School of Mathematics, Meteorology
and Physics

University of Reading

Department of Mathematics

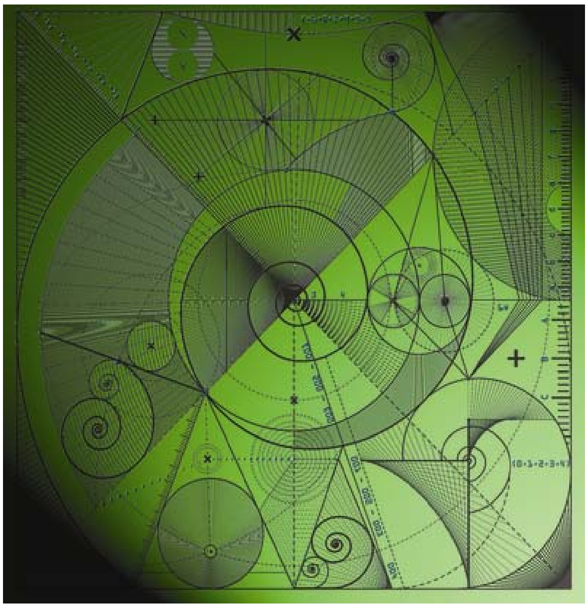
Preprint MPS_2010_18

21 April 2010

State Estimation using the Particle Filter with Mode Tracking

by

J.A. Pocock, S.L. Dance, A.S. Lawless



State Estimation using the Particle Filter with Mode Tracking

J.A. Pocock[1], S.L. Dance, A.S. Lawless

Department of Mathematics, University of Reading, Reading, Berkshire, RG6 6AX, UK

Abstract

A particle filter is a data assimilation scheme that employs a fully nonlinear, non-Gaussian analysis step. Unfortunately as the size of the state grows the number of ensemble members required for the particle filter to converge to the true solution increases exponentially. To overcome this Vaswani [Vaswani N., 2008, IEEE Transactions on Signal Processing, **56**:4583-4597] proposed a new method known as mode tracking to improve the efficiency of the particle filter. When mode tracking, the state is split into two subspaces. One subspace is forecast using the particle filter, the other is treated so that its values are set equal to the mode of the marginal pdf. There are many ways to split the state. One hypothesis is that the best results should be obtained from the particle filter with mode tracking when we mode track the maximum number of unimodal dimensions. The three dimensional stochastic Lorenz equations with direct observations are used to test this hypothesis. It is found that mode tracking the maximum number of unimodal dimensions does not always provide the best result. The best choice of states to mode-track depends on the number of particles used.

Keywords: Particle filter, Mode tracking, Stochastic Lorenz Equations

1. Introduction

Data assimilation is the incorporation of observational data into a numerical model to produce a model state which accurately describes the observed reality. It is applicable to many situations as it provides a complete set of accurate initial conditions for input into a numerical model.

[1]Corresponding author. email: j.a.pocock@student.reading.ac.uk

A particle filter (PF) is a data assimilation scheme that employs a fully non-linear, non-Gaussian analysis step [15]. Such filters approximate the conditional probability density function (pdf) of the model state given observations, using a weighted ensemble. As observations become available, the posterior weights for each ensemble member are updated using Bayes' rule. Consequently, the ensemble members with the largest weights correspond to the points in phase space with highest probability [4]. Reviews and tutorial introductions for the particle filter can be found, for example, in [15, 4, 5, 1].

There are a number of well-known issues with the practical application of the particle filter. These are discussed in detail by Snyder et al. [14]. The main difficulty is that straightforward Monte Carlo estimation of continuous pdfs is notoriously inefficient: the number of ensemble members needed grows exponentially with the dimension of the state [13]. For low dimensional systems the PF works well. However for problems such as numerical weather prediction, where the state space is of order 10^7, the PF scheme is too computationally costly.

There have been a number of attempts to deal with the difficulties of dimensionality. For example, several authors have made extra assumptions, such as spatially localized updates [2, 6]. For certain specialized scenarios, it may be possible to compute marginal pdfs by direct integration, thus reducing the dimension of the space that needs to be approximated by the particle filter [12].

