A Gaussian Resampling Particle Filter

By X. Xiong[1] and I. M. Navon[1*]
[1] *School of Computational Science and Department of Mathematics, Florida State University, Tallahassee, FL 32306, USA*

14 July 2005

ABSTRACT

Particle filter(PF) is a fully nonlinear filter with Bayesian conditional probability estimation, compared with the well known ensemble Kalman filter(EnKF). A Gaussian resampling(GR) method is proposed to generate the posterior analysis ensemble in an effective and efficient way. The Lorenz model is used to test the proposed method. The Gaussian resampling particle filter(GRPF) can approximate more accurately the Bayesian analysis. Moreover, it is applicable to systems with typical multimodal behavior, provided that certain prior knowledge is available about the general structure of posterior probability distribution. A simple scenario is considered to illustrate this point based on the Lorenz model attractors. The present work demonstrates that the proposed GRPF possesses good stability and accuracy and is potentially applicable to large-scale data assimilation problems.

1 INTRODUCTION

In recent years the ensemble filtering method has been the focus of increased interest in the meteorological community because it naturally fits in the ensemble prediction framework, coupled with ensemble representation of the initial atmospheric or oceanographic state conditions and ensemble forecast integration. The ensemble filter falls into the category of the stochastic approach under the general framework of Bayesian theory of posterior analysis estimation.

The ensemble Kalman Filter(EnKF) (see review by Evensen, 2003) combines ensemble sampling and integration with Kalman filtering method, providing an approximated least square estimation of underlying physical states based on Monte Carlo sampling theory. For the EnKF to be applicable to large scale atmospheric and oceanographic data assimilation problems it requires besides the usual assumption of linear dynamics and Gaussian distribution in the Kalman filter, the assumption that the dynamics are governed by only a small subset of the actual dynamical variables. This is due to the limitations of state of the art computational capability in presence of large scale characteristics of atmospheric or oceanographic state conditions. The practical difficulty is to track the subset degrees of freedom, and carry out Monte Carlo simulation with a sample size significantly smaller than the dimension of the random variables. Nevertheless the EnKF and its variants are currently under consideration for operational data assimilation implementation (see Bishop et al., 2001; Zupanski, 2005; Mitchell et al., 2002; Hunt et al., 2004, to cite but a few) with varied degrees of success.

Kalman filter estimation(KF) is a linear interpolation of the variance of the prior sample and the observation, with the posterior analysis variance minimized. KF has been shown to be equivalent to the mean or maximal mode estimation of the posterior analysis under the assumption of linearized dynamics and observations based on Bayesian's theory(see derivation by Cohn, 1997). It's well known that a particle filter(PF) generates a probability-weighted posterior sample through direct evaluation of the Bayesian's formula at each prior sample point. The evaluation does not restrict the probability distribution of the prior sample and the observation to be Gaussian.

However, the probability weights are computed based on the observation which normally has no correlation with the dynamics. Therefore the resulting weighted sample is unlikely to provide an efficient sampling of a continuous probability distribution like a standard Monte Carlo sampling. In a sequential application the estimation error increases as the filter is applied at every step. A large enough estimation error can induce a so called filter divergence or degeneracy

* Corresponding author.
e-mail: navon@csit.fsu.edu

problem, which refers to the fact that the ensemble sample diverges gradually from the true state and no longer produces a meaningful forecast. To avoid filter divergence in the PF squential Monte Carlo sampling method(Doucet et al., 2000) is used to convert the weighted sample to a Monte Carlo sample at each assimilation step of the PF. The challenge of applying the PF lies in the design of an efficient resampling method to recover an efficient Monte Carlo or quasi-Monte Carlo sample from the approximated weighted sample, which can alleviate filter divergence and stablize the the PF, even more so in operational forecast due to the limited ensemble size.

The PF is a fully nonlinear filter. Mean and variance properties of the error distribution are not directly used in the computation. Therefore the PF has potential applicability to non-Gaussian error probability distribution without linearization of dynamical and observational operators. Implementation of the PF aims to explore this advantage and obtain optimal estimation. The PF has been widely used in control applications(see tutorial by Arulampalam et al., 2002) for general state space estimation. Several authors(see Anderson and Anderson, 1999; Pham, 2001; Kim et al., 2003; van Leeuwen, 2003) have studied the PF for application in atmospheric or oceanography science. A kernel filter (Anderson and Anderson, 1999) resolves non-Gaussian behavior well with rather infrequent or large error observations, which generally constitute a challenge for the Kalman filter. Other statistical estimation technique, like parametric density function estimation and resampling(Kim et al., 2003), allows the PF to track the state transitions accurately, which the EnKF fails to do correctly(see Miller et al., 1999; Evensen and van Leeuwen, 2000). Most of the PFs are tested only in low dimensional models except a PF with a sequential importance resampling(SIR) variant(SIRPF) applied to a large-scale multilayer quasi-geostrophic model(van Leeuwen, 2003).

Although the PF showed varied degree of success in the above work, filter divergence remains a major concern in realistic application of the PF. Covariance inflation is the most common technique to stabilize the ensemble filter(Anderson and Anderson, 1999; Whitaker and Hamill, 2002). However, it can be argued that such techniques increase ensemble spread and may draw the analysis towards the observations which leads to loss of information implied by model dynamics. The inflation factor as a tuning parameter is also model and observation dependent, which can pose an extra layer of uncertainty in error sensitive filter applications, e.g. model error estimation. Other PF relies on the intrinsic smoothing capability of the model where the model noise and the nonlinear interactions among the growing modes may produce enough chaotic behavior to recover lost degrees of freedom in particle filtering(van Leeuwen, 2003).

This paper proposes an "a posteriori" Gaussian resampling(GR) method that aims to increase the stability of the PF and maintain the ensemble spread, while allowing for a potential generalization to higher-dimensional models. The rest of the paper is organized as follow: Section 2 reviews the PF method based on Bayesian analysis. The EnKF method will be used for comparison purposes. Section 3 introduces a Gaussian resampling particle filter(GRPF). Section 4 presents simulation results of an numerical test of the method using the Lorenz model comparing GRPF and EnKF. Section 5 concludes the work and discusses directions of future research effort.

