Computers & Geosciences 44 (2012) 70–77

Contents lists available at SciVerse ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Sequential Monte Carlo methods for parameter estimation in nonlinear state-space models

Meng Gao [a,*], Hui Zhang [b]

[a] Yantai Institute of Coastal Zone Research, CAS, Yantai 264003, China
[b] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China

ARTICLE INFO

Article history:
Received 20 November 2011
Received in revised form
9 March 2012
Accepted 13 March 2012
Available online 23 March 2012

Keywords:
Maximum likelihood
Expectation–Maximization
Markov Chain Monte Carlo
Bayesian inference

ABSTRACT

Stochastic nonlinear state-space models (SSMs) are prototypical mathematical models in geoscience. Estimating unknown parameters in nonlinear SSMs is an important issue for environmental modeling. In this paper, we present two recently developed methods that are based on the sequential Monte Carlo (SMC) method for parameter estimation in nonlinear SSMs. The first method, which belongs to classical statistics, is the SMC-based maximum likelihood estimation. The second method, belonging to Bayesian statistics, is Particle Markov Chain Monte Carlo (PMCMC). With a low-dimensional nonlinear SSM, the implementations of the two methods are demonstrated. It is concluded that these SMC-based parameter estimation methods are applicable to environmental modeling and geoscience.

1. Introduction

In the past few decades, more contemporary scientific methods have been adopted to conduct geoscience research (Zhao et al., 2009). For example, the revolution in high-performance computing and observing technologies allows mathematical models to describe the dynamic processes of the Earth system and the discovery of underlying mechanisms (Evensen, 2007). Given the complex natures of natural processes and the Earth System, mathematical models in geoscience are usually nonlinear with complex dynamical behaviors (Zhao et al., 2009). Because most mathematical models cannot be analyzed theoretically, numerical simulation models, which are discretized mathematical models, are usually used to find approximate solutions for geoscience problems. Stochastic elements also play a role in the Earth system, and these uncertainties are referred to as system noises. Moreover, measurements for state variables of mathematical models in geoscience are not free of errors. Therefore, system identification is a key issue for environmental modeling in geoscience research (Berliner et al., 2003; Wikle et al., 2003; Hansen and Penland, 2007).

Simultaneously considering the nonlinearity and uncertainty of Earth system processes, many discretized mathematical models in geoscience can be summarized as nonlinear state-space models (SSMs). SSMs, also known statistically as a hidden Markov models, provide a general framework for combining dynamic processes, system noise and measurement errors. A generic SSM consists of a state evolution model and an observation model, which can be expressed as the following:

$$x_{t+1} = f(x_t, v_t) \tag{1}$$

$$y_t = h(x_t, n_t) \tag{2}$$

where t is the time index, x_t is the state vector, and y_t is the measurement vector. v_t and n_t are independent and identically distributed random vectors representing the system noise and measurement error, respectively. When the model structure is well understood and the parameters are known, the main task of system identification is to estimate the "true" state variables x_t hiding behind the noisy observations y_t. State estimation, also known as data assimilation, is a classical research topic in geoscience. During the past few decades, various approaches to data assimilation have been developed (Kalman, 1960; Daley, 1991; Gordon et al., 1993; Evensen, 1994). The Bayesian paradigm provides a coherent probabilistic approach for data assimilation; however, the integration of Bayesian approaches into data assimilation is still in its infancy (Dowd, 2007; Wikle and Berliner, 2007). High-performance computing makes it possible to use the methods of computational statistics, especially the Monte Carlo method, to perform data assimilation. In the context of the hidden Markov model, the state transition density $p(x_{t+1}|x_t)$ and observation density $p(y_t|x_t)$ can be derived from Eqs. (1) and (2), respectively. This probabilistic framework provides the most complete and general solution to the state estimation problems.

* Corresponding author. Tel.: +86 535 2109197; fax: +86 535 210 9000.
E-mail address: mgao@yic.ac.cn (M. Gao).

0098-3004/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cageo.2012.03.013

From a Bayesian perspective, the aim of state estimation is to infer the probability function of the state variable x_t given the measurement sequence $y_{1:t}=\{y_1,y_2,\ldots,y_t\}$. It is worthwhile to note that many traditional data assimilation methods can be unified within a Bayesian framework (Wikle and Berliner, 2007).

In most geoscience models, there are a few parameters lacking a priori information; thus, it is necessary to estimate model states and unknown parameters simultaneously. However, compared with state estimation, parameter estimation is an important matter as model dynamics are usually sensitive to model parameters (Liu and West, 2001). In this study, we are concerned with estimating the static parameters of nonlinear SSMs. For the problem of parameter estimation, there exists a significant difference between classical and Bayesian statistics. Classical methods, sometimes referred to as frequentist methods in statistical literature, treat parameters as fixed, unknown constants. In this case, parameter estimation is based on the maximum likelihood method. However, maximum likelihood functions are difficult to construct and compute for a nonlinear non-Gaussian SSM (Poyiadjis et al., 2005). Maximum likelihood estimation in nonlinear SSM still remains an open problem until sequential Monte Carlo (SMC) is introduced to construct the maximum likelihood function (Poyiadjis et al., 2005; Wills et al., 2008). Bayesian methods treat parameters as random variables with prior distributions, and parameter estimation is implemented by deriving the posterior distributions. However, deriving the analytical expression of the posterior distributions is neither possible nor necessary (Andrieu et al., 2010). Recent research indicates that the SMC method also plays a very important role in Bayesian parameter estimation of nonlinear SSMs (Andrieu et al., 2010). In this paper, two recently developed batch parameter estimation methods, also known as off-line parameter estimation methods, are presented. The first method is referred to as the SMC-based maximum likelihood estimation because the output of SMC is used to compute the likelihood function. The second method involves the use of the Markov Chain Monte Carlo (MCMC) technique to implement a Bayesian inference of unknown parameters. In constructing the Markov Chain, SMC is also used. So the second method is referred to as Particle Markov Chain Monte Carlo (PMCMC). The basic ideas of these two methods belong to classical and Bayesian statistics, respectively.

The rest of this paper is organized as follows. In Section 2, we first formulate the problem of Bayesian inference in SSMs and introduce the SMC method. Then, a SMC-based maximum likelihood estimation method is presented. The basic idea and algorithms of PMCMC for parameter estimation are also described. Section 3 provides a numerical illustration of parameter estimation in a low-dimensional nonlinear SSMs using the two methods introduced in Section 2. Finally, we summarize this study in Section 4.

