Notes on data assimilation for nonlinear high-dimensional dynamics: stochastic approach

Lin WU

Abstract

This manuscript is devoted to the attempts on the design of new nonlinear data assimilation schemes. The variational and sequential assimilation methods are reviewed with emphasis on their performances on dealing with nonlinearity and high dimension of the environmental dynamical systems. The nonlinear data assimilation is based on Bayesian formulation and its approximate solutions. Sequential Monte Carlo methods, especially particle filters, are discussed. Several issues, i.e. variational formulation in the information viewpoint in the context of nonlinear data assimilation, efficient sampling techniques, and parallel implementation of the filters, might bring new ideas for the design of new nonlinear data assimilation schemes. In the end we briefly summarize the applications of data assimilation for air pollution.

Key words: Stochastic dynamics prediction, nonlinear data assimilation, ensemble Kalman filter, Particle filter

1 Introduction

Data assimilation aims at optimal estimations of system states from diverse sources of information, such as nonlinear dynamics governed by physical laws, statistics features, and observations. Subsequent forecast can thus be conducted based on the *assimilated* system states. Data assimilation plays important roles for numerical environmental predictions in meteorology and oceanography. The main challenging tasks are to deal with the nonlinearity and the high dimension of environmental dynamical systems.

The observations are sparse and insufficient in the phase space of state variables. In addition, there are observational errors due to the lack of measurement precisions. That is why the prediction problem can not be described to be deterministic, instead, stochastic dynamic prediction was proposed (Epstein 1969). It was assumed that the physical laws, which govern the system

dynamical behavior, are entirely deterministic, but the initial conditions are probabilistic. Therefore the state probability density evolves conditioned to the observations. The predictions are statistic items according to state probability density. Following this stochastic approach, Fleming (1971a, 1971b) extended the predictability studies, and Leith (1974) investigated the ensemble forecast based on a Monte Carlo sampling of the stochastic initial conditions.

In this stochastic context, the infinite degrees of freedom in realities are reduced to be the uncertainties of initial conditions. For instance the number of degrees of freedom modern primitive equations are of the order of 10^7 (Kalnay 2003). The numerical prediction is transformed to be an initial condition problem. However, the enormous computational load, which results from the evaluation of expected values of the state variables as well as their variances and covariances, impedes the immediate applications of stochastic dynamic prediction. Epstein stated in the end of his 1969 Tellus paper that this stochastic method was designed to be rather research tools than operational procedures at that moment.

One is often told that data assimilation schemes are catalogued into sequential and variational approaches. *Neither* of them are deterministic (for the equivalence between the two approaches, see Lorenc 1986, Li and Navon 2003). The variational approaches started by Sasaki (1958, 1970) alleviate the heavy computational load of the estimation problem by feasible iterative solution of variational systems. In the variational formulation of Sasaki (1970), the model dynamics was underemphasized, and the state *evolution* was tracked under a weak constraint of model dynamics. Initial conditions were not considered in the assimilation system, therefore, the original variational formulation of Sasaki is not adaptable to the Epstein stochastic approach. Le Dimet (1982, 1986 with Talagrand) is the first that realizes this and proposes the proper variational formulation to iteratively solve the estimation problem of initial conditions for high dimensional atmospheric system. The model dynamics is of great benefit for the efficient iteration in the form of adjoint model. The variational methods are then widely spread in data assimilation community in diverse applications, i.e. uncertainties estimation of boundary conditions (Le Dimet 1993) and model errors (Vidard 2001), sensitivity analysis (Ngodock 1996, Le Dimet 2002) and adaptive observations (Navon 2003). Further reductions can be achieved by optimal control in the subspace of the control variables (Sophie thesis 2001).

When considering model nonlinearities, representer methods are proposed to solve the generalized inverse problem by a sequence of linear iterates of Euler-Lagrange equations in a finite dimensional *data space* spanned by representers of measurements. However, both adjoint model and tangent linear model can be considered as linear operators on the direct model, thus the truncation errors can not be omitted for highly nonlinear systems. The cost function of

the variational methods will carry multiple optima (Gauthier 1992; Miller et al. 1994). In this case, one might resort to global optimization methods, such as simulated annealing (Kruger 1993). For high dimensional systems, global optimization methods are computationally prohibitive. Pires et al. (1996) propose the quasi-static variational assimilation algorithm for determining the global minimum based on successive small increments of assimilation period. In practice, regularization items can be added to the cost function to avoid the minimization process from getting trapped around local minima.

The main drawback of the variational approach is that no flow dependent error covariances are provided, whereas in sequential approach, there is a filtering process to estimate not only state estimation but also error statistics when new observations are collected. For review of sequential data assimilation techniques, we refer to Bertino et al. 2003, Eversen 2003 and Hamill 2004.

The theoretical backbones of sequential data assimilation methods are based on stochastic dynamic prediction (Epstein 1969) and optimal filter (Jazwinski 1970). The assimilation process is composed of two inter-evolutional steps: forecast (or prediction) and update (or analysis). In prediction step the estimation results of the latest analysis step are used as the initial conditions for the forecast at next time step. The resulting forecast is then filtered together with the new observations for an state analysis (optimal estimation). In the context of stochastic dynamic prediction, the forecast can be chosen as statistics items (usually mean) calculated according to the state probability density function (pdf).

The evolution of the state pdf is governed by continuity partial differential equations (PDE), such as Louiville equation (Epstein 1969; Ehrendorfer 1994) or Fokker-Planck equation (Jazwinski 1970; Miller et al. 1999). A general nonlinear filtering framework using Bayesian theorem can then be proposed to adjust the state prior pdf (solution of Louiville equation) by the observations to achieve the state posterior pdf (Jazwinski 1970; Lorenc 1986). The state estimation can thus be conducted either by maximizing the posterior pdf (maximum a posteriori, MAP) or by minimizing the variances (van Leeuwen 2003).

The practical problem is that neither pdf evolution PDEs nor Bayesian formula are computational feasible for high dimensional systems. Approximation has to be made. Note that the original formulation is actually nonlinear, and approximation methods defined in the low dimensional subspace usually relax this nonlinearity to some extent. We have to balance the nonlinearity constraint and approximation precision for problem-relevant realities.

The approximation approaches follow mainly simplification strategy and subspace concept, or both. For simplification methods, the original problem be-

comes relatively feasible by proper assumptions. For instance, if we assume that the model dynamics are linear, and if states and observations are Gaussian, the original problem will be reduced to a best linear unbiased estimation problem (BLUE). Kalman filter (KF) can thus be derived in this case (Kalman 1960). For nearly nonlinear systems, the truncation errors of the linearalization are acceptable, the Kalman filter can be extended (Extended KF – EKF, Jazwinski 1970, Ghil and Malanotte-Rizzoli 1991).