Vaswani [16] introduced the idea of mode tracking to try to reduce the cost of the particle filter in high dimensional state spaces. The key idea is to split the state space into two subspaces, such that the marginal posterior for one of the subspaces is unimodal and relatively narrow. We will call this the mode-tracking subspace. In this part of the state-space, the analysis step replaces each of the ensemble members with the mode of the pdf, which is found by minimizing a cost function. The cost function is not dissimilar to that seen in variational data assimilation [7], except that the control vector consists of only a part of the state. Thus Vaswani's method is related to ensemble-variational techniques such as the maximum likelihood ensemble filter [17] and those described by Buehner et al. [3]. Importantly however in Vaswani's method, the ensemble weights are not assumed equal, and the calculation of the weights takes into account the mode-tracking approximation. Mode tracking has been shown to give comparable results to the PF but with fewer numbers of particles, for a simple linear forecast model with a nonlinear observation operator [16].

Vaswani [16] suggests that the best results should be obtained from the particle filter with mode tracking when we mode track the maximum number of unimodal dimensions. In this paper we test this using the 3D stochastic Lorenz equations with a linear observation operator.

We begin, in section 2, by describing the algorithm for the particle filter with mode-tracking. We also consider how the state should be split to provide the best results from the particle filter with mode tracking. The simple nonlinear stochastic model used for our experiments is described in section 3. In section 4 we test one hypothesis for how to split the state to obtain the best results for the particle filter with mode tracking. We conclude in section 5 by summarising and discussing the main results.

2. The Particle Filter with Mode Tracking

In this section we consider the particle filter with mode tracking (PF-MT). We start by defining the notation used. Let $\psi_k \in \mathbb{R}^n$, $k = 0, 1, 2, 3, \ldots$, be a sequence of model states at discrete times k and assume the initial pdf of the state is given by $p(\psi_0)$. At subsequent times, ψ_k satisfies

$$\psi_k = M(\psi_{k-1}, w_k), \tag{1}$$

where $M : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is a possibly nonlinear function, and $w_k \in \mathbb{R}^n$ is a noise process sequence. This equation describes a Markov process with transition density $p(\psi_k|\psi_{k-1})$.

Let $d_k \in \mathbb{R}^p$ be the observation vector at time k which is related to the model state by the equation,

$$d_k = H(\psi_k, v_k), \tag{2}$$

where $H : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^p$ is a possibly nonlinear observation operator and $v_k \in \mathbb{R}^p$ is a noise process sequence. We assume that the observations are conditionally independent given the state, with observation likelihood $p(d_k|\psi_k)$.

We now consider the PF-MT [16]. The PF-MT is used to estimate a model state given observations and previous model states. Using Bayes' rule we show the pdf we are trying to approximate is,

$$p(\psi_k|d_k) = \frac{p(d_k|\psi_k)p(\psi_k)}{p(d_k)}. \tag{3}$$

We introduce mode tracking to reduce the computational cost of the particle filter in high dimensional systems. The essential idea behind mode-tracking is to split the state, ψ, into two parts, $\psi = [\Psi_s, \Psi_r]$. The Ψ_s part of the state, of dimension s, is treated using an ordinary particle filter. The Ψ_r part, of dimension r, is treated as if the associated marginal pdf is unimodal, and Ψ_r is set equal to the mode. It is hoped that reducing the dimension of the state treated by the PF may allow a given filter accuracy to be obtained using a smaller ensemble size.