2. Methods

2.1. Bayesian inference in state-space model

For the Bayesian inference in SSM, the state variables are denoted as $x_{1:T}\triangleq\{x_1,x_2,\ldots,x_T\}$ and the measurements as $y_{1:T}\triangleq\{y_1,y_2,\ldots,y_T\}$, where T indicates the length of the period of interest of the SSMs. Given the observations $y_{1:T}$, simply applying Bayes' rule yields the following:

$$p(x_{1:T}|y_{1:T})=\frac{p(y_{1:T}|x_{1:T})p(x_{1:T})}{p(y_{1:T})}\propto p(y_{1:T}|x_{1:T})p(x_{1:T}) \tag{3}$$

To explicitly distinguish the problem of state estimation and parameter estimation, we use two probability density functions (pdf), $p_\theta(\cdot)$ and $p(\theta,\cdot)$, corresponding to cases whose parameters are known and unknown. First, applying a Markov assumption to the prior pdf $p_\theta(x_{1:T})$ results in

$$p_\theta(x_{1:T})=p_\theta(x_1)\prod_{t=2}^{T}p_\theta(x_t|x_{t-1}) \tag{4}$$

where $p_\theta(x_t|x_{t-1})$ is the evolution distribution. Another critical assumption is that the observations are independent given that the true model states are known. Then, the likelihood function is

$$p_\theta(y_{1:T}|x_{1:T})=\prod_{t=1}^{T}p_\theta(y_t|x_t) \tag{5}$$

Combining Eqs. (4) and (5), the posterior pdf of states becomes

$$p_\theta(x_{1:T}|y_{1:T})\propto p_\theta(x_1)\prod_{t=2}^{T}p_\theta(x_t|x_{t-1})\prod_{t=1}^{T}p_\theta(y_t|x_t) \tag{6}$$

If the parameter θ is unknown, we ascribe a prior density $p(\theta)$ to θ; then we have

$$p(\theta,x_{1:T}|y_{1:T})\propto p(\theta)p_\theta(x_1)\prod_{t=2}^{T}p_\theta(x_t|x_{t-1})\prod_{t=1}^{T}p_\theta(y_t|x_t) \tag{7}$$

Eqs. (6)–(7) provide the mathematical basis for state and parameter estimation in SSMs respectively. In this study, the method for state estimation is SMC, while the problem of parameter estimation is solved by using SMC-based maximum likelihood estimation method and the PMCMC method.

2.2. Sequential Monte Carlo method

For non-linear non-Gaussian models, deriving the analytical expressions of $p_\theta(x_{1:T}|y_{1:T})$ is nearly impossible, making Bayesian inference difficult. It is therefore necessary to resort to approximations. A discrete weighted approximation to the true posterior pdf $p_\theta(x_{1:T}|y_{1:T})$ is

$$p_\theta(x_{1:T}|y_{1:T})\approx\sum_{i=1}^{N}\omega_T^i\delta(x_{1:T}-x_{1:T}^i) \tag{8}$$

where $\{x_{1:T}^i,\omega_T^i\}_{i=1}^N$ are referred to as support points and associated weights (Arulampalam et al., 2002). $\delta(\cdot)$ is the Dirac delta function.

Using the SMC method, the approximation of $p_\theta(x_{1:t}|y_{1:t})$ can be obtained sequentially (Doucet et al., 2001). At each time step, one has samples of $p_\theta(x_{1:t-1}|y_{1:t-1})$ and wants to approximate $p_\theta(x_{1:t}|y_{1:t})$ with a new set of samples. From Eqs. (3) and (6), it is easy to check that

$$p_\theta(x_{1:t}|y_{1:t})=p_\theta(x_{1:t-1}|y_{1:t-1})\frac{p_\theta(x_t|x_{t-1})p_\theta(y_t|x_t)}{p_\theta(y_t|y_{1:t-1})}\propto p_\theta(x_{1:t-1}|y_{1:t-1})p_\theta(x_t|x_{t-1})p_\theta(y_t|x_t) \tag{9}$$

Assuming the approximate samples $\{x_{1:t-1}^i\}_{i=1}^N$ of $p_\theta(x_{1:t-1}|y_{1:t-1})$ are available at time t, then we can draw samples $\{x_t^i\}_{i=1}^N$ from the proposal density $q_\theta(\cdot|y_t,x_{1:t-1}^i)$. The importance weight of x_t^i is defined as

$$\omega_t^i=\frac{p_\theta(x_t^i|x_{t-1}^i)p_\theta(y_t|x_t^i)}{q_\theta(\cdot|y_t,x_{1:t-1}^i)}$$

If only a filtered estimate $p_\theta(x_t|y_{1:t})$ is required at each time step, a simple importance density $q_\theta(\cdot|y_t,x_{t-1}^i)$ can be used. Then, the posterior pdf of x_t can be updated without calculating the pdf of all other states $x_{1:t-1}$. This sequential updating algorithm is referred to as a *particle filter*. Then the output of the SMC algorithm is filtered particles $\{x_t^i,\omega_t^i\}_{i=1}^N$ or $\{x_{1:t}^i,\omega_t^i\}_{i=1}^N$. In practice, normalized weights $\tilde{\omega}_t^i=\omega_t^i/\sum\omega_t^j$ are more commonly used in many variants of particle filter.

The byproduct of filtering is the estimate of marginal likelihoods $p_\theta(y_{1:t})$. From Eq. (9), we have

$$\begin{aligned} p_\theta(y_t|y_{1:t-1}) &= \int p_\theta(x_t|x_{t-1})p_\theta(y_t|x_t)p_\theta(x_{1:t-1}|y_{1:t-1})\,dx_{1:t} \\ &= \int \frac{p_\theta(x_t|x_{t-1})p_\theta(y_t|x_t)}{q_\theta(\cdot|y_t,x^i_{t-1})} q_\theta(\cdot|y_t,x^i_{t-1})p_\theta(x_{1:t-1}|y_{1:t-1})\,dx_{1:t} \\ &= \int \omega^i_t q_\theta(\cdot|y_t,x^i_{t-1})p_\theta(x_{1:t-1}|y_{1:t-1})\,dx_{1:t} \end{aligned} \tag{10}$$

Then, the estimate of $p_\theta(y_t|y_{1:t-1})$ becomes

$$\hat{p}_\theta(y_t|y_{1:t-1}) \coloneqq \frac{1}{N}\sum_{i=1}^{N}\omega^i_t \tag{11}$$

and

$$\hat{p}_\theta(y_1) \coloneqq \frac{1}{N}\sum_{i=1}^{N}\omega^i_1 \tag{12}$$

Multiplying the above estimates yields

$$\hat{p}_\theta(y_{1:T}) = \hat{p}_\theta(y_1)\prod_{t=2}^{T}\hat{p}_\theta(y_t|y_{1:t-1}) \tag{13}$$

2.3. SMC-based maximum likelihood estimation

The problem of parameter estimation in nonlinear SSMs (Eqs. (1) and (2)) using the maximum likelihood method can be stated as the classical maximum log-likelihood problem

$$\hat{\theta} \triangleq \arg\max_\theta L_\theta(Y), \quad L_\theta(Y) = \log p_\theta(y_{1:T}) \tag{14}$$

Eq. (13) already provides the estimate of marginal likelihood $\hat{p}_\theta(y_{1:T})$; however, the log-likelihood function $\hat{L}_\theta(Y)$ is discontinuous with respect to θ due to Monte Carlo variation (Kantas et al., 2009). Given this variation, it is nearly impossible to use a traditional iterative gradient-based search procedure to find the optimal θ^*. Poyiadjis et al. (2005) proposed a new approach to approximate the log-likelihood gradient, and then used a general gradient-ascent algorithm to find the optimal θ^*. Their method avoids the drawback of the increasing-variance of the general method that approximates the derivative directly based on SMC (Kantas et al., 2009).