2 Particle filter in Bayesian Framework

Dynamical evolution of discretized physical systems are described by

$$x_k = \mathcal{M}(x_{k-1}) + g(x_{k-1})\epsilon_{k-1}, \qquad (1)$$

where x_k represents the discretized true state of the system at time t_k, $\mathcal{M}$ is the evolution operator or propagator and $g(x_{k-1})\epsilon_{k-1}$ represents state dependent model error. For a detailed explanation of the discretization process and error term introduced please refer to Cohn (1997).

In the following the PF is derived as a sequential optimal estimator categorized as a stochastic approach to the data assimilation problem. A sequential estimator is optimal under the assumption of the decorrelation of both the observation and the model errors at different times. At any time t_{k-1}, the probability distribution of the state $p(x_{k-1})$ is approximated by a Monte Carlo sample, so called ensemble prepared for integration. The distinctive advantage of the ensemble method is that the integration can be simply applied to each individual member even in presence of a nonlinear dynamical operator $\mathcal{M}$. The ensemble after the integration to time t_k is a Monte Carlo sample of the prior probability distribution. Model error is introduced with a simple technique by perturbing each ensemble member once more with model error probability distribution, which is essentially a white noise decorrelated with the observation error. Model error estimation is not discussed in this paper. The EnKF proceeds to estimate the mean and variance from the Monte Carlo sample, then combines the observations available at t_k, updates the ensemble based on the famous Kalman formula for the posterior mean and variance estimation.

The PF does not use prior mean and variance estimation. Instead, the prior probability distribution at time t_k is approximated as

$$P^i(x) \approx \frac{1}{n}\sum_{j=1}^{n}\delta(x-\eta_j), \qquad (2)$$

where $\eta_j, j = 1, \ldots, n$ are the positions of the prior ensemble members. Using Bayesian analysis, with the observation

y the posterior conditional probability distribution can be computed as

$$P^a(x|y) = \frac{1}{N} P^i(x) P(y|x) = \frac{1}{N} P^i(x) P^o(y - \mathcal{H}(x)). \quad (3)$$

where $\mathcal{H}$ is the observation operator. $P^o(y - \mathcal{H}(x))$ is the observation probability distribution function. $N = \int dx P^i(x) P^o(y - \mathcal{H}(x))$ is the normalization constant. The posterior probability distribution can be used for mean or maximum likelihood estimation of the physical state at time t_k, or acts as an initial state probability distribution for future forecast. Combining eqn. (2) and (3), we have an estimate for the analysis probability distribution,

$$P^a(x|y) \approx \frac{1}{nN} \sum_{j=1}^{n} P^o(y - \mathcal{H}(\eta_j)) \delta(x - \eta_j) \quad (4)$$

The right hand side actually represents a probability weighted ensemble with unnormalized weight $P^o(y - \mathcal{H}(\eta_j))$ associated with each position η_j. It's well known that a weighted sample is inefficient compared with a true Monte Carlo sample in general. A resampling method is needed to locally smooth the weighted sample and recover a Monte Carlo or quasi-Monte Carlo sample. The simplest way of drawing a random sample from the η_j's based on the associated weights does not work very well. The problem is that high weighted points may be duplicated and low weighted points may be lost in the stochastic drawing. After repeating the integration and resampling for a few steps the effective ensemble size reduces and the ensemble fails to remain a valid approximation to continuous analysis probability distribution Thus, the filter would suffer from filter divergence problem because of insufficiency of local smoothing.

Kernel density estimation(Silverman, 1986) can be used to approximate the weighted discrete representation with a continuous function which can then be used for resampling(Anderson and Anderson, 1999; Pham, 2001). However, a tuning parameter is introduced to specify the kernel density function variance. Optimization and estimation of such tuning parameters may interfere with other error reduction and estimation objectives, e.g. model error estimation. Also the tuning parameter has to be small so that the assimilation does not increase the ensemble spread too much and draw the analysis towards the observations in subsequent steps. It has also been argued that the tuning parameter and the smoothing function may not needed provided that model noise and nonlinear dynamics can induce sufficient local expansion(van Leeuwen, 2003). The argument is certainly valid in some scenarios but again a filter with model independency is always more attractive.

In the following section we propose a posterior GR method that can update the ensemble and maintain PF stability with little model dependency, while allowing for a generalization to higher dimensional systems.

3 Posterior Gaussian resampling

The filter divergence problem is caused by the deviation of the whole ensemble sample too far away from the true state. If a Bayesian theoretical analysis (4) can be satisfactorily estimated then the filtering process can be carried out for a long run. The success of the Kalman filtering method indicates that mean and variance properties are often sufficient to define the error statistics for data assimilation purposes. The simplest way is to generate the updated ensemble with estimated mean and variance computed from the distribution (4), i.e, find $\xi_i, i = 1, \cdots, n$ so that

$$\bar{\xi}^a = \sum_j f_j \eta_j^a \quad (5)$$

$$\Sigma_\xi^{ab} = \sum_j f_j \eta_j^a \eta_j^b - \bar{\xi}^a \bar{\xi}^b, \quad (6)$$

where a, b are state space indices and j is the ensemble member indices. $\bar{\xi}$ and Σ_ξ represent the mean and the variance of η_i's. In particular Eqn. (6) can be considered as a nonlinear generalization of the well known Kalman filter analysis covariance. In higher order realistic systems, state space dimension may be much larger than the size of the ensemble. f_j's are the normalized weights,

$$f_j = \frac{P^o(y - \mathcal{H}(\eta_j))}{\sum_j P^o(y - \mathcal{H}(\eta_j))}. \quad (7)$$

In practice a large portion of f_j's are small enough to be ignored, and only a subset of the ensemble members are summed over in Eqn. (5) and (6).

