalue for the real part matrix of the linearized equation. In the strongly chaotic regime, low wavenumbers except 7-10 are unstable. Note that the eigenvector of (24) is a linear combination of x_i and Y_i. If we assume that there is time-scale separation between x_i and y_{ij}, the linear stability of y_{ij} (34) implies y_{ij} is linearly stable.

Table 2 shows the climatological properties of the three regimes. For the slow-fast and the strongly chaotic regimes, there are strongly non-Gaussian features (non-zero skewness and kurtosis away from 3). In the weakly chaotic regime, the decorrelation times of x_i and y_{ij} are inverted (y_{ij} has a longer decorrelation time than that of x_i) while the slow-fast and the strongly chaotic regimes have correct orders for decorrelation times; the presence of the small-scale advection makes the signal decorrelate rapidly in



(a) Strongly chaotic (b) Weakly chaotic

Figure 3: Strongly chaotic and weakly chaotic cases. Linear stability of x_i and $Y_i = \frac{1}{J}\sum_j y_{ij}$. Unstable wavenumbers are marked with squares while stable wavenumbers are marked with crosses. Solid line : $\overline{x}(\cos(\frac{2\pi k}{I}) - \cos(\frac{4\pi k}{I})) - d_1$. Dash-dot line : d_2. Dash line : $\lambda_1\lambda_2 J/d_2$

	Slow-fast system		Strongly chaotic		Weakly chaotic	
	x_i	y_{ij}	x_i	y_{ij}	x_i	y_{ij}
mean	0.022	0.033	1.69	-0.04	2.01	0.80
variance	0.009	0.021	5.71	6.80	8.51	0.75
skewness	0.261	-0.139	-0.02	-0.89	0.18	0.38
kurtosis	7.421	3.914	2.57	6.93	2.40	2.68
corr length	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1	≤ 1
corr time	1.91	0.88	0.92	0.22	2.93	3.52

Table 2: Climatological properties of the system (17).

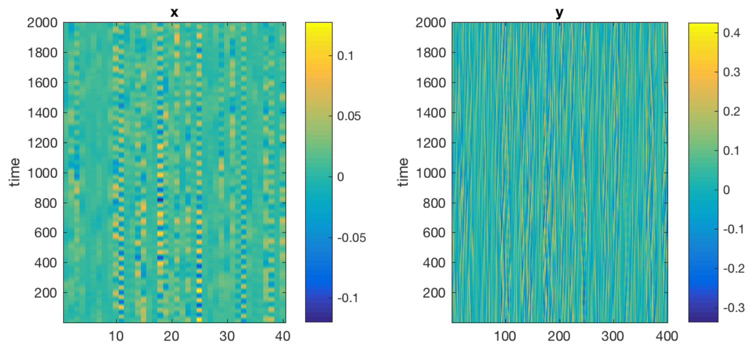
time.

Space-time diagrams of x_i and y_{ij} for all regimes are shown in Figure 4. In the slow-fast system case, there are random standing waves for $\mathbf{x}$ with intermittent local bursts and $\mathbf{y}$ is strongly mixing with no significant spatial structure. In the strongly chaotic case, $\mathbf{x}$ has westward moving waves and $\mathbf{y}$ has local bursts following the pattern of the moving waves of $\mathbf{x}$. In the weakly chaotic case, there are breaking waves while $\mathbf{y}$ has local bursts corresponding to the pattern of $\mathbf{x}$. Thus all three regimes have characteristics of turbulent flows, from strongly turbulent to weakly turbulent along with extreme events.

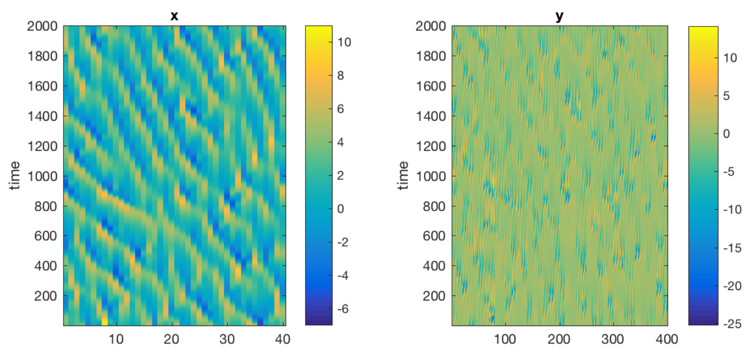
As a qualitative measure of non-Gaussian statistics, the stationary state PDFs of $x_i + y_{ij}$, x_i and y_{ij} of all regimes are shown in Figure 5 along with the Gaussian fits to the true. The top row of each figure shows the PDFs in log-scale (note that the log-scale of a Gaussian distribution is a parabola) while the bottom row of the figure shows the PDF without scaling. In all regimes, we can check that the system has strongly non-Gaussian statistics with fat-tails, which imply local extreme events.

Figure 6 shows the time series of x_i and y_{ij} at a grid point, $i = 2$ and $j = 5$. In the slow-fast system case, x_2 shows strong intermittency and $y_{2,5}$ has intermittent fast oscillation when there is intermittency in x_2. In the strongly and weakly chaotic cases, $y_{2,5}$ shows intermittent local bursts explaining the fat-tails of y_{ij}.

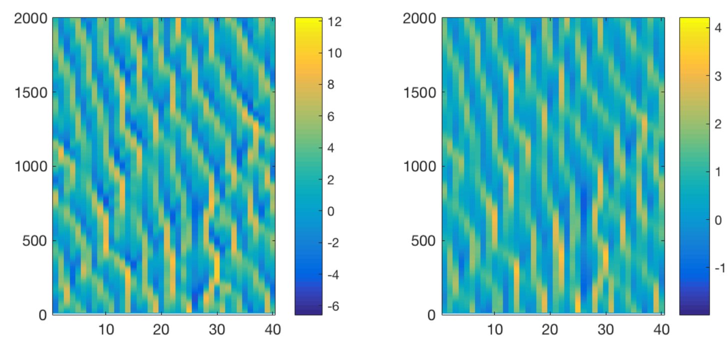
Another important statistical property of the turbulent system for data assimilation is decorrelation times and spatial correlation lengths. In Figure 7, the autocorrelation



