Tellus (2006), 58A, 456–460
Printed in Singapore. All rights reserved

TELLUS

A Note on the Particle Filter with Posterior Gaussian Resampling

By X. XIONG[1], I. M. NAVON[1,2]* and B. UZUNOGLU[1], [1]*School of Computational Science, Florida State University, Tallahassee, FL 32306, USA;* [2]*Department of Mathematics, Florida State University, Tallahassee, FL 32306,USA*

(Manuscript received 22 February 2005; in final form 6 February 2006)

ABSTRACT

Particle filter (PF) is a fully non-linear filter with Bayesian conditional probability estimation, compared here 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 PF with Gaussian resampling (PFGR) can approximate more accurately the Bayesian analysis. The present work demonstrates that the proposed PFGR 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. 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.

EnKF 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 is well known that, through direct evaluation of the Bayesian's formula at each prior sample point, a particle filter (PF) generates a probability-weighted posterior sample. 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 applies at every step. A large enough estimation error can induce a so-called filter divergence or degeneracy problem, which refers to the fact that the ensemble sample diverges gradually from the true state and no longer produces a meaningful forecast.

Although the PF showed varied degree of success, filter divergence remains a major concern in realistic application of the PF. Covariance inflation is the most common technique to stabilize the PF(Anderson and Anderson, 1999; Whitaker and Hamill, 2002). Inflation factors are introduced to offset a tendency of the ensemble forecast to become underdispersive. The cause of such underdispersion can be attributed to a failure to adequately represent model error, rank deficiency in the forecast error covariance model or misspecification of observation errors aspects of the algorithm. One of the primary concerns regarding inflation factors is that they do not address the root cause of ensemble underdispersion—suboptimality in the filter. 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, for example, model error estimation. Other PF relies on the intrinsic smoothing capability of the model where the model noise and the non-linear interactions among the growing modes may produce enough chaotic behaviour to recover lost degrees of freedom in PF(van Leeuwen, 2003).

This note 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 follows: Section 2 introduces a PF with the posterior Gaussian resampling (PFGR). Section 3 presents simulation results of a numerical test of the method using the Lorenz model comparing PFGR and EnKF. Section 4 concludes the work and discusses directions of future research effort.

*Corresponding author.
e-mail: navon@csit.fsu.edu
DOI: 10.1111/j.1600-0870.2006.00185.x

2. Particle Filter in Bayesian Framework

Dynamical evolution of discretized physical systems are described by

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

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).

We have an estimate for the analysis probability distribution,

$$P^a(x|y) \approx \frac{1}{nN}\sum_{j=1}^{n} P^o(y - \mathcal{H}(\eta_j))\delta(x - \eta_j), \tag{2}$$

where η_j, $j = 1, \ldots, n$ are the positions of the prior ensemble members, x is the state variable at time t_k with an observation y at time t_k and N is the normalization constant. 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 is 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.

In the following we propose the posterior GR method that generates the updated ensemble with estimated mean and variance (2) computed from the distribution (2) in matrix form, which is to find ξ_i, $i = 1, \ldots, n$ so that