To construct an ensemble update formula, first rewrite Eqn. (6) in matrix form with η,

$$\Sigma_\xi = \eta M \eta^T , \quad (8)$$

$\eta = [\eta_1, \eta_2, \cdots, \eta_n]$ is a $L \times n$ matrix where L is the number of system variables and n is the number of the ensemble members. M is a symmetric matrix with elements

$$M_{jk} = f_j \delta_{jk} - f_j f_k , \quad (9)$$

which can be factorized with a singular vector decomposition method,

$$M = V \Lambda V^T . \quad (10)$$

Then

$$\Sigma_\xi = \xi' \xi'^T \quad (11)$$

with $\xi' = \eta V \Lambda^{1/2}$. The row dimension of ξ' is state space dimension. The column dimension of ξ' is m, which is smaller than the prior ensemble size n due to the rank reduction of M with some f_j's (7) being very close to zero. Now randomly generate a $m \times n$ matrix X with all elements drawn from a one dimensional Gaussian sampling with mean zero and variance 1. Construct a matrix ξ, such that

$$\xi = \xi' X + \bar{\xi} . \quad (12)$$

The sample positions specified by the columns of ξ have an estimated mean $\bar{\xi}$ and variance Σ_ξ, i.e. ,

$$\sum_j \xi_j \approx \bar{\xi} , \quad (13)$$

$$\sum_j \xi_j \xi_j^T - \bar{\xi}\bar{\xi}^T \approx \Sigma_\xi . \quad (14)$$

It can be verified with standard techniques that the estimation error is proportional to $1/n\Sigma_\xi$, which decreases as the sample size n increases. The X matrix adjusts the mean, and acts as a smoothing factor. The updated sample is also an estimation of a Gaussian distribution with desired mean and variance.

A potential problem of the above enlargement and smoothing procedure by Eq. (11) is that the produced ensemble perturbations only lie within the vector sub-space of those obtained from the truncated SVD of Eq. (8). This fact has only little consequence in a low dimensional model(like the Lorenz model) where the number of ensemble members is sizeable smaller than the number of model variables. In a geographically relevant case where the number of the ensemble members is much less than the number of variables in the model the dimension reduction may lead to significant sub-optimality and eventually to filter divergence.

The cause of the dimension reduction in the PF is that certain prior ensemble members are weighted low by assimilation of the observations. The EnKF is less likely to have the dimension reduction problem due to the linear interpolation approach. In theory the EnKF is no better than the PF from Bayesian analysis derivation(Cohn, 1997). Given finite ensemble size, both the EnKF and the PF provide only approximate estimates of the Bayesian posterior analysis probability distribution. The posterior ensemble generated by the EnKF may span more dimensions but that generated by the PF may sample the large-weighted region more sufficiently. To claim which filter generates better posterior ensemble, one really asks the question which quasi-Monte Carlo sampling technique is more efficient and leads to more accurate ensemble forecast. The question is actually related to a famous issue raised in ensemble forecast and filtering research for operational forecast. It has been argued that only the growing error directions need to be adequately sampled to produce an accurate forecast(Toth and Kalnay, 1997). Singular vector(SV) construction and breeding method are well-known techniques to sample leading growing perturbations and generate an efficient ensemble (see Toth and Kalnay, 1997; Buizza and Palmer, 1995; Houtekamer et al., 1996, to cite but a few). Since the growing errors or the SVs are purely dynamic consequences(Toth and Kalnay, 1997; Buizza and Palmer, 1995; Hamill et al., 2003), the posterior ensemble generated by either the EnKF or the PF with the observations is not an efficient Monte Carlo sampling in general. As long as the ensemble size is sufficiently large compared to the dimensions of the system attractors, both the EnKF and the PF may avoid filter divergence with the nonlinear dynamics mixing the growing modes. (van Leeuwen, 2003). If the ensemble size is too small, then filter divergence can occur in both filters, in which case one might have to fall back on the SV or the breeding techniques.

4 Numerical Experiments with the Lorenz Model

Lorenz-63(Lorenz 1963) stochastic model, described by the following Eqns. (15), is used here to test the data assimilation performance of the GRPF.

$$\begin{aligned} dx &= -\sigma(x-y)dt + gdw_1, \\ dy &= (\rho x - y - xz)dt + gdw_2, \\ dz &= (xy - \beta z)dt + gdw_3. \end{aligned} \quad (15)$$

This model has become one of the mainstays for the study of chaotic systems due to its chaotic but well understood behavior. The well known three parameters of the Lorenz model are specified as follows: $\sigma = 10.0, \rho = 28.0$ and $\beta = 8/3$, enabling the dynamics to evolve within two distinctive attractors with appropriate time step dt. Model error variance per assimilation cycle can be adjusted as the stochastic forcing coefficient g changes. The initial ensemble is obtained as the perturbation of the true state(reference solution), with a 3×3 diagonal error covariance matrix, diag$(2, 2, 2)$. The size of the ensemble is set to either 1000 or 100 in the experiments. Model error is not estimated but simulated as a Gaussian random perturbation with variance varying from 0 to 10. Two types of measurements are considered. In one setup the measurement is performed on the state variable x only. In the other the measurement on x^2. In the last case the analysis probability distribution is multimodal in general due to the combination of the Lorenz attractors and the nonlinear observation operator. Measurement data is obtained as a perturbation of the reference solution at measurement times with variance 2.

Figure 1 compares data assimilation results from the GRPF and the EnKF methods with 40s run time and 800 time steps. The observation is measured on x available every 0.25s. The model error variance is 0. The ensemble size is 1000. The ensemble mean is computed as the prediction. One of the characteristics of the performance of the filter is the number of the spikes(mispredictions) that appear in the ensemble mean curve. Both filters yield similar performance and generally produce spikes at the same time(for example after t=31s). Simular results are obtained with the model error variance up to 10 and the ensemble size 100.

A quantitative measure of the filter performance is the root mean square(rms) error of the ensemble mean prediction of the reference solution. Figure 2 plots time series of the rms error of the EnKF and the GRPF through a entire run. Table 1 shows a comparison of the ensemble mean predic-

Table 1. Mean rms error of the ensemble mean as a function of the model error variance for 1000- or 100-member EnKF and GRPF assimilations of the Lorenz-63 system with measurement error variance 2.0.