(a) Slow-fast system



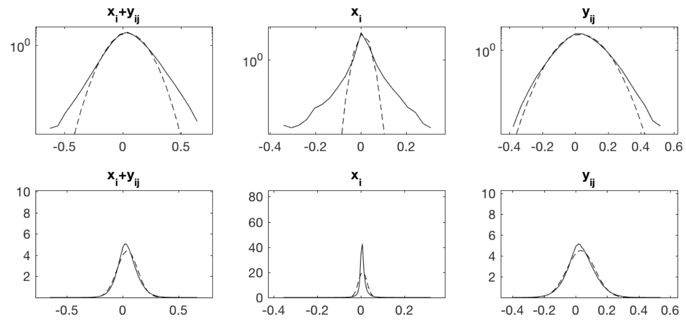
(b) Strongly chaotic



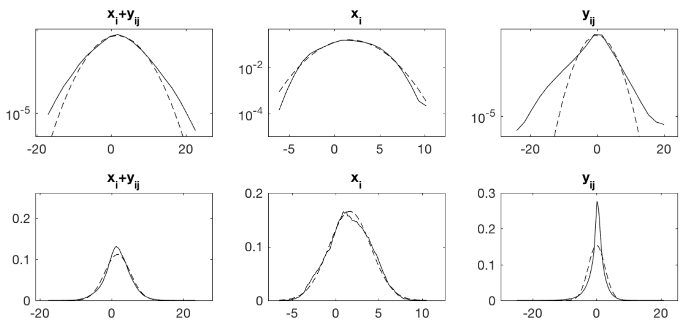
(c) Weakly chaotic

Figure 4: Space-time diagrams of $\mathbf{x}$ and $\mathbf{y}$ of the advective two-layer Lorenz-96 model (17) for all regimes.

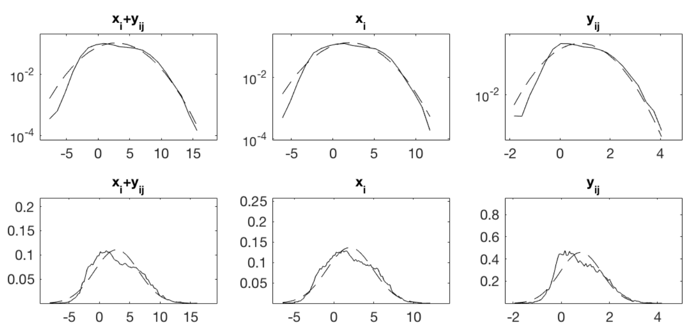
functions and spatial correlation functions are shown to analyze the decorrelation time and spatial correlation length. Except Regime 3, the decorrelation time of x_i is longer than that of y_{ij}, which are physical for slow-climate variable x_i and fast-weather variable



(a) Slow-fast system



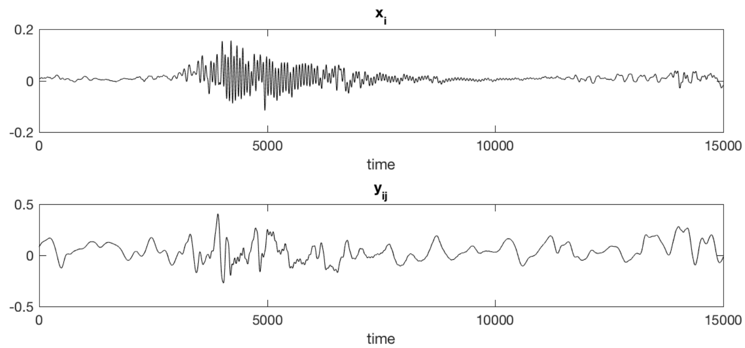
(b) Strongly chaotic



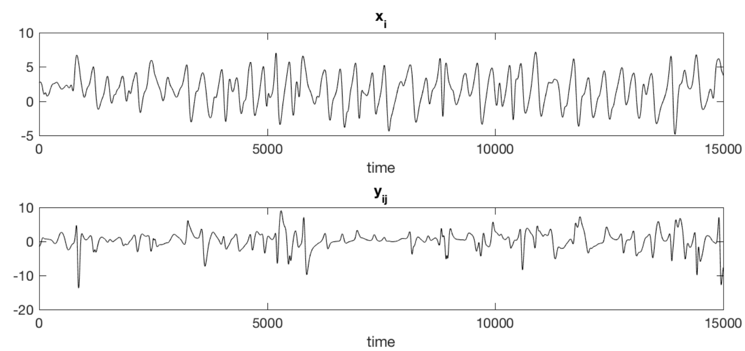
(c) Weakly chaotic

Figure 5: Stationary state PDFs of $x_i + y_{ij}$, x_i and y_{ij}. Log-scale (top) and without scaling (bottom). Dash lines are Gaussian fits. Note that the log-scale of a Gaussian distribution is a parabola.

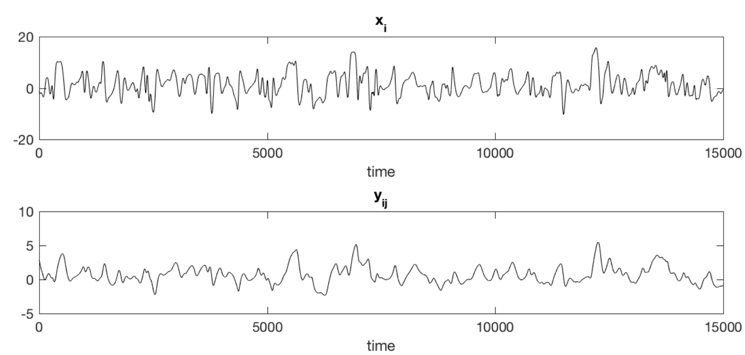
y_{ij}. Also, the spatial correlation length is less than 1 spatial grid point and thus all regimes are difficult test models for multiscale data assimilation.