The subspace concept is based on the studies of attractors of the nonlinear dynamics that lead to dominant low dimensional systems (Lorenz 1963, Lions et al. 1997). The state space is decomposed into subspaces that represent rapid and slow changes respectively (Lions 1997). The slow one can be considered as the climatic long-range tendence, whereas the rapid one characterizes the short-range oscillations like gravity waves. For short range prediction, since there is little change for slow-change subspace, the efficiency will be achieved by optimize *only* in the rapid-change space. However, special cautions have to be paid for the initialization and evolution of this subspace for best representation of the original model dynamics. For such reduction data assimilation methods in the framework of Kalman filter, we refer to multivariate empirical orthogonal functions (EOFs) representation (Cane et al. 1996), singular evolutive extended Kalman filter (SEEK, Pham et al. 1998) and reduced rank square root Kalman filter (RRSQRT, Verlaan and Heemink 1997).

Monte Carlo approximation can be considered as the projection of state space to a subspace that is spanned by random samples. The notorious $O(n^{\frac{1}{2}})$ convergence property (where n the number of samples, Casella and Berger 1990) of Monte Carlo methods impedes themselves from wide applications for data assimilation in early days. Things were changed when Evensen (1994) employed Monte Carlo methods to approximate the covariance for the Kalman filter update. In the applications of the so-called ensemble Kalman filter method (EnKF; Evensen 1994; Burgers et al. 1998; Evensen 2003), it is observed that 100 members of ensemble are sufficient for many real cases. The scientific interest are devoted to the generation of efficient ensembles. A notable process is to perform the sampling equipped with the concept of reduction methods (energy-based singular vector methods, Molteni et al. 1996; breeding methods, Toth and Kalnay 1997; singular evolutive interpolated Kalman filter – SEIK, Pham 2000; error subspace statistical estimation – ESSE, Lermusiaux and Robinson 1999).

Although EnKF solves partly the nonlinear problem at a moderate cost by Monte Carlo estimation of the mean and covariance of the prior governed by probability PDEs, the linear update of Kalman filter imposes the Gaussian assumption of prior and linearity of dynamics. Potentially (theoretically) there would be difficulties for EnKF to correctly track the state variables for non-Gaussian and strongly nonlinear high dimensional systems.

The nonlinear filtering, originally formulated by Jazwinski (1970), is brought to the scientific focus as a candidate of better assimilation scheme over linear update (van Leeuwen and Evensen 1996; Evensen 1997; Miller et al. 1999; Anderson and Anderson 1999; Pham 2001; Uzunoglu et al. 2005). The nonlinear filtering is supposed to better assimilate observations avoiding unbounded error growth within predictability limits. The techniques of sequential importance resampling (SIR, Liu and Chen, 1998, Doucet et al. 2001) and kernel approximation (Silverman 1986) in statistics community are imported for efficient Monte Carlo approximation of the nonlinear filtering problem. The sequential Monte Carlo technique – Particle filter – has drawn intensive attentions and becomes an active research subject (Pham 2001; Arulampalam et al. 2002; van Leeuwen 2003; Kivman 2003; Kim et al. 2003; Eyink and Kim 2005; Xiong and Navon 2005). Crucial issues are the methodologies on how to efficiently and adaptively generate the Monte Carlo samples for the complex dynamics and accumulation of observations (Bishop et al. 2001; Miller and Ehret 2002; Mitchel et al. 2002; Uzunoglu et al. 2005), and on how to adaptively perform the filtering given the relatively small ensembles (Lermusiaux 2002; Ott et al. 2004; Anderson 2004b). Much of these work are devoted to EnKF sampling, while efficient sampling for nonlinear particle filtering remains to be a scientific target.

This report is not for the purpose of a comprehensive state-of-art of the stochastic approach for data assimilation. We just summarize the main formulae for some important sequential schemes. The general nonlinear data assimilation framework is presented in section 2. The subsequent section is devoted to the linear suboptimal Kalman filter and its variants. Particle filter, as well as its applications in data assimilation, is introduced in section 4. While section 5 is devoted to several recent directions for particle filter design for data assimilation. Finally, we briefly summarize the applications of data assimilation to chemical atmospheric transport models.

2 Nonlinear data assimilation, a Bayesian framework

In the data assimilation context, the atmospheric, oceanic or other environmental systems are described as an equation from time t_{k-1} to t_k,

$$\mathbf{x}^t(t_k) = M_{k-1}[\mathbf{x}^t(t_{k-1})] + \eta_{k-1}, \tag{1}$$

where $\mathbf{x}$ is the state vector of n dimension, M corresponds to the (nonlinear) dynamics operator, and η_{k-1} is the system noise vector. Let $\mathbf{x}_k^t = \mathbf{x}^t(t_k)$.

At each time t_k, one observes,

$$\mathbf{y}_k^o = H_k[\mathbf{x}^t(t_k)] + \epsilon_k, \tag{2}$$

where H is the (nonlinear) observation operator, $\mathbf{y}_k^o \equiv \mathbf{y}^o(t_k)$ is the observation vector at time t_k of p dimension, and ϵ_k is the observation noise vector. Typically $p \ll n$. The system noise η_{k-1} and observation noise ϵ_k are supposed to be mutually independent. The initial state pdf is denoted $p(\mathbf{x}_0^t)$.

Let us denote the observation set up to t_k by $\mathbf{Y}_k \equiv \{\mathbf{y}_i^o, i = 1, ..k\}$. The data assimilation is thus an absorption process of the information from observation $\mathbf{Y}_k$ to decrease the model uncertainties for better state estimations of conditioned density. Note that if no observations at initial time t_0 we have $\mathbf{Y}_0 = \emptyset$, hence $p(\mathbf{x}_0|\mathbf{Y}_0) = p(\mathbf{x}_0)$. As indicated in introduction procession, the assimilation process follows the two steps of forecast and analysis. Suppose that our assimilation time interval is $[t_{k-1}, t_k]$, the objective is thus the estimation of $p(\mathbf{x}_k^t|\mathbf{Y}_k)$.

Forecast:

The prior pdf $p(\mathbf{x}_k^t|\mathbf{Y}_{k-1})$ can be obtained by theory of stochastic dynamics prediction in the form of Chapman-Kolmogorov equation (Jazwinski 1970),

$$p(\mathbf{x}_k^t|\mathbf{Y}_{k-1}) = \int p(\mathbf{x}_k^t|\mathbf{x}_{k-1}^t) p(\mathbf{x}_{k-1}^t|\mathbf{Y}_{k-1}) d\mathbf{x}_{k-1}^t \tag{3}$$

where $p(\mathbf{x}_k^t|\mathbf{x}_{k-1}^t)$ is defined by the dynamics model (1) and noise statistics of η_{k-1}.