1000-member

Model error	EnKF mean rms			GRPF mean rms		
variance	x	y	z	x	y	z
0	2.16	3.49	3.49	1.69	2.71	2.87
2	2.29	3.75	3.81	2.20	3.56	3.55
4	2.40	3.87	3.73	2.15	3.46	3.28
6	3.00	4.95	4.89	2.40	3.90	3.85
8	2.67	4.40	4.17	2.33	3.85	3.21
10	3.55	5.67	5.32	2.56	4.22	4.17

100-member

Model error	EnKF mean rms			GRPF mean rms		
variance	x	y	z	x	y	z
0	2.03	3.27	3.23	1.64	2.65	2.77
2	2.34	3.84	3.87	2.22	3.60	3.68
4	2.51	4.06	3.98	2.23	3.59	3.59
6	3.09	5.15	5.02	2.26	3.79	3.68
8	2.61	4.31	4.11	3.28	5.08	4.57
10	3.46	5.75	5.54	2.95	4.85	4.67

tion rms error between the GRPF and the EnKF(Evensen, 1994) averaged over time for each state variable. The EnKF used in the experiment induces observation perturbation to avoid an underestimated analyzed covariance(Burgers et al., 1998). More recent variants use square root methods instead(Tippett et al., 2003; Evensen, 2003). The model error variance per assimilation cycle is set to vary from 0 to 10, thus producing an increasing level of noise in the dynamical integration. An interesting result obtained is that the GRPF yields a lower mean square error most of the time. With the smaller ensemble size of $n = 100$ test one can find instances(for example, with model error variance 8) where the EnKF performs better than the GRPF. In practice the performance difference between two filters should be discussed on a case by case situation. Many factors, such as the model dynamics, the observations, the ensemble size and so on., could affect the performance of an ensemble filter.

Kernel density estimation technique(Silverman, 1986) can be used for detailed investigation of the data assimilation performance in the low dimension model, which basically constructs a smooth probability function based on the Monte Carlo sample. Figure 3 illustrates the estimated probability density functions of the prior and posterior ensemble sample obtained by the kernel density estimation technique. The level curves in the figure represent the 2-D probability density with the third state variable integrated out, i.e. $\int dzP(x, y, z)$ and $\int dyP(x, y, z)$.

The prior sample is selected from the data at assimilation instant of a particular data assimilation run. With the same prior sample and the measurement value $x = -3.884$, the posterior sample probability density estimation by EnKF, GRPF and direct Bayesian calculation are shown respectively. The prior sample probability density function shows typical non-Gaussian characteristics which is expected for the highly nonlinear dynamics of the Lorenz model. The outer surrounding curve and some small outliers represent a small probability density (less than 10 percent) contour. Direct computation through Bayesian analysis formula indicates that the region with the small prior probability density could be emphasized and yields larger likelihood. Both the EnKF and the GRPF can produce good posterior Gaussian estimation with the mean consistent with the Bayesian computation.

The EnKF method is known to be able to cope with nonlinear dynamics and non-Gaussian probability density distribution. However, Kalman filter as a least square estimation cannot be easily extended when the analysis probability distribution has multiple local maxima(Bürger and Cane, 1994). The GRPF directly estimates the probability density with Bayesian computation. In principle it should then be able to track the local maxima of the probability density function. It is well known that the stochastic Lorenz model has two attractors. In the case that the observation data is only available on x^2, the posterior analysis probability distribution is multimodal with two local maxima. Such probability distribution can be approximated with a two-component Gaussian mixture model. The ensemble members need to be separated into two samples based on which local attractor they reside. In the Lorenz model they can be simply distinguished through the sign of the system variable x or y. The relative probability weight between two samples can be computed through summing up the individual weights of each ensemble member within each sample. The GRPF can be carried out for two samples separately taking into account the relative weight during the computation. The actual procedure is as follows. First prior to the analysis the individual weight of each particle, f_j in Eqn. (7), can be computed based on the likelihood and the relative weight between the two samples. Second, the two samples need to be redivided due to the fact that some ensemble members may migrate from one attractor to the other. The relative weight also needs to be computed again. Finally the posterior GR is applied on the two samples respectively, which basically smoothes the samples and produces a two-component Gaussian mixture estimation.

Figure 4 illustrates the results from the data assimilation experiment following the above procedure. The parameters of the experiment are setup similarly as before. The observation interval or the assimilation cycle decreases to 0.0125s for 40s run. The ensemble mean prediction includes a high probability value and a low probability value, which represent two local maxima of the analysis. The high probability value agrees well with the reference solution for most of the time. At around $t = 7$s there exists a small interval during which the lower probability data actually yields the right prediction. The relative weight between two samples is around 6 : 4 during this period. This illustrates that the maximal mode prediction can fail sometimes for multimodal analysis distribution. Another interesting result appears at the time after $t = 38s$. During this period all the valid ensemble members essentially evolve within one attractor only. The analysis probability distribution reduces to the normal Gaussian form which has only one local maxima. The lower probability sample can be regarded as not participating in the filtering process. The filter still provides the right prediction with maximal probability.

5 Conclusion

The GRPF yields satisfactory results when tested in the framework of a low dimension Lorenz model. Extension to large dimensional applications should be relatively straightforward for the GRPF. The most computationally expensive part involves the singular decomposition of a matrix with dimensions of the ensemble sample size. The EnKF(Evensen, 2004; Zupanski, 2005) is also subject to similar computational constraints. We intend to test the GRPF in near future in high dimensional models.