(a) Slow-fast system



(b) Strongly chaotic

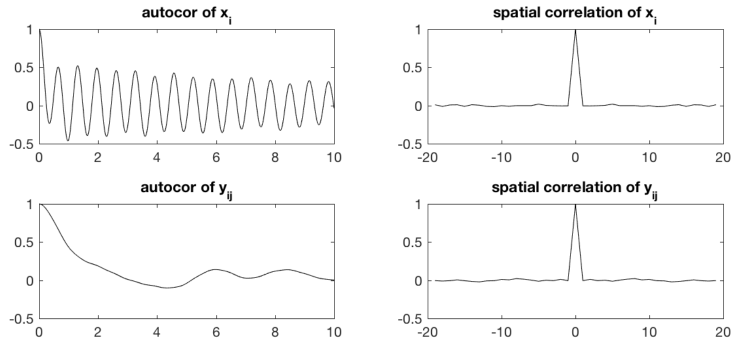


(c) Weakly chaotic

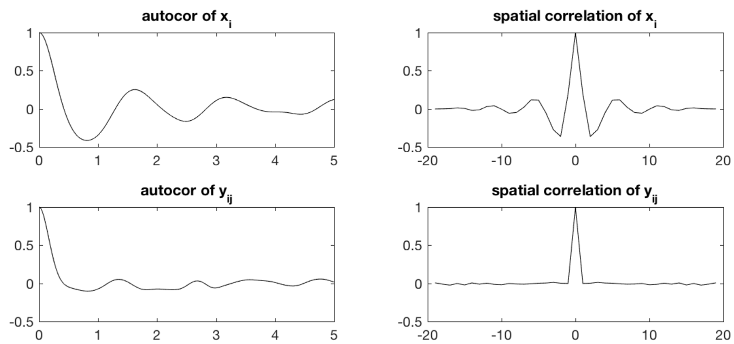
Figure 6: Time series at a grid point, x_2 and $y_{2,5}$

4. Numerical Experiments for Data Assimilation and Prediction using the Multiscale Particle Filter

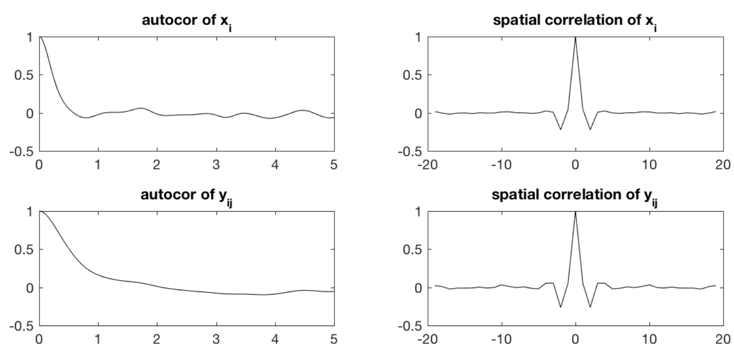
In this study, we are interested in the effect of the observation model error, i.e. the representation error, on the forecast skill for complex systems (see [11] for the study



(a) Slow-fast system



(b) Strongly chaotic



(c) Weakly chaotic

Figure 7: Autocorrelation (left) and spatial-correlation (right) functions of x_i (top) and y_{ij} (bottom)

of the effect of forecast model errors on the filter performance). To minimize the effect from the forecast model error, we use the perfect model as the forecast model. In the multiscale data assimilation setup, it is important to estimate the small-scale variance $\mathbf{R}'(\mathbf{x}_{l,k})$ for each large-scale variable. In our experiments, we approximate the

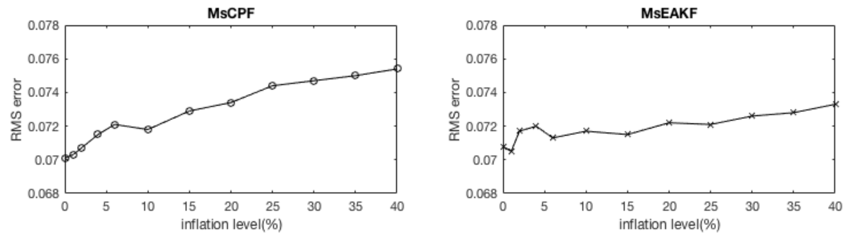


Figure 8: Slow-fast system. Time-averaged forecast RMS errors as functions of covariance inflation level. MsCPF (left) and MsEAKF (right). 20 observations.

small-scale covariance as a diagonal matrix whose diagonal components are given by the variance of $\{y_{ij}\}$ for each i. The original multiscale data assimilation framework provides a method to update the small-scale variables. However, this update is computationally expensive in real applications. Therefore, we update only the large-scale variables using the multiscale data assimilation method while the small-scale variables remain unchanged. This approximation is not optimal as it ignores information for the small-scale variables and thus there is an information barrier to get the optimal result. Although this is an interesting research topic, we do not investigate the barrier in the current study.

4.1. Experiment setup

We test the multiscale clustered particle filter (MsCPF) and the multiscale ensemble adjustment filter (MsEAKF) for the advective two-layer Lorenz 96 model. We first consider the expriments for the slow-fast and the strongly chaotic regimes. In each experiment, the true signal is given by one realization of the model. Both the true model and the forecast model use the same time integration method, the Euler-Maruyama method with a time step 10^{-3}. To mimic the incomplete partial observations in real applications, we test two scenarios, 40 full observations and 20 uniformly distributed observations that are available for each even i. Each observation component v_j directly observes the sum of x_j and $y_{j,5}$

$$v_j = x_j + y_{j,5} + \xi_i, \quad \xi_i : \text{iid random noise} \tag{35}$$

which has contributions from both the large-scale and the small-scale variables where the fifth component of y_{ij} contributes to the observation for each i. The observation interval varies from 0.1 to 0.8 for the slow-fast system case and from 0.05 to 0.1 for the strongly chaotic case, which are frequent compared with the decorrelation times of the large-scale variables in each regime. Observation error variance is only 1% of the total variance; however, the contribution from the unresolved small-scale variables, i.e., the representation error, is more than 50% of the total variance. Thus recovering accurate estimation and prediction skill for the resolved large-scale is difficult for both test regimes.

In each test, we run 5000 cycles and use the last 3000 cycles to measure the filter performance. Both MsCPF and MsEAKF use 50 samples and EAKF uses covariance localization using the smooth localization function by Gaspari and Cohn [40]. As the large-scale variable has a short decorrelation length (see Figure 7 and Table 2), we use a localization radius 2 that affects only the adjacent state variables. Covariance inflation plays an important role in recovering filter skill in the presence of model and sampling error [18, 19, 11]. In our multiscale data assimilation test, the covariance inflation plays no significant role in improving the filter performance. For the slow-fast system case, we tested several inflation levels and compare the time-averaged forecast RMS erros (Figure 8 shows the time-averaged forecast RMS errors as functions of the inflation level for both methods). Except the MsEAKF using a small inflation level and marginal gain, covariance inflation degrades the filter performance for both MsCpF and MsEAKF. Thus, the covariance inflation is not utilized in our tests.

4.2. Data Assimilation and Prediction

4.2.1. Slow-fast system regime

The slow-fast system system is typical in geophysical systems such as the atmosphere system where a slow advective vortical Rossby wave is coupled with fast inertia-gravity waves [30, 31]. Also more than two thirds of the total variance is carried by the unresolved small-scale variables, which is a difficult test problem for data assimilation as the unresolved small-scale variable plays an role of additional observation error in the estimation of x_i (i.e., the representation error).