In this study, we focus on an alternative method that applies the Expectation–Maximization (EM) algorithm to solve the maximum likelihood problem. The EM algorithm for parameter estimation in nonlinear SSMs has been widely applied due to the asymptotic consistency and efficiency of the resulting estimates (Chitralekha et al., 2010). The EM algorithm includes two steps: (1) computing the expectation and (2) the maximization step (Wills et al., 2008).

The first step is to compute the expectation (E-step)

$$Q(\theta,\theta_k) \triangleq \int L_\theta(X,Y)p_{\theta_k}(X|Y)\,dX \tag{15}$$

where θ_k is the current parameter estimate. Eq. (15) can be viewed as marginalization of the missing data, X. It is convenient to verify that (Wills et al., 2008)

$$\begin{aligned} Q(\theta,\theta_k) = &\int \log p_\theta(x_1)p_{\theta_k}(x_1|y_{1:T})\,dx_1 \\ &+ \sum_{t=1}^{T-1}\int \log p_\theta(x_{t+1}|x_t)p_{\theta_k}(x_{t+1},x_t|y_{1:T})\,dx_{t:t+1} \\ &+ \sum_{t=1}^{T}\int \log p_\theta(y_t|x_t)p_{\theta_k}(x_t|y_{1:T})\,dx_t \end{aligned} \tag{16}$$

where $p_{\theta_k}(x_t|y_{1:T})$ is the smoothed density. Given the current estimate θ_k, we first apply SMC to generate filtered particles $\{x^{(i)}_t,\omega^{(i)}_t\}^N_{i=1}$. Next, we set $x^{(i)}_{T|T} = x^{(i)}_T$ and $\omega^{(i)}_{T|T} = \omega^{(i)}_t$, and then generate smoothed particles $\{x^{(i)}_{t|T},\omega^{(i)}_{t|T}\}^N_{i=1}$ $(1 \le t < T)$ based on the following recursive rule:

$$\begin{aligned} p_\theta(x_{t+1},x_t|y_{1:T}) &= p_\theta(x_t|x_{t+1},y_{1:T})p_\theta(x_{t+1}|y_{1:T}) \\ &= \frac{p_\theta(x_{t+1}|x_t)}{p_\theta(x_{t+1}|y_{1:t})}p_\theta(x_t|y_{1:t})p_\theta(x_{t+1}|y_{1:T}) \end{aligned} \tag{17}$$

The above procedure is referred to as *particle smoothing*. There are also many algorithms to implement particle smoothing in practical applications, and two of them are used in the EM algorithm (Doucet et al., 2000; Tanizaki, 2001).

With smoothed particles and normalized weights $\tilde{\omega}^{(i)}_{t|T}$, we have

$$\begin{aligned} \hat{Q}(\theta,\theta_k) = &\sum_{i=1}^{N}\tilde{\omega}^{(i)}_{1|T}\log p_\theta(x^{(i)}_{1|T}) + \sum_{t=1}^{T-1}\sum_{i=1}^{N}\tilde{\omega}^{(i)}_{t+1|T}\log p_\theta(x^{(i)}_{t+1|T}|x_t) \\ &+ \sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{\omega}^{(i)}_{t|T}\log p_\theta(y_t|x^{(i)}_{t|T}) \end{aligned} \tag{18}$$

The second step of the EM algorithm is the maximization step (maximizing $\hat{Q}(\theta,\theta_k)$ with respect to θ). First, we need to calculate the gradient of $\hat{Q}(\theta,\theta_k)$

$$\begin{aligned} \nabla_\theta\hat{Q}(\theta,\theta_k) = &\sum_{i=1}^{N}\tilde{\omega}^{(i)}_{1|T}\frac{\partial \log p_\theta(x^{(i)}_{1|T})}{\partial\theta} \\ &+ \sum_{t=1}^{T-1}\sum_{i=1}^{N}\tilde{\omega}^{(i)}_{t+1|T}\frac{\partial \log p_\theta(x^{(i)}_{t+1|T}|x_t)}{\partial\theta} \\ &+ \sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{\omega}^{(i)}_{t|T}\frac{\partial \log p_\theta(y_t|x^{(i)}_{t|T})}{\partial\theta} \end{aligned} \tag{19}$$

With $\nabla_\theta\hat{Q}(\theta,\theta_k)$, we can use a classical gradient-based searching method, such as Quasi-Newton, to find $\theta_{k+1} \triangleq \arg\max_\theta \hat{Q}(\theta,\theta_k)$. Iterating these two steps produces the estimate of the parameters. In this study, we mainly describe the major procedures; and more details of this algorithm and its applications can be found in references such as Wills et al. (2008), Gopaluni (2008), Chitralekha et al. (2010) and Schön et al. (2011).

2.4. Particle Markov chain Monte Carlo

PMCMC originates from MCMC methods, which is a class of approaches for computational Bayesian statistics (Andrieu et al., 2010). The basic idea of an MCMC is to generate a Markov Chain with a stationary distribution (target distribution) that cannot be sampled directly (Metropolis et al., 1953; Hastings, 1970; Gilks et al., 1996). For SSMs, the target distribution of a Bayesian inference is $p(x_{1:T},\theta|y_{1:T})$ when the model parameters are unknown. However, $p(x_{1:T},\theta|y_{1:T})$ cannot be sampled directly. The key feature of PMCMC is using the approximations of $p_\theta(x_{1:T}|y_{1:T})$ produced by an SMC algorithm to construct the Markov Chain with the target distribution (Andrieu et al., 2010).

The first algorithm of PMCMC presented in this study is a particle marginal Metropolis-Hastings (PMMH) sampler, which is derived from classical Metropolis–Hastings algorithm. In PMMH, the Markov Chain of $(x_{1:T},\theta)$ can be constructed by iterating the following two steps: (1) generate a new sample $(x'_{1:T},\theta')$ from the proposal density $q((\cdot,\cdot)|(x_{1:T},\theta))$; and (2) accept $(x'_{1:T},\theta')$ as the next state of the Markov Chain with the probability

$$\min\left\{1, \frac{p(x'_{1:T},\theta'|y_{1:T})}{p(x_{1:T},\theta|y_{1:T})}\frac{q((x_{1:T},\theta)|(x'_{1:T},\theta'))}{q((x'_{1:T},\theta')|(x_{1:T},\theta))}\right\} \tag{20}$$

In practice, θ' and $x'_{1:T}$ are not updated simultaneously. Given current $(x_{1:T},\theta)$, we first use a proposal $q(\cdot|\theta)$ to generate a new sample θ', and then sample $x'_{1:T} \sim p_{\theta'}(\cdot|y_{1:T})$. Now, the proposal

density becomes

$$q((x'_{1:T},\theta')|(x_{1:T},\theta)) = q(\theta'|\theta)p_{\theta'}(x'_{1:T}|y_{1:T}) \tag{21}$$

and the acceptance ratio is

$$\frac{p(x'_{1:T},\theta'|y_{1:T})}{p(x_{1:T},\theta|y_{1:T})}\frac{q((x_{1:T},\theta)|(x'_{1:T},\theta'))}{q((x'_{1:T},\theta')|(x_{1:T},\theta))} = \frac{q(\theta|\theta')}{q(\theta'|\theta)}\frac{p_{\theta'}(y_{1:T})}{p_{\theta}(y_{1:T})}\frac{p(\theta')}{p(\theta)} \tag{22}$$

