State estimation and prediction using clustered particle filters

Yoonsang Lee[a] **and Andrew J Majda**[a]

[a]Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012

Particle filtering is an essential tool to improve uncertain model predictions by incorporating noisy observational data from complex systems including non-Gaussian features. A new class of particle filters, clustered particle filters, is introduced for high-dimensional nonlinear systems, which uses relatively few particles compared to the standard particle filter. The clustered particle filter captures non-Gaussian features of the true signal which are typical in complex nonlinear dynamical systems such as geophysical systems. The method is also robust in the difficult regime of high-quality sparse and infrequent observations. The key features of the clustered particle filtering are coarse-grained localization through the clustering of the state variables and particle adjustment to stabilize the method; each observation affects only neighbor state variables through clustering and particles are adjusted to prevent particle collapse due to the high-quality observations. The clustered particle filter is tested for the 40-dimensional Lorenz-96 model with several dynamical regimes including strongly non-Gaussian statistics. The clustered particle filter shows robust skill in both achieving accurate filter results and capturing non-Gaussian statistics of the true signal. It is further extended to the multiscale data assimilation which provides the large-scale estimation by combining a cheap reduced-order forecast model and mixed observations of the large- and small-scale variables. This approach enables use of a larger number of particles due to the computational savings in the forecast model. The multiscale clustered particle filter is tested for one-dimensional dispersive wave turbulence using a forecast model with model errors.

data assimilation | particle filter | uncertainty quantification

Data assimilation or filtering combines numerical forecast models with observational data to provide the best statistical estimation and prediction of complex systems. Due to the high-dimensionality of complex nonlinear systems such as geophysical systems, accurate and efficient estimation and prediction of such complex systems are formidable tasks as enormously large computational resources are required to run forecast models and observations are typically sparse and infrequent. As a Monte Carlo approach, ensemble based methods (1, 2) combined with covariance inflation and localization are indispensable tools as they allow computationally cheap, low-dimensional ensemble state approximation for the systems and have performed well for operational applications such as numerical weather prediction (3, 4). Nevertheless, as the ensemble based methods approximate the forecast distribution using Gaussian statistics, these methods lead to inaccurate estimation and prediction when the true signal has non-Gaussian statistics which are typical for a wide range of systems including many geophysical systems. Particle filtering captures non-Gaussian features using different weights for different samples (or particles) and is an established discipline for low-dimensional dynamical systems (5–7). Particle filtering does not require any assumption on the prior distribution and

leads to consistent Bayesian posterior statistics. The particle weights also determine which particles to remove and duplicate, i.e. resampling strategies (8), to prevent particle collapse where only a small fraction of particles have the most weight while the rest of the particles have minimal or zero weights. Despite the successful applications of the particle filter for low-dimensional systems, the particle filter suffers from the curse of dimensionality for high-dimensional systems, which requires exponentially increasing particle numbers with the dimension of the system (9, 10).

There are several attempts to overcome the curse of dimensionality and enhance the performance of the particle filter with a small particle size. The blended particle filter (11, 12) uses forecast models which have adaptively varying reduced-order models to capture non-Gaussian statistics using particle filtering while high-dimensional quasi-Gaussian statistics are maintained by Gaussian mixtures. The maximum entropy particle filter in Chapter 15 of (13) uses judicious use of partial marginal distributions to avoid particle collapse. Another method is to solve an optimal transport problem for the transition from the prior to the posterior (14, 15) to avoid the random sampling aspects of particle filters. Motivated by the success of the ensemble based method using covariance localization (16, 17), there are other methods which implement localization for non-Gaussian statistics including the rank histogram filter (18) and the localized particle filter (19). The localized particle filter uses vector particle weights instead of the scalar weights of the standard particle filter and provides a Bayesian update in regions near the physical locations of observations. This method shows successful results capturing

---

Significance Statement

Particle filtering is an essential tool for the estimation and prediction of complex systems including non-Gaussian features. A new class of particle filters, clustered particle filters, is introduced for high dimensional dynamical systems such as geophysical systems. The proposed method uses relatively few particles compared to the standard particle filter and captures the non-Gaussian features of the true signal which are typical in complex nonlinear systems. The method is also robust for the difficult regime of high-quality sparse and infrequent observations and does not show any filter divergence in our tests. In the clustered particle filter, coarse-grained localization is implemented through the clustering of state variables and particles are adjusted to stabilize the filter.

www.pnas.org/cgi/doi/10.1073/pnas.XXXXXXXXXX

PNAS | July 14, 2016 | vol. XXX | no. XX | 1–6

non-Gaussian statistics but is not robust in the difficult regime of high-quality (i.e., small error) infrequent observations.

