

Research Article

Baisen Liu, Liangliang Wang and Jiguo Cao*

Bayesian estimation of ordinary differential equation models when the likelihood has multiple local modes

<https://doi.org/10.1515/mcma-2018-0010>

Received November 15, 2017; accepted March 28, 2018

Abstract: Ordinary differential equations (ODEs) are popularly used to model complex dynamic systems by scientists; however, the parameters in ODE models are often unknown and have to be inferred from noisy measurements of the dynamic system. One conventional method is to maximize the likelihood function, but the likelihood function often has many local modes due to the complexity of ODEs, which makes the optimizing algorithm be vulnerable to trap in local modes. In this paper, we solve the global optimization issue of ODE parameters with the help of the Stochastic Approximation Monte Carlo (SAMC) algorithm which is shown to be self-adjusted and escape efficiently from the “local-trapping” problem. Our simulation studies indicate that the SAMC method is a powerful tool to estimate ODE parameters globally. The efficiency of SAMC method is demonstrated by estimating a predator-prey ODEs model from real experimental data.

Keywords: Dynamical model, stochastic approximation Monte Carlo, global optimization, system identification

MSC 2010: 62F15

1 Introduction

Ordinary differential equations are often used to model the rate of change of a dynamic process in time and/or space (expressed as derivatives). They are widely applied to describe complex dynamic systems in many areas of science and technology, such as engineering, physics, economics, pharmacokinetics, neurophysiology, and systems biology. The forms of ODEs are usually proposed based on the expert knowledge of the dynamic systems and scientific principles, such as conservation of mass and energy, and the parameters in these ODEs generally have scientific interpretations. On the other hand, the values of these parameters are typically unknown. One of the central problems in using ODEs is to estimate these parameters from the measurements of these dynamic systems in the presence of measurement errors.

Some challenges exist in estimating ODE parameters. First of all, most ODEs are nonlinear and, with a few exceptions, have no analytic solutions. Many methods have been developed to solve ODEs numerically, such as the Euler method and the Runge–Kutta method [5]. In addition, the ODE solutions are often sensitive

Baisen Liu, School of Statistics, Dongbei University of Finance and Economics, Dalian, 116025, P. R. China, e-mail: bsliu2014@dufe.edu.cn

Liangliang Wang, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, V5A1S6, Canada, e-mail: liangliang_wang@sfu.ca

***Corresponding author: Jiguo Cao**, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, V5A1S6, Canada, e-mail: jiguo_cao@sfu.ca

to the values of ODE parameters. Consequently, the likelihood surface has many local modes, which will be illustrated in our application and simulation studies. The goal of this paper is to solve the global optimization issue with the help of Bayesian approaches.

A host of statistical approaches have been proposed to estimate ODE parameters from noisy data. Bard [1] and Biegler, Damiano and Blau [3] introduced a nonlinear least squares method, which searched for the optimal values of ODE parameters by optimizing the fitting of the numerical solutions of the ODEs to the data. Optimization is usually carried out with gradient-based methods such as the Newton–Raphson method, but often suffers from convergence to local modes.

Varah [29] suggested a two-step estimating procedure based on some classic nonparametric smoothing methods, which was later developed by Ramsay and Silverman [24], Poyton [21], Chen and Wu [11], and Brunel [4]. This method avoids solving ODE numerically so that the burden of intensive-computations is reduced. However, the estimated ODE parameters often have a large bias because the ODEs did not be involved in when estimating the derivatives. Ramsay, Hooker, Campbell and Cao [23] developed a generalized profiling method, which used a nonparametric function to represent the dynamic process. This method was shown to be able to obtain good estimates of the ODE parameters with the low computational load. Qi and Zhao [22] proved that the generalized profiling estimates for ODE parameters are asymptotic efficient. Cao, Wang and Xu [10] proposed a robust estimation method for estimating ODE parameters. Cao, Huang and Wu [9] extended the generalized profiling method to estimate the time-varying parameters in ODEs. On the other hand, the generalized profiling method usually uses gradient-based methods and is hard to obtain globally optimized estimates for ODE parameters when the likelihood has multiple local modes.

Recently, Bayesian methodology is quickly developed and has been applied in numerous fields because it can answer complex questions cleanly and exactly and provide more intuitive and meaningful inference [14]. The classic Markov chain Monte Carlo (MCMC) method in Bayesian statistics is then applied to infer the ODE parameters from noisy data [13, 17]. However, it is well known that the classical MCMC method is vulnerable to be trapped in local modes when the likelihood surface is rugged, i.e., has a lot of local modes. Moreover, in high dimensions of many parameters to sample, the random walk becomes inefficient due to low rates of acceptance, poor mixing of the chain and highly correlated samples. To overcome these obstacles, Calderhead and Girolami [6] proposed a population-based MCMC sampling procedure called parallel tempering, which enabled the sampler to efficiently escape local posterior modes and hence worked well for sampling from multi-modal distributions. However, as stated in [7, Section 2.1], the posterior flattening strategies may lead to slower mixing and larger burn-in in the sampling process. Moreover, parallel tempering may fail if the prior information does not agree with the features of the observed data. As a remedy, Campbell and Steele [7] proposed a smooth functional tempering, which combines parallel tempering and model-based smoothing to define a sequence of approximations to the posterior. These methods treat the temperature as an auxiliary variable. For high temperatures, the proposal distribution is broadened so that the problem of convergence to local modes is mitigated by searching larger regions of the sampling space.