It is nearly impossible to sample exactly from $p_\theta(x_{1:T}|y_{1:T})$ and to compute the marginal likelihood $p_\theta(y_{1:T})$ and $p_{\theta'}(y_{1:T})$. However, Eq. (8) indicates that $p_\theta(x_{1:T}|y_{1:T})$ can be approximated using SMC methods. At that point it is possible to obtain a new sample $x'_{1:T}$. Moreover, approximations of the marginal likelihood $p_\theta(y_{1:T})$ and $p_{\theta'}(y_{1:T})$ are also available (Andrieu et al., 2010). With these approximations, a Markov Chain can be constructed that is as simple as the classical Metropolis–Hastings algorithm.

The second algorithm of the PMCMC is the particle Gibbs (PG) sampler, which also targets $p(x_{1:T},\theta|y_{1:T})$ but does not update θ and $x_{1:T}$ jointly. The PG sampler is more complicated than the classical Gibbs sampler because a conditional SMC algorithm is used to generate the sample $x_{1:T}$ from $p_\theta(x_{1:T}|y_{1:T})$. A conditional SMC algorithm is similar to standard SMC algorithm but is such that a pre-specified particle $\tilde{x}_{1:T}$ with ancestral lineage is ensured to survive all the resampling steps, while the other $N-1$ particles are generated in the usual way. Then, the particles generated in the next step are conditional on the current particle. In this study, we merely introduce the PG algorithm but omit the conditional SMC algorithm. Interested readers may refer to Andrieu et al. (2010). The pseudocode of the PG algorithm is as follows:

(a) initialize the Markov Chain ($i=0$) by setting $\theta(i)$, $x_{1:T}(i)$ and its ancestral lineage arbitrarily,
(b) set $i=i+1$ and sample $\theta(i)$ from $p(\theta|x_{1:T}(i-1),y_{1:T})$,
(c) run a conditional SMC algorithm targeting $p_{\theta(i)}(x_{1:T}|y_{1:T})$ conditional on $x_{1:T}(i-1)$ with its ancestral lineage returning an estimate $\hat{p}_{\theta(i)}(x_{1:T}|y_{1:T})$,
(d) sample $x_{1:T}(i)$ from $\hat{p}_{\theta(i)}(x_{1:T}|y_{1:T})$ and return its ancestral lineage,
(e) iterate steps (b–d) M times and record the Markov Chain $\theta(i)$ and $x_{1:T}(i)$ ($i=0,1,\ldots,M$).

In practical applications, the convergence of the Markov Chain should be checked to ensure that the samples drawn from the Markov Chain are truly representative of the target distribution. In general, a "burn-in" period is required and the samples in this period are discarded. In practice, it is unnecessary to calculate the length of the "burn-in" period if the total Markov Chain is sufficiently long (Dowd, 2007). Although there are many methods that can be used for convergence monitoring, one of the simplest to understand and implement is the autocorrelation function (ACF). The faster the ACF drops, the better the algorithm is. For a Markov Chain generated by the PMMH algorithm, acceptance rate is also a very simple indictor. A higher acceptance rate means that the Markov Chain mixes better (Andrieu et al., 2010).

3. Numerical illustrations

In this section, we choose a low-dimensional nonlinear dynamical model to illustrate the capability of EM and PMCMC approaches for parameter estimation. The low-dimensional nonlinear dynamical model is derived from the Van del Pol oscillator, which is described by a second-order differential equation

$$\frac{d^2x}{dt^2} - \alpha(1-x^2)\frac{dx}{dt} + x = 0 \tag{23}$$

A first-order Euler discretization of the differential equation of the Van del Pol oscillator yields

$$x_{1,t+1} = x_{1,t} + hx_{2,t}$$

$$x_{2,t+1} = x_{2,t} + h\alpha(1-x_{1,t}^2)x_{2,t} - hx_{1,t} \tag{24}$$

where h is the step size. First, we assume that the Van del Pol oscillator is driven by stochastic white noises with a zero mean and a covariance matrix $\mathbf{Q} \in R^{2\times 2}$. Moreover, for simplicity, we assume that either $x_{1,t}$ or $x_{2,t}$ can be measured, and the measurement error is in the form of additive white noise with a zero mean and a covariance matrix $\mathbf{R}$. Then, the stochastic Van del Pol oscillator can be restated as a nonlinear SSM

$$x_{t+1} = f(x_t) + w_t$$

$$y_{t+1} = x_t + v_t \tag{25}$$

where $w_t \sim N(0,\mathbf{Q})$, $v_t \sim N(0,\mathbf{R})$, and $x_t = (x_{1,t},x_{2,t})$. $N(\cdot,\cdot)$ represents the normal distribution. Let $\alpha=1$ and $h=0.1$ so that the discrete-time system without stochastic noise is stable. For simplicity we assume that only $x_{2,t}$ can be observed. The noise covariance of w_t is a diagonal matrix, while the variance of v_t is a scalar variable. In this study, we set

$$\mathbf{Q} = \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} = \begin{pmatrix} 0.0262 & 0 \\ 0 & 0.008 \end{pmatrix}, \quad \mathbf{R} = \sigma_3^2 = 0.003 \tag{26}$$

With these parameters, we simulate the system (25) from an arbitrary initial state $x_{1,0} \sim N(0,0.01)$ and $x_{2,0} \sim N(0,0.01)$. System (25) will iterate 1000 times, i.e., $T=1000$. Since h is the step length, estimating h is meaningless. Then, the interesting parameters that need to be estimated are α, σ_1^2, σ_2^2 and σ_3^2. Next, we will use EM and PMCMC methods to estimate these four parameters conditional on the observed time series $y_{1:T}$.

