

Article

Estimating Distributions of Parameters in Nonlinear State Space Models with Replica Exchange Particle Marginal Metropolis–Hastings Method

Hiroaki Inoue¹, Koji Hukushima^{2,3}  and Toshiaki Omori^{1,4,5,*} 

¹ Graduate School of Engineering, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan; 169t804t@stu.kobe-u.ac.jp

² Graduate School of Arts and Sciences, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan; k-hukushima@g.ecc.u-tokyo.ac.jp

³ Komaba Institute for Science, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8902, Japan

⁴ Organization for Advanced and Integrated Research, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan

⁵ Center for Mathematical and Data Sciences, Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Japan

* Correspondence: omori@eedept.kobe-u.ac.jp

Abstract: Extracting latent nonlinear dynamics from observed time-series data is important for understanding a dynamic system against the background of the observed data. A state space model is a probabilistic graphical model for time-series data, which describes the probabilistic dependence between latent variables at subsequent times and between latent variables and observations. Since, in many situations, the values of the parameters in the state space model are unknown, estimating the parameters from observations is an important task. The particle marginal Metropolis–Hastings (PMMH) method is a method for estimating the marginal posterior distribution of parameters obtained by marginalization over the distribution of latent variables in the state space model. Although, in principle, we can estimate the marginal posterior distribution of parameters by iterating this method infinitely, the estimated result depends on the initial values for a finite number of times in practice. In this paper, we propose a replica exchange particle marginal Metropolis–Hastings (REPMMH) method as a method to improve this problem by combining the PMMH method with the replica exchange method. By using the proposed method, we simultaneously realize a global search at a high temperature and a local fine search at a low temperature. We evaluate the proposed method using simulated data obtained from the Izhikevich neuron model and Lévy-driven stochastic volatility model, and we show that the proposed REPMMH method improves the problem of the initial value dependence in the PMMH method, and realizes efficient sampling of parameters in the state space models compared with existing methods.

Keywords: state space model; probabilistic graphical model; replica exchange particle Metropolis–Hastings method; replica exchange method; particle Metropolis–Hastings method; particle Markov chain Monte Carlo method



Citation: Inoue, H.; Hukushima, K.; Omori, T. Estimating Distributions of Parameters in Nonlinear State Space Models with Replica Exchange Particle Marginal Metropolis–Hastings Method. *Entropy* **2022**, *24*, 115. <https://doi.org/10.3390/e24010115>

Academic Editor: Leonardo Ricci

Received: 25 November 2021

Accepted: 7 January 2022

Published: 12 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Extracting latent nonlinear dynamics from observed time-series data is important for understanding the dynamic system against the background of the observed data. A state space model is a probabilistic graphical model for time-series data that assumes the existence of latent variables that cannot be observed directly [1–25]. State space models are used in various fields to forecast observation values [7,15,22] and to estimate latent variables [11,20,21,26]. In many cases, however, model parameters are unknown. Therefore, estimating the model parameters from observations is an important task for the state space models.

To estimate the parameters in the state space models from observations, a method combining the sequential Monte Carlo (SMC) method [2,4,8–12,14,16–20,23–25,27] with the expectation–maximization (EM) algorithm [8,28,29] has been proposed [9–11,14,20]. This method is based on a maximum likelihood estimation framework, and it estimates parameters by sequentially updating the parameters so that the likelihood increases. Although it is guaranteed that a local optimum can be estimated by iteratively updating the parameters, the global optimum may not be estimated depending on the initial values of parameters. Furthermore, since the EM algorithm is a point estimation method, it is not possible to identify whether converged values are local or global optima.

To estimate the distribution of parameters in a state space model, two kinds of particle Markov chain Monte Carlo (PMCMC) methods have been proposed: the particle Gibbs (PG) method and the particle marginal Metropolis–Hastings (PMMH) method [12]. Both methods combine Markov chain Monte Carlo (MCMC) methods with the SMC method, and the distribution of parameters is estimated by collecting samples. The PG method combines the SMC method with Gibbs sampling [8,30], and the PG method samples latent variables and parameters in a state space model from the joint posterior distribution of latent variables and parameters alternately. In the PG method, the SMC method is employed for sampling latent variables. The PMMH method, on the other hand, combines the SMC method with the Metropolis–Hastings (MH) algorithm [8,25,31,32], and the PMMH method samples parameters in a state space model from the marginal posterior distribution of parameters. In the PMMH method, the SMC method is employed for calculating the likelihood marginalized over the distribution of latent variables. Both the PG method and the PMMH method have been widely applied (for example, the PG method [33,34], the PMMH method [35–37]).

In recent years, some extended versions of PG methods have been proposed to improve the sampling efficiency and initial value dependence, such as the particle Gibbs with ancestor sampling (PGAS) method and the replica exchange particle Gibbs with ancestor sampling (REPGAS) method [16,18,19,24]. Thus far, however, the existing methods for such improvement have been proposed for only the PG method. Therefore, it is important to improve the PMMH method for the accurate estimation of parameters since the PMMH method may have the problem of the initial value dependence.