In this paper we introduce a new class of particle filters which is robust in the difficult regime of high-quality infrequent sparse observations capturing non-Gaussian features of the true signal. In our new particle filter, each observation at different spatial locations affects only the neighbor state variables through the clustering of the state variables, which implements coarse-grained localization. The clustering is based on the observation network and the data assimilation of the whole space is divided into smaller size data assimilation problems in each cluster which do not influence each other. In addition to the clustering, our method is different from the localized particle filter in that particles are adjusted so that the prior particles are sufficiently close to the observation when the observation is not in the span of the predicted observations by prior particles. Similar issues are discussed in (6) and Chapter 15 of (13). Our method is different from these methods in that we use a stable method similar to the ensemble square root filter (2) instead of using an ad hoc nudging technique of (6) or setting the posterior to be identically the observation likelihood (13).

Standard particle filter and localization

Throughout this paper, we consider the data assimilation of the true signal $\mathbf{x} = (x_1, x_2, ..., x_{N_{state}}) \in \mathbf{R}^{N_{state}}$ at discrete times $n\Delta T, n \in \mathbf{N}$ whose evolution is given by

$$\mathbf{x}^{n+1} = \psi(\mathbf{x}^n) \qquad [1]$$

for a nonlinear evolution operator ψ using observations $\mathbf{y} = \mathbf{y} = \{y_1, y_2, ..., y_{N_{obs}}\} \in \mathbb{R}^{N_{obs}}$ available at different physical locations. We assume a linear observation operator $H : \mathbb{R}^{N_{state}} \to \mathbb{R}^{N_{obs}}$

$$\mathbf{y} = H\mathbf{x} + \xi \qquad [2]$$

which observes N_{obs} uniformly distributed components of $\mathbf{x}^n$ perturbed by an observation error $\xi \in \mathbb{R}^{N_{obs}}$ which is independent at different spatiotemporal locations with mean zero and variance r_o.

Using K particles $\{\mathbf{x}_k, k = 1, 2, ..., K\}$ and scalar weights $\{\omega_k \geq 0, k = 1, 2, ..., K\}$, the standard particle filter represents a probability distribution of the state $\mathbf{x}$ of the form

$$p^f(\mathbf{x}) = \sum_{k=1}^{K} \omega_k^f \delta(\mathbf{x} - \mathbf{x}_k) \qquad [3]$$

where $\delta(\mathbf{x} - \mathbf{x}_0) = 1$ if $\mathbf{x} = \mathbf{x}_0$ and zero otherwise. The localized particle filter (19) implements localization which tapers the sampled covariances using vector-valued particle weights $\Omega_k = (\omega_{1,k}, \omega_{2,k}, ..., \omega_{N_{state},k})$ whose component $\omega_{i,k}$ is the weight for the corresponding state variable x_i. Using a correlation function $l(y_j, x_i, r)$ which has 1 for $y_j = x_i$ and decreases to zero as the distance between y_j and x_i increases and assuming that the prior weights are equally distributed, the analysis vector-valued particle weight for the k-th particle is given by

$$\omega_{i,k} = (p(y_j|x_{i,k}) - 1) \, l(y_j, x_i, r) + 1. \qquad [4]$$

where r is the localization radius which determines the localization length scale. Note that when $l = 1$, the analysis particle
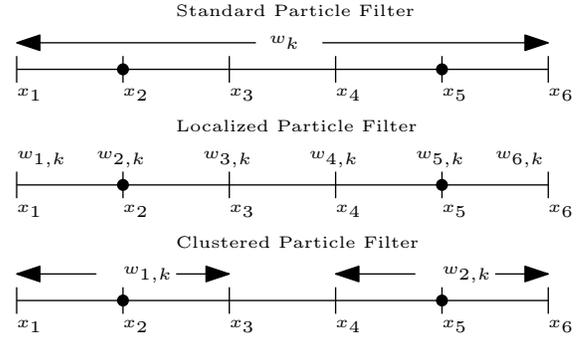
Fig. 1. Schematics of particle weight, w_k, for the k-th particle. Total dimension is 6 and there are two observations at x_2 and x_5 which yields two clusters in the clustered particle filter. The standard particle filter uses the same particle weight at different locations while the localized particle filter uses different weights at different locations. The clustered particle have different weights in different clusters but the weights are the same in the same cluster.

weight is fully influenced by the observation with the value $p(y_j|x_{i,j})$ while it remains unchanged when $l = 0$. Despite the successful applications of the localized particle filter for several interesting test regimes including non-Gaussian features (19), the method is not robust when the observation is infrequent and high-quality with a small observation error (see Fig. 2 for the performance of the localized particle filter applied to a standard data assimilation problem with high-quality observations). In the next section, we introduce a new particle filtering with localization which is robust for high-quality infrequent observations.