Different from the above approaches, in this article, we directly pursue the issue of maximizing the log-likelihood function of ODE parameters from the point view of Monte Carlo optimization. The Monte Carlo optimization approach enjoys a long history and numerous Monte Carlo optimization approaches have been developed, which include the gradient method, simulated annealing [15, 26], and the stochastic approximation Monte Carlo (SAMC) method [19]. The gradient method requires a precise knowledge of the target function and easy to trap in the local modes, while the main difficulty of using simulated annealing lies in choosing the cooling temperature schedule.

In contrast, the SAMC method utilizes the past samples and, essentially, is a dynamic importance sampling algorithm in which the trial distribution is learned dynamically from past samples. The SAMC method can self-adjust to escape from “trapping in” local multi-modes by partitioning sample space, setting desired sampling distribution, and choosing appropriate gain factors. It has been shown to be an extremely efficient tool to solve the “local-trap” optimization problem. In this paper, we develop the SAMC method to sample the posterior distribution of ODE parameters. In our simulation studies and real data analysis, we show that the SAMC method works very well for inference of complex nonlinear ODEs models and its implementation is very easy and convenient.

The remainder of this article is organized as follows. Section 2 introduces a Bayesian model for statistical inference of ODE parameters, which is estimated with the SAMC method. The SAMC method is then demonstrated in Section 3 by estimating a predator-prey dynamic model from real experimental data. Simulation studies are presented in Section 4 to illustrate the advantage of the SAMC method in comparison with the MCMC method. Conclusions are given in Section 5.

2 Methodology

Consider the following dynamic ODEs model:

$$\frac{dX_i(t)}{dt} = g_i(\mathbf{X}(t)|\boldsymbol{\beta}), \quad t \in [T_s, T_e], \quad i = 1, \dots, I, \quad (2.1)$$

where $\mathbf{X}(t) = (X_1(t), \dots, X_I(t))^T$ denotes the vector of ODE variables and $\boldsymbol{\beta}$ is an unknown vector of parameters of the ODEs model. Let $X_i(t|\boldsymbol{\theta})$, $i = 1, \dots, I$, denote the solutions of the ODEs of the form (2.1), where $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T \in \Theta$ with the ODE parameter values as $\boldsymbol{\beta}$ and the initial conditions as $\boldsymbol{\alpha}$. Let $\mathbf{y} = (y_{11}, \dots, y_{In_I})^T$ be the vector of all observations and let y_{ij} denote the observation for the i -th ODE variable at t_{ij} , $j = 1, \dots, n_i$, $i = 1, \dots, I$, which is assumed to follow some probability distribution $f(y|\boldsymbol{\theta})$, for example, the normal distribution with the mean $X_i(t_{ij}|\boldsymbol{\theta})$ and the variance σ_i^2 .

A popular approach to estimate $\boldsymbol{\theta}$ based on \mathbf{y} is to maximize the likelihood function

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^I \prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\theta}).$$

Under the Gaussian assumption of $y_{ij} \sim N(X_i(t_{ij}|\boldsymbol{\theta}), \sigma_i^2)$, $i = 1, \dots, I$, $j = 1, \dots, n_i$, the likelihood function of $\boldsymbol{\theta}$ based on \mathbf{y} is given by

$$L(\boldsymbol{\theta}) = \prod_{i=1}^I \prod_{j=1}^{n_i} (\sigma_i^2)^{-\frac{1}{2}} \exp\left\{-\frac{(y_{ij} - X_i(t_{ij}|\boldsymbol{\theta}))^2}{2\sigma_i^2}\right\}. \quad (2.2)$$

Define

$$U(\boldsymbol{\theta}) = \sum_{i=1}^I \sum_{j=1}^{n_i} \frac{\{y_{ij} - X_i(t_{ij}|\boldsymbol{\theta})\}^2}{\sigma_i^2}; \quad (2.3)$$

then finding the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$ based on (2.2) is equivalent to minimizing the function $U(\boldsymbol{\theta})$.

Adopting the idea of Monte Carlo optimization, instead of minimizing (2.3) by classical approaches (e.g., gradient-based methods), in this article, we suggest to apply the SAMC method to simulate a trial distribution which is proportional to $\exp\{-U(\boldsymbol{\theta})\}$. Firstly, we partition the domain space Θ into some subregions and seek to draw samples from each of the subregions with a pre-specified frequency. If this goal can be achieved, then the local-trap problem can be avoided successfully. Assume that the parameter space Θ is partitioned into M disjoint subregions, which are denoted by

$$\begin{aligned} E_1 &= \{\boldsymbol{\theta} : U(\boldsymbol{\theta}) \leq u_1\}, \\ E_2 &= \{\boldsymbol{\theta} : u_1 < U(\boldsymbol{\theta}) \leq u_2\}, \\ &\vdots \\ E_{M-1} &= \{\boldsymbol{\theta} : u_{M-2} < U(\boldsymbol{\theta}) \leq u_{M-1}\}, \\ E_M &= \{\boldsymbol{\theta} : U(\boldsymbol{\theta}) > u_{M-1}\}, \end{aligned}$$

where u_1, \dots, u_{M-1} are pre-specified real numbers by users. In practice, the maximum difference in each subregion should be bounded by a reasonable number, say, 2, which ensures that the local MH moves within the same subregion have a reasonable acceptance rate.