For the EM method, the major work is to compute the approximations $\hat{Q}(\theta,\theta_k)$ and $\nabla_\theta\hat{Q}(\theta,\theta_k)$. In this study, the state variable is a vector $(x_{1,t},x_{2,t})$ and it can be easily verified that

$$p_\theta(x_{t+1}|x_t) = p(x_{1,t+1}|x_t)p(x_{2,t+1}|x_t) \tag{27}$$

and we have $p(x_{1,t+1}|x_t) \sim N(x_{1,t}+hx_{2,t},\sigma_1^2)$, $p(x_{2,t+1}|x_t) \sim N(x_{2,t}+h\alpha(1-x_{1,t}^2)x_{2,t}-hx_{1,t},\sigma_2^2)$ and $p(y_t|x_t) \sim N(x_{2,t},\sigma_3^2)$. With these normal distributions, the numerical approximations $\hat{Q}(\theta,\theta_k)$ and $\nabla_\theta\hat{Q}(\theta,\theta_k)$ can be directly computed given the smoothed particles

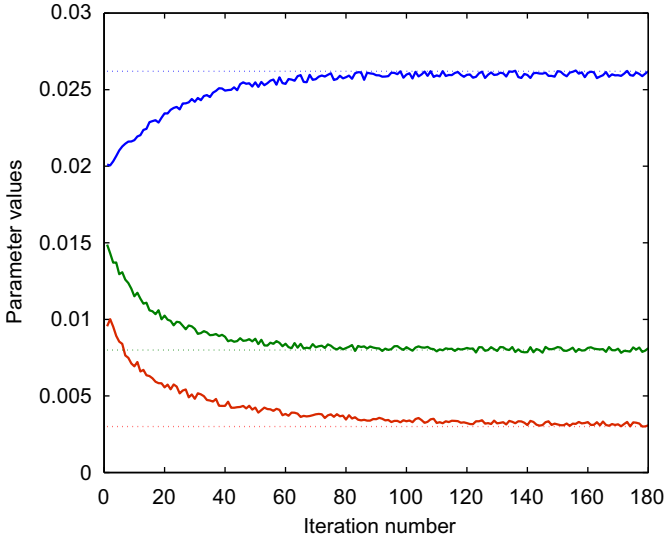

Fig. 1. Evolution of the parameter values using SMC-based maximum likelihood method (EM algorithm). From top to bottom: σ_1^2, σ_2^2 and σ_3^2. The true values $\theta^* = (0.0262,0.008,0.003)$ are marked by the dotted lines.

and their associated weights. The number of particles is chosen as $N=500$. The optimization method is the standard Quasi-Newton method. In Fig. 1, the evolution of parameters with respect to EM iterations are presented. We also replicate this numerical experiment for 100 times with different initial conditions, and the results are summarized in Table 1. It is clear that the EM method gives a satisfactory estimate.

To test the PMCMC methods, we use the same synthetic data and initial setting. Moreover, we specify the prior distribution for the unknown parameters $\alpha \sim U(0.5,1.5)$, $\sigma_1^2 \sim IG(0.5,0.01)$, $\sigma_2^2 \sim IG(0.5,0.002)$ and $\sigma_3^2 \sim IG(0.5,0.001)$. $U(c,d)$ represents a continuous uniform distribution in interval $[c,d]$, and $IG(a,b)$ is the inverse Gamma distribution with shape parameter a and scale parameter b. In the PMMH, we use a normal random-walk proposal with a diagonal covariance matrix

$$q(\theta'|\theta) \sim N(\theta,\mathbf{C}) \tag{28}$$

where $\theta=(\alpha,\sigma_1^2,\sigma_2^2,\sigma_3^2)$ represents the current parameter estimate. The diagonal elements of $\mathbf{C}$ are $(10^{-3},10^{-5},10^{-6},10^{-6})$. As the value parameters in this SSM must be positive, negative values generated by the normal random-walk proposal are omitted. In the PG algorithm, we first initialize $\theta(0)$ using the prior distribution, and run SMC to obtain a sample $x_{1:T}(0)$ from the particles ensemble $\{x_{1:T}^{(i)}\}_{i=1}^N$. Then, we use the full-conditional distributions to obtain samples of unknown parameters. The derivation of all full-conditional distributions are shown in the Appendix, and we list only the results here

$$p(\sigma_1^2|-\sigma_1^2,x_{1:T},y_{1:T}) \sim IG\left(a+\frac{T-1}{2},b+S_1\right) \tag{29}$$

Table 1
True and estimated parameter values for system (25) using the SMC-based maximum likelihood method (EM algorithm). The mean value and standard deviations are shown for the estimates based on 100 Monte Carlo runs. In each Monte Carlo simulation, the estimated parameter values is the average value of θ in the last 20 iterations.

Parameters	True values	Estimated
α	1	1.02 ± 0.037
σ_1^2	0.0262	$0.026 \pm 2.8 \times 10^{-3}$
σ_2^2	0.008	$0.008 \pm 3.3 \times 10^{-4}$
σ_3^2	0.003	$0.0031 \pm 9.2 \times 10^{-5}$