Clustered particle filter

We now describe our proposed particle filtering method, clustered particle filter (CPF), which utilizes the clustering of the state variables and particle adjustment. One of the key differences between the localized particle filter and the clustered particle filter is the particle adjustment which updates the prior particle values closer to the observation instead of reweighing the particles when the prior is too far from the observation likelihood. In the case of sparse observations, not all state variables are observed and thus adjacent state variables of the observed variable must have the same particle weights for cross covariances. For this purpose, the clustered particle filter partitions the state variables into non-overlapping clusters $\{C_l, l = 1, 2, ..., N_{obs}\}$ where each cluster is centered at each observation point. This yields N_{obs} clusters corresponding to N_{obs} observation locations. Instead of using different weights at different locations as in the localized particle filter, the clustered particle filter uses scalar particle weights $\{\omega_{l,k}\}$ for the state variables in the same cluster C_l (See Fig. 1 for the schematics of particle weights of the standard, localized, and clustered particle filers for a six-dimensional system with two observations at x_2 and x_5).

For the sub-state vector $\mathbf{x}_{C_l} = \{x_i|x_i \in C_l\} \in \mathbb{R}^{N_{state}/N_{obs}}$ corresponding to cluster C_l, the clustered particle filter considers the marginalized probability distribution function (PDF)

$$p(\mathbf{x}_{C_l}) = \sum_{k}^{K} \omega_{l,k}\delta(\mathbf{x}_{C_l} - \mathbf{x}_{C_l,k}) \qquad [5]$$

and each observation y_j updates only the marginalized PDF of the corresponding cluster which implements coarse-grained

localization. That is, for forecast particle weights $\{\omega_{l,k}^f\}$ of each cluster C_l, the analysis particle weights $\{\omega_{l,k}^a\}$ are given by

$$\omega_{l,k}^a = \begin{cases} \dfrac{\omega_{l,k}^f p(y_j|\mathbf{x}_k)}{\sum_m^K \omega_{l,m}^f p(y_j|\mathbf{x}_m)} & l = j, \\ \omega_{l,k}^f & l \neq j. \end{cases} \quad [6]$$

Thus the assimilation of the full state vector is decomposed into N_{obs} independent assimilation problems for each cluster of a dimension $\frac{N_{state}}{N_{obs}}$ smaller than the full state dimension N_{state}. Note that in contrast to the localization using a smoothly varying correlation function with a localization radius parameter, the clustered particle filter has no adjustable parameter to tune localization.

Particle adjustment. One of the central issues of particle filtering which collapses the particles is when the observations are not in the span of the predicted observations by prior particles $\{H\mathbf{x}_k^f\}$ which is highly possible for high-quality (i.e., small error) observations. This issue is discussed in Chapter 15 of (13) under the name of 'Dynamic Range A issue' which implies that the prior distribution is in the tail of the observation likelihood. In (13), their remedy for this issue is to simply match the posterior distribution to the observation likelihood function. Also a similar issue is addressed in (6) using an ad hoc nudging technique to constrain the prior particles to be sufficiently close to the observations. Our proposed method, CPF, is different from these approaches in that the particles are adjusted in a stable way similar to the ensemble square root filters (2).

As the observation error ξ is independent and identically distributed at different different locations, the clustered particle filter uses a sequential update of the observations $(y_1, y_2, ..., y_{N_{obs}})$. In the assimilation of observation y_j, the forecast particles in the cluster C_j, $\{\mathbf{x}_{C_j,k}^f\}$ with weights $\{\omega_{j,k}\}$, are adjusted by an adjustment matrix A

$$\mathbf{x}_{C_j,k}^a = \mathbf{x}_{C_j}^a + A(\mathbf{x}_{C_j,k}^f - \mathbf{x}_{C_j}^f) \quad [7]$$

so that the mean and covariance of the adjusted particles satisfy the Kamlan analysis mean $\mathbf{x}_{C_j}^a$ and covaraince $R_{C_j}^a$ which are given as

$$\mathbf{x}_{C_j}^a = \mathbf{x}_{C_j}^f + G(y_j - H\mathbf{x}_{C_j}^f) \quad [8]$$

and

$$R_{C_j}^a = (I - GH)R_{C_j}^f \quad [9]$$

respectively. Here G is the Kalman gain matrix $G = R^f H^T (HR^f H^T + r_o I)^{-1}$, $\mathbf{x}_{C_j}^f$ is the forecast mean $\mathbf{x}_{C_j}^f = \sum_k^K \omega_{j,k} \mathbf{x}_{C_j,k}^f$ and $R_{C_j}^f$ is the forecast covariance $\sum_k^K \omega_{j,k} (\mathbf{x}_{C_j,k}^f - \mathbf{x}_{C_j}^f)(\mathbf{x}_{C_j,k}^f - \mathbf{x}_{C_j}^f)^T$ (See SI for the method to find the adjustment matrix A). The particle adjustment helps to avoid the particle collapse. However the particle filter still can have particle collapse due to many other factors such as insufficient instability in forecast models and sampling errors. Thus as in the standard particle filter, the clustered particle filter uses resampling (8) which discard low weight particles and duplicate large weight particles when the effective particle number $K_{eff} = \frac{1}{\sum_k (\omega_k^a)^2}$ drops below a threshold value. In our study, we trigger resampling with a threshold value $K/2$.

