Bayesian reconstruction of chaotic dynamical systems

Renate Meyer[1,*] and Nelson Christensen[2,†]
[1]Department of Statistics, The University of Auckland, Auckland, New Zealand
[2]Physics and Astronomy, Carleton College, Northfield, Minnesota 55057

(Received 31 March 2000; revised manuscript received 23 May 2000)

We present a Bayesian approach to the problem of determining parameters of nonlinear models from time series of noisy data. Recent approaches to this problem have been statistically flawed. By applying a Markov chain Monte Carlo algorithm, specifically the Gibbs sampler, we estimate the parameters of chaotic maps. A complete statistical analysis is presented, the Gibbs sampler method is described in detail, and example applications are presented.

PACS number(s): 05.45.Tp, 02.70.Lq, 06.20.Dk, 02.60.Pn

I. INTRODUCTION

Many observed time series stemming from physical laboratory experiments or ''real world'' systems exhibit a very complex and apparently random time behavior that may be explained by an underlying chaotic process. By a chaotic process we mean a nonlinear dynamical system [1–4], i.e., a discrete time series of unknown (due to noise) system states x_i, $i=1,\ldots,N$, that are nonlinear functions of previous states $x_i=g(x_{i-1})$. Various statistical approaches have been suggested to reconstruct the underlying nonlinear dynamics from a time series of noisy observations [5–9]. These are based on estimating the unknown parameters that define the nonlinear function, and comprise maximum-likelihood methods, Bayesian techniques, and approaches based on minimizing a certain cost function.

On the one hand, likelihood techniques aim at finding the values of the parameters that maximize the likelihood function, the joint probability density function (PDF) of the observations given the unknown parameters. The Bayesian approach is also based on the likelihood function but treats the parameters as random variables and assumes a joint prior distribution that summarizes the available information about the parameters before observing the data. In the light of the observations, the information about the unknown parameters is then updated via the Bayes theorem to the posterior distribution, which is proportional to the product of likelihood and prior density [10].

On the other hand, heuristic approaches based on minimization of a certain cost function make no distributional assumptions. The most prominent are least squares (LS) methods [11,12] that minimize the sum of squared one-step prediction errors, and total least squares (TLS) techniques [13,14], and modifications thereof [15]. In a recent paper [5], McSharry and Smith give an overview of various cost functions that have been used in reconstructing nonlinear dynamics and display their shortcomings in simulation studies. It is well known that LS estimates give systematically wrong results due to ignoring the facts that (i) the values of the ''independent variable'' are subject to measurement error, the so-called errors-in-variables bias, and (ii) there is serial correlation between successive observations, the so-called time-series bias.

To reduce errors-in-variables bias, the errors-in-variables regression or TLS cost function was first proposed by Kostelich [16]. Instead of minimizing the sum of squared vertical distances, the TLS technique aims at minimizing the sum of squared perpendicular distances between two consecutive observations to corresponding pair of points on the hypersurface defined by the nonlinear dynamics. From a statistical point of view, this is justified only if pairs are independent. But this is clearly not the case in a time series.

McSharry and Smith [5] propose a different cost function which they somewhat misleadingly call the ''maximum-likelihood (ML) cost function.'' They demonstrate that the estimator based on minimizing this cost function outperforms the LS and TLS methods and gives nearly unbiased parameter estimates even for large noise levels. Notwithstanding, our main criticism is that the derivation of this estimator is based on yet another *ad hoc* cost function instead of a sound statistical paradigm. Furthermore, we point out major flaws in its derivation. We suggest a Bayesian approach instead, by integrating the problem into the framework of nonlinear state-space modeling [17,18]. This alleviates both problems (i) and (ii) by incorporating the known serial correlation as prior information in a complete probability model for the observations and the unknown states. We even consider the more realistic generalization where the underlying dynamic evolution is not assumed deterministic but is subject to unpredictable external or environmental effects, so-called dynamic noise. Difficulties with posterior computations are overcome using Markov chain Monte Carlo (MCMC) techniques, in particular the Gibbs sampler in conjunction with the Metropolis-Hastings (MH) algorithm [19].

The paper is organized as follows. In Sec. II we give a motivation for this Bayesian state-space approach by considering the recently proposed approach via maximization of a cost function and its shortcomings. In the third section, the general Bayesian approach to statistical inference for state-space models is presented. After that, we point out similarities between McSharry and Smith's ML cost function and the posterior density employed in the Bayesian approach, thus offering an explanation for the good performance of McSharry and Smith's estimator. The use of MCMC simu-

*Email address: meyer@stat.auckland.ac.nz
†Email address: nchriste@carleton.edu

lation techniques, specifically of the Gibbs sampler, is explained in Sec. IV. In Sec. V, we show the superior performance of this Bayesian technique using the same examples as in Ref. [5] for comparative purposes. We conclude with a discussion of the flexibility of a Bayesian state-space approach.

