

Ensemble Particle Filter with Posterior Gaussian Resampling

By X. Xiong¹ and I. M. Navon^{1*}

¹*School of Computational Science and Department of Mathematics, Florida State University, Tallahassee, FL 32306, USA*

3 March 2005

ABSTRACT

An ensemble particle filter(EnPF) was recently developed as a fully nonlinear filter of Bayesian conditional probability estimation, along with the well known ensemble Kalman filter(EnKF). A Gaussian resampling 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. With the posterior Gaussian resampling method the EnPF 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 EnPF 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. Kalman filter estimation is equivalent to the mean or maximal mode estimation of the posterior analysis under the assumption of linearized dynamics and observations (see derivation by Cohn, 1997).

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 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 less than the dimension of the random variables. Nevertheless the EnKF and its variants are currently under consideration

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

for operational data assimilation implementation (see Bishop, 2001; Zupanski, 2004; Mitchell et al., 2002; Hunt & Kalnay et al., 2004, to cite but a few) with varied degrees of success.

An alternative filtering method, called ensemble particle filter(EnPF), has been previously investigated by several authors(see Anderson & Anderson, 1999; Pham, 2001; Kim et al., 2002; Kivman, 2003) also for ensemble data assimilation purpose. It also relies on the assumption of small ensemble subset representation of large scale systems. The particle filter, more generally known as a sequential Monte-Carlo method(Doucet et al., 2002), is a fully nonlinear filter. Mean and variance properties of the error distribution are not directly used in the computation. Therefore the particle filter has potential applicability to non-Gaussian error probability distribution even without linearization of dynamical and observational operators. Implementation of the EnPF is intended to explore this advantage when applied on nonlinear dynamical models with multi-modal behavior. A kernel filter (Anderson & 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., 2002), allows the particle filter to track the state transitions accurately, which the EnKF fails to do correctly(see Miller et al., 1999; Evensen & van Leeuwen, 1999). Also there are many well known variants of particle filtering methods(see tutorial by Arulampalam et al., 2002) applied to the general state space estimation problems. Among them the sequential importance resampling(SIR) method (Gordon et al., 1993) is the best known and can yield better performance than the EnKF and even provide a model parameter estimation(Kivman, 2003; Pham, 2001). However more research should be carried out prior to these results being claimed to be of practical value for realistic high-dimensional systems due to both stability and computational burden issues. One problem with the EnPF is the filter divergence or the so called degeneracy problem which refers to the fact that due to size reduction through filtering the ensemble sample diverges gradually from the true state and no longer produces a meaningful forecast. The problem can be understood in two ways. On one hand it indicates that the filter technique may need to be improved so that it becomes robust enough to make predictions even when observations are scarce. On the other hand the dynamical model or the error parameters which feed into the filter may need to be constructed or estimated more accurately.

This paper proposes an ‘‘a posteriori’’ resampling method that aims at increasing the stability of the EnPF and applies it to systems with typical multi-modal behavior, while allowing for a potential generalization to higher-dimensional models. The rest of the paper is organized as follow: Section 2 reviews the EnPF method based on Bayesian analysis. The EnKF method will be used for comparison purposes. Section 3 introduces the posterior Gaussian resampling method used in the analysis stage of the EnPF. Section 4 presents simulation results of an numerical test of the method using the Lorenz model comparing EnKP and EnKF. Section 5 concludes the work and discusses directions of future research effort.

2 Ensemble particle filter

Dynamical evolution of discretized physical systems are described by

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

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 EnPF 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 , update the ensemble based on the famous Kalman formula for the posterior mean and variance estimation.