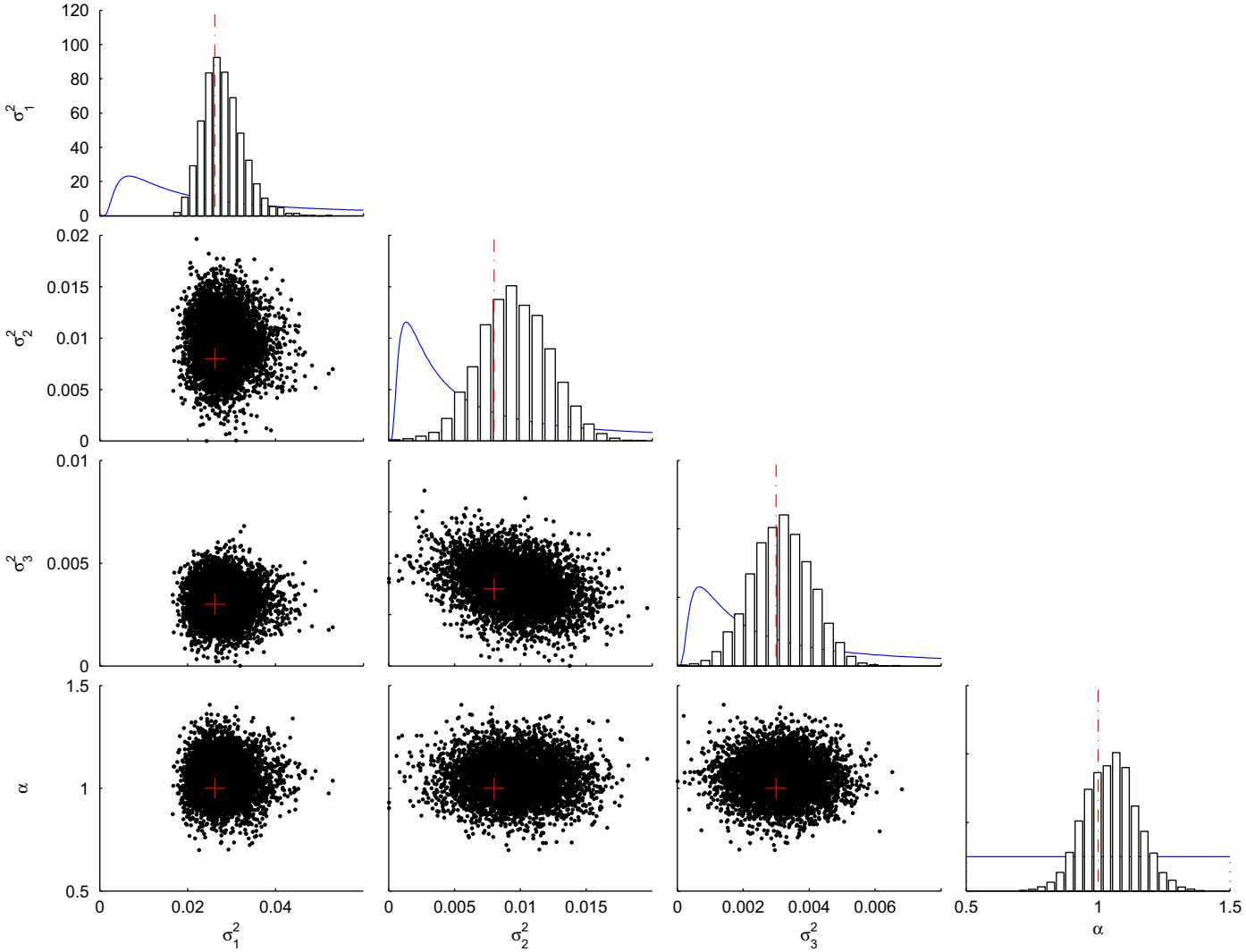

Fig. 2. Histogram approximations of the posterior densities (diagonal plots) and samples (scatter plots) of model parameters based on the output of the PMMH algorithm. In the diagonal plots, the solid lines are the prior densities $p(\theta)$, and the dash-dotted lines indicate the true values of parameters. In the scatter plots, the red crosses indicate the true values. The number of particles is 2000. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$p(\sigma_2^2|-\sigma_2^2,x_{1:T},y_{1:T}) \sim IG\left(a+\frac{T-1}{2},b+S_2\right) \tag{30}$$

$$p(\sigma_3^2|-\sigma_3^2,x_{1:T},y_{1:T}) \sim IG\left(a+\frac{T}{2},b+S_3\right) \tag{31}$$

$$p(\alpha|-\alpha,x_{1:T},y_{1:T}) \sim N_{[c,d]}\left(\frac{\sum_{t=1}^{T-1}A_tB_t}{\sum_{t=1}^{T-1}A_t^2},\frac{\sigma_2^2}{\sum_{t=1}^{T-1}A_t^2}\right) \tag{32}$$

where $N_{[c,d]}(\cdot,\cdot)$ is a truncated normal distribution within interval $[c,d]$ and the minus before a parameter indicates taking out this parameter from the parameter set θ. Other terms in Eqs. (29)–(32) are

$$S_1 = \frac{1}{2}\sum_{t=1}^{T-1}(x_{1,t+1}-x_{1,t}-hx_{2,t})^2$$

$$S_2 = \frac{1}{2}\sum_{t=1}^{T-1}(x_{2,t+1}-x_{2,t}-\alpha h(1-x_{1,t}^2)x_{2,t}+hx_{1,t})^2$$

$$S_3 = \frac{1}{2}\sum_{t=1}^{T}(y_t-x_{2,t})^2$$

$$A_t = h(1-x_{1,t}^2)x_{2,t}$$

$$B_t = x_{2,t+1}-x_{2,t}-hx_{1,t}$$

The PG algorithm will be implemented using these full-conditional distributions. Both the PMMH algorithm and the PG algorithm are run for 25,000 steps, and the first 5000 steps are discarded as burn-in steps. Andrieu et al. (2010) recommend to choose N in the same order as T. In this study, the numbers of particles are chosen as $N=1000$, 1500, 2000, 3000 in the SMC and the conditional SMC algorithms.

The marginal posterior distributions of the four parameters based on the Markov Chain (the final 5000 values) generated by the PMMH and the PG algorithms are shown in Figs. 2 and 3, respectively. Obviously, the posterior densities are different from the prior ones but close to the true values. For the PMMH, the overall acceptance rates are 0.29, 0.37, 0.39, and 0.40 when $N=1000$, 1500, 2000, 3000, respectively. Additionally, we present the ACFs of the Markov Chain of parameter α in Fig. 4. It is also verified that the performance improves as N increases. Moreover, we find that the PG algorithm performs better than the PMMH algorithm.

4. Conclusion

This paper presents two recently developed methods for parameter estimation in nonlinear state-space models. Estimating model parameters in nonlinear SSMs is a difficult task. Due to measurement errors, the true state variables can only be treated as missing values in constructing the likelihood functions. In

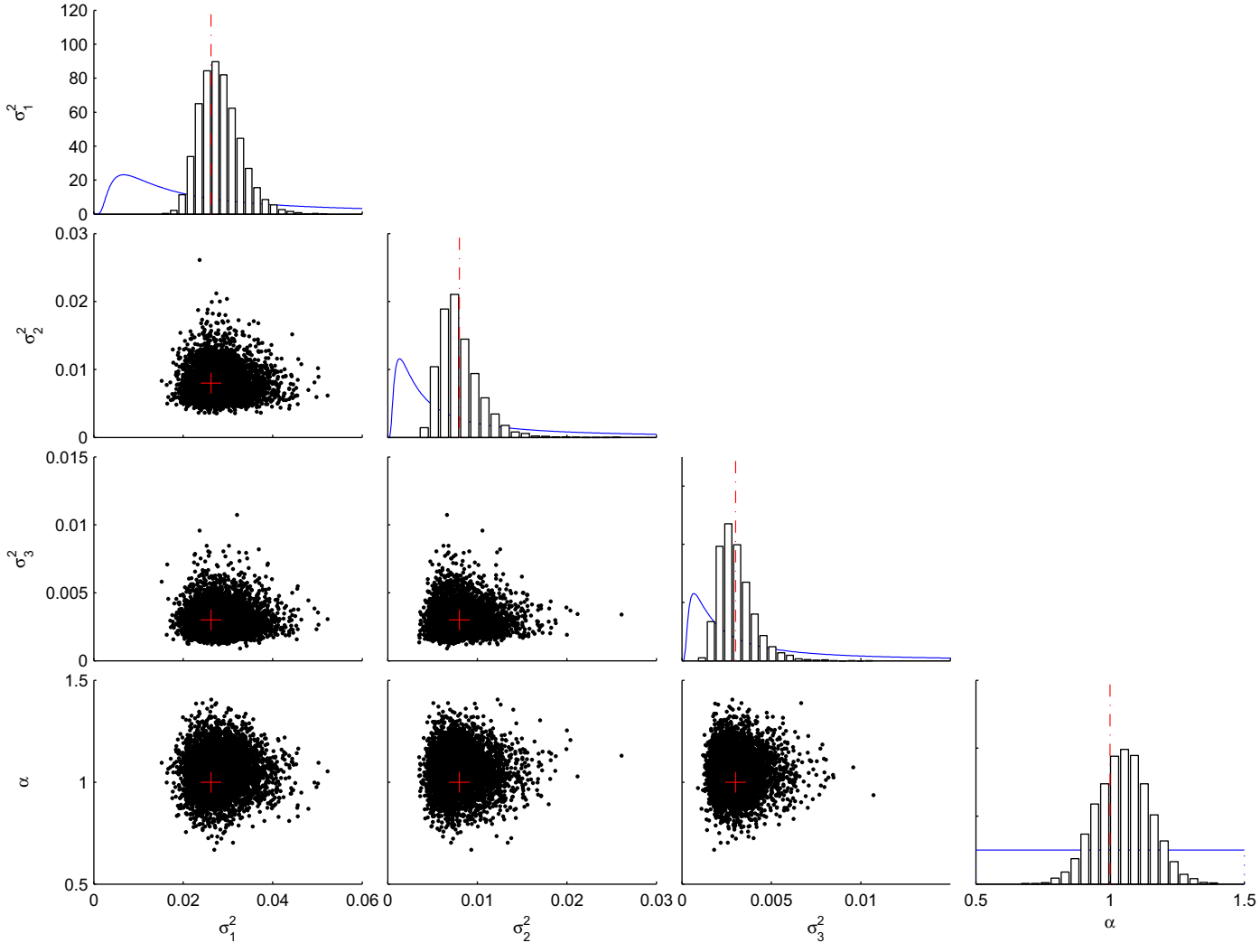

Fig. 3. Histogram approximations of the posterior densities (diagonal plots) and samples (scatter plots) of model parameters based on the output of the PG algorithm. The symbols and parameter setting are the same as that in Fig. 2.