II. A STATE-SPACE APPROACH

Like McSharry and Smith [5], we consider the situation where we are given a time series of noisy observations y_i, $i=1,\ldots,N$. These are modeled as *conditionally* independent (given the underlying unknown true system states x_i) and normally distributed random variables, i.e.,

$$y_i|x_i=x_i+v_i, \quad v_i \overset{iid}{\sim} N(0,\epsilon^2), \ \ i=1,\ldots,N, \tag{1}$$

with known error variance ϵ^2. The time evolution of the system states is described by a nonlinear function $x_i = f(x_{i-1},a), i=1,\ldots,N$, where a is a p-dimensional parameter and x_0 a starting value. As an example consider the logistic map $x_i=1-ax_{i-1}^2$ with one-dimensional parameter a. The likelihood function, i.e., the joint PDF of all observables $\mathbf{y}=(y_1,\ldots,y_N)$ given all unknowns, is thus

$$L(a,x_0|\mathbf{y})=p(\mathbf{y}|a,x_0)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\epsilon}} \times\exp\left(-\frac{1}{2\epsilon^2}[y_i-f^i(x_0,a)]^2\right), \tag{2}$$

where $f^i(x_0,a)=x_i$ is the i-fold composition of f. Here and in the following we use $p(\cdot)$ as a generic symbol for a PDF.

McSharry and Smith give the conditional bivariate PDF of a pair of two consecutive observations (y_i,y_{i+1}) as $p(y_i,y_{i+1}|a,x_i)=(1/2\pi\epsilon^2)\exp(-(1/2\epsilon^2)\{(y_i-x_i)^2+[y_{i+1}-f(x_i,a)]^2\})$. This is perfectly correct, since, with independent noise v_i and v_{i+1} in Eq. (1), y_i and y_{i+1} are also independent, despite the authors' statement in [5] that "... the weakest link in the derivation of the ML cost function is the assumption that the y_i and y_{i+1} are independent; while they may be linearly uncorrelated, they cannot be independent. Attempts to relax this assumption will be presented in later work... ." No attempts are needed. However, the authors then incorrectly assume that the conditional *joint* PDF of a sequence of $N-1$ pairs $\mathbf{Y}=\{(y_i,y_{i+1})\}_{i=1,\ldots,N-1}$, given parameter a and a sequence of system states $\mathbf{x}=(x_1,\ldots,x_N)$, is $p(\mathbf{Y}|a,\mathbf{x})=\Pi_{i=1}^{N-1}p(y_i,y_{i+1}|a,x_i)$, even though every consecutive pair (y_i,y_{i+1}) and (y_{i+1},y_{i+2}) is obviously violating the independence assumption. In contrast to Eq. (2), they suggest a "pseudo"-likelihood function given by

$$L(a|\mathbf{Y})=\prod_{i=1}^{N-1}\int_x p(y_i,y_{i+1}|a,x)d\mu(x,a) \tag{3}$$

obtained by integrating over the system's invariant measure $\mu(x,a)$. They proceed by minimizing what they call the ML cost function,

$$C_{ML}(a)=-\sum_{i=1}^{N-1}\ln\int\exp\left(-\frac{1}{2\epsilon^2}\{(y_i-x)^2+[y_{i+1}-f(x,a)]^2\}\right)d\mu(x,a), \tag{4}$$

where the integral in practice is replaced by a sum over a model trajectory.

It should be pointed out that the idea behind integrating the dependency on the x_i's out of the pseudo-likelihood function is very similar to calculating the marginal posterior PDF of a by integrating the joint posterior PDF over all unknown states $x_0,x_1,\ldots,x_N$. To develop this idea within a proper statistical paradigm requires treating the system states as stochastic instead of deterministic. We therefore consider the more realistic case that the system dynamics are subject to random disturbances. This casts the problem into the general framework of a Bayesian state-space model [17,18], one of the most powerful tools for dynamic modeling and forecasting. State-space models relate time-series observations to unobserved states by a stochastic observation model. The states are assumed to follow a stochastic transition over time, given by the state equations. The state equations, i.e., the *conditional* distribution of the system state at time i, given the previous states and unknown parameters, are

$$x_i|x_{i-1},a=f(x_{i-1},a)+u_i, \quad u_i\overset{iid}{\sim}N(0,\tau^2), \ \ i=1,\ldots,N \tag{5}$$

and the observation equations are given by (1). This state-space approach eliminates the errors-in-variables bias and time-series bias mentioned before. It takes the temporal dependencies of the observations into account through a conditional modeling of the observations, given unknown states, and specification of Markovian transition of states. Via the Bayesian paradigm, both process and observation errors are explicitly captured and quantified through posterior distributions of the parameters, as described in the next section.

III. BAYESIAN INFERENCE FOR STATE-SPACE MODELS

The starting point of the Bayesian approach to statistical inference is setting up a full probability model that consists of the joint probability distribution of all observables, denoted by $\mathbf{z}=(z_1,\ldots,z_n)$, and unobservable quantities, denoted by $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_d)$. Using the notion of conditional probability, this joint PDF $p(\mathbf{z},\boldsymbol{\theta})$ can be decomposed into the product of the PDF of all unobservables, $p(\boldsymbol{\theta})$, referred to as the *prior* PDF of $\boldsymbol{\theta}$, and the conditional PDF of the observables given the unobservables, $p(\mathbf{z}|\boldsymbol{\theta})$, referred to as the sampling distribution or *likelihood*, i.e.,

$$p(\mathbf{z},\boldsymbol{\theta})=p(\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta}).$$

The prior PDF contains all the information about the unobservables that is known from substantive knowledge and expert opinion before observing the data. All the information about the $\boldsymbol{\theta}$ that stems from the experiment is contained in the likelihood. In the light of the data, the Bayesian paradigm then updates the prior knowledge about $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$, to the *posterior* PDF of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}|\mathbf{z})$. This is done via an application of Bayes theorem through conditioning on the observations

$$p(\boldsymbol{\theta}|\mathbf{z})=\frac{p(\boldsymbol{\theta},\mathbf{z})}{m(\mathbf{z})}\propto p(\boldsymbol{\theta})p(\mathbf{z}|\boldsymbol{\theta}),$$

where $m(\mathbf{z})=\int p(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the marginal PDF of $\mathbf{z}$, which can be regarded as a normalizing constant as it is independent of $\boldsymbol{\theta}$.