Analysis:

Applying Bayes rule, we have (see van der Merwe et al. 2000)

$$p(\mathbf{x}_k^t|\mathbf{Y}_k) = \frac{p(\mathbf{y}_k^o|\mathbf{x}_k^t) p(\mathbf{x}_k^t|\mathbf{Y}_{k-1})}{p(\mathbf{y}_k^o|\mathbf{Y}_{k-1})}, \tag{4}$$

where $p(\mathbf{x}_k^t|\mathbf{Y}_{k-1})$ is the *prior* pdf provided by the forecast step, $p(\mathbf{x}_k^t|\mathbf{Y}_k)$ is the *posterior* pdf, $p(\mathbf{y}_k^o|\mathbf{x}_k^t)$ is called *likelihood* and defined by observation model (2) and noise statistics of ϵ_k, and $p(\mathbf{y}_k^o|\mathbf{Y}_{k-1})$ is the normalization constant sometimes called *evidence* or *marginal likelihood*,

$$p(\mathbf{y}_k^o|\mathbf{Y}_{k-1}) = \int p(\mathbf{y}_k^o|\mathbf{x}_k^t) p(\mathbf{x}_k^t|\mathbf{Y}_{k-1}) d\mathbf{x}_k^t \tag{5}$$

Note that there is an implicit assumption that model (1) is a Markov process of order one. When model dynamics are described as Itô stochastic differential equation (SDE), one can derive the pdf evolution PDE as Fokker-Planck equation (Jazwinski 1970; Miller et al. 1999) instead of the integral form of Chapman-Kolmogorov (3). Furthermore when system error item is omitted in eqn. (1), the pdf evolution will be governed by Louiville equation (Epstein 1969; Ehrendorfer 1994).

3 Kalman Filter and its Variants

3.1 Extended Kalman filter

Given Gaussian $p(\mathbf{x}_{k-1}^f|\mathbf{Y}_{k-1})$, if noise ϵ_k and η_{k-1} are mutually independent Gaussian with known moments, and if dynamics operator M_{k-1} and observation operator H_k are linear with respect to state vector $\mathbf{x}^t$ and noises, the forecast and analysis process of (3) and (4) equals to that of Kalman filter (Kalman 1960).

For nonlinear cases, denote the linear operator $\mathbf{M}_{k-1}$ and $\mathbf{H}_k$ the linear parts of M_{k-1} and H_k, that is,

$$\begin{aligned} \mathbf{M}_{k-1} &\equiv \frac{\partial M_{k-1}}{\partial \mathbf{x}_{k-1}^t} \\ \mathbf{H}_k &\equiv \frac{\partial H_k}{\partial \mathbf{x}_k^t} \end{aligned} \tag{6}$$

The approximate process of Kalman filter is then,

Forecast:

$$\begin{aligned} \mathbf{x}^f(t_k) &= M_{k-1}[\mathbf{x}^a(t_{k-1})] \\ \mathbf{P}^f(t_k) &= \mathbf{M}_{k-1}\mathbf{P}^a(t_{k-1})\mathbf{M}_{k-1}^T + \mathbf{Q}(t_{k-1}) \end{aligned} \tag{7}$$

Analysis:

$$\begin{aligned} \mathbf{x}^a(t_k) &= \mathbf{x}^f(t_k) + \mathbf{K}_k(\mathbf{y}_k^o - H_k[\mathbf{x}^f(t_k)]) \\ \mathbf{P}^a(t_k) &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}^f(t_k) \end{aligned} \tag{8}$$

where $\mathbf{K}_k$ is the Kalman gain,

$$\mathbf{K}_k = \mathbf{P}^f(t_k)\mathbf{H}_k^T[\mathbf{H}_k\mathbf{P}^f(t_k)\mathbf{H}_k^T + \mathbf{R}_k]^{-1} \tag{9}$$

Here $\mathbf{Q}_k$ and $\mathbf{R}_k$ are the covariance matrices of the zero mean noises η_k and ϵ_k respectively.

In this case, the posterior pdf $p(\mathbf{x}_k^f|\mathbf{Y}_k)$ is approximate by a Gaussian,

$$p(\mathbf{x}_k^t|\mathbf{Y}_k) \simeq \mathcal{N}(\mathbf{x}_k^t; \mathbf{x}_k^a, \mathbf{P}_k^a) \tag{10}$$

under the assumptions of Gaussian prior,

$$p(\mathbf{x}_k^t|\mathbf{Y}_{k-1}) \simeq \mathcal{N}(\mathbf{x}_k^t; \mathbf{x}_k^f, \mathbf{P}_k^f) \tag{11}$$

When dynamics are highly nonlinear, these assumptions will no longer be valid and prediction errors can be encountered. For instance the storm Lothar is misforecasted in 1999, as causes enormous damages in Europe (Mackenzie 2003).

3.2 Reduced Kalman filters

The Kalman filter is computational infeasible for high dimensional systems, for instance the primitive equations with state dimension $n = 10^7$ (Kalnay 2003). A single storage of the covariance matrix is of size $O(n^2)$, and the covariance calculation formula of P^f in (7) requires $2n$ model integrals. For current settings it will cost CPU times for years. In addition to the computational load, unbounded error growth of EKF was reported (Evensen 1992).

Several methods have been proposed to approximate the huge covariance matrix $\mathbf{P}^f$ based on subspace concept. Cane et al. (1996) were the first to apply Kalman filter in the reduced state space spanned by r principal components, namely multivariate empirical orthogonal functions (EOFs). Cohn and Todling (1996) propose three reduced schemes for Kalman filter: the coarse grid approximation of $\mathbf{P}^f$, the singular value decomposition of the tangent linear model and the approximation eigendecomposition of $\mathbf{P}^f$. Typical such methods are for example SEEK (Pham 1998) and RRSQRT (Verlaan and Heemink 1997).

We briefly list the data assimilation scheme of SEEK as follows.

Initialization:

$$\begin{aligned} \mathbf{x}_0^a &= \mathbf{x}_0 \\ \mathbf{P}_0^a &= \mathbf{L}_0\mathbf{U}_0\mathbf{L}_0^T \end{aligned} \tag{12}$$

Forecast:

$$\begin{aligned} \mathbf{x}_k^f &= M_{k-1}[\mathbf{x}_{k-1}^a] \\ \mathbf{L}_k &= \mathbf{M}_{k-1}\mathbf{L}_{k-1} \\ \mathbf{P}_k^f &= \mathbf{L}_k\mathbf{U}_{k-1}\mathbf{L}_k^T \end{aligned} \tag{13}$$

Analysis:

$$\begin{aligned} \mathbf{U}_k^{-1} &= \mathbf{U}_{k-1}^{-1} + \mathbf{L}_k^T\mathbf{H}_k^T\mathbf{R}_k^{-1}\mathbf{L}_k \\ \mathbf{x}_k^a &= \mathbf{x}_k^f + \mathbf{K}_k(\mathbf{y}_k^o - H_k[\mathbf{x}_k^f]) \\ \mathbf{P}_k^a &= \mathbf{L}_k\mathbf{U}_k\mathbf{L}_k^T \end{aligned} \tag{14}$$

where the filter gain $\mathbf{K}_k$ is as,

$$\mathbf{K}_k = \mathbf{L}_k\mathbf{U}_k\mathbf{L}_k^T\mathbf{H}_k^T\mathbf{R}_k^{-1} \tag{15}$$

where the columns of $\mathbf{L}_k$ are r dominant eigenvectors of P_k^f, and $\mathbf{U}_k$ is a diagonal matrix of corresponding eigenvalues. The Kalman filter is performed in the subspace spanned by the r eigenvectors.