$$\Sigma_\xi = \eta M \eta^T \tag{3}$$

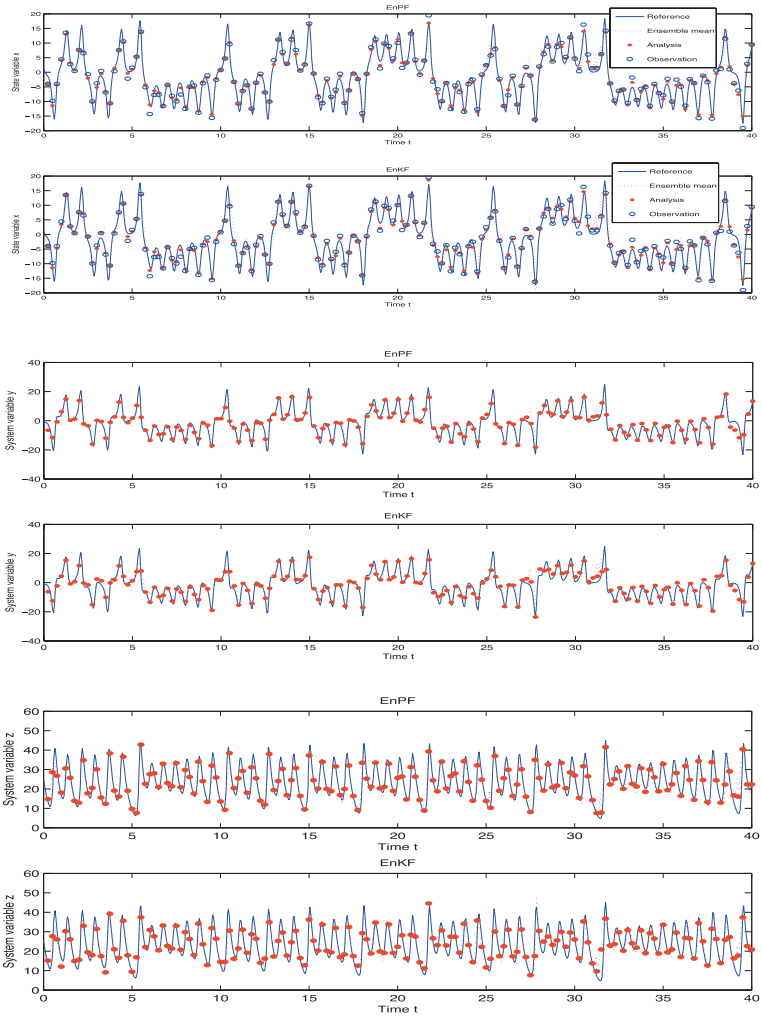

Fig. 1. Results of data assimilation experiments with EnKF and PFGR. System variables x, y, z, reference solution, observations and ensemble mean prediction. 160 observations within 40 s run time. Zero model error variance. Measurements on x only with observation variance 2.0. Ensemble size 1000. The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

$\eta = [\eta_1, \eta_2, \ldots, \eta_n]$ is a $L \times n$ matrix where L is the number of system variables and n is the number of the ensemble members and j is the ensemble member indices where Σ_ξ represent the the variance of ξ_i's associated with each position η_j. M is a symmetric matrix with elements

$$M_{jk} = f_j \delta_{jk} - f_j f_k \tag{4}$$

where f_j's are the normalized weights,

$$f_j = \frac{P^o(y - \mathcal{H}(\eta_j))}{\sum_j P^o(y - \mathcal{H}(\eta_j)}. \tag{5}$$

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 eq. (3).

In particular eq. (3) can be considered as a non-linear 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.

M is a symmetric matrix that can be factorized with a singular vector decomposition method,

$$M = V \Lambda V^T. \tag{6}$$

Then

$$\Sigma_\xi = \xi' \xi'^T, \tag{7}$$

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 (5) 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

$$\xi = \xi' X + \bar{\xi}. \tag{8}$$

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

$$\sum_j \xi_j \approx \bar{\xi} \tag{9}$$

$$\sum_j \xi_j \xi_j^T \approx \Sigma_\xi. \tag{10}$$

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.

The GR method has some similarity with the ensemble transform Kalman filter (ETKF) (Bishop et al., 2001). The ETKF proposes a sampling method that preserves the mean and variance derived from the EnKF, which allows rapid generation of posterior ensembles.

3. Numerical Experiments with the Lorenz Model

Lorenz-63(Lorenz 1963) stochastic model, described by eqs. (11), is used here to test the data assimilation performance of the PFGR, with three parameters specified as follows: $\sigma = 10.0$, $\rho = 28.0$ and $\beta = 8/3$.

$$\begin{aligned} dx &= -\sigma(x - y)dt + g dw_1, \\ dy &= (\rho x - y - xz)dt + g dw_2, \\ dz &= (xy - \beta z)dt + g dw_3. \end{aligned} \tag{11}$$

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 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. The measurement is performed on the state variable x only. 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 PFGR and the EnKF methods with 40 s run time and 800 time-steps. The observation is measured on x available every 0.25 s. 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

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

1000-member Model error	EnKF mean rms			PFGR 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
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