In the state-space model defined through the observation equations (1) and the state equations (5) with $f(x_{i-1},a)=1-ax_{i-1}^2$ given by the logistic map, the observables corresponding to $\mathbf{z}$ are $\mathbf{y}=(y_1,\ldots,y_N)$ with $n=N$, and the unobservables corresponding to $\boldsymbol{\theta}$ are $(x_0,x_1,\ldots,x_N,a,\tau^2)$ with $d=N+3$. To elicit a joint prior distribution for the unobservables, $x_0,x_1,\ldots,x_N,a,\tau^2$, we make use of successive conditioning and the Markovian nature of state transitions by writing

$$p(x_0,x_1,\ldots,x_N,a,\tau^2)=p(a,\tau^2)p(x_0)\prod_{i=1}^{N}p(x_i|x_{i-1},a,\tau^2),$$

where the conditional prior density of $x_i|x_{i-1},a,\tau^2$ is defined through Eq. (5) with some small but unknown error variance τ^2. We assume a noninformative prior distribution for x_0, i.e., $x_0\sim$ Uniform$[-1,1]$. We assume prior independence of a and τ^2 and use a noninformative prior for a. Reflecting our prior expectation that there is only small dynamic noise, we assume a vague prior inverse-gamma ($\alpha=2.01,\beta=0.005\,05$) distribution for τ^2 which has mean 0.005 and standard deviation 0.05. By Bayes theorem, the joint posterior density of all unobservables is proportional to prior $\times$ likelihood:

$$\begin{aligned}p(x_0,\mathbf{x},a,\tau^2|\mathbf{y})&\propto\frac{1}{\tau^{2(N+1)}}\exp\left[-\frac{1}{2\tau^2}\left(x_0^2+\sum_{i=1}^{N}[x_i-f(x_{i-1},a)]^2\right)\right]\\&\quad\times\tau^{-2(\alpha+1)}\exp(-\beta/\tau^2)\\&\quad\times\exp\left(-\frac{1}{2\epsilon^2}\sum_{i=1}^{N}(y_i-x_i)^2\right).\end{aligned}\tag{6}$$

It is worthwhile to compare the posterior density in Eq. (6) with McSharry and Smith's ML cost function in Eq. (4). Terms in the exponential of Eq. (6) corresponding to likelihood contributions for (y_i,y_{i+1}) and prior contributions for (x_{i+1},x_{i+2}),

$$\cdots-\frac{1}{2\epsilon^2}[(y_i-x_i)^2+(y_{i+1}-x_{i+1})^2]-\frac{1}{2\tau^2}\{[x_{i+1}-f(x_i,a)]^2+[x_{i+2}-f(x_{i+1},a)]^2\}-\cdots,$$

have their counterparts in Eq. (4) given by

$$-\frac{1}{2\epsilon^2}\{\cdots+(y_i-x_i)^2+(y_{i+1}-x_{i+1})^2+[y_{i+1}-f(x_i,a)]^2+[y_{i+1}-f(x_{i+1},a)]^2+\cdots\}.$$

It now becomes evident that by considering pairs of observations terms in the ML cost function are artificially blown up to mimic terms in the posterior PDF. Thus one would expect the minimum of C_{ML} to be close to the posterior mode. These considerations give an explanation for the good performance of the ML cost function suggested by McSharry and Smith.

IV. BAYESIAN POSTERIOR COMPUTATION VIA MCMC

The main difficulty with the Bayesian approach is high-dimensional integration. To calculate the normalizing constant of the joint posterior PDF, for instance, requires d-dimensional integration. Having obtained the joint posterior PDF of $\boldsymbol{\theta}$, the posterior PDF of a single parameter θ_i of interest can be obtained by integrating out all the other components, i.e.,

$$p(\theta_i|\mathbf{z})=\int\cdots\int p(\boldsymbol{\theta}|\mathbf{z})d\theta_1\cdots d\theta_{i-1}d\theta_{i+1}\cdots d\theta_d.$$

Calculation of the posterior mean of θ_i necessitates a further integration, e.g., $E[\theta_i|\mathbf{z}]=\int\theta_i p(\theta_i|\mathbf{z})d\theta_i$.

In the example above, finding the marginal posterior PDF of the parameter a, for instance, would require integration over $x_0,x_1,\ldots,x_N,\tau^2$, i.e., the calculation of an $(N+2)$-dimensional integral. For large N, numerical integration becomes unfeasible, so that the only alternative is simulating vectors $(a,x_0,x_1,\ldots,x_N,\tau^2)^j$, $j=1,\ldots,J$, for some large J from the posterior PDF. The first component of each sampled vector $(a,x_0,x_1,\ldots,x_N,\tau^2)^j$ constitutes a sample from the marginal posterior of a. Then any characteristic of the marginal posterior distribution of a can be approximated by its sample equivalent, e.g., its mean by the sample mean.