3.3 Ensemble Kalman filter

The idea of ensemble Kalman filter is to use Monte Carlo methods for the approximation of forecast covariance matrix $\mathbf{P}^f$ in the forecast step, whereas linear update of Kalman filter is kept in the analysis step.

We list the data assimilation schemes of EnKF following the notations of Nerger (2003).

Initialization:

Given initial pdf $p(\mathbf{x}_0^t)$, an ensemble of r members are generated randomly,

$$\{\mathbf{x}_0^{a(\alpha)}, \quad , \alpha = 1, \ldots, r\} \tag{16}$$

The approximate mean and covariance of initial pdf is,

$$\begin{array}{l} \bar{\mathbf{x}}_0^a = \frac{1}{r} \sum_{\alpha=1}^{r} \mathbf{x}_0^{a(\alpha)} \\ \tilde{\mathbf{P}}_0^a = \frac{1}{r-1} \sum_{\alpha=1}^{r} \left(\mathbf{x}_0^{a(\alpha)} - \bar{\mathbf{x}}_0^a\right) \left(\mathbf{x}_0^{a(\alpha)} - \bar{\mathbf{x}}_0^a\right)^T \end{array} \tag{17}$$

Forecast:

$$\begin{array}{l} \mathbf{x}_k^{f(\alpha)} = M_{k-1}[\mathbf{x}_{k-1}^{a(\alpha)}] + \eta_{k-1}^{(\alpha)} \\ \tilde{\mathbf{P}}_k^f \quad = \frac{1}{r-1} \sum_{\alpha=1}^{r} \left(\mathbf{x}_k^{f(\alpha)} - \bar{\mathbf{x}}_k^f\right) \left(\mathbf{x}_k^{f(\alpha)} - \bar{\mathbf{x}}_k^f\right)^T \end{array} \tag{18}$$

where $\bar{\mathbf{x}}_k^f$ is the mean of ensemble $\{\mathbf{x}_k^{f(\alpha)}, \alpha = 1, .., r\}$ defined similar to $\bar{\mathbf{x}}_0^a$ in (17).

Analysis:

$$\begin{array}{l} \mathbf{x}_k^{a(\alpha)} = \mathbf{x}_k^{f(\alpha)} + \tilde{\mathbf{K}}_k \left(\mathbf{y}_k^{o(\alpha)} - H_k[\mathbf{x}_k^{f(\alpha)}]\right) \\ \mathbf{x}_k^a = \frac{1}{r} \sum_{\alpha=1}^{r} \mathbf{x}_k^{a(\alpha)} \\ \tilde{\mathbf{P}}_k^a = \frac{1}{r-1} \sum_{\alpha=1}^{r} \left(\mathbf{x}_k^{a(\alpha)} - \mathbf{x}_k^a\right) \left(\mathbf{x}_k^{a(\alpha)} - \mathbf{x}_k^a\right)^T \end{array} \tag{19}$$

where

$$\tilde{\mathbf{K}}_k = \tilde{\mathbf{P}}_k^f \mathbf{H}_k^T \left(\mathbf{H}_k \tilde{\mathbf{P}}_k^f \mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \tag{20}$$

Here $\mathbf{y}_k^{o(\alpha)}$ is usually considered as random variables consistent with observation error covariance $\mathbf{R}_k$ (Burgers et al. 1998).

4 Particle filter

4.1 Theoretical aspects

In addition to the application of covariance matrix approximation for ensemble Kalman filter, Monte Carlo simulation can also be employed to estimate the posterior pdf $p(\mathbf{x}_k^t|\mathbf{Y}_k)$ for dynamical systems, as is often referred to sequential Monte Carlo (Liu and Chen 1998; Doucet et al. 2001).

4.1.1 Sequential Importance Sampling (SIS)

High dimensional distributions, say $\pi(\mathbf{x})$, is usually hard to be achieved. We can instead draw samples from another appropriate distribution $g(\mathbf{x})$, and then weight the resulting samples to accommodate $\pi(\mathbf{x})$. This is the so-called importance sampling.

Suppose that we are interested in some expectation $E_\pi(h(\mathbf{x}))$,

$$\begin{aligned} E_\pi(h(\mathbf{x})) &= \int h(\mathbf{x})\pi(\mathbf{x})d\mathbf{x} \\ &= \int h(\mathbf{x})\frac{\pi(\mathbf{x})}{g(\mathbf{x})}g(\mathbf{x})d\mathbf{x} \\ &= E_g\left(h(\mathbf{x})\frac{\pi(\mathbf{x})}{g(\mathbf{x})}\right) \end{aligned} \tag{21}$$

Let $\omega(\mathbf{x}) = \frac{\pi(\mathbf{x})}{g(\mathbf{x})}$, we have

$$E_\pi(h(\mathbf{x})) = E_g(h(\mathbf{x})\omega(\mathbf{x})) \tag{22}$$

The expectation $E_\pi(h(\mathbf{x}))$ can then be approximate by samples that taken from density $g : \{\mathbf{x}^{(i)} \sim g(\mathbf{x}), i = 1, .., r\}$ weighted by important weights $\omega^{(i)} = \frac{\pi(\mathbf{x}^{(i)})}{g(\mathbf{x}^{(i)})}$,

$$E_\pi\left(h(\mathbf{x})\right) = \lim_{r\to\infty}\frac{\sum_{i=1}^r h(\mathbf{x}^{(i)})\omega^{(i)}}{\sum_{i=1}^r \omega^{(i)}} \tag{23}$$

One can consider π as being approximated by the discrete distribution supported on the $\mathbf{x}^{(i)}$ with probabilities proportional to the weights $\omega^{(i)}$. That is

$$\pi(\mathbf{x}) \simeq \frac{1}{A}\sum_{i=1}^{r}\omega^{(i)}\delta(\mathbf{x}-\mathbf{x}^{(i)}) \tag{24}$$

where $A = \sum_{i=1}^{r}\omega^{(i)}$ is the normalization parameter.

Remember that in the Bayesian framework, we are interested in the posterior $p(\mathbf{x}_k^t|\mathbf{Y}_k)$. Let $\mathbf{X}_k \equiv \{x_j^t, j = 0, .., k\}$. Applying the importance sampling concept, we draw samples $\{\mathbf{X}_k^{(i)}, i = 1, .., r\}$ from some importance function $g(\mathbf{X}_k|\mathbf{Y}_k)$. The expectation of posterior $p(\mathbf{X}_k|\mathbf{Y}_k)$ with respect to some interest $h_k(\mathbf{X}_k)$ can thus be approximated as

$$E_{p(\mathbf{X}_k|\mathbf{Y}_k)}(h_k) \simeq \frac{1}{A}\sum_{i=1}^{r} h_k\left(\mathbf{X}_k^{(i)}\right)\cdot\omega_k^{(i)} \tag{25}$$

where $A = \sum_{i=1}^{r}\omega^{(i)}$ is the normalization parameter, and

$$\omega_k^{(i)} = \frac{p\left(\mathbf{X}_k^{(i)}|\mathbf{Y}_k\right)}{g\left(\mathbf{X}_k^{(i)}|\mathbf{Y}_k\right)} \tag{26}$$

In the sequential case when observations are available successively, if we assign the following factorized form for the importance function,

$$g\left(\mathbf{X}_k|\mathbf{Y}_k\right) = g\left(\mathbf{x}_k|\mathbf{X}_{k-1},\mathbf{Y}_k\right)\cdot g\left(\mathbf{X}_{k-1}|\mathbf{Y}_{k-1}\right) \tag{27}$$

The samples $\{\mathbf{X}_k^{(i)}\} \sim g\left(\mathbf{X}_k|\mathbf{Y}_k\right)$ are thus composed of previous samples $\{\mathbf{X}_{k-1}^{(i)}\} \sim g\left(\mathbf{X}_{k-1}|\mathbf{Y}_{k-1}\right)$ and new samples $\{\mathbf{x}_k^{(i)}\} \sim g((\mathbf{x}_k|\mathbf{X}_{k-1},\mathbf{Y}_k)$.