Before we describe the PF-MT algorithm we make some additional assumptions about the models and observations used. These assumptions are not necessary conditions, but they are relevant to the experiments carried out in this work. For the purposes of this paper, the general class of stochastic forecast models given by (1) are restricted to the form,

$$\psi_k = f(\psi_{k-1}) + \epsilon_k, \tag{4}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ and $\epsilon_k \sim N(0, Q)$. Taking the splitting $\psi = [\Psi_s, \Psi_r]$ into account our model (4) becomes

$$\begin{pmatrix} \Psi_{k,s} \\ \Psi_{k,r} \end{pmatrix} = \begin{pmatrix} f_s(\psi_{k-1}) \\ f_r(\psi_{k-1}) \end{pmatrix} + \begin{pmatrix} \epsilon_{k,s} \\ \epsilon_{k,r} \end{pmatrix}, \tag{5}$$

where $f_s : \mathbb{R}^n \to \mathbb{R}^s$ and $f_r : \mathbb{R}^n \to \mathbb{R}^r$. The model error matrix Q can be written in block form,

$$Q = \begin{pmatrix} Q_{ss} & Q_{sr} \\ Q_{rs} & Q_{rr} \end{pmatrix}, \tag{6}$$

where $Q_{ss} \in \mathbb{R}^{s\times s}$, $Q_{sr} \in \mathbb{R}^{s\times r}$, $Q_{rr} \in \mathbb{R}^{r\times r}$, $Q_{rs} \in \mathbb{R}^{r\times s}$ and Q is symmetric positive definite.

Similarly we restrict the general observation equation (2) to,

$$d_k = \psi_k^t + \eta_k, \tag{7}$$

where ψ_k^t represents the true solution at time k, η_k is random Gaussian noise, $\eta_k \sim N(0, R)$ with $R = \sigma_o^2 I$.

Table 1 summarises the general PF-MT algorithm. For the remainder of this section we shall consider each stage of the algorithm to see how PF-MT works in detail. Here we assume that the state splitting $\psi = [\Psi_s, \Psi_r]$ has been predetermined. How we make this choice in practice is discussed at the end of this section.

Table 1: The PF-MT Algorithm

The PF-MT Algorithm [16]

Initialization:

- Set $k = 0$ and sample N times from the importance function $\pi(\psi_0)$ to give the initial ensemble $\{\psi_0^i\}_{i=1}^N$.
- Set the weights, $w_0^i = 1/N$.

For times $k = 1, 2, ...$

1. Importance Sample $\Psi_{k,s}$: For $i = 1, 2, \ldots, N$, sample

$$\Psi_{k,s}^i \sim p(\Psi_{k,s}^i|\psi_{k-1}^i)$$

.

2. Mode track $\Psi_{k,r}$: For $i = 1, 2, \ldots, N$, set $\Psi_{k,r} = m_k^i$ where

$$m_k^i(\psi_{k-1}^i, \Psi_{k,s}^i, d_k) = \arg\min_{\Psi_{k,r}}[-\log p(d_k|\Psi_{k,s}^i)p(\Psi_{k,r}|\psi_{k-1}^i, \Psi_{k,s}^i)]$$

.

3. Weight: For $i = 1, 2, \ldots N$, compute $w_k^i = \frac{\omega_k^i}{\sum_{j=1}^N \omega_k^j}$ where

$$\omega_k^i = w_{k-1}^i p(d_k|\psi_k^i)p(\Psi_{k,r}^i|\psi_{k-1}^i, \Psi_{k,s}^i)$$

and $\psi_k^i = [\Psi_{k,s}^i, \Psi_{k,r}^i]$.

4. Resample: Replicate particles in proportion to their weights and reset the weights $\omega_k^i = 1/N$.

Step 1. Once the filter has been initialized, the first step shown in Table 1 is the importance sampling step for the Ψ_s dimensions. This step involves forecasting the Ψ_s dimensions forward for one timestep. In practice, it may be necessary to evolve the full stochastic numerical model, (4), forward in time from $k-1$ to k, and then simply restrict to the relevant dimensions.