As the joint posterior is too complex to sample from directly, we propose to use the MCMC technique [19,20]. In MCMC, a Markov chain is constructed with the joint posterior as its equilibrium distribution. Thus, after running the Markov chain for a certain ''burn-in'' period, one obtains (correlated) samples from the limiting distribution, provided that the Markov chain has reached convergence. We use the Gibbs sampler, a specific MCMC method that samples iteratively from each of the univariate full conditional posterior distributions

$$p(\theta_i|\mathbf{z},\theta_1,\ldots,\theta_{i-1},\theta_{i+1},\ldots,\theta_d).\tag{7}$$

Given an arbitrary set of starting values $\theta_1^{(0)},\ldots,\theta_d^{(0)}$ the algorithm proceeds as follows:

$$\begin{aligned}&\text{simulate } \theta_1^{(1)}\sim p(\theta_1|\mathbf{z},\theta_2^{(0)},\ldots,\theta_d^{(0)})\\&\text{simulate } \theta_2^{(1)}\sim p(\theta_2|\mathbf{z},\theta_1^{(1)},\theta_3^{(0)},\ldots,\theta_d^{(0)})\\&\quad\vdots\\&\text{simulate } \theta_d^{(1)}\sim p(\theta_d|\mathbf{z},\theta_1^{(1)},\ldots,\theta_{d-1}^{(1)})\end{aligned}\tag{8}$$

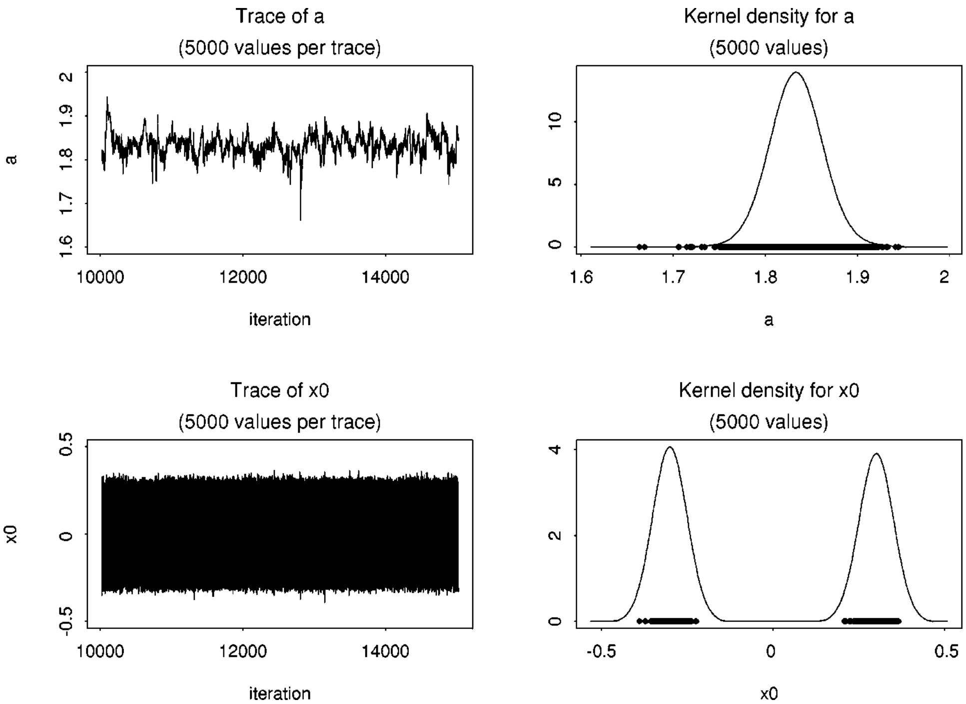

FIG. 1. Trace and kernel density plots of the marginal posterior distributions for the parameters a and x_0 based on 100 observations from the logistic map with true parameters $a=1.85$, $x_0=0.3$, and noise level 0.6.

and yields $\boldsymbol{\theta}^{(m)}=(\theta_1^{(m)},\ldots,\theta_d^{(m)})$ after m such cycles. This defines a Markov chain with transition kernel

$$k(\boldsymbol{\theta}^{(m+1)},\boldsymbol{\theta}^{(m)})=\prod_{i=1}^{d} p(\theta_i^{(m+1)}|\mathbf{y},\times\theta_1^{(m+1)},\ldots,\theta_{i-1}^{(m+1)},\theta_{i+1}^{(m)},\ldots,\theta_d^{(m)}), \tag{9}$$

which converges to the joint posterior as its equilibrium distribution [19]. Consequently, if all the full conditional posterior distributions are available, all that is required is sampling iteratively from these. Thereby, the problem of sampling from a d-variate PDF is reduced to sampling from d univariate PDFs.