One of the problem associated with the application of the GRPF to high dimensional model is that the GR procedure, Eqn.(12), can lead to sampling errors when a smaller size (say $n = 100$) is used. A possible solution is to use the method put forward by (Evensen, 2004), in which the eigenvectors of a larger Gaussian matrix are chosen. Specifically, One can enlarge the matrix X in Eqn. (12) to a $m \times \beta n$ matrix, and then perform SVD on the product $\xi' X$ and retain only leading n singular vectors. Certain spatial correlations have to be preserved in the resampling and truncation in order to avoid spurious gravity waves(Szunyogh et al., 2005) when the GRPF is applied to global models such as a shallow water model. Initialization techniques(Zupanski et al., 2005) can also be used to impose correlations directly on the uncorrelated random fields.

It suffices to illustrate the difference between the better-known SIR methods (Anderson and Anderson, 1999; van Leeuwen, 2003; Kivman, 2003) and the GR. Both the SIR and the GR are resampling procedures applied to the posterior sample of the PF. The SIR and its variants attempt to capture non-Gaussian characterics reflected in the posterior weighted probability distribution with or without an extra reweighting to smooth out the posterior sample. While the SIR and its variants are not subject to Gaussian assumption on posterior analysis probability distribution, the price they pay is either in the form of a more complicated algorithm or a less smoothed posterior ensemble. For example, the SIR variant (van Leeuwen, 2003) replicates large-weighted members and generates small-weighted members stochastically. The posterior ensemble is not smoothed by reweighting and the small weighted members may be ignored because of low probability.

Compared to the EnKF, the SIRPF is more sensitive to filter divergence, especially with limited sample size. This is probably the reason why the SIRPF receives less popularity in large scale applications than in control applications. The GR assumes a posterior Gaussian distribution which resembles the second order exact method (Pham, 2001). Unlike the EnKF, the GRPF imposes no Gaussian assumption on the probability distribution of the prior sample and the observations. The GRPF attempts to retain as many ensemble members as possible and smoothes them out regardless of their weights. Therefore the GRPF is expected to exhibit better stability than the SIRPF. Like the EnKF, the GRPF also has a relatively straightforward and spread-preserving algorithm. The GRPF can be suited for applications in complex data assimilation scenarios where the observations are sparse and ill-estimated or the model noise is unknown and significant in which case the first two moments of the analysis error are primary forecast objectives.

Multimodal probability distribution can arise in simple systems (Li, 1991), but is considered as a less desirable feature in an operational forecast. In principle observations should be carefully selected to filter out other local maxima and leave only the prediction with maximal probability. However, due to the complexity of the operational forecast and scarcity of observations, mispredictions in operational forecast still occur. Applications of the GRPF as an efficient tool for multimodal probability distribution should be further explored.

6 Acknowledgements

The research of Prof. Navon and X. Xiong is sponsored by the National Science Foundation(NSF), Grant ATM-0327818. The authors would like to thank two anonymous reviewers for the in-depth comments which contribute to improving paper presentation.

REFERENCES

Anderson, J. L. and Anderson, S. L. 1999. A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilation and forecasts. *Mon. Wea. Rev.* **127**, 2741–2758.

Arulampalam, M. S., Maskell, S., Gordon, N. and Clapp, T. 2002. A tutorial on particle filters for online nonlinearnon-Gaussian Bayesian tracking. *IEEE Trans. Sig. Pro.* **50**, 174–188.

Bishop, C. H., Etherton, B. J. and Majumdar, S. J. 2001. Adaptive sampling with the ensemble transform Kalman filter. *Mon. Wea. Rev.* **129**,420–436.

Buizza, R. and Palmer, T.N. 1997. The Singular-Vector Structure of the Atmospheric Global Circulation. *J. Atoms. Sci.* **52**, 1434–1456.

Bürger, G. and Cane, M. A. 1994. Interactive Kalman filtering. *J. Geophys. Res. Oceans.* **99**, 8015–8031.

Burgers, G., van Leeuwen, P. J. and Evensen, G. 1998. Analysis Scheme in the Ensemble Kalman Filter. *Mon. Wea. Rev.* **126**, 1719-1724.

Cohn, S. E., 1997. An Introduction to Estimation Theory. *J. Met. Soc. Japan* **75**,257–288.

Doucet, A., Godsill, S. J. and Andrieu, C. 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statist. Comp.* **10**, 197–208.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10143–10162.

Evensen, G. and van Leeuwen, P. J. 2000. An Ensemble Kalman Smoother for Nonlinear Dynamics. *Mon. Wea. Rev.* **128**, 1852–1867.

Evensen, G., 2003. The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.* **53**, 343–367.

Evensen, G., 2004. Sampling strategies and square root analysis schemes for the EnKF. *Ocean Dyn.* **54**, 539–560.

Hamill, T. M., Snyder, C. and Whitaker, J. S. 2003. Ensemble Forecasts and the Properties of Flow-Dependent Analysis-Error Covariance Singular Vectors. *Mon. Wea. Rev.* **131**, 1741–1758.

Houtekamer, P.L., Lefaivre, L., Derome, J., Ritchie, H. and Mitchell, H.L. 1996. A System Simulation Approach to Ensemble Prediction. *Mon. Wea. Rev.* **124**, 1225–1242.

Hunt, B. R., Kalnay, E., Kostelich, J., OTT, E., PATIL, D. J., Sauer, T., Szunyogh, I., Yorke, J. A. and Zimin, A. V. 2003. Four-dimensional ensemble Kalman filtering. *Tellus* **56A**, 273–277.

Kim, S., Eyink, G., Restrepo, A. and Johnson, G. 2003. Ensemble filtering for nonlinear dynamics. *Mon. Wea. Rev.* **131**, 2586–2594.

Kivman, G. A. 2003. Sequential parameter estimation for stochastic systems. *Non. Process Geo.* **10**,253–259.

van Leeuwen, P. J., 2003. A variance-minimizing filter for large-scale applications. *Mon. Wea. Rev.* **131**,2071–2084.

Li, Y., 1991. A note on the uniqueness problem of variational adjustment approach to four-dimensional data assimilation. *J. Met. Soc. Japan* **69**,581–585.

Lorenz, E. N., 1963. Deterministic nonperiodic flow. *J. Atoms. Sci.* **20**, 130–141.