Step 2. The second step in the PF-MT algorithm is mode tracking on $\Psi_{k,r}$, where we set $\Psi_{k,r}^i$ equal to the mode of the pdf conditioned on the Ψ_s

part of the subspace. The conditional pdf for $\Psi_{k,r}$ may be written as

$$p(\Psi_{k,r}|\psi^i_{k-1}, \Psi^i_{k,s}, d_k) \propto p(d_k|\psi^i_{k-1}, \Psi^i_{k,s})p(\Psi_{k,r}|\psi^i_{k-1}, \Psi^i_{k,s}), \tag{8}$$

using Bayes' rule, since ψ_k is a Markov process and the observations are conditionally independent of the model state [16]. Following [16] we let

$$\begin{aligned} J^i(\Psi^i_{t,s}, \Psi_{t,r}) &\overset{def}{=} -\log p(\Psi_{k,r}|\psi^i_{k-1}, \Psi^i_{k,s}, d_k) & (9)\\ &= -\log p(d_k|\psi^i_{k-1}, \Psi^i_{k,s}) \\ &\quad -\log p(\Psi_{k,r}|\psi^i_{k-1}, \Psi^i_{k,s}) + \text{const.} & (10)\end{aligned}$$

If the pdf is unimodal we can set the constant term to zero and find the mode by minimizing the cost function J^i with respect to $\Psi_{t,r}$. The cost function for the example used in this paper follows on from (10) and can be written as

$$J^i(\Psi^i_{k,s}, \Psi_{k,r}) = \frac{1}{2}\left[J_o(\Psi^i_{k,s}, \Psi_{k,r}) + J_q(\Psi_{k,r})\right], \tag{11}$$

where, using (7),

$$\begin{aligned} J^i_o(\Psi^i_{k,s}, \Psi_{k,r}) &= -\log p(d_k|\psi^i_{k-1}, \Psi^i_{k,s}) \\ &= \left(d_k - \begin{pmatrix} \Psi^i_{k,s} \\ \Psi_{k,r} \end{pmatrix}\right)^T R^{-1}\left(d_k - \begin{pmatrix} \Psi^i_{k,s} \\ \Psi_{k,r} \end{pmatrix}\right), \quad (12)\end{aligned}$$

and, using (5),

$$\begin{aligned} J^i_q(\Psi_{k,r}) &= \log p(\Psi_{k,r}|\psi^i_{k-1}, \Psi^i_{k,s}) \\ &= (\Psi_{k,r} - f_r(\psi^i_{k-1}))^T[Q_{rr} - Q_{rs}Q^{-1}_{rr}Q^T_{rs}]^{-1}(\Psi_{k,r} - f_r(\psi^i_{k-1})). \quad (13)\end{aligned}$$

Step 3. At this stage in the algorithm, we have available updated values for each part of the state. In this step, the weights for the resulting ensemble members are calculated using the observation likelihood and probability of the mode tracking subspace at time k given $\Psi_{k,s}$ and the state at the previous time. The derivation of weights is discussed in [16].

Step 4. Now the particles have been weighted the PF-MT continues with a resampling step. The idea behind resampling is to produce a new ensemble of particles with the same pdf, but equal weights on each particle. In practice, particles with low weights in the original ensemble are discarded, and the sample size returned to N by creating copies of particles with higher weights. There are many resampling algorithms. Here we use a resampling scheme known as stratified resampling [8] as it is efficient and simple to implement [1]. We return to the forecast step to continue the assimilation cycle. We repeat the iteration until the final forecast time is reached.

We must now consider how to choose the state-space splitting. For the most efficient algorithm, $\Psi_{k,r}$ ostensibly should contain the maximum number of dimensions such that $p(\Psi_{k,r}|\psi^i_{k-1}, \Psi^i_{k,s}, d_k)$ is unimodal, so that the subspace that is modelled by the standard PF is as small as possible [16]. In principle, the unimodality of the mode-tracking subspace might change every time k and for each ensemble member. Vaswani [16] argues that the method should be successful if unimodality holds for most particles at most times. For our case the full pdf of the state (3), may not be unimodal at all times due to the nonlinear deterministic part of the forecast model. However the conditional pdf (8), is unimodal regardless of the choice of state splitting as the model and observation errors are Gaussian. This suggests that we should obtain the best results from the PF-MT when two states are mode tracked. In our work, we have tested a range of different choices of subspace to test this hypothesis.