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

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

where $\eta_j, j = 1, \dots, n$ are the positions of the prior ensemble members. Using Bayesian analysis, with the observation y the posterior conditional probability distribution can be computed as

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

where \mathcal{H} is the observation operator. $P^o(y - \mathcal{H}(x))$ is the observation probability distribution function. $N = \int dx P^p(x) P^o(y - \mathcal{H}(x))$ 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 weighted ensemble with unnormalized weight $P^o(y - \mathcal{H}(\eta_j))$ associated with each position η_j . To generate the Monte-Carlo sample at time t_k , one can simply draw a random sample from the η_j 's based on the associated weights. However, since the updated sample is restricted to the n fixed points, after repeating the integration and resampling for a few steps the number of distinctive points reduces and the ensemble deviates far from the true state. Thus, the filter would suffer from severe filter divergence problem, since the weighted ensemble is just an approximated discrete representation of the analysis probability distribution, which fails to remain a valid approximation as filter step increases. One remedy is to approximate the discrete representation with a continuous function which can then be used for resampling. Kernel density estimation (Silverman, 1986) replaces the delta function $\delta(x - \eta_j)$ by a smoothing kernel, e.g. a Gaussian function kernel $\phi_\Sigma(x - \eta_j)$, defined as Gaussian distribution function of x with mean η_j and variance Σ . Resampling can be done by drawing a random sample of kernel centers followed by Gaussian perturbations with variance Σ (Anderson & Anderson, 1999; Pham, 2001). Kernel density estimation is a local smooth fit of the original probability function, but increases the overall variance of the distribution by the kernel variance Σ . Kernel function with large Σ would cause inaccurate estimation of the theoretical analysis distribution $P^a(x|y)$, but a small Σ may not be able to provide enough of a smoothing factor to avoid the filter divergence.

A tuning factor is introduced to carefully adjust the variance Σ , which is a less satisfying solution and the stability of the filter remains a concern. Also if Σ is set to be proportional to the overall variance of ensemble sample, resampling is a computational concern especially for higher order dimensions. Other typical statistical techniques, like parametric density estimation and resampling (Kim et al., 2002), may also encounter similar problems when generalized to realistic higher order dimension systems.

In the following section we propose a posterior resampling method that can update the ensemble and maintain the filter stability, 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 $\xi_i, i = 1, \dots, n$ 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 phase space indices and j is the ensemble member indices. $\bar{\xi}$ and Σ_{ξ} represent the mean and the variance of ξ_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, phase 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 ,

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

where $\eta = [\eta_1, \eta_2, \dots, \eta_n]$, and

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

M is a symmetric matrix, 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 phase 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. ,

$$\frac{1}{N} \sum_j \xi_j \approx \bar{\xi}, \quad (13)$$

$$\frac{1}{N} \sum_j \xi_j \xi_j^T \approx \Sigma_{\xi}. \quad (14)$$

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.

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 EnPF.

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

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

Table 1. Mean rms error of the ensemble mean as a function of the stochastic model error volatility for 1000-member EnKF and EnPF assimilations of the Lorenz-63 system with measurement error 2.0

Model error variance(g^2)	EnKF ensemble mean rms error			EnPF ensemble mean rms error		
	x	y	z	x	y	z
0	2.16	3.47	3.48	1.68	2.70	2.86
2	2.29	3.75	3.80	2.20	3.56	3.54
4	2.40	3.87	3.73	2.15	3.45	3.28
6	2.99	4.94	4.88	2.40	3.89	3.83
8	2.67	4.39	4.15	2.33	3.84	3.20
10	3.52	5.62	5.29	2.56	4.21	4.14

dynamics to evolve within two distinctive attractors with appropriate time step dt . Model error variance per step(g^2) may change. The initial ensemble is obtained as the perturbation of the true state(reference solution), with a 3×3 diagonal error covariance matrix, $\text{diag}(2, 2, 2)$. The size of the ensemble is set to 1000 in all experiments. 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 EnPF and the EnKF methods with 40s run time and 800 time steps. The observation is measured on x available every 0.25s. The stochastic forcing is switched off. 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 $t=31s$).

A quantitative measure of the filter performance is the root mean square(rms) error of the ensemble mean prediction of the reference solution. Table 1 shows a comparison of the ensemble mean prediction rms error between the EnPF and the EnKF(Evensen 1994). The model error variance per time step 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 EnPF always yields a lower mean square error.