Miller, R. N., Carter, E. F. Jr. and Blue, S. T. 1999. Data assimilation into nonlinear stochastic models. *Tellus* **51A**, 167–194.

Mitchell, H. L., Houtekamer, P. L. and Pellerin, G. 2002. Ensemble size, and model-error representation in an ensemble Kalman filter. *Mon. Wea. Rev.* **130**, 2791–2808.

Pham, D. T., 2001. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Mon. Wea. Rev.* **129**,1194–1207.

Silverman, B. W. 1986. Density estimation for statistics and data analysis. Chapman and Hall, 175pp.

Szunyogh, I., Kostelich, E. J., Gyarmati, G., Patil, D. J., Hunt, B. R., Kalnay, E., Ott, E. and Yorke, J. A. 2005. Assessing a local ensemble Kalman filter: perfect model experiments with the National Centers for Environmental Prediction global model *Tellus* **57A**, 528-545.

Tippett, M. K., Anderson, J. L., Bishop, C. H., Hamill, T. M. and Whitaker, J. S. 2003. Notes and Correspondence: Ensemble Square Root Filters. *Mon. Wea. Rev.* **131**, 1485–1490.

Toth, Z. and Kalnay, E. 1997. Ensemble Forecasting at NCEP and the Breeding Method. *Mon. Wea. Rev.* **125**, 3297–3319.

Whitaker, J. S. and Hamill, T. M. 2002. Ensemble Data Assimilation without Perturbed Observations. *Mon. Wea. Rev.* **130**, 1913–1924.

Zupanski, M., 2005. Maximum Likelihood Ensemble Filter: Theoretical Aspects. *Mon. Wea. Rev.* **133**, 1710-1726.

Zupanski, M., Fletcher, S. J., Navon, I. M., Uzunoglu, B., Heikes, R. P., Randall, D. A., Ringler, T. D. and Daescu, D. 2005. Initiation of ensemble data assimilation. *Manuscript submitted to Tellus.*

Figure Captions

Fig. 1. Results of data assimilation experiments with EnKF and GRPF. System variables x, y, z, reference solution, observations, and ensemble mean prediction. 160 observations within 40s run time. Zero model error variance. Measurements on x only with observation variance 2.0. Ensemble size 1000. The performance of the two filters is comparable, producing similar number of spikes(mispredictions).

Fig. 2. Rms error time series of the EnKF and the GRPF with ensemble size 1000(upper) and 100(lower). 160 observations within 40s run time. Zero model error variance. Measurements on x only with observation variance 2.0. The rms error between the mean prediction and the reference solution is comparable for two filters.

Fig. 3. Kernel estimate of the prior and posterior probability density function. Observation value $x = -3.884$. EnKF, GRPF and Bayesian Analysis. The prior and posterior ensemble data obtained from the same run as of Fig. 1 at t = 34.5s. The probability profiles of the EnKF and the GRPF posterior ensembles show similarity.

Fig. 4. Results of data assimilation with observation operator x^2. System variables x, y, z, reference solution, high probability prediction, low probability prediction, output of data assimilation scheme. The high probability prediction yields correct forecast most of the time. At around t = 7s the low probability prediction yields correct forecast instead.