Now we summarize the hard threshold version clustered particle filter which triggers the particle adjustment if the following condition is satisfied

Hard threshold criterion for particle adjustment :

$$y_j \notin \text{Range of predicted observations}$$
$$= \text{Range}\{H\mathbf{x}_{C_j,k}^f | k = 1, 2, ..., K\} \quad [10]$$

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

   Given :

1) N_{obs} observations $\{y_1, y_2, ..., y_{N_{obs}}\}$
2) prior K particles $\{\mathbf{x}_{C_j,k}^f, k = 1, 2, ..., K\}$ and weight vectors $\{\omega_{l,k}^f, k = 1, 2, ..., K\}$ for each cluster $C_l, l = 1, 2, .., N_{obs}$

    For y_j from $j = 1$ to N_{obs}

      If The hard threshold criterion Eq. (10) is satisfied

        Update the prior particles using Eq. (7) to match the Kalman update Eq. (8) and Eq. (9)

      Else Use particle filtering

        Update $\{\omega_{j,k}^f\}$ using Eq. (6)

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

        Do resampling

        Add additional noise to the resampled particles

$$\mathbf{x}_{C_l,Resample(k)} \leftarrow \mathbf{x}_{C_l,Resample(k)} + \epsilon \quad [11]$$

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

      End If

      End If

    End For

Note that, in addition to the hard threshold criterion, other criteria can be used to trigger the particle adjustment utilizing additional information such as innovation statistics (see SI for a soft threshold version clustered particle filter using innovation statistics). Therefore there can be many other variants of the clustered particle filter incorporating a different criterion for the particle adjustment and other factors such as several resampling strategies and inflation techniques. Throughout this study, we use the hard threshold clustered particle filter with residual resampling (8) which shows robust and accurate filtering skills in our tests including the 40-dimensional Lorenz 96 (20) model and the MMT wave turbulence model (21, 22).

Numerical Tests

We test the filter performance of the clustered particle filter and compare the result with other methods, the localized particle filter and an ensemble based method EAKF (23). The methods are applied to the following 40-dimensional Lorenz 96 model (20)

$$\frac{du_i}{dt} = (u_{i+1} - u_{i-2})u_{i-1} - u_i + F, \quad i = 1, 2, ..., J = 40. \quad [12]$$

in a periodic domain. This system is a popular model for turbulent systems mimicking baroclinic turbulence in the mid-latitude atmosphere with energy conserving nonlinear advection and dissipation (13, 20). As the external constant force F varies, this model generates a wide variety of dynamical regimes ranging from weakly chaotic to strongly chaotic to strongly turbulent (12, 13). In our numerical experiments, we test the filter performance in the difficult regime using high-quality infrequent observations. The observations are available at every time interval 0.2 with an observation error variance 0.05 which is about or less than 1% of the true signal
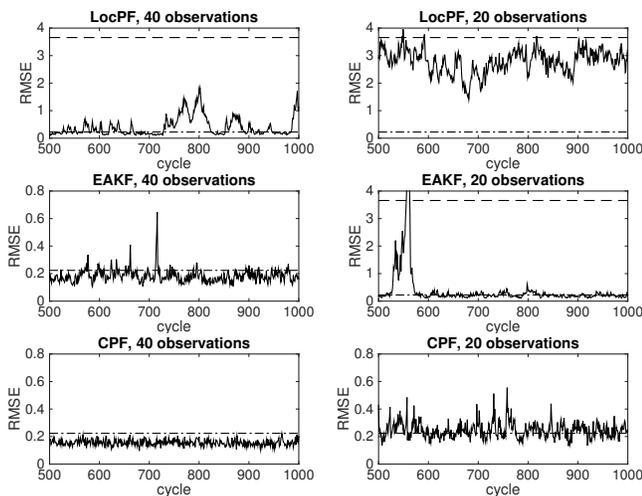
Fig. 2. Lorenz 96 F=8. Time series of RMS errors of the localized particle filter (LocPF), EAKF and the clustered particle filter (CPF). 40 observations and observation error variance 0.05. Dash-dot line is the observation error 0.22 and dash line is the climatological error 3.64. 50 particle and ensemble members for the particle filter and EAKF respectively.