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. Figure3 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, EnPF 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 most outer surrounding curve and those small circles represent small probability density (less than 10 percent). 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 EnPF can produce good posterior Gaussian estimation with the mean consistent with the Bayesian computation. The upper row($\int dzP(x, y, z)$) shows that EnPF has better smoothing capability and produce larger variance, and thus more closely approximates the results of Bayesian analysis. This suggests the reason that the EnPF yields more accurate results than the EnKF in general.

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 & Cane, 1994). The EnPF 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 state 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 EnPF 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 Gaussian resampling is applied on the two samples respectively, which basically smoothes the samples and produces a two-component Gaussian mixture estimation.

Figure(6) 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 $t = 7s$ 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 EnPF with posterior Gaussian resampling yields satisfactory results when tested in the framework of a low dimension Lorenz model. Compared to other particle filter methods proposed in the literature, the posterior Gaussian resampling method preserves the ensemble spread and allows relatively straightforward extension to large dimensional applications. The most computationally expensive part involves the singular decomposition of a matrix with dimensions of the ensemble sample size. The EnKF(Evensen, 2004; Zupanski, 2004) is also subject to similar computational constraints. Low dimension simulation results show that the EnPF may still suffer from filter divergence problem if the ensemble sample size is too small to represent the error perturbation. A Kalman filter is less likely to suffer from filter divergence as it only interpolates the error between the prior ensemble and the measurement. It is more stable however it can still yield sub par analysis results if the prior ensemble diverges from the true state. The EnPF proposed here allows non-Gaussian probability distribution for both the prior ensemble and the measurement but requires the estimation of their probability distributions to be accurate enough to avoid filter divergence.

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 EnPF as an efficient tool for multimodal probability distribution can be further explored.

6 Acknowledgements

The research of Prof. Navon and X. Xiong is sponsored by the National Science Foundation(NSF), Grant ATM-0327818.

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 nonlinear non-Gaussian Bayesian tracking. *IEEE Trans. Sig. Pro.* **50**, 174–188.
- Bishop, C. H., 2001. Adaptive sampling with the ensemble transform Kalman filter. *Mon. Wea. Rev.* **129**,420–436.

- Bürger, G. and Cane, M. A. 1994. Interactive Kalman filtering. *J. Geophys. Res. Oceans*. **99**, 8015–8031.
- 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.
- Gordon, N. J., Salmond, D. J. and Smith, A. F. M., 1993. Novel approach to nonlinear-non-Gaussian Bayesian state estimation. *IEEE-Proceedings-F* **140**, 107–113.
- 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. *Nonlinear Process in Geophysics* **10**,253–259.
- 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. Atmos. 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.
- Zupanski, M., 2004. Maximum Likelihood Ensemble Filter: Theoretical Aspects. *In press with Mon. Wea. Rev.*