In many applications where the prior PDF is conjugate to the likelihood, the full conditionals in fact reduce analytically to closed-form PDFs and we can use highly efficient special purpose Monte Carlo methods for generating from these (see, e.g., [21]). In general, however, we need a fast and efficient black-box method to sample from an arbitrarily complex full conditional posterior distribution in each cyclic step of the Gibbs sampler. Such an all-purpose algorithm, called *adaptive rejection sampling* (ARS), was developed by Gilks and Wild [22,23] for the rich class of distributions with

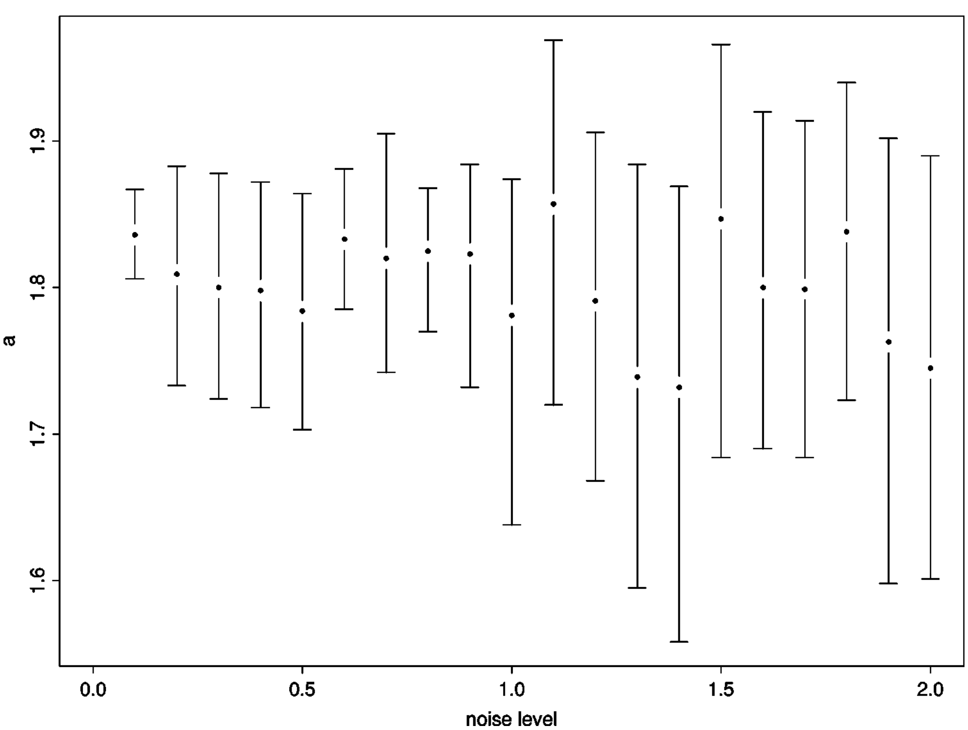

FIG. 2. Posterior means and 95% posterior probability intervals for increasing noise levels based on 100 observations from the logistic map with true parameters $a=1.85$ and $x_0=0.3$.

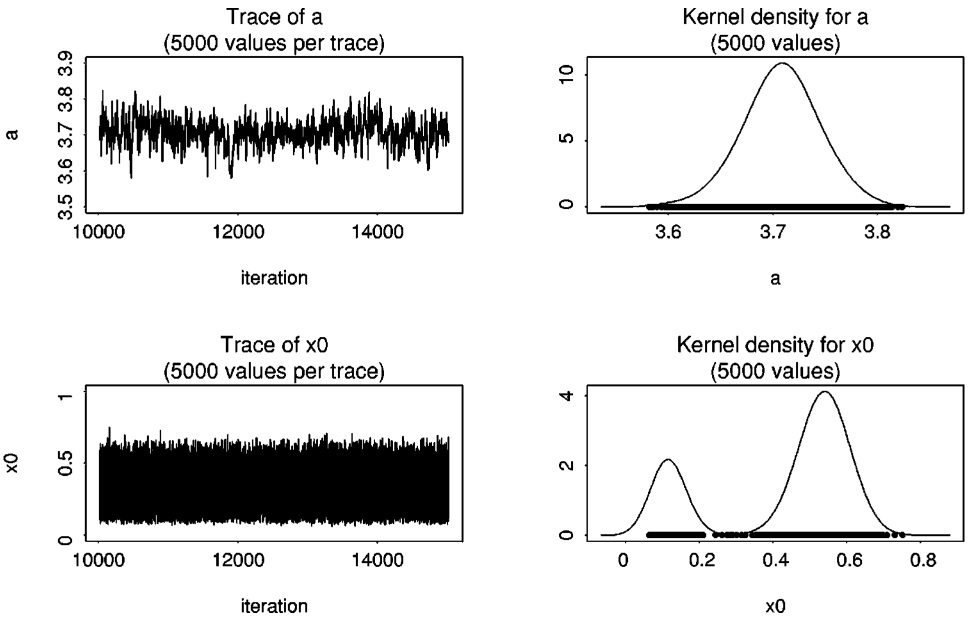

FIG. 3. Trace and kernel density plots of the marginal posterior distributions for the parameters a and x_0 based on 100 observations from the Moran-Ricker map with true parameters $a=3.7$, $x_0=0.5$, and noise level 0.2.

log-concave densities. Due to nonlinearity in the state equations, however, the full conditional densities of the x_i and a are not log-concave. But we can use a recently developed ''Metropolized'' version of adaptive rejection sampling (ARMS) for non-log-concave distributions [24,25]. C subroutines of ARS and ARMS are available [24] and can thus be tailored to the full conditional posteriors of a nonlinear state-space model.

Significant progress has been made in facilitating the routine implementation of the Gibbs sampler with the help of BUGS (Bayesian inference using Gibbs sampling), a recently developed software package [26] by the Medical Research Council Biostatistics Unit, Institute of Public Health, Cambridge, England.[1] BUGS samples from the joint posterior distribution by using the Gibbs sampler. For reviews on BUGS the reader is referred to [27–29].