Define $\mathbf{g} = (g_1, \dots, g_M)^T$, where $g_m = \int_{E_m} \exp\{-U(\boldsymbol{\theta})\} d\boldsymbol{\theta}$ for $m = 1, \dots, M$. To present the idea clearly, we temporarily assume that $g_m > 0$ for all $m = 1, \dots, M$, but it is allowed for some subregions with $g_m = 0$ in practice. Define a pre-specified frequency, say $\boldsymbol{\pi} = (\pi_1, \dots, \pi_M)$ with $0 < \pi_m < 1$ and $\sum_{m=1}^M \pi_m = 1$. Generally,

a uniform sequence of $\pi_m = \frac{1}{M}$, $m = 1, \dots, M$, is chosen, as in all examples of this article. If this goal can be achieved, then the local-trap problem is avoided essentially. To achieve this goal, we try to sample θ from the following trial distribution:

$$f_g(\theta) \propto \sum_{m=1}^M \frac{\pi_m \exp\{-U(\theta)\}}{g_m} I(\theta \in E_m),$$

where $I(\cdot)$ is an indicator function.

Obviously, the value of g_m affects the probability of θ being sampled in the subregion E_m at each iteration. If a subregion is visited, say, E_m , then \mathbf{g} will be updated according to some mechanism (see the details discussed later) such that the subregion E_m has a smaller probability to be revisited and other subregions have larger probabilities to be visited in the next iteration. This mechanism enables the algorithm to escape from local multi-mode very quickly.

In practice, g_m is always unknown in sampling implementation, but it can be estimated together with sampling θ iteratively. Hence, the whole sampling procedure consists of two steps: sampling in Step 1 and updating weights in Step 2. More detailed, let $g_m^{(k)}$ denote the estimate of g_m at the k -th iteration, $\mathbf{g}^{(k)} = (g_1^{(k)}, \dots, g_M^{(k)})^T$, and $\theta^{(k)}$ denote the sample of θ at the k -th iteration; then we perform the following procedures in the $(k+1)$ -st iteration:

(a) Sampling: Sample $\theta^{(k+1)}$ by a single Metropolis–Hastings update from the distribution

$$f_g(\theta) \propto \sum_{m=1}^M \frac{\pi_m \exp\{-U(\theta^{(k)})\}}{g_m^{(k)}} I(\theta \in E_m)$$

in the following three steps:

- (1) Generate θ^* in the sample space Θ according to a proposal distribution $q(\theta^*; \theta^{(k)})$.
- (2) Calculate the ratio

$$r = \frac{f_g(\theta^*)}{f_g(\theta^{(k)})} \frac{q(\theta^{(k)}; \theta^*)}{q(\theta^*; \theta^{(k)})}.$$

- (3) Set

$$\theta^{(k+1)} = \begin{cases} \theta^*, & \text{with the probability } \min(1, r), \\ \theta^{(k)}, & \text{otherwise.} \end{cases}$$

(b) Weight update: Set

$$\mathbf{g}^{(k+1)} = \mathbf{g}^{(k)} \exp\{\gamma^{(k)}(\mathbf{e}^{(k)} - \boldsymbol{\pi})\},$$

where $\gamma^{(k)}$ is called the gain factor in the context of stochastic approximation [25]. In practice, as in this article, we often choose

$$\gamma^{(k)} = \frac{T_0}{\max(T_0, k)}, \quad k = 1, 2, \dots, \quad (2.4)$$

for some specified value of $T_0 > 1$. The indicator vector $\mathbf{e}^{(k)} = (e_1^{(k)}, \dots, e_M^{(k)})^T$ with $e_m^{(k)} = 1$ if $\theta^{(k+1)} \in E_m$ and 0 otherwise.

A large value of T_0 will force the sampler to reach all subregions quickly, even in the presence of multiple local modes. Therefore, T_0 should be set to a large value for a complex problem. For the nonempty subregion E_m , let f_m be the realized sampling frequency, and let \bar{f} be the average sampling frequency. Define

$$\epsilon_f = \min \left\{ \frac{f_m}{\bar{f}} : m = 1, \dots, M, E_m \neq \emptyset \right\}.$$

An appropriate choice of T_0 and the total iteration number N are chosen such that the sampling frequency of each nonempty subregion is not less than 80 % of the average sampling frequency, that is, $\epsilon_f \geq 80\%$. Once a run was checked not to converge, we re-run the above iterations with a larger value of N and/or a larger value of T_0 . In this article, the following scheme is adopted to update T_0 and N : the number of total iteration N is increased to $2N$, and T_0 defined in (2.4) is increased to $1.5T_0$.

Our method can give some estimation of the mean square error which cannot be made smaller. Our method first partitions the parameter space according to the range of the mean square errors, $U(\theta)$, and then seeks to draw samples from each of the subregions with a pre-specified frequency. At the same time, our

method allows for some subregions to be empty, i.e., some subregions are never visited in a long sampling run. For example, if the subregions

$$\begin{aligned} E_1 &= \{\boldsymbol{\theta} : U(\boldsymbol{\theta}) \leq u_1\}, \\ E_2 &= \{\boldsymbol{\theta} : u_1 < U(\boldsymbol{\theta}) \leq u_2\}, \\ &\vdots \\ E_K &= \{\boldsymbol{\theta} : u_{K-1} < U(\boldsymbol{\theta}) \leq u_K\} \end{aligned}$$

are never visited in a long run, and $E_{K+1} = \{\boldsymbol{\theta} : u_K < U(\boldsymbol{\theta}) \leq u_{K+1}\}$ is visited with some samples, then we can claim that the minimum mean square error is probable between u_K and u_{K+1} .

3 Application

It is of great interest in ecology to study the predator-prey interactions among species [18, 28]. Nonlinear ODE models display a similar set of dynamic behaviors as ecological populations, such as coexistence at an equilibrium and a limit cycle [2], and hence are popularly used to model the predator-prey dynamic systems [20].