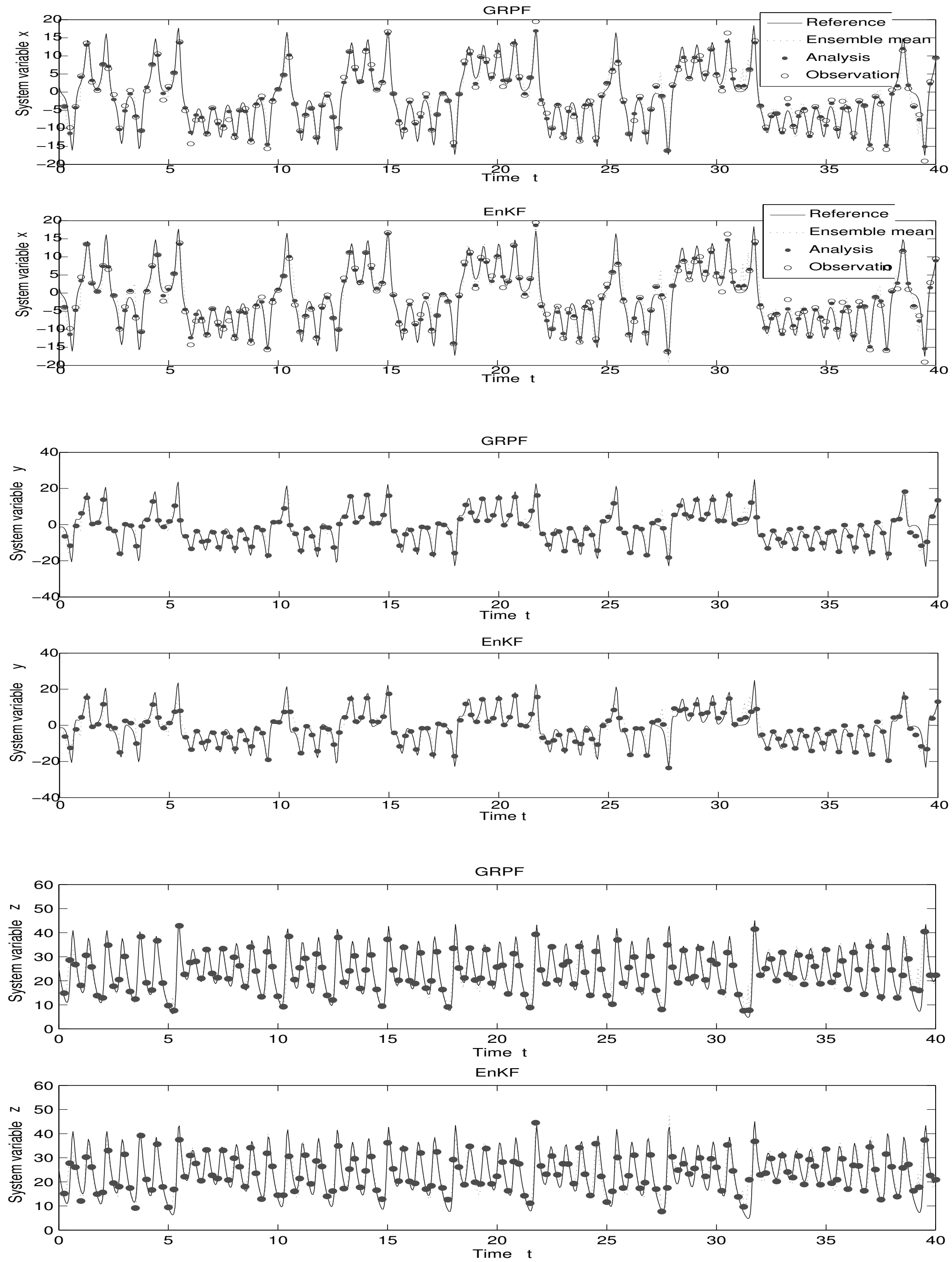

Figure 1. Results of data assimilation experiments with EnKF and GRPF. System variables x, y, z, reference solution, observations, and ensemble mean prediction. 160 observations within 40s run time. Zero model error variance. Measurements on x only with observation variance 2.0. Ensemble size 1000. The performance of two filters is comparable, producing similar number of spikes(mispredictions).

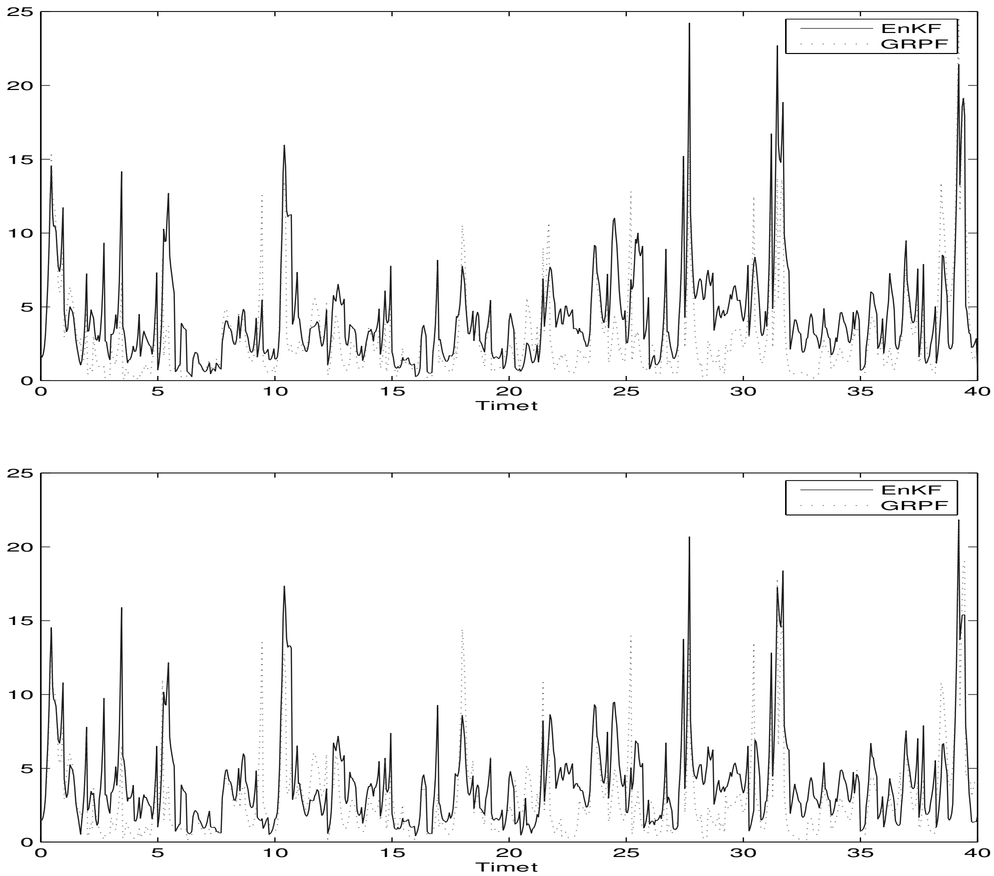

Figure 2. Rmse of the EnKF and the GRPF with ensemble size 1000(upper) and 100(lower). 160 observations within 40s run time. Zero model error variance. Measurements on x only with observation variance 2.0. The rmse between the mean prediction and the reference solution is comparable for two filters.