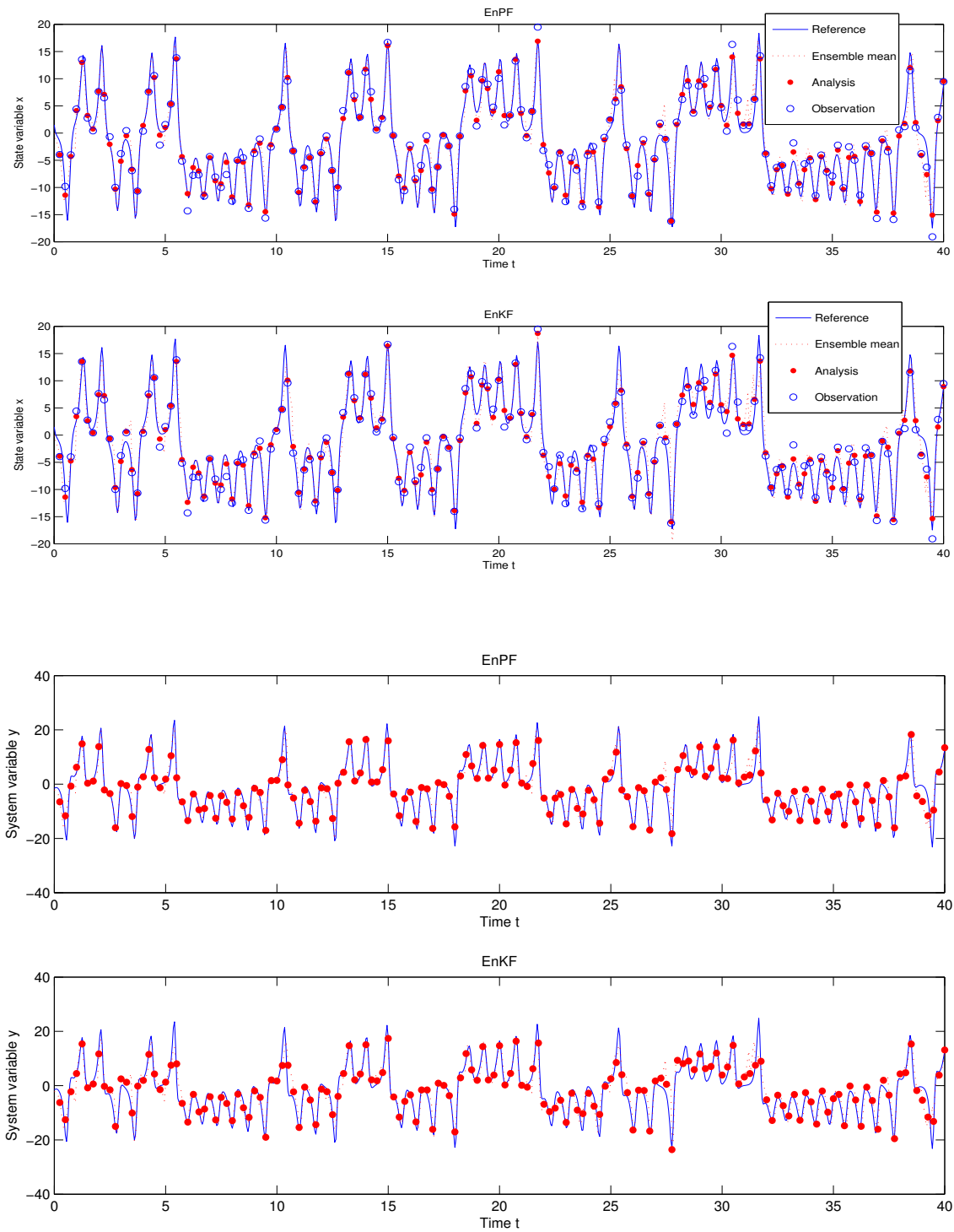


Figure 1. Results of data assimilation experiments with EnKF and EnPF. State variables, reference solution, observations, and ensemble mean prediction