Fig. 3. Lorenz 96 F=5. Time series of RMS errors of EAKF and the clustered particle filter (CPF). Observation error variance 0.05. Dash-dot line is the observation error 0.22. 50 particle and ensemble members for the particle filter and EAKF respectively.

variance. The forecast model is the perfect model with a time step 10^{-3} using a fourth-order Runge-Kutta time integration method and thus each assimilation cycles requires 200 time integrations of the forecast model. Both EAKF and the clustered particle filter use 50 ensemble members and particles. EAKF uses a localization radius 3 so that each observation updates four adjacent state variables.

Lorenz 96 with $F = 8$, standard test regime. The Lorenz 96 with $F = 8$ is a standard dynamical regime to test data assimilation methods with strongly chaotic features (12, 19). Fig. 2 shows the time series of the RMS errors of the three methods applied to the Lorenz 96 model with $F = 8$ with 40 plentiful and 20 sparse observations. The localized particle filter with 40 observations has RMS errors larger than the observation error (dash-dot line) and fails to achieve meaningful filtering skill; if observations become sparser the performance degrades and RMS errors are comparable to the climatological error (dash line). EAKF uses no covariance inflation and shows better results than the localized particle filter. However the method suffers from local bursts of RMS errors which becomes larger than the climatological error with 20 observations. Our proposed method, CPF, which uses localization and particle adjustment, has robust filter performance with RMS errors comparable to the observation error without significant local bursts in the errors. When the observation becomes very sparse with 10 observations, the soft threshold version clustered particle filter which uses innovation statistics to trigger particle adjustment has a better performance than the hard threshold version method. See SI for the results of both the hard and soft versions with 10 observations. Note that the performance of EAKF can be improved by tuning localization (See Fig. S1 in SI for the result of EAKF with tuned parameters). However the clustered particle filter has robust performance without tuning the localization.

Lorenz 96 with $F = 5$, strongly non-Gaussian test regime. The 40-dimensional Lorenz 96 with $F = 8$ is a standard
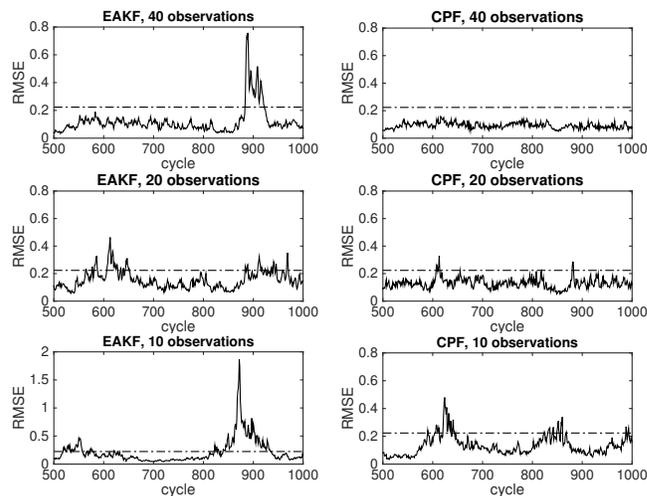
test case for data assimilation methods but has weakly non-Gaussian statistics (12). To compare the filter performance in capturing non-Gaussian features, we compare the results of EAKF and the clustered particle filter applied to the Lorenz 96 with $F = 5$ with strongly non-Gaussian statistics (11, 12) (see SI for the results of the localized particle filter which does not show accurate filtering skill). Fig. 3 shows the time series of RMS errors of EAKF and the clustered particle filter using 40 (top), 20 (middle) and 10 (bottom) observations. EAKF has RMS errors comparable to the clustered particle filter. However EAKF has local bursts of large errors. As the observation becomes sparse, the clustered particle filter also has local bursts of errors but their magnitudes are much less than ones of EAKF. Note that, as in the case with $F = 8$, the RMS errors of EAKF can be reduced by tuning parameters. See Fig. S4 in SI for the tuned results of EAKF.