For example, an aquatic laboratory community containing two microbial species has studied in [12, 27, 30]. This dynamic system is a nutrient-based predator-prey food chain, in which the growths of unicellular green algae, *Chlorella vulgaris*, are limited by the supply of nitrogen, and *Chlorella* are eaten by planktonic rotifers, *Brachionus calyciflorus*. The prey, *Chlorella*, and the predator, *Brachionus*, are growing together in replicated, experimental flow-through cultures, called chemostats. Nitrogen continuously flows into the system with the concentration N^* at the dilution rate δ , and all components of the dynamic system are removed from the chemostats at the same rate δ .

Fussmann, Ellner, Shertzer and Hairston Jr. [12] proposed a set of nonlinear ODEs to model consumer-resource interactions between *Chlorella*, *Brachionus*, and the nitrogen resource. The nonlinear ODEs model can be expressed as follows:

$$\begin{aligned} \frac{dN(t)}{dt} &= \delta(N^* - N(t)) - F_C(N(t))C(t), \\ \frac{dC(t)}{dt} &= F_C(N(t))C(t) - \frac{F_B(C(t))B(t)}{\epsilon} - \delta C(t), \\ \frac{dR(t)}{dt} &= F_B(C(t))R(t) - (\delta + m + \alpha)R(t), \\ \frac{dB(t)}{dt} &= F_B(C(t))R(t) - (\delta + m)B(t), \end{aligned} \quad (3.1)$$

where $N(t)$, $C(t)$, $R(t)$, $B(t)$ are concentrations of nitrogen, *Chlorella*, reproducing *Brachionus*, and total *Brachionus*, respectively,

$$F_C(N) = \frac{b_C N}{k_C + N} \quad \text{and} \quad F_B(C) = \frac{b_B C}{k_B + C}$$

are two functional responses (with b_C and b_B the maximum birth rates of *Chlorella* and *Brachionus*; k_C and k_B the half-saturation constants of *Chlorella* and *Brachionus*), and ϵ , α , and m are the assimilation efficiency, the decay of fecundity, and the mortality of *Brachionus*, respectively. The seven parameters ϵ , α , m , b_C , b_B , k_C , and k_B in the above ODEs model all have interesting biological interpretations, but their values are unknown and need to be estimated from measurements of the dynamic system.

Figure 1 displays the experimental measurements for the concentrations of *Chlorella* and *Brachionus* in the predator-prey dynamic system collected by [30] when the dilution rate $\delta = 0.65$, and the inflow nitrogen concentration $N^* = 80$. Both populations of *Chlorella* and *Brachionus* show oscillation behavior, and it is interesting to estimate the parameters in the ODEs model (3.1) from these data. Notice that we have no observations for two variables, the concentrations of nitrogen and reproducing *Brachionus*, in the ODEs model (3.1), which increases the challenge for parameter estimation.

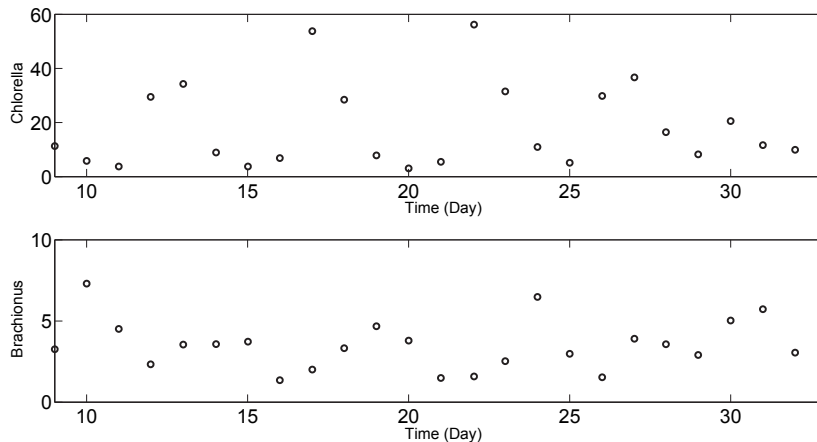


Figure 1: The experimental measurements for the concentrations of Chlorella and Brachionus in the predator-prey dynamic system when the dilution rate $\delta = 0.65$ and the inflow nitrogen concentration $N^* = 80$.

Let

$$\boldsymbol{\beta} = \log(\epsilon, \alpha, m, b_C, b_B, k_C, k_B)^T$$

be the logarithms of the vector of ODE parameters and let

$$\boldsymbol{\alpha} = \log(N(t_0), C(t_0), R(t_0), B(t_0))^T$$

be the logarithms of the vector of initial conditions of the ODEs model (3.1), where t_0 is the starting time point in the predator-prey experiment. Let $X_i(t|\boldsymbol{\theta})$, $i = 1, \dots, 4$, denote the solution of the ODEs model (3.1) in the time domain $[t_0, t_n]$, where $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)$ is a vector of length 11, and denote θ_k , $k = 1, \dots, 11$, as the element of $\boldsymbol{\theta}$. Let y_{2j} and y_{4j} , $j = 1, \dots, n$, denote the measured concentrations of Chlorella and Brachionus, respectively. We assume y_{ij} , $i = 2, 4$, is distributed with $N(X_i(t_{ij}|\boldsymbol{\theta}), \sigma_i^2)$. Then the log-likelihood function of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2 = (\sigma_2^2, \sigma_4^2)^T$ is given by

$$\ell(\boldsymbol{\theta}, \boldsymbol{\sigma}^2) = - \sum_{i=2,4} \frac{n}{2} \log(\sigma_i^2) - \frac{1}{2} \sum_{i=2,4} \sigma_i^{-2} \sum_{j=1}^n \{y_{ij} - X_i(t_{ij}|\boldsymbol{\theta})\}^2.$$