BUGS can handle the two main tasks necessary for implementation of the Gibbs sampler. These tasks are (i) to construct and (ii) to sample from the full conditional posterior densities. BUGS can perform these tasks for a variety of complex models such as random effects, generalized linear, proportional hazards, latent variable, and frailty models. As shown by Meyer and Millar [30,31], state-space models are also amenable to a Bayesian analysis via BUGS. This has the major advantage that no one-off program in a low-level programming language such as C or FORTRAN needs to be written for every new analysis. Only the prior and sampling distributions for unobservables and observables, respectively, have to be specified in a BUGS program. The tedious task of constructing the full conditionals is automated by BUGS using directed acyclic graphs [32]. Sophisticated routines such as adaptive rejection sampling to sample from log-concave full conditionals and MH algorithms based on slice sampling to sample from non-log-concave full conditional densities have been implemented in BUGS and are continuously being refined. Furthermore, various methods to assess convergence, i.e., methods used for establishing whether a MCMC algorithm has converged and whether its output can be regarded as samples from the target distribution of the Markov chain, have been developed and implemented in CODA [33]. CODA is a menu-driven collection of SPLUS functions for analyzing the output obtained from BUGS. As well as trace plots and the usual tests for convergence, CODA calculates statistical summaries of the posterior distributions and kernel density estimates.[2]

V. EXAMPLES

A. Logistic map

We simulated $N=100$ observations from Eq. (1) and the underlying system evolution given by the logistic map $x_i = 1 - a x_{i-1}^2$ with starting value $x_0=0.3$, parameter $a=1.85$, and noise levels $l=\sigma_{noise}/\sigma_{signal}$ ranging from 0 to 2. To obtain a sample from the posterior distribution in the logistic map example, we performed 110 000 cycles of the Gibbs sampler, used a burn-in period of 10 000 iterations, and thinned the chain by taking every 20th observation to avoid highly correlated values. This yielded a final sample size of 5000 and took 10 min on a Pentium III PC. Convergence diagnostics [33] confirmed that the Markov chain had converged toward its equilibrium distribution. Figure 1 shows an exemplary trace and kernel density plot for the parameters a and x_0 for an error level of 0.6. Figure 2 displays the posterior mean of the parameter a together with 95% credibility intervals for varying degrees of noise levels. A comparison with Fig. 2 of McSharry and Smith [5] shows an equivalent performance of the Bayesian estimator compared to the one obtained by minimizing the ML cost function. One should note, however, that in order to obtain 95% confidence inter-

[1] BUGS is available free of charge from http://www.mrc-bsu.cam.ac.uk/bugs/Welcome.html for the operating systems UNIX, LINUX, and WINDOWS, among others. It comes with complete documentation and two example volumes.

[2] CODA is maintained and distributed by the same research group responsible for BUGS.

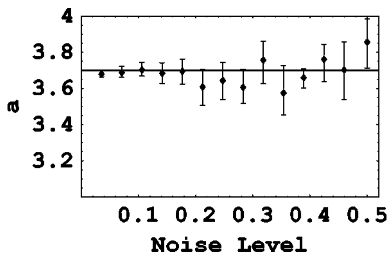

FIG. 4. Posterior means and 95% posterior probability intervals for increasing noise levels based on 100 observations from the Moran-Ricker map with true parameters $a=3.7$ and $x_0=0.5$.

vals, the approach via minimization of an *ad hoc* cost function requires rerunning the optimization algorithm some large number of times, say 1000 times as did McSharry and Smith for the logistic map, and then using the 2.5 and 97.5 percentiles of the minima as lower and upper limits. In contrast, 95% credibility intervals as well as other measures of dispersion such as the standard deviation can be extracted from the MCMC output simultaneously with the posterior mean, for statistical theory ensures that averaging of a function of interest over realizations from a single run of the Markov chain provides a consistent estimate of its expectation. Thus, the Bayesian approach provides a parameter estimate *and* an assessment of its uncertainty at the same time.

The BUGS code for estimating the parameters of the logistic map can be downloaded from [34].

B. Moran-Ricker map

The versatility and ease of usage becomes even more evident when one wants to fit a different nonlinear map or use a different data set. Only one or two lines need to be changed in the BUGS code. We simulated 100 observations from the Moran-Ricker map $f(x,a)=x\exp[a(1-x)]$ with true values of $a=3.7$, $x_0=0.5$, and noise levels ranging from 0 to 0.5. We used noninformative priors again to make the Bayesian results comparable to the frequentist estimates. For instance, with a noise level of 0.2, we obtain a posterior mean of 3.707, and a 95% credibility interval of [3.634,3.777] which is basically identical to the one in [5]. Figure 3 shows the corresponding trace and kernel density plots of a and x_0. Figure 4 displays the posterior mean of the parameter a together with 95% credibility intervals for varying degrees of noise levels. A comparison with Fig. 3 of McSharry and Smith [5] shows an equivalent performance of the Bayesian estimator compared to the one obtained by minimizing the ML cost function.