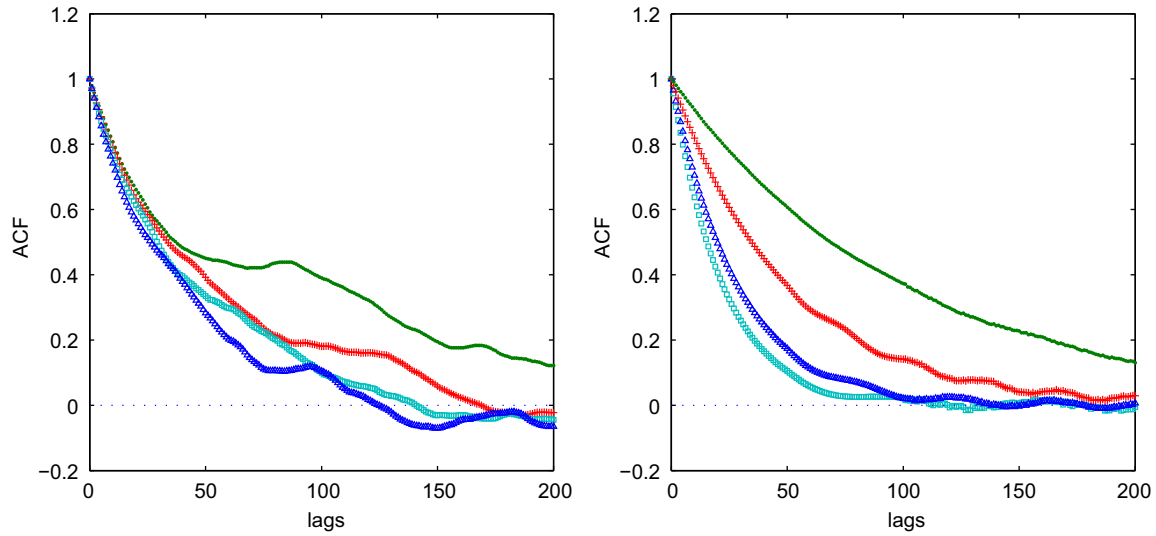

Fig. 4. Autocorrelation functions (ACFs) of α. The left panel corresponds to the PMMH algorithm, and the right panel corresponds to PG algorithm. Symbols: $N=1000$,'$\cdots$'; $N=1500$, '+'; $N=2000$, '△'; $N=3000$, '□'.

other words, parameter estimation in SSMs relies on state estimation. SMC is a standard approach to state estimation in nonlinear SSMs, and further provides the basis for parameter estimation. The two methods presented in this paper both rely on SMC. The first method uses SMC to compute the approximation of maximum likelihood, and then uses the Expectation–Maximization algorithm to find the optimal values in the global parameter space (Wills et al., 2008; Schön et al., 2011). The second method is a Bayesian inference that uses MCMC to approximate the posterior density of unknown parameters. Because SMC is used to build an efficient high-dimensional proposal distribution in each MCMC step, this method is referred to as particle Markov Chain Monte Carlo (Andrieu et al., 2010). The performance of these two methods for parameter estimation was examined with the stochastic Van del Pol oscillator. The results indicate that the two methods both perform well, although the underlying statistical framework uses frequentist and Bayesian methods. Because SMC is needed in each iteration, the computational expense of these two methods for high-dimensional state-space model in geoscience remains a limiting factor. Developing parallel simulation method on the utility of modern computing architectures, such as graphics processing units, is necessary.

Acknowledgments

This work was supported by the Knowledge Innovation Project of CAS (No. KZCX2-EW-QN209), as well as NNSF of China (No. 31000197). The authors also wish to thank Dr. Jef Caers (Editor-in-Chief) and two anonymous reviewers for their useful comments and suggestions relating to this manuscript.

Appendix A. Derivation of Eqs. (29)–(32)

We denote the state variables and process noise as vectors $x_t=(x_{1,t},x_{2,t})$ and $w_t=(w_{1,t},w_{2,t})$, then rewrite system (25) as

$$x_{1,t+1}=x_{1,t}+hx_{2,t}+w_{1,t} \tag{A.1}$$

$$x_{2,t+1}=x_{2,t}+\alpha h(1-x_{1,t}^2)x_{2,t}-hx_{1,t}+w_{2,t} \tag{A.2}$$

$$y_t=x_{2,t}+v_t \tag{A.3}$$

where $w_{1,t}\sim N(0,\sigma_1^2)$, $w_{2,t}\sim N(0,\sigma_2^2)$ and $v_t\sim N(0,\sigma_3^2)$. The prior distribution assigned to σ_i^2 $(i=1,2,3)$ is inverse gamma distribution $IG(a,b)$, then we have

$$p(\sigma_i^2)\propto(\sigma_i^2)^{-a-1}\exp\left(-\frac{b}{\sigma_i^2}\right). \tag{A.4}$$

We first show how the posterior distribution of σ_1^2 is derived,

$$\begin{aligned}p(\sigma_1^2|-\sigma_1^2,x_{1:T},y_{1:T})&=p(\sigma_1^2|x_{1:T})\propto p(\sigma_1^2)p(x_{1:T}|\sigma_1^2)\\&=p(\sigma_1^2)p(x_1)\prod_{t=1}^{T-1}p(x_{t+1}|x_t,\sigma_1^2)\\&=p(\sigma_1^2)p(x_1)\prod_{t=1}^{T-1}\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x_{1,t+1}-x_{1,t}-hx_{2,t})^2}{2\sigma_1^2}\right\}\\&\propto(\sigma_1^2)^{-a-1}\exp\left(-\frac{b}{\sigma_1^2}\right)\sigma_1^{-(T-1)}\exp\left(-\frac{S_1}{\sigma_1^2}\right)\\&\propto(\sigma_1^2)^{-(a+\frac{T-1}{2}+1)}\exp\left(-\frac{b+S_1}{\sigma_1^2}\right)\end{aligned} \tag{A.5}$$

where $S_1=\frac{1}{2}\sum_{t=1}^{T-1}(x_{1,t+1}-x_{1,t}-hx_{2,t})^2$. From Eq. (A.5), we find that the posterior distribution of σ_1^2 is an inverse gamma distribution,

$$p(\sigma_1^2|-\sigma_1^2,x_{1:T},y_{1:T})\sim IG\left(a+\frac{T-1}{2},b+S_1\right). \tag{A.6}$$

Similarly, we can derive the posterior distribution of σ_2^2 and σ_3^2.

$$p(\sigma_2^2|-\sigma_1^2,x_{1:T},y_{1:T})\sim IG\left(a+\frac{T-1}{2},b+S_2\right) \tag{A.7}$$

$$p(\sigma_3^2|-\sigma_1^2,x_{1:T},y_{1:T})\sim IG\left(a+\frac{T}{2},b+S_3\right) \tag{A.8}$$

where $S_2=\frac{1}{2}\sum_{t=1}^{T-1}(x_{2,t+1}-x_{2,t}-\alpha h(1-x_{1,t}^2)x_{2,t}+hx_{1,t})^2$ and $S_3=\frac{1}{2}\sum_{t=1}^{T}(y_t-x_{2,t})^2$.