The accuracy of the filtering methods relies on the forecast models' skill to obtain the right statistics of the true signal. We now check the filter performance of the two methods in capturing the non-Gaussian features of the true signal. Due to the ergodicity of the Lorenz 96 system Eq. (12), the PDFs of the model can be obtained by running an ensemble of the model in statistical steady state for a long time. The PDFs for the physical space variable x and the two most energetic 7th and 8th Fourier modes are shown in the first row of Fig. 4. The PDF of x is far from a Gaussian distribution and the ones of the absolute values of $\hat{x}_7$ and $\hat{x}_8$ are far from Rayleigh distributions, which show strong non-Gaussian behaviors of the true signal. The other two lines, which are the PDFs of the clustered particle filter (real) and EAKF (dash-dot) with 20% covariance inflation both using 10 observations, show a significant performance difference between the two methods. The forecast PDF of the clustered particle filter is on top of the true signal while EAKF has poor performance in capturing the non-Gaussian features; as EAKF uses a Gaussian assumption for assimilation, the forecast PDFs of EAKF look very similar to the Gaussian distribution for x and Rayleigh distributions for $|\hat{x}_7|$ and $|\hat{x}_8|$. The bottom row of Fig. 4 shows the forecast error PDFs of the state variable and the real parts of the 7th and 8th Fourier modes. In comparison
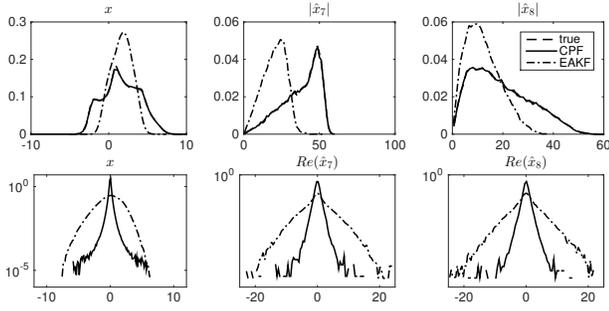
Fig. 4. Lorenz 96 F=5. Forecast error PDFs (top row) and forecast PDFs (bottom row) of x and the two most energetic modes $\hat{x}_7$ and $\hat{x}_8$. EAKF uses 20% covariance inflation.
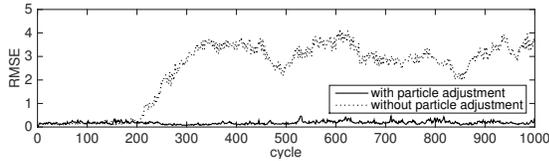
Fig. 5. Lorenz 96 F=5. Time series of RMS errors of the clustered particle with and without particle adjustment. 10 observations and observation error variance is 0.05.

with EAKF, the clustered particle filter has much tighter error bounds than EAKF which implies better forecast estimation skill in predicting the true signal.

In our experiments, the particle adjustment is triggered less than 30% on average and at most 70% in assimilating clusters at each observation time. However the particle adjustment plays a crucial role in stabilizing the filter. Fig. 5 shows two time series of RMS errors of the clustered particle filter with and without particle adjustment. At the beginning the clustered particle filter without particle adjustment shows stable filter performance. However, at the 200th cycle, its performance quickly degrades and the filter does not show skillful performance. On the other hand, the clustered particle filter with particle adjustment shows robust and stable results throughout the long assimilation cycles.

Application to multiscale data assimilation

Tremendously large computational costs to run forecast models for high-dimensional complex dynamical systems limit the number of samples, which degrades the filter performance. The multiscale data assimilation method (24) aims at increasing the sample numbers using computationally cheap but robust reduced-order coarse grid forecast models and provides the resolved large scale estimation and prediction using mixed observations of the large and small scales. The reduced-order forecast model resolves the large-scale variables while the small-scale variables are approximated using conditional Gaussian distributions which can represent non-Gaussian statistics through the interaction between the large and small scales. The multiscale filtering also provides a mathematical framework for representation errors, the errors due to the contribution of unresolved scales (25, 26) in the observation. Under this framework, the multiscale data assimilation method has been applied for an ensemble based method and shown successful applications in several stringent multiscale problems including wave turbulence (27) and baroclinic turbulence (28). In addi-

tion to the ensemble based approach, the multiscale particle filter in (24) uses particle filtering for the resolved coarse large scales while the unresolved small scales are updated using the Kalman update which has significantly low computational costs compared to the standard particle filter. Here we use the clustered particle filter algorithm, CPF, for multiscale particle filtering of high-dimensional resolved large-scale variables without the Gaussian assumption and test it for a one-dimensional wave turbulence model. As the reduced-order forecast method has model errors, the experiment in the next section also serves as a data assimilation test for the clustered particle filter with model errors.

Multiscale filtering of wave turbulence, test with model errors.
We apply the multiscale cluster particle filter for the MMT model introduced in (21, 22) as a computationally tractable model of wave turbulence. The model is described by the following one-dimensional PDE for a complex scalar ψ

$$i\partial_t \psi = |\partial_x|^{1/2}\psi - |\psi|^2\psi + iF + iD\psi \qquad [13]$$

in a periodic domain of length L with large-scale forcing set to $F = 0.0163\sin(4\pi x/L)$ and dissipation D for both the large and small scales. It has several features of wave turbulence which make it a difficult test problem for data assimilation. The model has a shallow energy spectrum proportional to $k^{-5/6}$ for wavenumber k and inverse cascade of energy from small to large scales. It also has non-Gaussian extreme event statistics caused by intermittent instability and breaking of solitons. As the unresolved small scales carry more than two-thirds of the total variance, it is a difficult filtering problem to estimate and predict the resolved large scales using mixed observations of the large-and small-scale components.