Accordingly we can factorize posterior $p\left(\mathbf{X}_k|\mathbf{Y}_k\right)$ in terms of $p\left(\mathbf{X}_{k-1}|\mathbf{Y}_{k-1}\right)$, $p\left(\mathbf{y}_k^o|\mathbf{x}_k^t\right)$ and $p\left(\mathbf{x}_k|\mathbf{x}_{k-1}\right)$, the weight update formula thus takes the recursive form (Arulampalam et al. 2002),

$$\omega_k^{(i)} = \omega_{k-1}^{(i)}\cdot\frac{p\left(\mathbf{y}_k^o|\mathbf{x}_k^{(i)}\right)\cdot p\left(\mathbf{x}_k^{(i)}|\mathbf{x}_{k-1}^{(i)}\right)}{g\left(\mathbf{x}_k^{(i)}|\mathbf{X}_{k-1}^{(i)},\mathbf{Y}_k\right)} \tag{28}$$

In summary, the sequential importance sampling is as follows (Doucet et al. 2000) :

Algorithm: Sequential Importance Sampling (SIS).

At each time $k = 0, 1, 2, ...$, for $i = 1, .., r$,

- Draw samples $\mathbf{x}_k^{(i)}$ from $g\left(\mathbf{x}_k|\mathbf{X}_{k-1}, \mathbf{Y}_k\right)$, and form the samples $\mathbf{X}_k^{(i)} \equiv \{\mathbf{x}_k^{(i)}, \mathbf{X}_{k-1}^{(i)}\}$.
- Compute importance weight $\omega_k^{(i)}$ according to formula (28)
- Normalize the importance weight

$$\tilde{\omega}_k^{(i)} = \frac{\omega_k^{(i)}}{\sum_{i=1}^{r} \omega_k^{(i)}} \tag{29}$$

4.1.2 Resampling

Kong, Liu and Wong (1994) proved that with importance function taking the form as (27), the variance of the importance weights increases inevitably over time (called degeneracy phenomenon). When the weights variance is considerably large, these weights become very skewed. The weights of many samples are very small, thus they have almost no contributions to the posterior approximations. In this case the effective sample size N_{eff} will be much less than r. Kong, Liu and Wong (1994) define N_{eff} as

$$N_{eff} = \frac{r}{1 + v_k^2}, \tag{30}$$

to measure the degeneracy phenomenon. Denoting

$$\omega(\mathbf{X}_k) = \frac{p\left(\mathbf{X}_k|\mathbf{Y}_k\right)}{g\left(\mathbf{X}_k|\mathbf{Y}_k\right)}, \tag{31}$$

then v_k^2 is the variance of $\omega(\mathbf{X}_k)$ with respect to $g\left(\mathbf{X}_k|\mathbf{Y}_k\right)$,

$$v_k^2 = VAR_{g(\mathbf{X}_k|\mathbf{Y}_k)}\left(\omega(\mathbf{X}_k)\right) \tag{32}$$

An estimation of N_{eff} is

$$\hat{N}_{eff} = \frac{1}{\sum \tilde{\omega}_k^{(i)}} \tag{33}$$

A resampling process is to eliminate samples with small weights and to enhance the samples with great weights, hence "rejuvenate" the sampler to an effective ensemble that better represents the system evolution. The resampling process minimizes the ensemble variation, therefore, we have larger effective sample size N_{eff} and the degeneracy effect is reduced.

Liu and Chen (1998) summarize three resampling methods: random resampling, residual resampling and local Monte Carlo resampling. The random

resampling is essentially i.i.d sampling from discrete density (24). In SIS algorithm, a new ensemble is generated by drawing r times the samples from $\{\mathbf{X}_k^{(i)}, i = 1, .., r\}$ according to their probability $\{\tilde{\omega}_k^{(i)}\}$. The weights are set to $\frac{1}{r}$ after the sampling.

The residual sampling scheme is as follows:

- Copy $r_i = \lfloor r\tilde{\omega}_k^{(i)} \rfloor$ times of sample $\mathbf{X}_k^{(i)}$ into the new ensemble. Let $r' = r - \sum_1^r r_i$.
- Process r' i.i.d sampling from $\{\mathbf{X}_k^{(i)}, i = 1, .., r\}$ according to their probability $\{r\tilde{\omega}_k^{(i)} - r_i\}$.
- Set $\omega_k^{(i)}$ to 1.

Liu and Chen (1998) compare the two resampling methods, and suggest to process residual sampling whenever possible. In local Monte Carlo sampling, the importance function is chosen to be $p\left(\mathbf{x}_k|\mathbf{X}_{k-1}, \mathbf{Y}_k\right)$, which is approximated based on the discrete a priori distribution of $\mathbf{X}_{k-1}$. The resampling is thus automatically achieved in SIS process of local Monte Carlo sampling (for details, see Liu and Chen 1998).

In summary, the sequence importance resampling is as follows (Doucet et al. 2000):

Algorithm: Sequence Importance Resampling (SIR).

At each time $k = 0, 1, 2, ...$, for $i = 1, .., r$,

- Draw samples $\tilde{\mathbf{x}}_k^{(i)}$ from $g\left(\mathbf{x}_k|\mathbf{X}_{k-1}, \mathbf{Y}_k\right)$, and form the samples $\tilde{\mathbf{X}}_k^{(i)} \equiv \{\tilde{\mathbf{x}}_k^{(i)}, \mathbf{X}_{k-1}^{(i)}\}$.
- Compute importance weight $\omega_k^{(i)}$ according to formula (28)
- Normalize the importance weight

$$\tilde{\omega}_k^{(i)} = \frac{\omega_k^{(i)}}{\sum_{i=1}^r \omega_k^{(i)}} \tag{34}$$

- Evaluate $\hat{N}_{eff}$ according to (33).
- If $\hat{N}_{eff} \geq \eta$, where η is some threshold, then

$$\mathbf{X}_k^{(i)} = \tilde{\mathbf{X}}_k^{(i)}$$

  otherwise resampling is performed to generate $\mathbf{X}_k^{(i)}$ based on $\tilde{\mathbf{X}}_k^{(i)}$ (either random or residual resampling).