3. The numerical model

The model used for our numerical experiments is a discretization of the stochastic Lorenz equations [10],

$$dX = \sigma(Y - X)dt + B_X dW_X, \tag{14}$$

$$dY = (X(\rho - Z) - Y)dt + B_Y dW_Y, \tag{15}$$

$$dZ = (XY - \beta Z)dt + B_Z dW_Z, \tag{16}$$

where X, Y and Z represent random variables and $\sigma = 10$, $\rho = 28$ and $\beta = \frac{8}{3}$. W_X, W_Y, W_Z represent independent Wiener processes.

The constants B_X, B_Y and B_Z are parameters governing the magnitude of the noise in the system. For simplicity we set each of these values equal, $B_X = B_Y = B_Z = B$. In our experiments we have set $B = 0.1$ as it allows relatively smooth solutions [11].

In our experiments, an Euler-Maruyama scheme [9] is used to solve (14)-(16) numerically, using a timestep, $\Delta = 0.01$. We run a 'twin' experiment using the first run of the model to generate a true solution and observations. We add random Gaussian noise with variance σ_o^2 to our observations giving the observation error matrix $R = \sigma_o^2 I$ where I is the 3×3 identity matrix and $\sigma_o^2 = 0.04$ is the observation error variance. These observations are then used in subsequent runs in the PF-MT. For the following experiments (14) to (16) are solved for $t \in [0, 1]$ from initial conditions $X = 0$, $Y = 0$ and $Z = 2$. These initial conditions lie in an unstable region of phase space so different solutions with the same initial conditions may evolve on different sides of the attractor due to the stochastic nature of the equations. Our implementation of the model was validated in [11].

4. Splitting the State

We wish to test the hypothesis that the best results should be obtained when we mode track the maximum number of unimodal dimensions. Following [16], we measure the performance of the filter using the root mean squared error (RMSE) of the mean of the ensemble compared to the truth run. We test the hypothesis by considering the RMSE obtained from the PF-MT after splitting the state in each possible way. We compare these RMSEs to determine which states to mode track. We also see if the best states to mode track are dependent on the number of particles used.

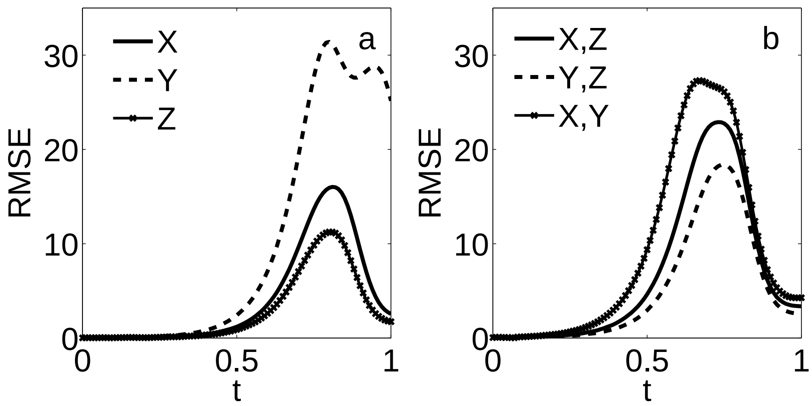


Figure 1: RMSE for different parts of the state being mode tracked with $N = 5$. a) shows the RMSE when one state is mode tracked, b) shows RMSE when two states are mode tracked.

Figure 1 shows the RMSE of the PF-MT when different states are mode tracked with the number of particles set to $N = 5$. We see that for each possible splitting the RMSE initially grows slowly. This is then followed by a period of rapid growth then decay. From Figure 1 we see that when $N = 5$ it is best to mode track $\Psi_r = Z$ and set $\Psi_s = (X, Y)$. In Table 2 we summarise

Number of Particles	Best States to Mode Track	Worst States to Mode Track
5	Z	Y
50	Z	X,Y
500	X,Z	X,Y

Table 2: The best and worst states to mode track for different numbers of particles as measured by the RMSE.

the best and worst states to mode track for various numbers of particles as measured by the RMSE. The hypothesis we are testing suggested that the best results should be obtained when the maximum number of unimodal dimensions was mode tracked. For our model the conditional pdf (8) is always unimodal suggesting that the best results should occur when we mode track two states. We can see from Table 2 that this is not always the case. For small numbers of particles the best results are obtained when only one state is mode tracked. This is perhaps due to the complicated nonlinear structure of the full problem as, while the conditional pdf is unimodal, the full pdf (3) is not. It is also possible that for small ensemble sizes the number of particles is not large enough to resolve the pdf sufficiently for mode tracking to be sensibly applied.