C. Henon map

Similarly, the results for the two-dimensional Henon map $f(x_i,x_{i-1})=1-ax_i^2+bx_{i-1}$ with true parameters $a=1.4$, $b=0.3$, and noise level 0.05 based on 500 observations, compare with those of McSharry and Smith with posterior means and standard deviations of $1.392\pm0.017\,26$, and $0.296\pm0.016\,93$ for a and b, respectively. Figure 5 shows the corresponding trace and kernel density plots for the parameters a and b.

D. Other Applications

MCMC techniques have been applied in numerous areas, from science to economics. For higher-level tasks (such as multivariate state-space models with hundreds to thousands of parameters [35]), BUGS reaches its limits, and software specifically designed for the problem in question needs to be written. It should be noted that applications of state-space modeling in finance, e.g., stochastic volatility models applied to time series of daily exchange rates or returns of stock exchange indices, easily have 1000–5000 parameters and the Gibbs sampler shows slow convergence due to high posterior correlations [32,36,37]. Specially tailored MCMC algorithms, like multimove Gibbs samplers or Metropolis-Hastings algorithms, can markedly improve the speed of convergence [38]. We are also continuing our work on the implementation of a MCMC scheme for identifying the parameters from coalescing binary stars [20] as observed from the continuous data output of the soon to be completed Laser Interferometric Gravity Wave Observatory [39].

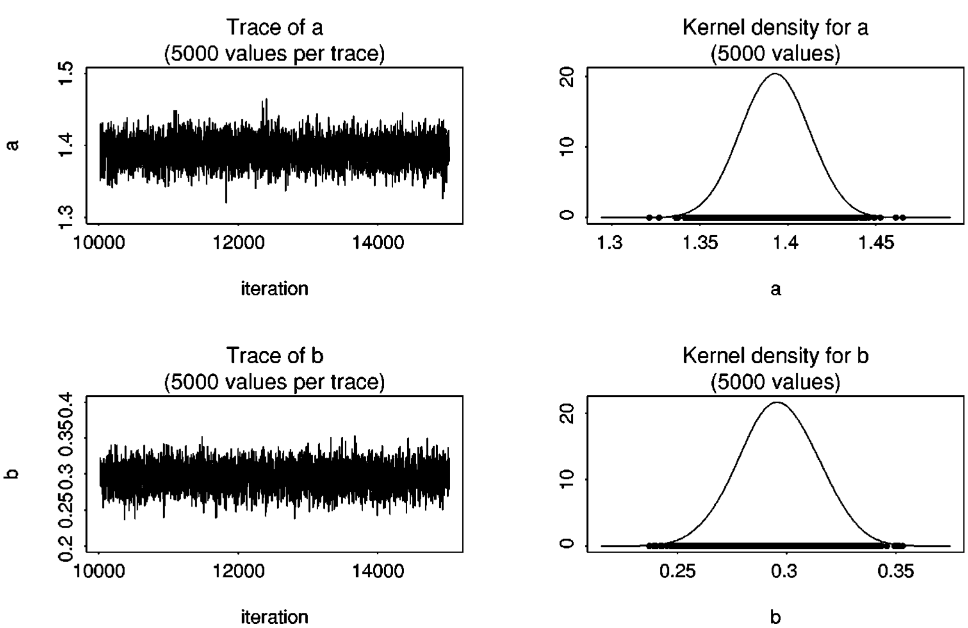

FIG. 5. Trace and kernel density plots of the marginal posterior distributions for the parameters a and b based on 500 observations from the Henon map with true parameters $a=1.4$, $b=0.3$, and noise level 0.05.

Model comparison is another application of MCMC methods. If we desire to compare two models M_0 and M_1, then the Bayes factor $B_{01}=p(\mathbf{z}|M_1)/p(\mathbf{z}|M_0)$ used for model determination is the ratio of the marginal likelihoods $p(\mathbf{z}|M_k)=\int p(\mathbf{z}|\boldsymbol{\theta}_k,M_k)p(\boldsymbol{\theta}_k|M_k)d\boldsymbol{\theta}_k$, for the data $\mathbf{z}$, parameters $\boldsymbol{\theta}_k$ of the model M_k, and prior density $p(\boldsymbol{\theta}_k|M_k)$ [40]. The multidimensional integrals required to calculate the marginal likelihoods are best accomplished via the MCMC technique when the number of parameters gets large [41].

VI. DISCUSSION

When using noninformative prior distributions, the Bayesian approach to parameter estimation in nonlinear models from time series of noisy data gives similar results to those obtained by McSharry and Smith via the maximization of a certain cost function. Note, however, that the Bayesian state-space approach provides various advantages. First of all, it is based on an unassailable statistical paradigm rather than optimization of an *ad hoc* cost function. Due to recent revolutionary advances in Bayesian posterior computation via computer-intensive MCMC simulation techniques, difficulties with posterior computations can be overcome. A Bayesian state-space model is readily implemented using standard Bayesian software such as BUGS. One can therefore avoid writing one-off programs in a low-level language. Any modifications, such as different prior distributions, applications to different data sets, or the use of different sampling distributions, require the change of just a single line in the code. In addition to the ease of implementation and its flexibility, state-space modeling in BUGS is far more versatile in that it does not require the unrealistic assumption of *known* error variance ϵ^2. The state-space approach allows one to simultaneously estimate both dynamic and measurement noise. Furthermore, it is not restricted to the assumption of Gaussian noise. A heavy-tailed observation error distribution such as a Student *t* distribution with large degrees of freedom might be more appropriate to allow for crude measurement errors and ensure that resulting estimates will be robust against additive outliers. Non-Gaussian error distributions are readily incorporated in BUGS. We therefore strongly advocate the Bayesian state-space approach via Gibbs sampling in practical analyses of chaotic time series.