Define

$$U(\boldsymbol{\theta}) = \sum_{i=2,4} \sum_{j=1}^n \hat{w}_i \{y_{ij} - X_i(t_{ij}|\boldsymbol{\theta})\}^2,$$

where the weights $\hat{w}_i = \frac{1}{\text{var}(\mathbf{y}_i)}$, $\mathbf{y}_i = (y_{i1}, \dots, y_{in})^T$ for $i = 2, 4$. Figure 2 displays the surface of $-U(\boldsymbol{\theta})$ by varying the values of m and ϵ in $[0.4, 0.8]$ and $[0.1, 0.3]$, respectively, while setting other parameter values as estimates obtained in [8]. It can be seen that the surface of $-U(\boldsymbol{\theta})$ has multiple local modes, and it is not easy to arrive at the global mode using traditional optimization approaches.

Instead, we will develop the SAMC method to search the minimum of $U(\boldsymbol{\theta})$ in this article. Let $\boldsymbol{\Theta}$ be the sample space for the ODE parameters, which is partitioned according to the values of $U(\boldsymbol{\theta})$ into the following subregions: $E_1 = \{\boldsymbol{\theta} : U(\boldsymbol{\theta}) < u_1\}$, $E_2 = \{\boldsymbol{\theta} : u_1 \leq U(\boldsymbol{\theta}) < u_2\}$, \dots , $E_m = \{\boldsymbol{\theta} : U(\boldsymbol{\theta}) > u_{m-1}\}$ with an equal bandwidth, where u_1 and u_{m-1} are pre-specified real numbers. We set $u_1 = 20$ and $u_{m-1} = 36$ with $m = 50$ and T_0 defined in (2.4) is set as $T_0 = 5,000$.

After burning in the first 10^5 iterations, the SAMC method is continued to run for 10^5 iterations. Table 1 displays the summary of SAMC estimation for the seven parameters in the ODEs model (3.1). The mean and standard deviation of posterior samples are used as the parameter estimates and standard errors with the SAMC method, respectively. The parameter estimates with the SAMC method is consistent to the generalized profiling estimates obtained in [8]. On the other hand, the SAMC method has larger standard errors for parameter estimates than the generalized profiling method, because the latter are given for a fixed value of the smoothing parameter. The SAMC method can also easily provide the 95 % posterior credible intervals for the

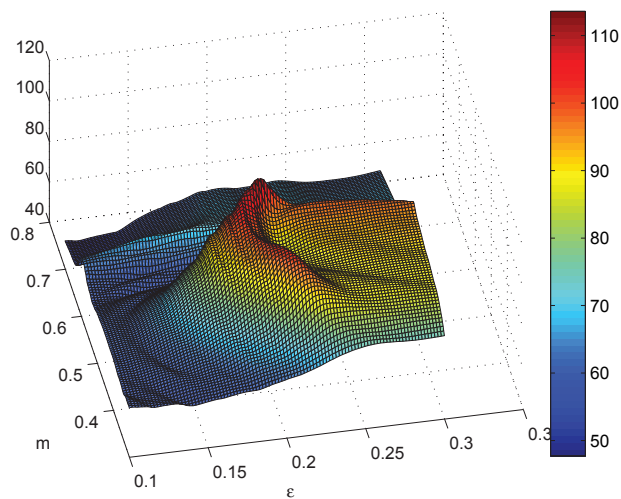


Figure 2: The surface of $-U(\theta)$ by varying the values of m and ϵ in $[0.4, 0.8]$ and $[0.1, 0.3]$ while setting other parameter values as estimates obtained in [8].

	Estimates	Standard Errors	95 % C.I.
ϵ	0.192	0.015	(0.162, 0.222)
α	0.796	0.022	(0.753, 0.838)
m	0.459	0.027	(0.406, 0.513)
b_C	3.876	0.404	(3.084, 4.667)
b_B	4.72	0.323	(4.083, 5.372)
κ_C	7.065	1.778	(3.579, 10.550)
κ_B	28.461	4.173	(20.282, 36.640)

Table 1: The summary of Bayesian estimation for the seven parameters in the ODEs model (3.1), where C.I. denotes the posterior credible interval of the ODE parameters.

ODE parameter estimates based on the posterior sampling sequence, which is the most appealing feature of the SAMC method in comparison with the generalized profiling method.

4 Simulation

A simulation study is implemented to illustrate the advantage of the SAMC method in comparison with the Metropolis–Hastings method when estimating parameters in an ODEs model from the noisy measurements of the dynamic system.

Goodwin [16] introduced a mathematical model that was able to simulate physiological oscillation on the basis of a negative feedback in cellular systems such as circadian rhythms and enzymatic regulation. A simple Goodwin model can be expressed in form of a set of ODEs

$$\begin{cases} \frac{dX_1(t)}{dt} = \frac{72}{36 + X_2(t)} - \kappa_1, \\ \frac{dX_2(t)}{dt} = \kappa_2 X_1(t) - 1, \end{cases} \quad (4.1)$$

where $X_1(t)$ and $X_2(t)$ are the levels of mRNA and protein in the system, respectively. The parameter κ_1 is the degradation rate constant, and κ_2 is the synthesis rate constant.