As a quantitative path-wise measure, we check the RMS error of the forecast estimates. Figure 9 shows the time series of forecast RMS errors with 20 observations and observation time 0.1 by MsCPF and MsEAKF along with two benchmark values. The dash line is the climatological error given by the standard deviation of the resolved scale x_i, which is the error when we use the steady state mean. The other line, dash-dot line, is the effective observation error, which is the square root of the unresolved small-scale variance in addition to the raw observation error variance, which accounts for the representation error from the unresolved scale variables. From the figure, both MsCPF and MsEAKF have RMS errors staying below the climatological error except intermittent times, which shows filter skill from the noisy observational data both from the raw instrumental observation error and the unresolved scale error. Table 3 shows the time-averaged RMS errors and pattern correlation in parenthesis for several observation times and 40 full and 20 partial observations. As the observation time increases and the observation number decreases, the RMS error increases. However both methods are comparable and the RMS errors are smaller than the climatological error, which show filter skill.

One of the important measures in filtering high-dimensional systems is the recovery of the true PDF, which assess the lack of information in the filtered estimation and prediction. The RMS error and pattern correlation, which are path-wise measures of filter performance and are related to the Shannon entropy and the mutual information in information theory [3], fail to assess the lack of information in the filter estimates

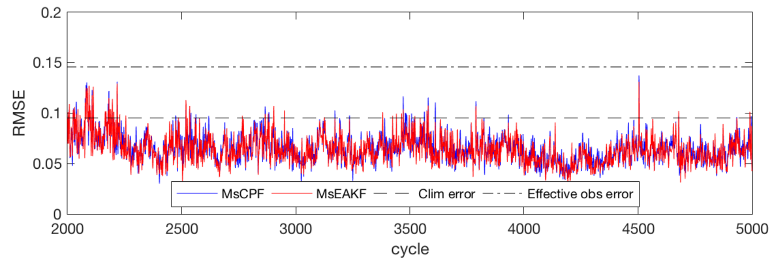


Figure 9: Slow-fast system. Time series of x-estimation RMS errors by MsCPF (blue) and MsEAKF (red). 20 observations and observation time 0.1. Dash line : climatological error. Dash-dot line : effective observation error.

	40 observations		20 observations	
obs time	MsCPF	MsEAKF	MsCPF	MsEAKF
0.1	0.061 (0.781)	0.061 (0.758)	0.079 (0.467)	0.078 (0.451)
0.3	0.062 (0.727)	0.060 (0.736)	0.080 (0.453)	0.078 (0.449)
0.5	0.072 (0.633)	0.069 (0.643)	0.085 (0.413)	0.085 (0.421)
0.8	0.075 (0.600)	0.071 (0.606)	0.087 (0.397)	0.085 (0.406)

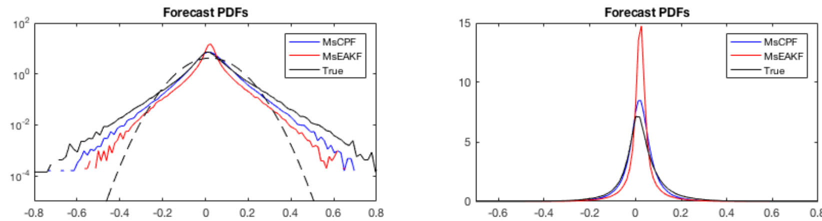
Table 3: Slow-fast system. Time averaged RMS errors and pattern correlation in parenthesis. Climatological error is 0.095. Effective observation error is 0.145.

and the predicted states [41, 42]. It is shown in [42] that two filtered trajectories with disparate amplitudes can have the same RMS error and pattern correlation. Especially in complex high-dimensional systems, which show extreme events and non-Gaussian statistics, it is important to quantify the ability of filters in capturing extreme events and non-Gaussian statistics. Figure 10 shows the climatological PDFs ((a) in log-scale and (b) without scaling) of the forecast estimates of x_i. The true PDF of x_i shows a strongly non-Gaussian PDF with fat-tails (see Figure 10 (a)). Both MsCPF and MsEAKF have fat-tails but MsCPF has a better fit to the true PDF than MsEAKF. From the PDFs without scaling (Figure 10 (b)), we can check more significant difference between MsCPF and MsEAKF; MsCPF has a comparable PDF with the true PDF with marginal misfit but MsEAKF has a very sharp peak and shallow tails with significant misfit from the true PDF.

The relative entropy, which is also called Kullback-Leibler divergence in probability theory and information theory, is defined as follows

$$\mathcal{P}(\pi, \pi^{filter}) = \int \pi(\mathbf{x}) \ln \frac{\pi(\mathbf{x})}{\pi^{filter}(\mathbf{x})} d\mathbf{x} \tag{36}$$

where $\pi(\mathbf{x})$ and $\pi^{filter}(\mathbf{x})$ are the true and filtered forecast PDFs of $\mathbf{x}$ respectively. The relative entropy measures the lack of information in estimating the true PDF π using the filtered forecast PDF π^{filter} and this has been successfully applied in quantifying the filter performance in several contexts [5, 43]. Note that if we have $\pi^{filter} = \pi$ the



(a) in log-scale (b) without scaling

Figure 10: Slow-fast system. Forecast PDFs of $\mathbf{x}$ by MsCPF (blue) and MsEAKF (red) along with the true value (black). Dash-line is the Gaussian fit to the true PDF. 20 observations

relative entropy is 0 and a large value means much lack of information of the filtered PDF. The forecast relatively entropy using the forecast PDFs by MsCPF and MsEAKF are shown in Table 4 for 40 and 20 observations and observation times from 0.1 to 0.8. As we use the forecast PDFs for the relative entropy, a smaller relative entropy means better prediction and forecast skill than a larger relative entropy. As expected from the recovery of the true PDF, the forecast relative entropy of MsCPF is smaller than one of MsEAKF, the relative entropy of MsEAKF is about four times larger than that of MsCPF. As the number of observations and the observation interval increase, the lack of information in the forecast filter estimate increases, that is, the relative entropy increases. However, the ratio between MsCPF and MsEAKF does not change.

	40 observations		20 observations	
obs time	MsCPF	MsEAKF	MsCPF	MsEAKF
0.1	0.0365	0.1647	0.0398	0.1701
0.3	0.0383	0.1783	0.0403	0.1795
0.5	0.0410	0.1812	0.0421	0.1819
0.8	0.0437	0.1841	0.0438	0.1881

Table 4: Slow-fast system. Forecast relative entropy using the forecast estimate PDFs by MsCPF and MsEAKF.