In general, resampling process arouses both theoretical and practical prob-

lematics. The theoretical convergence results to the optimal filter needs to be investigated in this context. And such resampling procedure of mixing samples limits the practical parallel implementation of particle filter. When model dynamics is deterministic or has small errors, the resampling procedure selects identical samples, as made the resulting ensemble losing diversity (also named sample impoverishment problem). Remedies for this case can be either sample perturbation methods (singular vector or bred vector; Miller and Ehret 2002) or kernel methods in which samples are drawn from an approximate continuous distribution of kernel mixture (Silverman 1986, Pham 2001).

4.1.3 Importance function

One crucial concern for sequential Monte Carlo methods is the determination of the appropriate importance function $g\left(\mathbf{x}_k|\mathbf{X}_{k-1}^{(i)}, \mathbf{Y}_k\right)$ (Kong, Liu and Wong 1994). One nature selection of importance function is the posterior $p\left(\mathbf{x}_k|\mathbf{X}_{k-1}^{(i)}, \mathbf{Y}_k\right)$, with the weight update formula as

$$\omega_k^{(i)} = \omega_{k-1}^{(i)} \cdot p\left(\mathbf{y}_k^o|\mathbf{X}_{k-1}^{(i)}, \mathbf{Y}_{k-1}\right) \tag{35}$$

The posterior importance function is optimal in the sense of variance minimization for importance weights, as limits the degeneracy phenomenon. For simple cases, $p\left(\mathbf{x}_k|\mathbf{X}_{k-1}^{(i)}, \mathbf{Y}_k\right) = p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{y}_k^o\right)$, and the importance weights are updated by $\omega_k^{(i)} = \omega_{k-1}^{(i)} \cdot p\left(\mathbf{y}_k^o|\mathbf{x}_{k-1}^{(i)}\right)$. However, for practical applications it is usually the case that there is no analytical evaluations of importance function $p\left(\mathbf{x}_k|\mathbf{X}_{k-1}, \mathbf{Y}_k\right)$ and marginal likelihood $p\left(\mathbf{y}_k^o|\mathbf{X}_{k-1}^{(i)}, \mathbf{Y}_{k-1}\right)$. In this case, one usually resorts to simple prior importance function $p\left(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}\right)$ if the model possesses Markovian feature (bootstrap filter, Gordon, Salmond and Smith 1993; Kitagawa 1996). The corresponding weight update formula is

$$\omega_k^{(i)} = \omega_{k-1}^{(i)} \cdot p\left(\mathbf{y}_k^o|\mathbf{x}_k^{(i)}\right). \tag{36}$$

We have simpler forms for importance function and weight update formula. The main drawback of the prior importance function is that no information from latest observation $\mathbf{y}_k^o$ is considered. The new ensemble is thus less representative to the system evolutions and has large variations, when parts of the samples are drawn from the the prior distribution tails, where the posterior possibly has spikes because of large likelihood. That is why a resampling step is obligatory.

It is a nature concept to design the importance function based on not only prior information but also the accumulated observations. The difficulty of an-

alytical evaluations for importance function $g\left(\mathbf{x}_k|\mathbf{X}_{k-1}^{(i)}, \mathbf{Y}_k\right)$ can be partially alleviated by approximation methods which profit from either previous resulting samples or empirical information (knowledge). In this case, the importance function can be denoted by some parametric form $g\left(\mathbf{x}_k|\theta(\mathbf{X}_{k-1}^{(i)}, \mathbf{Y}_k)\right)$, where θ is a deterministic mapping from $\{\mathbf{X}_{k-1}^{(i)}, \mathbf{Y}_k\}$ to some parameter set Θ. In a typical exampleposterior Gaussian is employed as importance function (Doucet et al. 2000). The Gaussian mean and variances are parameters derived by linearizations of model dynamics and optimal importance function.

Another example is the auxiliary particle filter (Pitt and Shephard 2001). Here the importance function is chosen to be some joint distribution $q\left(\mathbf{x}_k, i|\mathbf{Y}_k\right)$, which is an approximation of $p\left(\mathbf{x}_k, i|\mathbf{Y}_k\right)$. The auxiliary variable i is the index of samples at time $k-1$. Essentially it differs from bootstrap approach in that not only prior information but also current observations are considered in the importance function of auxiliary particle filter. The prior in this case is the discrete distribution of sample mixture. Whereas in the parametric resampling filter (Kim et al. 2003), the prior distribution is represented by a mixture Gaussian. A parametric distribution $P(\mathbf{x}_k; \lambda, \Lambda)$ approximates the posterior. Note that in this approach, the observations are processed in a manner similar to ensemble Kalman filter. The Bayes rule is applied to update the parameter λ, Λ given current observations, and no sequential importance sampling (weighting) scheme is employed.

4.2 Particle filter applications in data assimilation

The first attempt of particle filter in data assimilation community is perhaps the work of van Leeuwen and Evensen (1996). The posterior is approximated by importance sampling, in which the importance function is chosen to be prior (or forecast) $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. However, there is no explicit formulation of sequential importance sampling, and resampling process is omitted. It is therefore not surprising that this preliminary attempt suffers from severe degeneracy phenomenon. Approximately 10,000 samples are needed for satisfactory estimation of posterior in this direct ensemble approach for a two-layer quasi-geostrophic ocean model.

Pham (2001) brings particle filter back to focus by furnishing a resampling step whenever great discrepancies among importance weights are observed. The importance weight discrepancy is evaluated according to entropy difference between two discrete distributions represented by different sets of importance weights. The importance function is still the prior. Kernel methods (Silverman 1986) are employed to approximate the discrete posterior distribution by the continuous distribution, such that the diversity within ensemble preserves when the model dynamics is deterministic or the system noise is

small. The state of chaotic Lorenz system is tracked back by particle filter based on observations of part of the states. Approximately 10-50 samples are needed for satisfactory assimilation results. Kivman (2003) enriches the results by comparing EnKF and particle filter with the conclusion that particle filter is superior to EnKF especially when observation is sparse for stochastic Lorenz system.

van Leeuwen (2003) realizes the importance of resampling, and extends the preliminary attempt (van Leeuwen and Evensen 1996) by introducing residual resampling for the realistic large-scale (2×10^5 dimension) quasi-geostrophic ocean model. There are no special treatments (say kernel methods) for the maintenances of ensemble diversity because of large errors of model dynamics. The importance function in this case is still the prior, and the ensemble size is 495. It is perhaps the first successful realistic application of particle filter in data assimilation community. Several challenging issues remain, for example particle filtering with smaller ensemble size (< 100), and the ensemble collapse problem.