The posterior distribution of α is also simple to obtain. We have assumed that the prior distribution of α is a continuous uniform distribution $U(c,d)$, and the posterior distribution is

$$p(\alpha|-\alpha,x_{1:T},y_{1:T})\propto p(\alpha)p_\theta(x_{1:T},y_{1:T})$$

$$
\begin{aligned}
&= p(\alpha)p(x_1)\prod_{t=1}^{T-1} p_\theta(x_{t+1}|x_t)\prod_{t=1}^{T} p_\theta(y_t|x_t)\\
&= p(\alpha)p(x_1)\prod_{t=1}^{T-1} p_\theta(x_{1,t+1}|x_t)p_\theta(x_{2,t+1}|x_t)\prod_{t=1}^{T} p_\theta(y_t|x_t)\\
&\propto p(\alpha)\prod_{t=1}^{T-1} p_\theta(x_{2,t+1}|x_t)\\
&\propto p(\alpha)(\sigma_2^2)^{-(T-1)}\exp\left\{-\frac{1}{2\sigma_2^2}\sum_{t=1}^{T-1}[A_t\alpha-B_t]^2\right\} \qquad \text{(A.9)}
\end{aligned}
$$

where $A_t = h(1-x_{1,t}^2)x_{2,t}$ and $B_t = x_{2,t+1}-x_{2,t}-hx_{1,t}$. A simple mathematical manipulation further gives

$$
p(\alpha|-\alpha,x_{1:T},y_{1:T}) \propto p(\alpha)\exp\left\{-\frac{\sum_{t=1}^{T-1}A_t^2}{2\sigma_2^2}\left[\alpha-\frac{\sum_{t=1}^{T-1}A_tB_t}{\sum_{t=1}^{T-1}A_t^2}\right]^2\right\}. \qquad \text{(A.10)}
$$

Then, we have

$$
p(\alpha|-\alpha,x_{1:T},y_{1:T}) \sim N_{[c,d]}\left(\frac{\sum_{t=1}^{T-1}A_tB_t}{\sum_{t=1}^{T-1}A_t^2}, \frac{\sigma_2^2}{\sum_{t=1}^{T-1}A_t^2}\right) \qquad \text{(A.11)}
$$

where $N_{[c,d]}(\cdot,\cdot)$ is a truncated normal distribution.

Appendix B. Supplementary data

Supplementary data associated with this article can be found in the online version at http://dx.doi.org.10.1016/j.cageo.2012.03.013.

References

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. Journal of the Royal Statistical Society: Series B 72, 269–342.

Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Transactions on Signal Processing 50, 174–188.

Berliner, L.M., Milliff, R.F., Wikle, C.K., 2003. Bayesian hierarchical modeling of air–sea interaction. Journal of Geophysical Research 108, 3104.

Chitralekha, S.B., Prakash, J., Raghavan, H., Gopaluni, R.B., Shah, S.L., 2010. A comparison of simultaneous state and parameter estimation schemes for a continuous fermenter reactor. Journal of Process Control 20, 934–943.

Daley, R., 1991. Atmospheric Data Analysis. Cambridge University Press, London.

Doucet, A., Godsill, S.J., Andrieu, C., 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. Statistics and Computation 10 (3), 197–208.

Doucet, A., de Freitas, N., Gordon, N. (Eds.), 2001. Sequential Monte Carlo in Practice. Springer-Verlag, New York.

Dowd, M., 2007. Bayesian statistical data assimilation for ecosystem models using Markov Chain Monte Carlo. Journal of Marine Systems 68, 439–456.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. Journal of Geophysical Research 99, 10143–10162.

Evensen, G., 2007. Data Assimilation: The Ensemble Kalman Filter. Springer, Berlin.

Gilks, W.R., Richardson, S., Spiegelhalter, D.J., 1996. Markov Chain Monte Carlo in Practice. Chapman and Hall, London.

Gordon, N., Salmond, D., Smith, A.F.M., 1993. Novel approach to nonlinear and non-Gaussian Bayesian state estimation. Proceedings of the Institute of Electrical and Electronics Engineers F 140, 107–113.

Gopaluni, R.B., 2008. Identification of nonlinear processes with known model structure using missing observations, in Proceedings of the 17th IFAC World Congress, Seoul, South Korea.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57, 97–109.

Hansen, G.A., Penland, C., 2007. On stochastic parameter estimation using data assimilation. Physica D 230, 88–98.

Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Transactions of the ASME: Journal of Basic Engineering 82, 35–45.

Kantas, N., Doucet, A., Singh, S., Maciejowski, J., 2009. An overview of sequential Monte Carlo methods for parameter estimation in general state-space models. In: Proceedings of the IFAC Symposium on System Identification (SYSID).

Liu, J., West, M., 2001. Combined parameter and state estimation in simulation-based filtering. In: Doucet, A., de Freitas, N., Gordon, N. (Eds.), Sequential Monte Carlo in Practice. Springer-Verlag, New York, pp. 197–223.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087–1091.

Poyiadjis, G., Doucet, A., Singh, S.S., 2005. Maximum likelihood parameter estimation using particle methods, in: Proceedings of the Joint Statistical Meeting.

Schön, T.B., Wills, A., Ninness, B., 2011. System identification of nonlinear state-space models. Automatica 47 (1), 39–49.

Tanizaki, H., 2001. Nonlinear and non-Gaussian state space modeling using sampling techniques. Annals of the institute of Statistical Mathematics 53 (1), 63–81.

Wikle, C.K., Berliner, L.M., Milliff, R.F., 2003. Hierarchical Bayesian approach to boundary value problems with stochastic boundary conditions. Monthly Weather Review 131, 1051–1062.

Wikle, C.K., Berliner, L.M., 2007. A Bayesian tutorial for data assimilation. Physica D 230, 1–16.

Wills, A.G. , Schön, T.B., Ninness, B., 2008. Parameter estimation for discrete-time nonlinear systems using EM, in: Proceedings of the 17th IFAC World Congress, Seoul, South Korea.

Zhao, C., Hobbs, B.E., Ord, A., 2009. Fundamentals of Computational Geoscience: Numerical Methods and Algorithms. Springer.