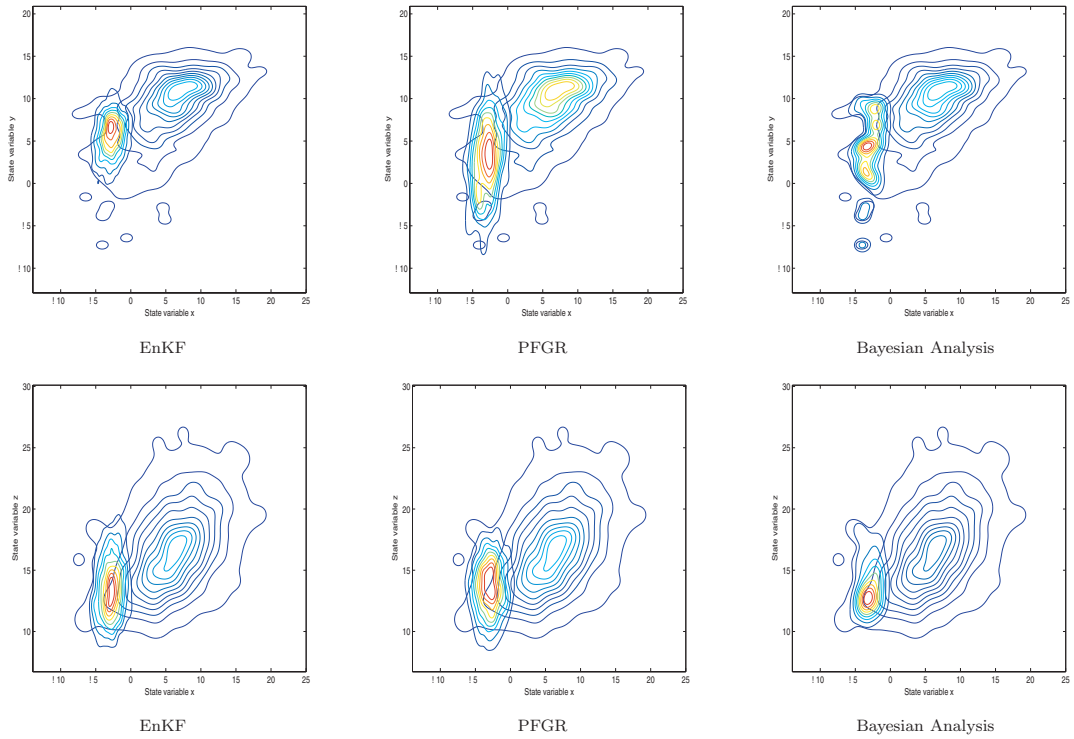

Fig. 2. Kernel density estimate of the prior and posterior probability density function integrated over z direction (upper row) and y direction (lower row). Observation value $x = -3.884$. EnKF, PFGR and Bayesian Analysis. The prior and posterior ensemble data obtained from the same run as of Fig. 1 at $t = 34.5$ s. The probability profiles of the EnKF and the PFGR posterior ensembles show similarity.

at the same time (e.g. after t = 31 s). Similar 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. Table 1 shows a comparison of the ensemble mean prediction rms error between the PFGR and the EnKF (Evensen, 1994) The EnKF used in the experiment induces observation perturbation to avoid an underestimated analysed covariance (Bürger and Cane, 1994). 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 PFGR yields a lower mean-square error most of the time. In practice the performance difference between two filters should be discussed on a case by case situation. Many factors, such as the 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 2 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, that is, $\int dz P(x, y, z)$ and $\int dy P(x, y, z)$.

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, PFGR and direct Bayesian calculation are shown, respectively. The prior sample probability density function shows typical non-Gaussian characteristics, which is expected for the highly non-linear dynamics of the Lorenz model. The most outer surrounding curve and the small circles represent small probability density (less than 10 %). 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 PFGR can produce good posterior Gaussian estimation with the mean consistent with the Bayesian computation.

4. Conclusion

The PFGR yields satisfactory results when tested in the framework of a low dimension Lorenz model. 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. It seems likely that the GRPF approach,

however, would not solve the problems of long-distance spurious error correlations issue associated small ensemble size seen in EnKFs.

The GR procedure, eq. (8), 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 eq. (8) to a $m \times \beta n$ matrix, and then perform SVD on the product $\xi' X$ and retain only leading n singular vectors.

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 characteristics 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 paid is either 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.

5. Acknowledgments

The research of Prof. I. M. Navon, X. Xiong and B. Uzunoglu 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.

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.

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.

Evensen, G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**, 10 143–10 162.

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.

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.

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

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

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.

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.