The filer performance between MsCPF and MsEAKF in capturing the non-Gaussian statistics also can be investigated from the time series of the forecast estimate of x_{10} shown in Figure 11. The true value of x_{10} stays bounded but it shows amplified fast oscillations extreme events beginning from time 2100. Both methods capture the beginning of fast oscillations; however the amplitude of MsEAKF is less than half of the true amplitude at time around 2700 while MsCPF has a comparable amplitude of the true value, which explains the narrower tail bounds of MsEAKF and the sharp peak in the forecast PDF of x_i (Figure 10).

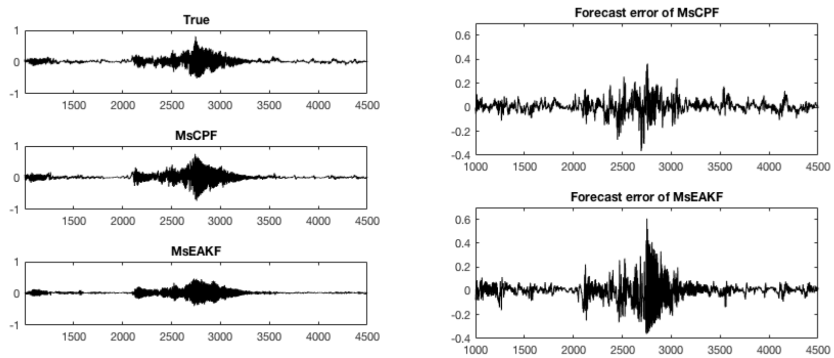


Figure 11: Slow-fast system. Time series of x_{10} forecast estimates (left) and forecast error (right) by MsCPF and MsEAKF. 20 observations

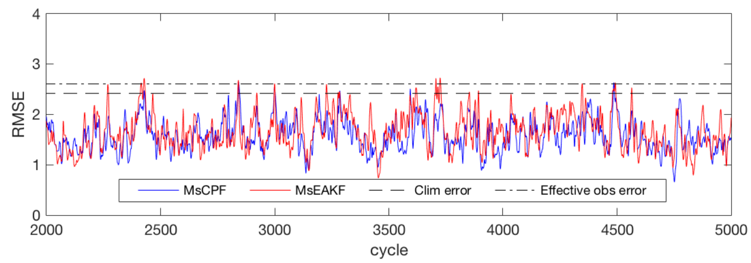


Figure 12: Strongly chaotic case. Time series of x-estimation RMS errors by MsCPF (blue) and MsEAKF (red). Dash line : climatological error. Dash-dot line : effective observation error.

4.2.2. Strongly chaotic regime

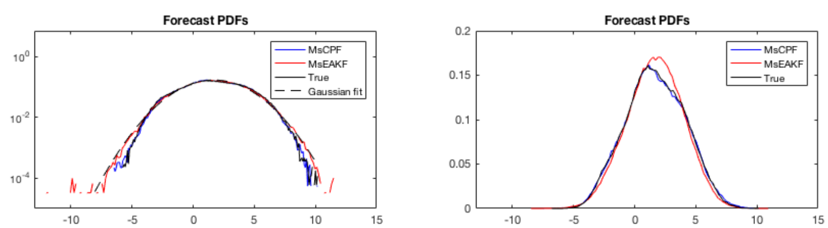
We now investigate the filter performance of MsCPF and MsEAKF applied for the second test regime, which has both the large- and small-scale advection to the small-scale dynamics ($a_L \neq 0, a_S \neq 0$). The westward moving waves seen in x_i is typical in the midlatitude atmosphere, i.e., the Rossby waves and x_i has non-Gaussian statistics, which is of our interest to recover using the multiscale data assimilation method.

As in Slow-fast system, we compare the filter performance using the path-wise measures, RMS error and pattern correlation. Figure 12 shows the forecast estimate RMS errors of x_i as a function of time (the blue line is MsCPF and the red line is MsEAKF along with the climatological error (dash line) and the effective observation error (dash-dot line)). Both methods have filter skill and have comparable RMS errors that are smaller than both the climatological and the effective observation errors. Table 5 shows the time averaged forecast RMS errors and pattern correlations in parenthesis for frequent observation times 0.05 and 0.1 and 40 full and 20 sparse observations. Sparse observation and long observation time degrade the filter performance but both methods show skillful filter performance with RMS errors smaller than the climatological error along with pattern correlations larger than 88% and 73% for 40 and 20 observations respectively.

	40 observations		20 observations	
obs time	MsCPF	MsEAKF	MsCPF	MsEAKF
0.05	1.12 (0.885)	1.17 (0.896)	1.66 (0.747)	1.67 (0.747)
0.10	1.17 (0.879)	1.21 (0.890)	1.76 (0.732)	1.79 (0.731)

Table 5: Strongly chaotic case. Time averaged RMS errors and pattern correlation in parenthesis. Climatological error is 2.39. Effective observation error is 2.620.

In the slow-fast system, the filter performance between MsCPF and MsEAKF is observed in quantifying the lack of information in the filter estimates and the predicted states, that is, the recovery of the true PDF. The climatological PDFs of the forecast estimates of x_i by both methods along with the true PDF are shown in Figure 13. In the log-scale plot (Figure 13 (a)), we can check that the forecast PDF of MsCPF is on top of the true PDF, which has sub-Gaussian tails. On the other hand, the PDF of the ensemble-based method, MsEAKF, is a Gaussian fit to the true PDF. Without scaling, we can check more significant performance difference between MsCPF and MsEAKF. In Figure 13 (b), the PDF of MsCPF is on top of the true PDF capturing the non-symmetric peak of the true PDF. However, the PDF of MsEAKF fails to capture the non-symmetric peak of the true PDF.



(a) in log-scale (b) without scaling

Figure 13: Strongly chaotic case. Forecast PDFs of $\mathbf{x}$ by MsCPF (blue) and MsEAKF (red) along with the true value (black). Dash-line is the Gaussian fit to the true PDF.

As in the slow-fast system, the forecast relatively entropy using the forecast estimate PDFs by MsCPF and MsEAKF are shown in Table 6, which measure the prediction skill and the lack of information in the forecast. The lack of information in the forecast prediction is much larger for MsEAKF; the forecast relatively entropy of MsCPF is about four times smaller than the relative entropy of MsEAKF. This result implies that the filter prediction can have significant performance difference in quantifying the uncertainty although they have comparable performance measured by path-wise measures such as the RMS error and pattern correlation [17, 43].