Here we compare the filtering results of the ensemble based multiscale data assimilation method (27) and the multiscale clustered particle filter for the MMT model. As the forecast model for both the methods, we use the stochastic superparameterization multiscale method (29, 30) which is a seamless multiscale coarse-grid method using conditional Gaussian statistics for the unresolved small scales. The forecast model uses only 128 grid points while the full resolution uses 8192 grid points, which yields about 250 times cheaper computational savings (considering savings in the time step). As the forecast model has a low computational cost compared to the full resolution signal, the forecast model has significant model errors. Observations of the full-scale variables are available at uniformly distributed 64 grid points (which are extremely sparse compared to the full resolution 8192 grid points) with an observation error variance corresponding to 3% of the total variance at every time interval of 0.25. The ensemble based method uses the tuned parameters in (27), i.e., a short localization radius 1 and 2% covariance inflation. For the hard threshold version clustered particle filter, the particle adjustment is triggered if either real or imaginary parts are not in the range of the predicted observation as we observes both the real and imaginary parts of the true signal. Both methods use 129 samples.

Time series of the large-scale estimation RMS errors of the ensemble based filter and clustered multiscale particle filter are shown in Fig. 6. The dash-dot line is the effective observation error 0.34 which is defined as $\sqrt{\text{observation error variance} + \text{small-scale variance}}$ by treating the small-scale contribution as an additional error, i.e.
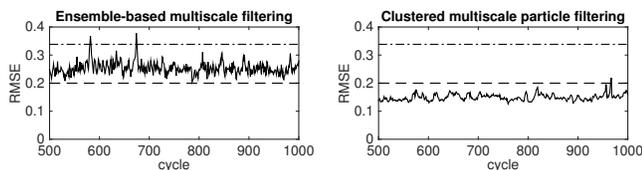
Fig. 6. MMT. Time series of the large-scale estimation RMS errors of EAKF and clustered particle filter. 64 observations. Dash line is the climatological error 0.20. Dash-dot line is the effective observation error $\sqrt{\text{observation error variance} + \text{small-scale variance}} = 0.34$
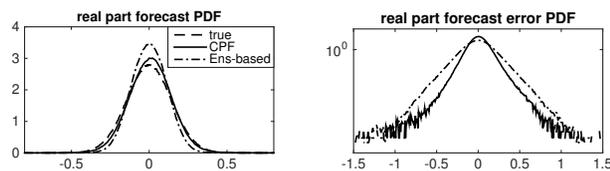
Fig. 7. MMT. Large-scale real part forecast PDF (left) and forecast error PDF (right) using 64 observations.

the representation error (25, 26). The dash line is the climatological error 0.20 which is the standard deviation of the large-scale variables. The ensemble based method has RMS errors smaller than the effective observation error but larger than the climatological error. The clustered particle filter, on the other hand, shows skillfull filter performance with RMS errors smaller than the climatological error. The forecast PDFs and forecast error PDFs of the real part which show the prediction skill of the method are shown in Fig. 7 (see Fig. S5 in SI for the result of the imaginary part). The clustered particle filter has a better forecast PDF fit to the true signal and a narrower peak in the forecast error PDF than the ensemble-based method.

Concluding Discussions

The clustered particle filter introduced in this paper shows robust state estimation and prediction skill for high-dimensional systems using high-quality sparse observations and relatively few particles. In addition to its accurate filter results measured by RMS errors, the clustered particle filter captures the non-Gaussian features of the true signal while the ensemble based EAKF has incorrect Gaussian features. The clustered particle filter is also extended to multiscale particle filtering with model errors. Using a cheap reduced-order forecast model with a model error, our method is successfully applied to the stringent one-dimensional wave turbulence MMT model. The clustered particle filter uses clustering of the state variables based on the observation network to implement coarse-grained localization. It is also shown that the particle adjustment is an essential step to stabilize the filter and achieve accurate estimation and prediction skill in addition to localization. This suggests that it is an interesting research topic to investigate various criteria to trigger the particle adjustment which can affect the performance of the clustered particle filter and combine the clustered particle filter with several covariance inflation techniques (31, 32) to account for variance underestimation and sampling errors and stabilize the filter. Here we have considered only one-dimensional uniformly distributed linear observations. It would be natural to speculate if the clustered particle filter can be extended to two- or three-dimensional non-uniformly distributed and nonlinear observations for operational applications.