The simulated data of Goodwin ODEs model are generated as follows. The Goodwin ODEs model is first numerically solved at 120 equally-spaced time points in the time interval $[0, 60]$ by setting the true ODE parameters $\kappa_1 = 2$ and $\kappa_2 = 1$, and the initial conditions $X_1(0) = 7$ and $X_2(0) = -10$. The simulated data are

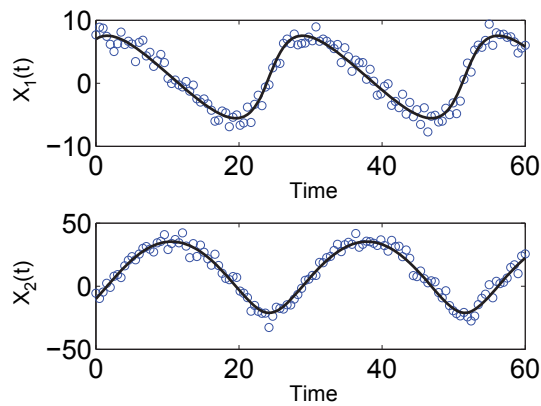


Figure 3: The data simulated by adding white noises to the numerical solutions of the Goodwin ODE (4.1). The solid lines are the ODE solutions by setting parameter, $\kappa_1 = 2$ and $\kappa_2 = 1$, and initial conditions $X_1(0) = 7$, $X_2(0) = -10$.

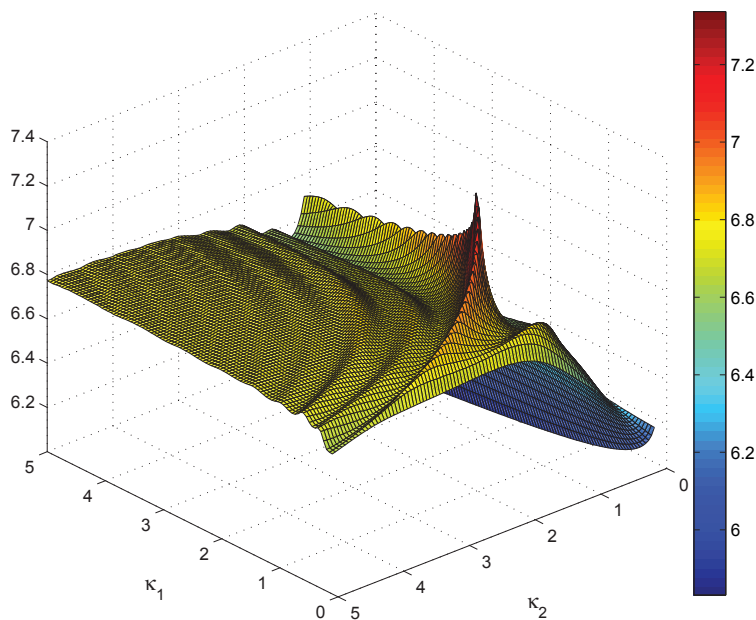


Figure 4: The surface of the log-likelihood of $(\kappa_1, \kappa_2)^T$ given in (4.2).

then generated by adding white noises from $N(0, \text{diag}(1, 16))$ to the ODE numerical solutions. Figure 3 shows one set of simulated data along with the ODE solutions. The ODE solutions display the cyclic behavior and have roughly two cycles in the whole time interval $[0, 60]$.

Let $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})^T$, $i = 1, 2$, denote the simulated data at the time point $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})^T$ with $n_i = 120$. Define a vector of ODE parameters, $\boldsymbol{\theta} = (\kappa_1, \kappa_2)^T$, and denote the ODE solutions to be

$$\mathbf{X}(t|\boldsymbol{\theta}) = (X_1(t|\boldsymbol{\theta}), X_2(t|\boldsymbol{\theta}))^T.$$

The data \mathbf{y}_i is assumed to follow a normal distribution with the mean $X_i(\mathbf{t}_i|\boldsymbol{\theta})$ and the variance-covariance matrix $\sigma_i^2 \mathbf{I}_{n_i}$, $i = 1, 2$. The log-likelihood function of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2)^T$ based on observations $\mathbf{y}_1, \mathbf{y}_2$ is given by

$$\ell(\boldsymbol{\theta}, \boldsymbol{\sigma}^2) = -\frac{120}{2} \sum_{i=1}^2 \log(\sigma_i^2) - \sum_{i=1}^2 \sum_{j=1}^{120} \frac{1}{2\sigma_i^2} \{y_{ij} - X_i(t_{ij}|\boldsymbol{\theta})\}^2. \quad (4.2)$$

Figure 4 displays the surface of the log-likelihood function (4.2) when $\sigma_1^2 = 1$ and $\sigma_2^2 = 16$. This surface has several strong ripples which form a number of ridges on the log-likelihood surface. These ridges will raise the difficulty of global optimization of the likelihood function.