The space-time diagrams of the forecast estimates of x_i along with the true x_i are shown in Figure 14. Both methods have wave patterns comparable to the true state however the wave of MsEAKF has artificial local intermittency (for example, check the

	40 observations		20 observations	
obs time	MsCPF	MsEAKF	MsCPF	MsEAKF
0.05	0.0016	0.0069	0.0018	0.0078
0.10	0.0018	0.0072	0.0019	0.0089

Table 6: Strongly chaotic case. Forecast relative entropy using the forecast estimate PDFs by MsCPF and MsEAKF.

time around 350). This comparison also shows that there is no significant evidence of dynamic imbalance in MsCPF although MsCPF uses the coarse-grained localization. As another qualitative measure of filter performance, Figure 15 shows the time series

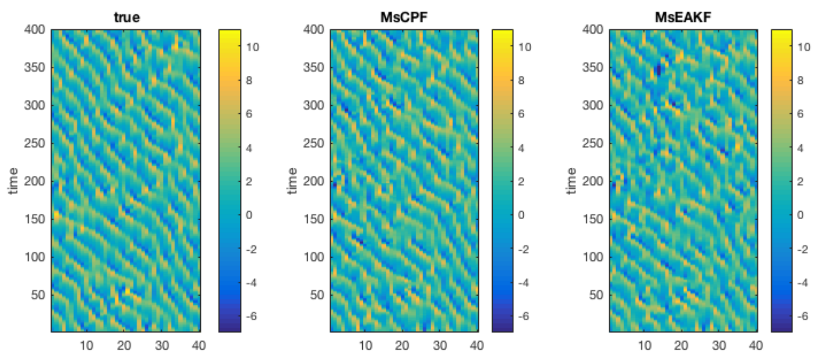


Figure 14: Snapshots of the forecast estimates of $\mathbf{x}$ by MsCPF (middle) and MsEAKF (right) along with the true value (left)

of x_{10}. At the 3320th and 3430th cycles, MsCPF captures the correct local peaks but the ensemble-based multiscale filter fails to capture the comparable peaks.

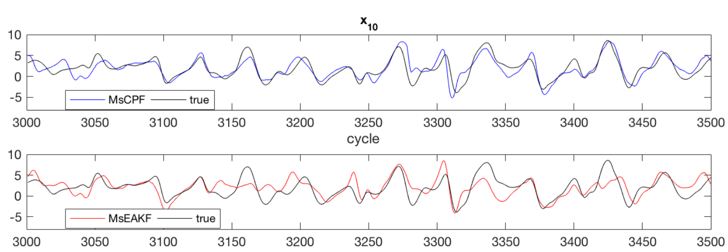


Figure 15: Strongly chaotic case. Time series of x_{10} forecast estimates by MsCPF (top) and MsEAKF (bottom) along with the true value.

4.3. Weakly chaotic regime : prediction of the large-scale of y_{ij}

In the previous two test regimes, we were interested in the estimation and prediction of the slow resolved variable x_i, which has a longer decorrelation time than the one of

y_{ij}. In the weakly chaotic regime, the decorrelation times of x_i and y_{ij} are reversed and thus it is a non-physical and uninteresting test to predict x_i instead of y_{ij} as the unresolved y_{ij} is easier to predict than x_i and thus this setup is not a typical situation of data assimilation in real applications. In this section, we change the role of x_i and y_{ij}, that is, we compare the multiscale filtering methods in the estimation and prediction of y_{ij} instead of x_i.

More precisely, we use the following observation $\mathbf{v} = \{v_1, v_2, ..., v_j\}$

$$v_j = x_j + Y_j + \xi_i, \quad \xi_i : \text{iid random noise} \tag{37}$$

where Y_j is the local average of y_{ij}

$$Y_j = \frac{1}{J}\sum_j y_{ij} \tag{38}$$

so that there are equal number of variables for x_i and Y_i. This setup is not artificial but pratical in that in real applications, many observations have collective information of different locations or variables such as radiation information from satellites [44]. This coupled observation test and its mathematical analysis has already been studied in Chapter 7 of [1]. Our experiment setup is comparable to the setup in [1] but our test in this study is different from them as we test computationally efficient and cheap multiscale data assimilation methods instead of single-scale standard data assimilation methods.

Except the new observation operator (37), the other setup parameters are the same as in the previous two tests. We test 40 and 20 full and sparse observations with frequent observation intervals 0.05 and 0.10. Observation error variance is only 1% of the total variance and thus most of the observation error comes from the unresolved scale, i.e., the representation error. Both MsCPF and MsEAKF use 50 samples and run 5000 assimilation cycles and use the last 3000 cycles to measure the filter performance.

4.3.1. Data assimilation and prediction in the weakly chaotic regime

The time series of the forecast RMS errors by MsCPF (blue) and MsEAKF (red) with 20 observations and an observation interval 0.05 are shown in Figure 16 along with the climatological (dash) and effective observation (dash-dot) errors. In contrast to the previous two test regimes, there is significant performance difference in the RMS error, a path-wise filter measure; the RMS error of MsCPF stays below the climatological error, which shows significant filter skill but the RMS error of MsEAKF is larger than the climatological error without any filter skill. For other test scenarios (40 observations and an observation interval 0.10), the time averaged RMS errors and pattern correlations are shown in Table 7. For all possible observation scenarios, the RMS errors of MsCPF is at least 30% less than the climatological error while MsEAKF has errors larger than the climatological error. Regarding the forecast pattern correlations, which explains how much of the spatial variation is explained by the forecast, the pattern correlations

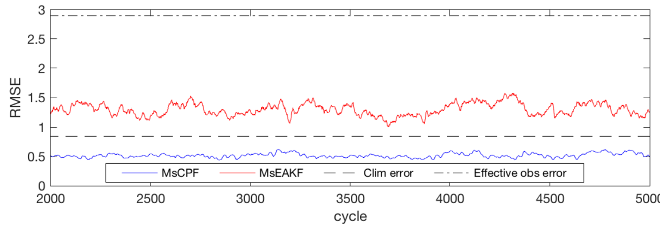


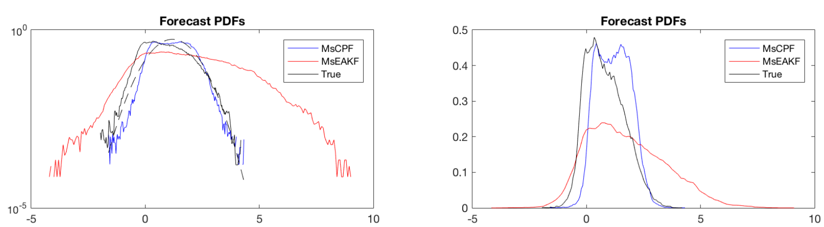
Figure 16: Weakly chaotic case. Time series of forecast Y-estimation RMS errors by MsCPF (blue) and MsEAKF (red). Dash line : climatological error. Dash-dot line : effective observation error.