5 Recent directions for efficient particle filter design in data assimilation

5.1 Information point of view of data assimilation

Two fundamental questions for data assimilation systems are how to represent the information from observations, and how such information propagates during the assimilation process. In linear case, we can define the observability or information matrix (Zupanski et al 2004). In the nonlinear context, it is a difficult problem that has no satisfactory answer. Once we know better this information propagation process, it would lead to efficient DA schemes that particle filter can benefit from. We notice that the latest work of Eyink and Kim follows this idea by defining relative entropy function for the resampling process. It is usually related to variational formulation (Eyink 2001; Eyink et al. 2002). One possible approach is to pay more attention to information representation related to the work of Eyink and Kim Note that such similar idea has been introduced by Pham (2002) for a mutual information approach to blind sources separation.

5.2 Efficient generation of ensemble

For ensemble data assimilation, efficient generation of samples has been intensively investigated. It is usually based on subspace concept of reduction methods, say singular vectors, breeding methods and empirical orthogonal functions. Adaptive, hierarchical, and multiscale ensemble data assimilation schemes have been introduced. One question is thus how to apply these sample generation techniques in particle filter. The work of Xiong and Navon follows this idea. But they come back to Gaussian posterior assumption. Singular vector decomposition is employed to represent only the samples with significant weight. It is conceptually similar to SEIK (Pham 2001), where 3 samples are found to be sufficient to track back the state evolution of the stochastic Lorenz system.

In recent work of van Leeuwen (2005), the ensemble size has been reduced to 32 by Guided Sequential Importance Resampling (GSIR) algorithm for the realistic quasi-geostrophic ocean model. The ensemble in this algorithm is resampled before the current observations arrive, and therefore guided towards the current observation. It is not clear at the moment the relationship between the GSIR algorithm and auxiliary particle filter (Pitt and Shephard 2001).

5.3 Parallelism consideration

Particle filtering is essentially parallel. For parallel implementation of Kalman filter and its variants, we refer to the thesis of Nerger (2004) and references therein. In particle filter case, the main difficulty is that the resampling techniques impede the parallelism.

6 Brief summary of data assimilation application for chemical atmospheric transport models

Data Assimilation for chemical atmospheric transport models There are applications of variational DA (Elbern and Schmidt 2001, Qulo 2004) and Kalman filter (Segers 2002). Particle filter for DA of air quality demands both practical implementations and theoretical investigations.

References

[1] Anderson J.L. and Anderson S.L., 1999, A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts, Monthly Weather Review, vol. 127, pp. 2741-1758

[2] Anderson J.L., 2004a, A Scalable Least Squares Implementation of Ensemble (Kalman) Filter, Submitted to Journal of Atmospheric and Oceanic Technology -A May

[3] Anderson J.L., 2004b, A Hierarchical Ensemble Filter for Data Assimilation, submitted to Monthly Weather Review

[4] Arulampalam S., Maskell S., Gordon N., and Clapp T., 2002, A tutorial on particle filters for online nonlinear non-gaussian Bayesian Tracking, IEEE Trans. on Signal Processing, Vol. 50, No. 2, pp. 174-188

[5] Bertino L., Evensen G., and Wackernagel H., Sequential Data Assimilation Techniques in Oceanography, International Statistical Review, 71, pp. 223-241, 2003

[6] Bishop C.H., Etherton B.J., and Majumdar S.J., 2001, Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects, Mon. Wea. Rev. 129, pp. 420-436.

[7] Burgers G., van Leeuwen P.J., and Evensen G., 1998, On the analysis scheme in the Ensemble Kalman Filter, Monthly Weather Rev. 126, 1719-1724

[8] Cane M. A., Kaplan A., Miller R.N., Tang B.Y., Hackert E.C., and Busalacchi A.J., 1996, Mapping tropical Pacific sea level: Data assimilation via a reduced state space Kalman filter. Journal of Geophysical Research-Oceans, 101(C10): 22599-22617

[9] Casella G. and Berger R.L., 1990, Statistical Inference, Duxbury Press, 650 pp.

[10] Cohn S. and Todling R., 1996, Appropriate data assimilation schemes for stable and unstable dynamics, J. Meteorol. Soc. Jpn. 74(1), pp. 63-75

[11] Doucet A., Godsill S., Andrieu C., 2000, On sequential Monte Carlo sampling methods for Bayesian filtering, Statistics and Computing, 10, pp. 197-208

[12] Doucet A., de Freitas N., Gordon N. (Eds), 2001, Sequential Monte Carlo Methods in Practice, Springer-Verlag

[13] Ehrendorfer M., 1994, The Liouville Equation and its Potential Usefulness for the Prediction of Forecast Skill. Part I: Theory, Monthly Weather Review, Vol. 122, pp. 703-713

[14] Epstein, E S, 1969. Stochastic dynamic prediction. Tellus A, 21, 739-759.

[15] Epstein, E S, 1969. The role of initial uncertainties in prediction, J. Appl. Meteor. 8, 190 - 198

[16] Evensen G., 1992, Using the extended Kalman filter with a multilayer quasi-geostrophic ocean model. J. Geophys. Res., Vol. 97, No. c11, pp. 17,905-17,924.

[17] Evensen G., 1994, Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res., Vol. 99, No. c5, pp. 10,143-10,162

[18] Evensen G., 1997, Advanced data assimilation for strongly nonlinear dynamics, Mon. Weather Rev., 125, pp. 1342-1354

[19] Evensen G., 2003, The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation, Ocean Dynamics, 53, pp. 343-367

[20] Eyink G., 2001, A Variational Formulation of Optimal Nonlinear Estimation, submitted to Methodology and Computating in Applied Probability

[21] Eyink G., Restrepo J. and Alexander F.J., 2002, A statistical-mechanical approach to data assimilation for nonlinear dynamics I-III, submitted to Phys. Rev. E

[22] Eyink G.L. and Kim S., 2005, Particle Filtering Method by Resampling from Maximum-Entropy Distributions, Submitted

[23] Fleming, R J, 1971, On stochastic dynamic prediction. I: the energetics of uncertainty and the question of closure. Mon. Wea. Rev., 99, pp. 851-872.

[24] Fleming, R J, 1971. On stochastic dynamic prediction. II: predictability and utility. Mon. Wea. Rev., 99, pp. 927-938.

[25] Gauthier P., 1992, Chaos and quadri-dimensional data assimilation a study based on the Lorenz model, Tellu 44A, pp. 2-17

[26] Ghil M. and Malanotte-Rizzoli P., 1991, Data Assimilation in Meteorology and Oceanography, Adv. Geophys., 33, pp. 141-266

[27] Gordon N.J., Salmond D.J., Smith A.F.M., 1993, Novel approach to nonlinear/non-Gaussian Bayesian state estimation, IEE-Proceedings-F 140, pp. 107-113

[28] Hamill T.M., Ensemble-Based Atmospheric Data Assimilation, 2004

[29] Huang Y. and Dijurć, 2004, A hybrid importance function for particle filtering, IEEE Signal Processing Letters, vol. 11, no. 3, pp. 404-406

[30] Jazwinski A.H., 1970, Stochastic Processes and Filtering theory, Academic Press, 376 pp.