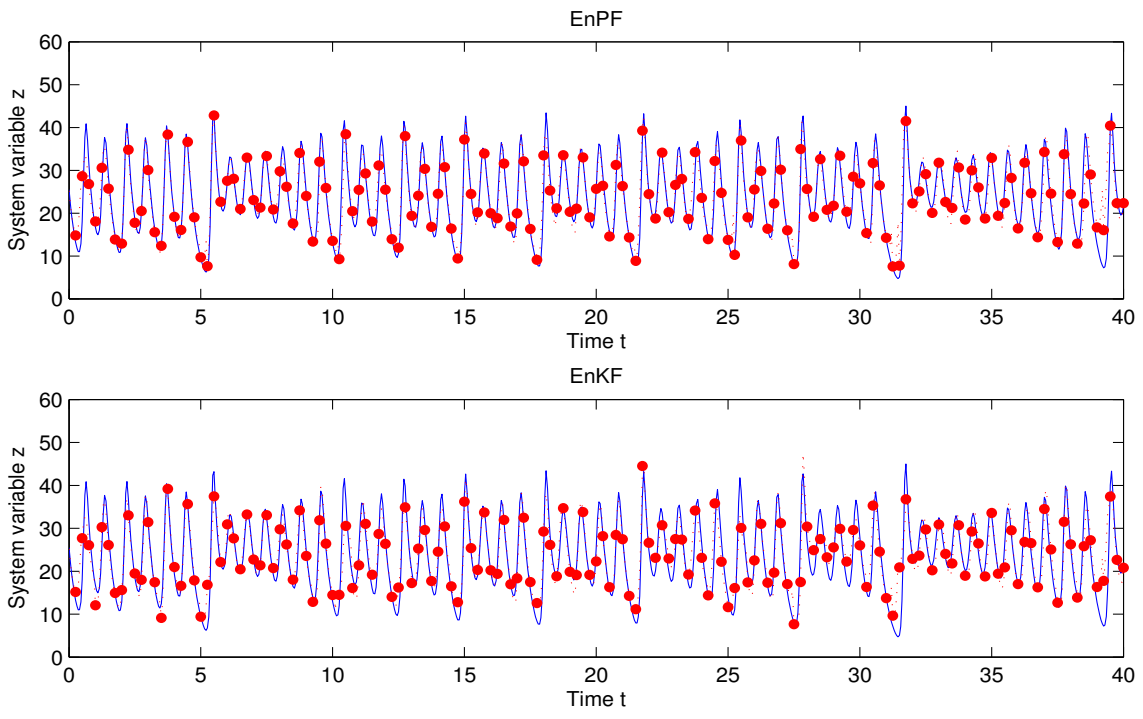
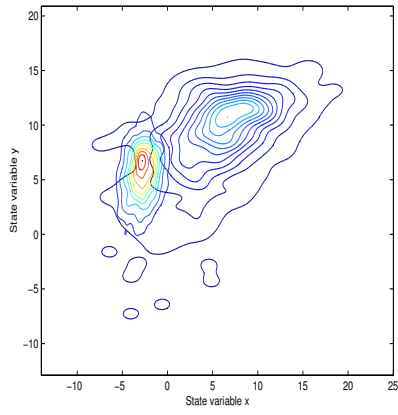
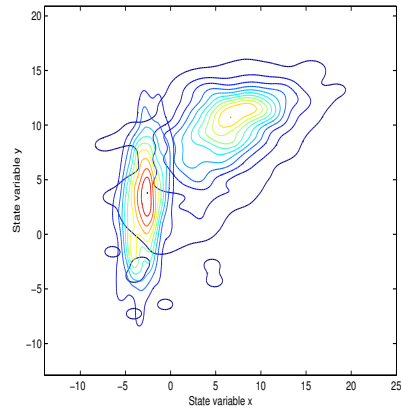


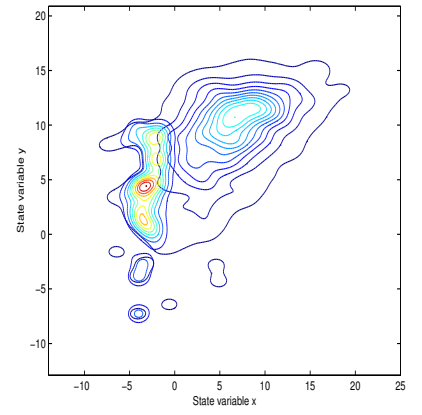
Figure 2.