	40 observations		20 observations	
obs time	MsCPF	MsEAKF	MsCPF	MsEAKF
0.05	0.52 (0.90)	1.30 (0.64)	0.55 (0.83)	1.46 (0.52)
0.10	0.54 (0.87)	1.32 (0.63)	0.61 (0.81)	1.53 (0.50)

Table 7: Weakly chaotic case. Time averaged RMS errors and pattern correlation in parenthesis. Climatological error is 0.844. Effective observation error is 2.900.

of MsCPF is at least 80% but the forecast pattern correlation of MsEAKF is less than 65% for all scenarios and is marginally above 50% for the toughest test scenario.

Next we consider the recovery of the true PDF using the forecast estimates and the relative entropy to assess the lack of information in the forecast estimates and predictions. The forecast PDFs of Y_i (blue : MsCPF, red : MsEAKF) using 20 observations and an observation time 0.05 along with the true PDF of Y_i (black) are shown in Figure 17. The PDF of MsCPF captures the comparable variance and shape of the true PDF although it is not on the top of the true PDF compared to the previous two test regimes. In contrary, the PDF of MsEAKF has a too large variance compared to the true PDF. This result shows that forecast using MsEAKF is inadequate as it provides incorrect weights on large deviated values while MsCPF has comparable weights to the true PDF. As a quantitative measure of the lack of information, the relative entropy for



(a) in log-scale

(b) without scaling

Figure 17: Weakly chaotic case. Forecast PDFs of $\mathbf{x}$ by MsCPF (blue) and MsEAKF (red) along with the true value (black). Dash-line is the Gaussian fit to the true PDF.

several scenarios are shown in Table 8. As discussed before, a smaller relative entropy

	40 observations		20 observations	
obs time	MsCPF	MsEAKF	MsCPF	MsEAKF
0.05	0.1631	0.3024	0.1787	0.4328
0.10	0.1791	0.3234	0.1891	0.4523

Table 8: Weakly chaotic case. Forecast relative entropy using the forecast estimate PDFs by MsCPF and MsEAKF.

implies a better prediction or less lack of information. In comparison between MsCPF and MsEAKF, it is obvious that MsCPF has a superior prediction skill with relative entropies half of those of MsEAKF. As the number of observation decreases or the observation interval increases, the relative entropy decreases, which implies performance degradation. However, the relative entropies of MsEAKF never becomes smaller than those of MsCPF.

5. Conclusions

In the data assimilation of high-dimensional complex systems such as turbulent geophysical systems, it is indispensable to use coarse-resolution forecast models as it is computationally prohibitive to resolve all active spatiotemporal scales. To mitigate the problem related to the incorporation of coarse-resolution forecast models, i.e., mixed contributions from both the resolved and unresolved scales, we have proposed and tested the multiscale clustered particle filter (MsCPF). MsCPF follows the single-scale clustered particle filter [29] that use coarse-grained localization and particle adjustment while the update in each cluster follows the general multiscale particle filter [22] instead of the standard particle filter update.

To test the multiscale algorithm under effect of the observation model error, we proposed and developed an advective two-layer Lorenz-96 system. Using several combination of large- and small-scale advection on the small-scale equation, the model can mimic several different test regimes including the standard slow-fast system that is typical in atmosphere where a slow advective vortical Rossby wave is coupled with fast inertia-gravity waves. All different regimes we considered in this study have important features of turbulent systems such as non-Gaussian statistics including fat-tails and intermittent extreme events. The multiscale clustered particle filter shows robust skill in recovering the true non-Gaussian PDF using a relatively few particles while an ensemble-based method fails to capture the non-Gaussian feature. In the weakly chaotic test regime with collective observation of the slow variables, which mimics one of the difficult test scenario in real-applications such as radiation observation from satellites, MsCPF shows superior performance to the ensemble based multiscale methods, MsEAKF, in both the path-wise measure, RMS errors and pattern correlations and the information theoretic measure, recovery of true PDF and relative entropy.

In this study, we focused on the investigation of the effect of the observation model error, which is indispensable in the multiscale data assimilation as the forecast model provides only the resolved large-scale components. For this purpose, only the perfect forecast model has been tested in our study to minimize the error from the forecast model error, which is another important factor for filter performance. Thus it is natural to extend the current study to the investigation of the forecast model error, especially from the coarse-resolution model error. Also we believe that the information barrier related to the ignored small-scale update in our study could hinder further performance improvement of the multiscale clustered particle filter. In the near future, we plan to investigate the effects of the information barrier and the forecast model error on the multiscale filter performance.

Acknowledgements

The research of Y. Lee is supported as a postdoctoral fellow by DARPA 25-74200-F4414. A.J. Majda is partially supported by Office of Naval Research grant ONR MURI N00014-12-1-0912 and DARPA 25-74200-F4414.

[1] A. J. Majda, J. Harlim, Filtering complex turbulent systems, Cambridge University Press, 2012.

[2] A. Doucet, N. de Freitas, N. Gordon, An Introduction to Sequential Monte Carlo Methods, Springer New York, New York, NY, 2001, pp. 3–14.

[3] A. Majda, X. Wang, Nonlinear dynamics and statistical theories for basic geophysical flows, Cambridge University Press, 2006.

[4] A. J. Majda, J. Harlim, B. Gershgorin, Mathematical strategies for filtering turbulent dynamical systems, Discrete Contin. Dyn. Syst 27 (2) (2010) 441–486.

[5] A. J. Majda, M. Branicki, Lessons in uncertainty quantification for turbulent dynamical systems, Discrete Cont. Dyn. Systems 32 (9).

[6] J. Berner, G. Branstator, Linear and nonlinear signatures in the planetary wave dynamics of an agcm: Probability density functions, Journal of the Atmospheric Sciences 64 (1) (2007) 117–136.

[7] J. D. Neelin, B. R. Lintner, B. Tian, Q. Li, L. Zhang, P. K. Patra, M. T. Chahine, S. N. Stechmann, Long tails in deep columns of natural and anthropogenic tropospheric tracers, Geophysical Research Letters 37 (5).