[31] Kalnay E., 2003, Atmospheric Modeling, Data Assimilation and Predictability, Cambridge Univ. Press

[32] Keppenne C.L., 2000, Data Assimilation into a Primitive Equation model with a Parallel Ensemble Kalman Filter, Monthly Weather Review, pp. 1971-1981

[33] Kim S., Eyink G.L., Restrepo J.M., Alexander F.J., and Johnson G., 2003, Ensemble Filtering for Nonlinear Dynamics. Monthly Weather Review, vol. 131, no. 11, pp. 2586-2594

[34] Kitagawa G., 1996, Monte Carlo filter and smoother for non-Gaussian nonlinear state space models, Journal of Computational and Graphical Statistics, vol 5, pp. 1-25

[35] Kivman G.A., 2003, Sequential parameter estimation for stochastic systems, Nonlinear Process in Geophysics 10, pp. 253-259

[36] Kong A., Liu J.S., and Wong W.H., 1994, Sequential imputations and Bayesian missing data problem, Journal of the American Statistical Association, vol. 89, pp. 278-288

[37] Kruger J., Simulated Annealing: A Tool for Data Assimilation into an Almost Steady Model State, Journal of Physical Oceanography, Vol. 23, No. 4, pp. 679C688, 1993

[38] Le Dimet F.-X., A General Formalism of Variational Analysis, CIMMS Report No. 22, 1982

[39] Le Dimet F.-X., Talagrand O., Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects, Tellus, 38A, pp. 97-110, 1986

[40] Lermusiaux P.F.J. and Robinson A.R., 1999, Data assimilation via error subspace statistical estimation. Part I: Theory and schemes, Mon. Wea. Rev. 127, pp. 1385-1407

[41] Lermusiaux P.F.J., 2002, On the Mapping of Multivariate Geophysical Fields: Sensitivities to Size, Scales, and Dynamics, J. of Atmospheric and Oceanic Technology, Vol 19, pp. 1602-1637

[42] Lions J.-L., 1997, Modélisation mathématique et environnement - Quelques remarques, In *Tendances nouvelles en modélisation pour l'environnement*, Journées du Programme Environnement, Vie et Sociétés du CNRS, Coordinateur: Blasco F., pp. 79-90

[43] Lions J.-L., Manley O.P., Teman R., and Wang S., 1997, Physical Interpretation of the Attractor Dimension for the Primitive Equatios of Atmospheric Circulation, Journal of Atmospheric Sciences, 54(9)

[44] Liu J.S. and Chen R., 1998, Sequential Monte Carlo Methods for Dyanmic Systems, Journal of the American Statistical Association, Vol. 93, No. 443, pp. 1032-1044

[45] Lorenz E.N., 1963, Deterministic Nonperiodic Flow, Journal of the Atmospheric Sciences, Vol. 20, No. 2, pp. 130-148

[46] Mackenzie D., Ensemble Kalman Filters Bring Weather Models Up to Date, SIAM News, Vol. 36, No. 8, 2003

[47] Miller R.N., Ghil M., and Gauthiez F., 1994, Advanced data assimilation in strongly nonlinear dynamical systems, J. Atmos. Sci. 51, pp. 1037-1056

[48] Miller R.N., Carter E.F., and Blue S.T., 1999, Data assimilation into nonlinear stochastic models, Tellus 51A, pp. 167-194

[49] Miller R.N. and Ehret L.L., 2002, Ensemble Generation for Models of Multimodal Systems, Monthly Weather Review 130, pp. 2313-2333

[50] Mitchell H.L., Houtekamer P.L., Pellerin G., 2002, Ensemble Size, Balance, and Model Error Representation in an ensemble Kalman Filter, Monthly Weather Review 130, pp. 2791-2808

[51] Molteni F., Buizza R., Palmer T.N., Petroliagis T., 1996, The ECMWF Ensemble Prediction System: Methodology and Validation, Quart. J. Roy. Meteor. Soc., 122, pp. 73-119

[52] Nerger L., 2004, Parallel Filter Algorithms for Data Assimilation in Oceanography, Ph.D thesis, Berichte zur Polar- und Meeresforschung, Universitat Bremen, 174 pp.

[53] Ott E., Hunt B.R., Szunyogh I., Zimin A.V., Kostelich E.J., Corazza M., Kalnay E., Patil D.J., and Yorke J.A., 2004, A Local Ensemble Kalman Filter for Atmospheric data assimilation, Tellus 56A, pp. 415-428

[54] Pham D.T., Verron J., and Roubaud M., 1998, A Singular Evolutive Extended Kalman Filter for Data Assimilation in Oceanography, J. Marine System, 16(3-4), pp. 323-340

[55] Pham D.T., 2001, Stochastic Methods for Sequential Data Assimilation in Strongly Nonlinear Systems, Monthly Weather Review, vol. 129, pp. 1194-1207

[56] Pham D.T., 2002, Mutual information approach to blind seperation of stationary sources, IEEE Transactions on Information Theory, vol. 48, no. 7, pp. 1935-1946

[57] Pires C., Vautard R. and Talagrand O., On Extending the Limits of Variational Assimilation in Nonlinear Chaotic Systems, Tellus, 48A, pp. 96-121, 1996

[58] Pitt M. and Shephard N., 2001, Auxiliary varible based particle filter, in Doucet A., de Freitas N., and Gordon N. (Eds), Sequential Monte Carlo Methods in Practice

[59] Sasaki Y., An Objective Analysis Based on the Variational Method, Journal of the Meteorological Society of Japan, Ser. 2, Vol. 36, No. 3, pp. 77-88, 1958

[60] Sasaki Y., Numerical Variational Analysis with Weak Constraint and Application to Surface Analysis of Severe Storm Gust, Monthly Weather Review, pp. 899-910, 1970

[61] Toth Z. and Kalnay E., 1997, Ensemble Forecasting at NCEP and the Breeding Method, Monthly Weather Review, pp. 3297-3319

[62] Uzunoglu B., Navon I.M., Zupanski M., 2005, Adaptive ensemble size reduction and preconditioning, submitted to Tellus

[63] van der Merwe R., Doucet A., de Freitas N., and Wan E., 2000, Technical Report CUED/F-INFENG/TR 380, Cambridge University Engineering Department

[64] van Leeuwen P. and Evensen G., 1996, Data Assimilation and Inverse Methods in Terms of a Probabilistic Formulation, Mon. Weather Rev., 124, pp. 2898-2913

[65] van Leeuwen P., 2003, A variance-minimizing filter for large-scale applications, Monthly Weather Review 131, pp. 2071-2084

[66] van Leeuwen P., 2005, Ensemble Kalman filters, sequential importance resampling and beyond, preprint

[67] Verlaan M. and Heemink A.W., 1997, Tidal Flow Forecasting using Reduced Rank Square Root Filters, Stoch. Hydrol. Hydraul. 11(5), pp. 349-368

[68] Xiong X. and Navon I.M., 2005, Ensemble Particle Filter with Posterior Gaussian Resampling, submitted

[69] Zupanski D., Zupanski M., De Maria M., Grasso L., 2005, Critical Issues of Ensemble Data Assimilation in Application to Geos-R risk Reduction Program, Ninth Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface (IOAS-AOLS), Jan 9-13, 2005