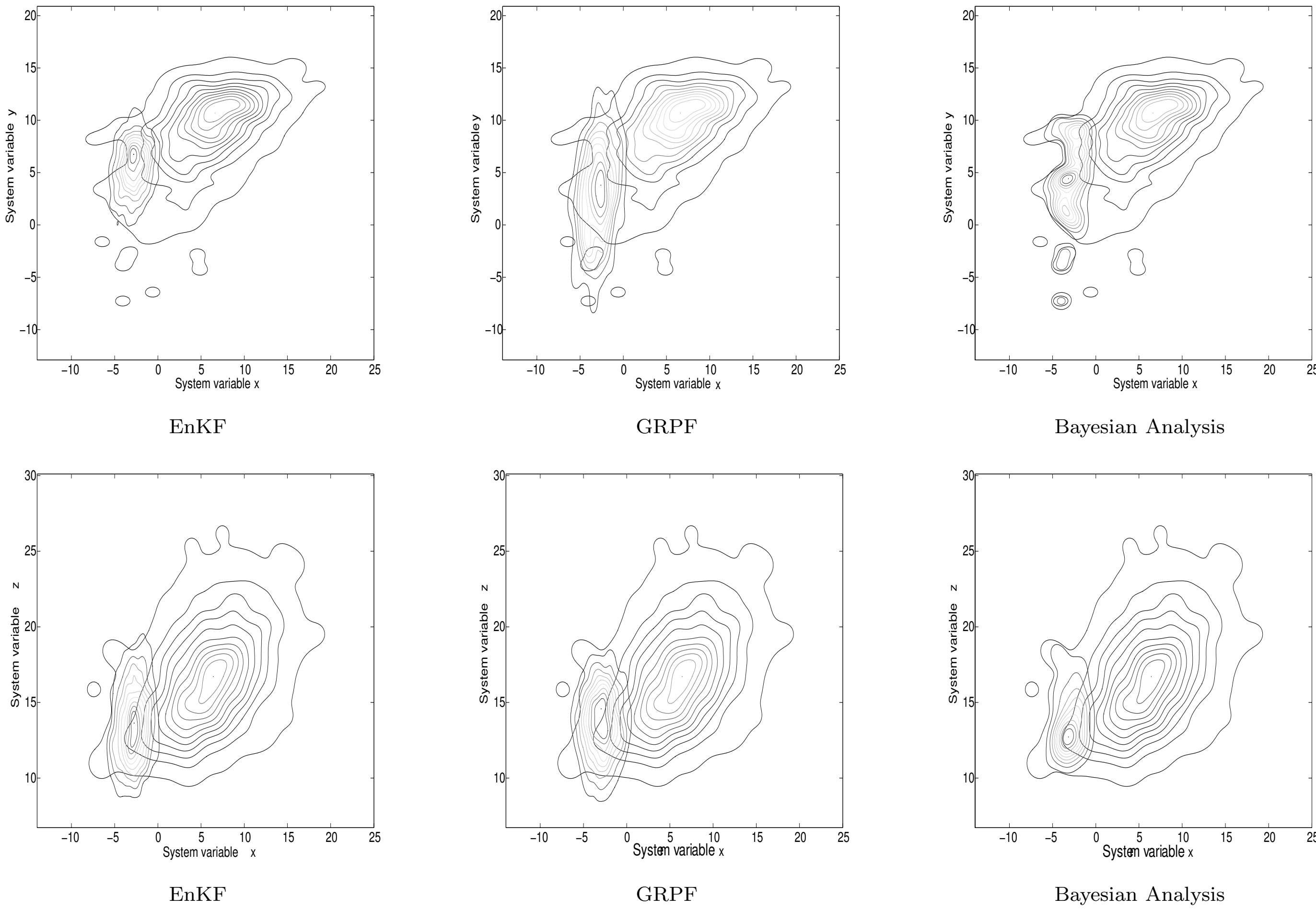

Figure 3. Kernel density estimate of the prior and posterior probability density function integrated over z direction(upper row) and y direction (lower row). Observation value $x = -3.884$. EnKF, GRPF and Bayesian Analysis. The prior and posterior ensemble data obtained from the same run as of Fig. 1 at $t = 34.5$s. The probability profiles of the EnKF and the GRPF posterior ensembles show similarity.

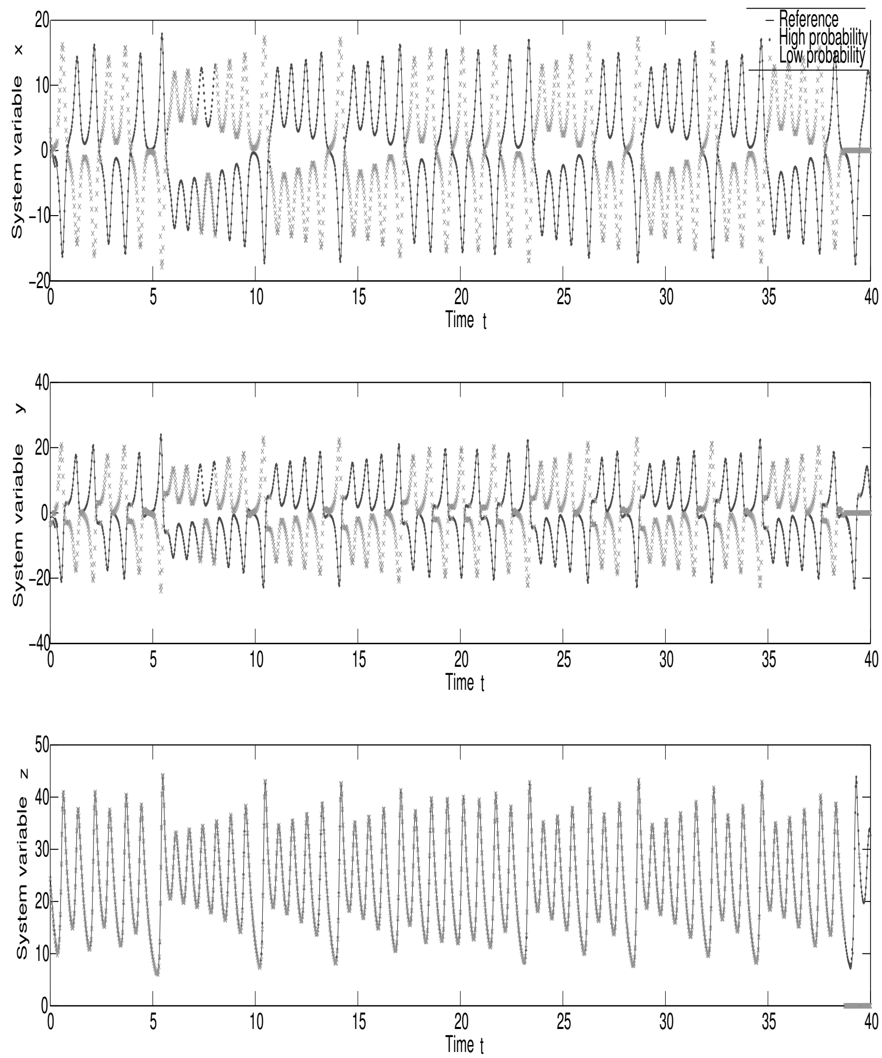

Figure 4. Results of data assimilation with observation operator x^2. System variables x, y, z, reference solution, high probability prediction, low probability prediction, output of data assimilation scheme. The high probability prediction yields correction forecast most of the time. At around $t = 7$s the low probability prediction yields correct forecast instead.