Additional experiments were carried out using varying numbers of particles and different initial conditions. It was found the best states to mode track depends on both the numbers of particles used and the initial conditions. For initial conditions in an unstable region of phase space, the choice of state splitting appears important as different choices for the mode tracking subspace provide results with a large range of RMSEs. However the PF-MT seems less sensitive to the choice of state splitting when the initial conditions lie in a stable region of phase space. These results suggest that both the initial conditions and number of particles must be considered when deciding how to split the state.

5. Conclusion

In this paper we have considered Vaswani's idea of mode tracking [16] that was introduced to reduce the computational cost of the particle filter. Our main focus was on how to split the state to obtain the best possible result when using the PF-MT. We tested the hypothesis that the best results should be obtained from the PF-MT when we mode track the maximum number of unimodal dimensions. When using a nonlinear model it was found that the best results from the PF-MT were not always obtained when the maximum number of unimodal dimensions was mode tracked. This was possibly due to the complicated nonlinear structure of the full problem. It was found that the best states to mode track depended on the number of particles used. It is possible that for some of the small ensemble sizes considered here, the pdf is too under-resolved for mode-tracking to be applied sensibly. Our results suggest that more research is required into how to split the state before mode tracking can be successfully used in large-scale systems.

References

[1] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Trans. Sig. Proc.*, 50(2):174–188, 2002.

[2] T. Bengtsson, C. Snyder, and D. Nychka. Toward a nonlinear ensemble filter for high-dimensional systems. *J. Geophys. Res.*, 108(D24):8775, 2003. doi:10.1029/2002JD002900.

[3] M. Buehner, P. Houtekamer, C. Charette, H. L. Mitchell, and B. He. Intercomparison of variational data assimilation and the ensemble Kalman filter for global deterministic NWP. Part I: Description and single-observation experiments. *Mon. Wea. Rev.*, 2010. doi:10.1175/2009MWR3157.1.

[4] A. Doucet, N. de Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*, chapter 1. Springer, 2000.

[5] A. Doucet, S. Godshill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, 10: 197–208, 2000.

[6] J. Harlim and B. R. Hunt. A non-Gaussian ensemble filter for assimilating infrequent noisy observations. *Tellus*, 59A:225–237, 2007.

[7] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 2002.

[8] G. Kitagawa. Monte carlo filter and smoother for non-gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5:1–25, 1996.

[9] P. E. Kloden and E. Platen. *Numerical Solution of Stochastic Differential Equations*. Springer, 1999.

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

[11] J. A. Pocock. Ensemble data assimilation: How many members do we need? Master's thesis, Department of Mathematics, University of Reading, 2009.

[12] H. Salman. A hybrid grid/particle filter for Lagrangian data assimilation. I: Formulating the passive scalar approximation. *Q. J. R. Meteorol. Soc.*, 134:1539–1550, 2008.

[13] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.

[14] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson. Obstacles to high dimensional particle filtering. *Monthly Weather Review*, pages 4629–4640, 2008.

[15] P. van Leeuwen. Particle filtering in geophysical systems. *Mon. Wea. Rev.*, 137:4089–4114, 2009.

[16] N. Vaswani. Particle filtering for large-dimensional state spaces with multimodal observation likelihoods. *IEEE Transactions on Signal Processing*, 56:4583–4597, 2008.

[17] M. Zupanski. Maximum likelihood ensemble filter: Theoretical aspects. *Mon. Wea. Rev.*, 133:1710–1726, 2005.