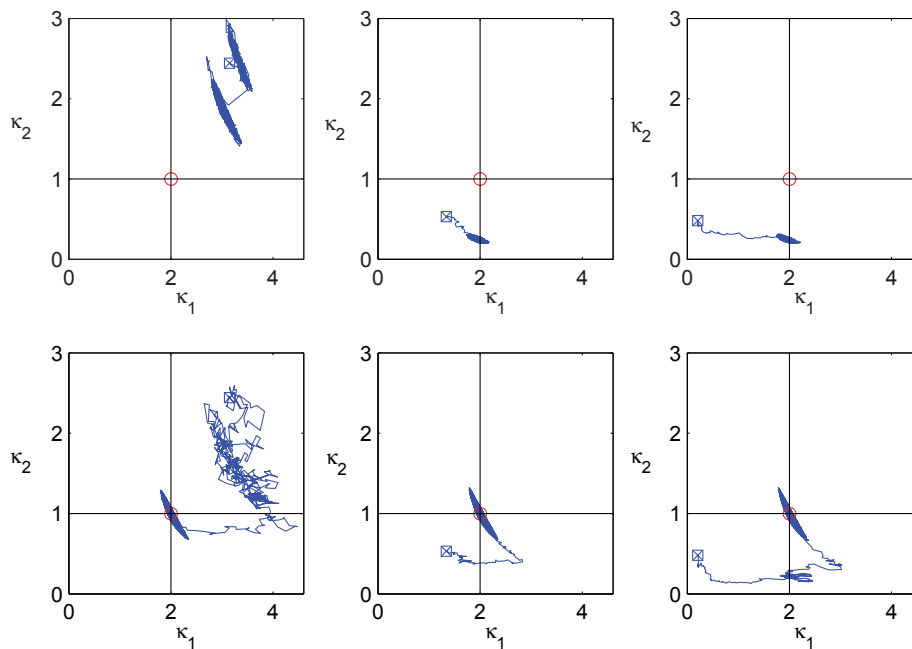


Figure 5: The sampling sequences for $(\kappa_1, \kappa_2)^T$ in the ODEs model (4.2) using the Metropolis–Hastings method (upper panels) and the SAMC method (lower panels) with three starting values chosen randomly. The true values of the two ODE parameters are $\kappa_1 = 2.0$ and $\kappa_2 = 1.0$, marked with circles. The panels from left to right correspond to three starting values (marked with squares) of the two ODE parameters chosen randomly as: $(3.1, 2.4)$, $(1.3, 0.5)$, $(0.2, 0.5)$.

Alternatively, we can maximize (4.2) by Monte Carlo optimization. However, these ridges are still suspected to cause the poor mixing and local trapping of the Metropolis–Hastings method which is shown in the next.

Define

$$U(\boldsymbol{\theta}) = \sum_{i=1,2} \sum_{j=1}^n \hat{w}_i \{y_{ij} - X_i(t_{ij}|\boldsymbol{\theta})\}^2,$$

where the weights $\hat{w}_i = \frac{1}{\text{var}(y_i)}$, $i = 1, 2$. To minimize $U(\boldsymbol{\theta})$, we construct a trial distribution $\exp\{-U(\boldsymbol{\theta})\}$. We sample $\boldsymbol{\theta}$ from $\exp\{-U(\boldsymbol{\theta})\}$ via the Metropolis–Hastings method and SAMC method, respectively.

The Metropolis–Hastings method is used to sample 50,000 iterations for the two ODE parameters $(\kappa_1, \kappa_2)^T$ based on the trial distribution $\exp\{-U(\boldsymbol{\theta})\}$, in which the starting value of $\boldsymbol{\theta}$ was randomly chosen. The sampling sequences for the two ODE parameters are displayed in the upper panels of Figure 5, which shows that the sampling sequences are trapped at a local mode and are hard to converge to the true parameter values. In contrast, the SAMC method is also applied to sampling 50,000 iterations for the two ODE parameters with the same starting value of $\boldsymbol{\theta}$ using the same simulated data. Let $\boldsymbol{\Theta}$ be the sample space for the two ODE parameters. We partition $\boldsymbol{\Theta}$ according to the values of the objective function $U(\boldsymbol{\theta})$ into the subregions $E_1 = \{\boldsymbol{\theta} : U(\boldsymbol{\theta}) < u_1\}$, $E_2 = \{\boldsymbol{\theta} : u_1 \leq U(\boldsymbol{\theta}) < u_2\}$, \dots , $E_m = \{\boldsymbol{\theta} : U(\boldsymbol{\theta}) > u_{m-1}\}$, where $u_\ell = 7 + 0.592\ell$, $\ell = 1, \dots, m$, and $m = 50$. The sampling sequences for the two ODE parameters are displayed in the lower panels of Figure 5. It shows that the SAMC method is able to escape the local trapping mode efficiently and converge to the true ODE parameters quickly.

5 Conclusions and discussion

Ordinary differential equations are popular models to elucidate complex dynamic system. Parameters in ODEs usually have important scientific interpretations, but require to be estimated from noisy measurements of the dynamic system. Most ODEs have no analytic solutions and can only be solved using numeric methods,

such as Runge–Kutta methods. Besides some frequentist methods, such as the two-step method and generalized profiling method, Bayesian methods are gained extensive attentions to estimating ODE parameters, which have the natural advantage in making statistical inferences for ODE parameters such as confidence intervals and hypothesis tests.

One popular Bayesian method is using the MCMC method to sample the posterior distribution of ODE parameters, which is easy to understand and implement. However, it is well known that the classical MCMC method is easy to be trapped in local modes of posterior distributions. Because ODE solutions are sensitive to ODE parameters, the posterior distribution of ODE parameters often has many local modes. Therefore, the MCMC method is found to often be stuck in local modes when sampling for ODE parameters.

In this paper, we suggest a Bayesian approach to solve the global optimization problem of the ODEs when the likelihood function has multiple local modes. To sample the posterior distributions of ODE parameters, we develop the stochastic approximation Monte Carlo (SAMC) method which is a self-adjusting mechanism and can update automatically the probabilities of subregions being visited in the sampling process. By performing numerical simulations, the advantage of the SAMC method is illustrated, in which the SAMC method more efficiently escapes from local modes than the classical MCMC method. The SAMC method is also demonstrated by estimating a popular nutrient-based predator-prey dynamic model from the experimental data.

Acknowledgment: We thank Professor Gregor F. Fussmann for providing us the predator-prey data set. The authors are also very grateful for the suggestions of Professor Faming Liang.

Funding: The research was supported by the Liaoning Provincial Education Department (No. LN2017ZD001) to Baisen Liu and the discovery grants (RGPIN 356044-2013, 435713-2013) from the Natural Science and Engineering Research Council of Canada to Jiguo Cao and Liangliang Wang.