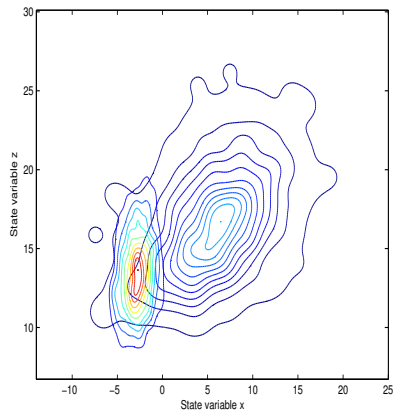
EnKF



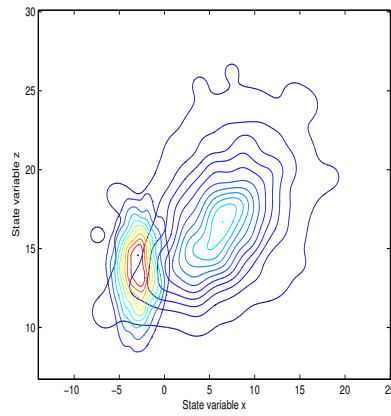
EnPF



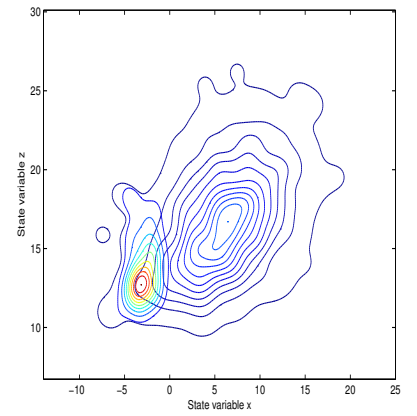
Bayesian Analysis



EnKF



EnPF



Bayesian Analysis

Figure 3. Kernel estimate of the prior and posterior probability density function. Observation value $x = -3.884$. EnKF, EnPF and Bayesian Analysis.

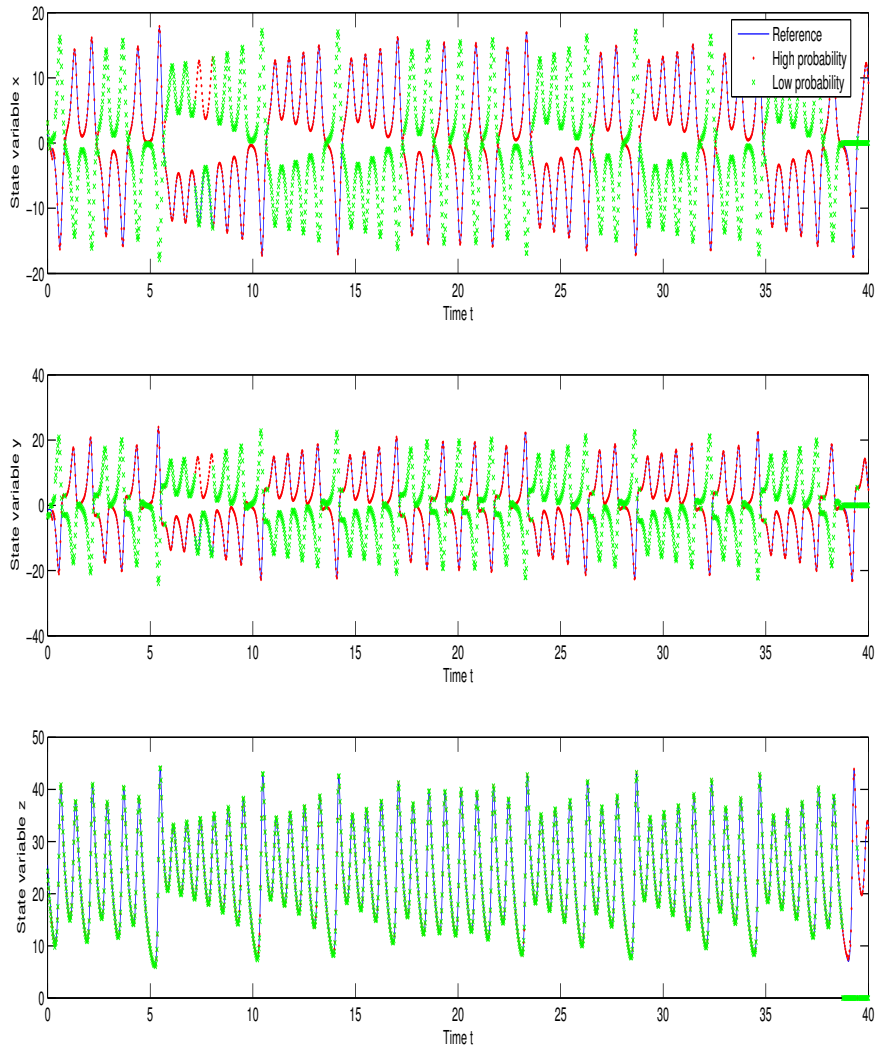


Figure 4. Results of data assimilation with observation operator x^2 . State variables, reference solution, high probability prediction, low probability prediction, output of data assimilation scheme.