1. Evensen G (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans* 99(C5):10143–10162.
2. Tippett MK, Anderson JL, Bishop CH, Hamill TM, Whitaker JS (2003) Ensemble square root filters. *Monthly Weather Review* 131(7):1485–1490.
3. Kalnay E (2003) *Atmospheric modeling, data assimilation, and predictability.* (Cambridge University Press).
4. Buehner M (2005) Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational nwp setting. *Quarterly Journal of the Royal Meteorological Society* 131(607):1013–1043.
5. Doucet A, de Freitas N, Gordon N (2001) An introduction to sequential monte carlo methods. sequential monte carlo methods in practice in *Sequential Monte Carlo Methods in Practice.* (Springer Verlag), pp. 3–14.
6. van Leeuwen P (2009) Particle filtering in geophysical systems. *Monthly Weather Review* 137:4089–4114.
7. Bain A, Crisan D (2009) *Fundamentals of Stochastic Filtering*, Stochastic Modeling and Applied Probability. (Springer, New York) Vol. 60.
8. Moral PD, Jacod J (2001) Interacting particle filtering with discrete observations in *Sequential Monte Carlo Methods in Practice.* (Springer Verlag), pp. 43–75.
9. Bengtsson T, Bickel P, Li B (2008) *Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems*, Collections, eds. Nolan D, Speed T. (Institute of Mathematical Statistics) Vol. Volume 2, pp. 316–334.
10. Snyder C, Bengtsson T, Bickel P, Anderson J (2008) Obstacles to high-dimensional particle filtering. *Monthly Weather Review* 136:4629–4640.
11. Majda AJ, Qi D, Sapsis TP (2014) Blended particle filters for large-dimensional chaotic dynamical systems. *Proc National Acad Sci USA* 111(21):7511–7516.
12. Qi D, Majda AJ (2015) Blended particle methods with adaptive subspaces for filtering turbulent dynamical systems. *Physica D: Nonlinear Phenomena* 298–299:21 – 41.
13. Majda AJ, Harlim J (2012) *Filtering complex turbulent systems.* (Cambridge University Press).
14. Reich S (2013) A nonparametric ensemble transform method for bayesian inference. *SIAM Journal on Scientific Computing* 35(4):A2013–A2024.
15. Chustagulprom N, Reich S, Reinhardt M (2016) A hybrid ensemble transform particle filter for nonlinear and spatially extended dynamical systems. *SIAM/ASA Journal on Uncertainty Quantification* 4(1):592–608.
16. Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble kalman filter technique. *Monthly Weather Review* 126(3):796–811.
17. Hamill TM, Whitaker JS, Snyder C (2001) Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review* 129(11):2776–2790.
18. Anderson JL (2010) A non-gaussian ensemble filter update for data assimilation. *Monthly Weather Review* 138(11):4186–4198.
19. Poterjoy J (2016) A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review* 144:59–76.
20. Lorenz E (1996) Predictability: A problem partly solved. *Proc Semin Predict* 1(1):1–18.
21. Majda AJ, McLaughlin D, Tabak E (1997) A one-dimensional model for dispersive wave turbulence. *J. Nonlinear Sci.* 7:9–44.
22. Cai D, Majda AJ, McLaughlin DW, Tabak EG (2001) Dispersive wave turbulence in one dimension. *Physica D: Nonlinear Phenomena* 152–153:551 – 572. Advances in Nonlinear Mathematics and Science: A Special Issue to Honor Vladimir Zakharov.
23. Anderson JL (2001) An ensemble adjustment kalman filter for data assimilation. *Monthly Weather Review* 129(12):2884–2903.
24. Lee Y, Majda AJ (2015) Multiscale methods for data assimilation in turbulent systems. *Multiscale Modeling and Simulatio* 13(2):691–713.
25. Daley R (1993) Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophys.* 11(7):634–647.
26. Janjic T, Cohn SE (2006) Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Monthly Weather Review* 134(10):2900–2915.
27. Grooms I, Lee Y, Majda AJ (2014) Ensemble kalman filters for dynamical systems with unresolved turbulence. *Journal of Computational Physics* 273(15):435–452.
28. Lee Y, Majda AJ, Qi D (2016) Stochastic superparameterization and multiscale filtering of turbulent tracers, preprint. submitted to Multiscale Modeling and Simulations.
29. Majda AJ, Grooms I (2014) New perspectives on superparameterization for geophysical turbulence. *Journal of Computational Physics* 271:60 – 77. Frontiers in Computational Physics Modeling the Earth System.
30. Grooms I, Majda AJ (2012) Stochastic superparameterization in a one-dimensional model for wave turbulence. *Communications in Mathematical Sciences* 12(3).
31. Tong X, Majda AJ, Kelly D (2016) Nonlinear stability of the ensemble kalman filter with adaptive covariance inflation. *Commun. Math. Sci.* 14(5):1283–1313.
32. Luo X, Hoteit I (2014) Ensemble kalman filtering with residual nudging: an extension to the state estimation problems with nonlinear observations. *Monthly Weather Review* 142:3696–3712.