References

- [1] Y. Bard, *Nonlinear Parameter Estimation*, Academic Press, New York, 1974.
- [2] L. Becks, F. M. Hilker, H. Malchow, K. Jürgens and H. Arndt, Experimental demonstration of chaos in a microbial food web, *Nature* **435** (2005), 1226–1229.
- [3] L. T. Biegler, J. J. Damiano and G. E. Blau, Nonlinear parameter estimation: A case study comparison, *AIChE J.* **32** (1986), 29–45.
- [4] N. J.-B. Brunel, Parameter estimation of ODE's via nonparametric estimators, *Electron. J. Stat.* **2** (2008), 1242–1267.
- [5] J. C. Butcher, *Numerical Methods for Ordinary Differential Equations*, 2nd ed., John Wiley & Sons, Chichester, 2008.
- [6] B. Calderhead and M. Girolami, Estimating Bayes factors via thermodynamic integration and population MCMC, *Comput. Statist. Data Anal.* **53** (2009), no. 12, 4028–4045.
- [7] D. Campbell and R. J. Steele, Smooth functional tempering for nonlinear differential equation models, *Stat. Comput.* **22** (2012), no. 2, 429–443.
- [8] J. Cao, G. F. Fussmann and J. O. Ramsay, Estimating a predator-prey dynamical model with the parameter cascades method, *Biometrics* **64** (2008), no. 3, 959–967.
- [9] J. Cao, J. Z. Huang and H. Wu, Penalized nonlinear least squares estimation of time-varying parameters in ordinary differential equations, *J. Comput. Graph. Statist.* **21** (2012), no. 1, 42–56.
- [10] J. Cao, L. Wang and J. Xu, Robust estimation for ordinary differential equation models, *Biometrics* **67** (2011), no. 4, 1305–1313.
- [11] J. Chen and H. Wu, Efficient local estimation for time-varying coefficients in deterministic dynamic models with applications to HIV-1 dynamics, *J. Amer. Statist. Assoc.* **103** (2008), no. 481, 369–384.
- [12] G. F. Fussmann, S. P. Ellner, K. W. Shertzer and N. G. Hairston Jr., Crossing the Hopf bifurcation in a live predator-prey system, *Science* **290** (2000), 1358–1360.
- [13] A. Gelman, F. Bois and J. Jiang, Physiological pharmacokinetic analysis using population modeling and informative prior distributions, *J. Amer. Statist. Assoc.* **91** (1996), 1400–1412.
- [14] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin, *Bayesian Data Analysis*, 2nd ed., Texts Statist. Sci. Ser., Chapman & Hall/CRC, Boca, 2004.
- [15] C. J. Geyer, Estimation and optimization of functions, in: *Markov Chain Monte Carlo in Practice*, Chapman & Hall, London (1996), 241–258.

- [16] B. Goodwin, Oscillatory behavior in enzymatic control processes, *Adv. Enzyme Regul.* **3** (1965), 425–438.
- [17] Y. Huang, D. Liu and H. Wu, Hierarchical Bayesian methods for estimation of parameters in a longitudinal HIV dynamic system, *Biometrics* **62** (2006), no. 2, 413–423.
- [18] B. E. Kendall, C. J. Briggs, W. W. Murdoch, P. Turchin, S. P. Ellner, E. McCauley, R. M. Nisbet and S. N. Wood, Why do populations cycle? A synthesis of statistical and mechanistic modeling approaches, *Ecology* **80** (1999), 1789–1805.
- [19] F. Liang, C. Liu and R. J. Carroll, Stochastic approximation in Monte Carlo computation, *J. Amer. Statist. Assoc.* **102** (2007), no. 477, 305–320.
- [20] W. W. Murdoch, C. J. Briggs and R. M. Nisbet, *Consumer-Resource Dynamics*, Princeton University Press, New York, 2003.
- [21] A. Poyton, *Application of principal differential analysis to parameter estimation in fundamental dynamics models*, Master's thesis, Queen's University, Kingston, 2005.
- [22] X. Qi and H. Zhao, Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations, *Ann. Statist.* **38** (2010), no. 1, 435–481.
- [23] J. O. Ramsay, G. Hooker, D. Campbell and J. Cao, Parameter estimation for differential equations: A generalized smoothing approach, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** (2007), no. 5, 741–796.
- [24] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, 2nd ed., Springer Ser. Statist., Springer, New York, 2005.
- [25] H. Robbins and S. Monro, A stochastic approximation method, *Ann. Math. Statistics* **22** (1951), 400–407.
- [26] C. P. Robert and G. Casella, *Monte Carlo statistical methods*, 2nd ed., Springer Ser. Statist., Springer, New York, 2004.
- [27] K. W. Shertzer, S. P. Ellner, G. F. Fussmann and N. G. Hairston, Predator-prey cycles in an aquatic microcosm: Testing hypotheses of mechanism, *J. Animal Ecol.* **71** (2002), 802–815.
- [28] P. Turchin, *Complex Population Dynamics: A Theoretical/Empirical Synthesis*, Monogr. Population Biol. 35, Princeton University Press, Princeton, 2003.
- [29] J. M. Varah, A spline least squares method for numerical parameter estimation in differential equations, *SIAM J. Sci. Statist. Comput.* **3** (1982), no. 1, 28–46.
- [30] T. Yoshida, L. E. Jones, S. P. Ellner, G. F. Fussmann and N. G. Hairston, Rapid evolution drives ecological dynamics in a predator-prey system, *Nature* **424** (2003), 303–306.