[8] G. Evensen, Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics, Journal of Geophysical Research: Oceans 99 (C5) (1994) 10143–10162. doi:10.1029/94JC00572.

[9] M. K. Tippett, J. L. Anderson, C. H. Bishop, T. M. Hamill, J. S. Whitaker, Ensemble square root filters, Monthly Weather Review 131 (7) (2003) 1485–1490.

[10] C. Snyder, T. Bengtsson, P. Bickel, J. Anderson, Obstacles to high-dimensional particle filtering, Monthly Weather Review 136 (12) (2008) 4629–4640.

[11] I. Grooms, Y. Lee, A. J. Majda, Ensemble filtering and low-resolution model error: Additive inflation, stochastic parameterization, and model numerics, Mon. Wea. Rev. 143 (2015) 3912–3924.

[12] I. Grooms, Y. Lee, A. J. Majda, Numerical schemes for stochastic backscatter in the inverse cascade of quasigeostrophic turbulence, Multiscale Modeling and Simulation 13 (2015) 1001–1021.

[13] J. Harlim, A. J. Majda, et al., Catastrophic filter divergence in filtering nonlinear dissipative systems, Communications in Mathematical Sciences 8 (1) (2010) 27–43.

[14] G. A. Gottwald, A. Majda, A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks, Nonlinear Processes in Geophysics 20 (5) (2013) 705–712.

[15] D. Kelly, A. J. Majda, X. T. Tong, Concrete ensemble kalman filters with rigorous catastrophic filter divergence, Proceedings of the National Academy of Sciences 112 (34) (2015) 10589–10594.

[16] Y. Lee, A. J. Majda, D. Qi, Preventing catastrophic filter divergence using adaptive additive inflation for baroclinic turbulence, Monthly Weather Review 145 (2) (2017) 669–682.

[17] M. Branicki, A. J. Majda, Quantifying uncertainty for predictions with model error in non-gaussian systems with intermittency, Nonlinearity 25 (9) (2012) 2543.

[18] J. Anderson, S. Anderson, A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts, Mon. Wea. Rev. 127 (1999) 2741–2758.

[19] H. L. Mitchell, P. Houtekamer, An adaptive ensemble kalman filter, Monthly Weather Review 128 (2) (2000) 416–433.

[20] P. L. Houtekamer, H. L. Mitchell, A sequential ensemble kalman filter for atmospheric data assimilation, Monthly Weather Review 129 (1) (2001) 123–137.

[21] R. Daley, Estimating observation error statistics for atmospheric data assimilation, Annales geophysicae 11 (7) (1993) 634–647.

[22] Y. Lee, A. J. Majda, Multiscale methods for data assimilation in turbulent systems, Multiscale Modeling and Simulation 13 (2015) 691–173.

[23] A. J. Majda, D. Qi, T. P. Sapsis, Blended particle filters for large-dimensional chaotic dynamical systems, Proceedings of the National Academy of Sciences 111 (21) (2014) 7511–7516.

[24] I. Grooms, Y. Lee, A. J. Majda, Ensemble kalman filters for dynamical systems with unresolved turbulence, Journal of Computational Physics 273 (2014) 435–452.

[25] Y. Lee, A. J. Majda, D. Qi, Stochastic superparameterization and multiscale filtering of turbulent tracers, Multiscale Modeling & Simulation 15 (1) (2017) 215–234.

[26] J. Harlim, A. J. Majda, Test models for filtering with superparameterization, Multiscale Modeling & Simulation 11 (1) (2013) 282–308.

[27] A. J. Majda, Y. Lee, Conceptual dynamical models for turbulence, Proceedings of the National Academy of Sciences 111 (18) (2014) 6548–6553.

[28] B. L. T. Bengtsson, P. Bickel, Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems, IMS Collections, eds Nolan D, Speed T 2 (2008) 316–334.

[29] Y. Lee, A. J. Majda, State estimation and prediction using clustered particle filters, Proceedings of the National Academy of Sciences (2016) 201617398.

[30] R. Salmon, Lectures on geophysical fluid dynamics, Oxford University Press, 1998.

[31] A. Majda, Introduction to PDEs and Waves for the Atmosphere and Ocean, Vol. 9, American Mathematical Soc., 2003.

[32] S. Reich, A nonparametric ensemble transform method for bayesian inference, SIAM Journal on Scientific Computing 35 (4) (2013) A2013–A2024.

[33] N. Chustagulprom, S. Reich, M. Reinhardt, A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems, SIAM/ASA Journal on Uncertainty Quantification 4 (1) (2016) 592–608.

[34] J. Poterjoy, A localized particle filter for high-dimensional nonlinear systems, Monthly Weather Review 144 (1) (2016) 59–76.

[35] A. C. Lorenc, The potential of the ensemble kalman filter for nwp a comparison with 4d-var, Quarterly Journal of the Royal Meteorological Society 129 (595) (2003) 3183–3203.

[36] J. D. Kepert, Covariance localisation and balance in an ensemble kalman filter, Quarterly Journal of the Royal Meteorological Society 135 (642) (2009) 1157–1176.

[37] J. Anderson, An ensemble adjustment kalman filter for data assimilation, Mon. Wea. Rev. 129 (2001) 2884–2903.

[38] E. N. Lorenz, Predictability: A problem partly solved, in: Proc. Seminar on predictability, Vol. 1, 1996.

[39] A. J. Majda, R. Abramov, B. Gershgorin, High skill in low-frequency climate response through fluctuation dissipation theorems despite structural instability, Proceedings of the National Academy of Sciences 107 (2) (2010) 581–586.

[40] G. Gaspari, S. E. Cohn, Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society 125 (554) (1999) 723–757. doi:10.1002/qj.49712555417.

[41] M. Branicki, A. J. Majda, Quantifying bayesian filter performance for turbulent dynamical systems through information theory, Comm. Math. Sci 12 (5) (2014) 901–978.

[42] N. Chen, A. J. Majda, Predicting the real-time multivariate maddenjulian oscillation index through a low-order nonlinear stochastic model, Monthly Weather Review 143 (6) (2015) 2148–2169.

[43] N. Chen, A. J. Majda, Filtering nonlinear turbulent dynamical systems through conditional gaussian statistics, Monthly Weather Review 144 (12) (2016) 4885–4917.

[44] R. H. Reichle, Data assimilation methods in the earth sciences, Advances in Water Resources 31 (11) (2008) 1411 – 1418, hydrologic Remote Sensing.