ACKNOWLEDGMENTS

This work was supported by the Royal Society of New Zealand Marsden Fund, the University of Auckland Research Committee, and Carleton College.

[1] P. Collet and J.-P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser, Berlin, 1980).

[2] J.-P. Eckmann and D. Ruelle, Rev. Mod. Phys. **57**, 617 (1985).

[3] D. Ruelle, *Chaotic Evolution and Strange Attractors* (Cambridge University Press, New York, 1989).

[4] R.L. Devaney, *Introduction to Chaotic Dynamical Systems* (Benjamin-Cummings, Menlo Park, CA, 1989).

[5] P.E. McSharry and L.A. Smith, Phys. Rev. Lett. **83**, 4285 (1999).

[6] M.E. Davies, Chaos **8**, 775 (1998).

[7] E.J. Kostelich and J.A. Yorke, Phys. Rev. A **38**, 1649 (1988).

[8] S.M. Hammel, Phys. Lett. A **148**, 421 (1990).

[9] M. Berliner, J. Am. Stat. Assoc. **86**, 938 (1991).

[10] B.P. Carlin and T.A. Louis, *Bayes and Emperical Bayes Methods for Data Analysis* (Chapman and Hall, London, 1996).

[11] P. Grassberger, T. Schreiber, and C. Schaffrath, Int. J. Bifurcation Chaos Appl. Sci. Eng. **1**, 521 (1991).

[12] H.D.I. Abarbanel, R. Brown, J.J. Sidorowich, and L.Sh. Tsimring, Rev. Mod. Phys. **65**, 1331 (1993).

[13] E.J. Kostelich and T. Schreiber, Phys. Rev. E **48**, 1752 (1993).

[14] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem* (SIAM, Philadelphia, 1991).

[15] L. Jaeger and H. Kantz, Chaos **6**, 440 (1996).

[16] E.J. Kostelich, Physica D **58**, 138 (1992).

[17] L. Fahrmeir and G. Tutz, *Multivariate Statistical Modelling Based on Generalized Linear Models* (Springer, New York, 1994).

[18] B.P. Carlin, N.G. Polson, and D.S. Stoffer, J. Am. Stat. Assoc. **87**, 493 (1992).

[19] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).

[20] N.L. Christensen and R. Meyer, Phys. Rev. D **58**, 082001 (1998).

[21] L. Devroye, *Non-Uniform Random Variate Generation* (Springer, New York, 1986).

[22] W.R. Gilks and P. Wild, Appl. Stat. **41**, 337 (1992).

[23] W.R. Gilks, in *Bayesian Statistics 4*, edited by J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith (Oxford University Press, Oxford, 1992), pp. 89–114.

[24] W.R. Gilks, N.G. Best, and K.K.C. Chan, Appl. Stat. **44**, 455 (1995).

[25] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).

[26] D.J. Spiegelhalter, A. Thomas, N. Best, and W.R. Gilks, *BUGS 0.5, Bayesian Inference Using Gibbs Sampling. Manual (version ii)* (MRC Biostatistics Unit, Cambridge, England, 1996).

[27] A. Thomas, D.J. Spiegelhalter, and W.R. Gilks, in *Bayesian Statistics 4* (Ref. [23]), pp. 837–842.

[28] W.R. Gilks, A. Thomas, and D.J. Spiegelhalter, The Statistician **43**, 169 (1994).

[29] R. Gentleman, Chance **10**, 48 (1997).

[30] R. Meyer and R.B. Millar, Can. J. Fish. Aquat. Sci. **56**, 1078 (1999).

[31] R.B. Millar and R. Meyer, Appl. Stat. **49**, 327 (2000).

[32] R. Meyer, in *Data Analysis, Classification, and Applications*, edited by W. Gaul, O. Opitz, and M. Schader (Springer, Berlin, in press).

[33] N.G. Best, M.K. Cowles, and S.K. Vines, *CODA Manual Version 0.30* (MRC Biostatistics Unit, Cambridge, England, 1995).

[34] http://www.stat.auckland.ac.nz/meyer/rmeyer.html

[35] R.B. Millar and R. Meyer, Can. J. Fish. Aquat. Sci. **57**, 43 (2000).

[36] E. Jacquier, N.G. Polson, and P.E. Rossi, J. Bus. Econ. Stat. **12**, 371 (1994).
[37] R. Meyer and J. Yu (unpublished).
[38] S. Kim, N. Shephard, and S. Chib, Rev. Econ. Stud. **65**, 361 (1998).
[39] A. Abramovici, W.E. Althouse, R.W.P. Drever, Y. Grsel, S. Kawamura, F.J. Raab, D. Shoemaker, L. Sievers, R.E. Spero, K.S. Thorne, R.E. Vogt, R. Weiss, S.E. Whitcomb, and M.E. Zucker, Science **256**, 325 (1992).
[40] A.E. Raftery, *W.R. Gilks, S. Richardson, and D.J. Spiegelhalter* (Chapman and Hall, London, 1996).
[41] C. Han and B.P. Carlin, e-print http://www.maths.surrey.ac.uk/personal/st/S.Brooks/MCMC/.