In this paper, we propose the replica exchange particle marginal Metropolis–Hastings (REPMMH) method, which combines the PMMH method with the replica exchange method [24,38–40] in order to improve the problem of initial value dependence in the PMMH method. Combining the replica exchange method with the PMMH method makes it possible to estimate the parameters governing the dynamics for very complex and non-linear time-series data. We first describe the state space models and explain the PMMH method as a conventional method. Then, after explaining our proposed method, we conduct experiments to compare the proposed method with the conventional methods, the PMMH method and the REPGAS method.

2. Methods

In this section, we propose our replica exchange particle marginal Metropolis–Hastings (REPMMH) method. First, we describe a state space model for time-series data using a probabilistic graphical model. Next, we describe the conventional particle marginal Metropolis–Hastings (PMMH) method to estimate the marginal posterior distribution of parameters obtained by marginalization over the distribution of latent variables in a state space model. After this, we propose the REPMMH method, which combines the PMMH method with the replica exchange method to improve the problem of initial value dependence in the PMMH method.

2.1. State Space Model

We show a state space model as a probabilistic graphical model in Figure 1. Note that there are two type of variables, latent variables $z_{1:N} = \{z_1, z_2, \dots, z_N\}$ and observations

$\mathbf{y}_{1:N} = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$, at the time steps $\{1, 2, \dots, N\}$ in the state space model. The latent variables $\mathbf{z}_{1:N}$ cannot be observed directly and only the observations $\mathbf{y}_{1:N}$ are observable. At a time step n , the state space model is represented as follows:

$$\mathbf{z}_n \sim f(\mathbf{z}_n \mid \mathbf{z}_{n-1}, \boldsymbol{\theta}_f), \tag{1}$$

$$\mathbf{y}_n \sim g(\mathbf{y}_n \mid \mathbf{z}_n, \boldsymbol{\theta}_g), \tag{2}$$

where $f(\mathbf{z}_n \mid \mathbf{z}_{n-1}, \boldsymbol{\theta}_f)$ and $g(\mathbf{y}_n \mid \mathbf{z}_n, \boldsymbol{\theta}_g)$ are called a system model and an observation model, respectively. $\boldsymbol{\theta}_f$ are the parameters of the system model, and $\boldsymbol{\theta}_g$ are the parameters of the observation model. The system model $f(\mathbf{z}_n \mid \mathbf{z}_{n-1}, \boldsymbol{\theta}_f)$ represents the process of updating the latent variables \mathbf{z}_n from the previous latent variables \mathbf{z}_{n-1} . Moreover, the observation model $g(\mathbf{y}_n \mid \mathbf{z}_n, \boldsymbol{\theta}_g)$ represents the process of obtaining observations \mathbf{y}_n from the latent variables \mathbf{z}_n .

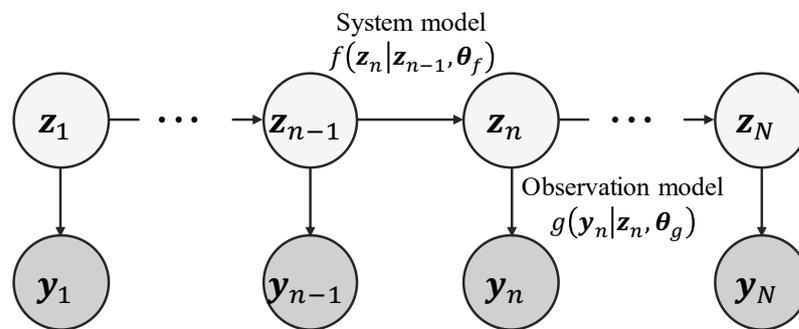


Figure 1. Probabilistic graphical model of a state space model. $\mathbf{z}_{1:N} = \{z_1, z_2, \dots, z_N\}$ and $\mathbf{y}_{1:N} = \{y_1, y_2, \dots, y_N\}$, respectively, represent latent variables and observations for time step $n = 1, 2, \dots, N$. The arrow to the latent variables \mathbf{z}_n at the time step n from the latent variables \mathbf{z}_{n-1} at the previous time step $n - 1$ represents a system model $f(\mathbf{z}_n \mid \mathbf{z}_{n-1}, \boldsymbol{\theta}_f)$, and the arrow to the observations \mathbf{y}_n at the time step n from the latent variables \mathbf{z}_n at the time step n represents an observation model $g(\mathbf{y}_n \mid \mathbf{z}_n, \boldsymbol{\theta}_g)$. $\Theta = \{\boldsymbol{\theta}_f, \boldsymbol{\theta}_g\}$ are parameters to be estimated.

The goal of this paper is to estimate the posterior distribution of the parameters $p(\Theta \mid \mathbf{y}_{1:N})$ for the given observations $\mathbf{y}_{1:N}$, where Θ is represented as $\Theta = \{\boldsymbol{\theta}_f, \boldsymbol{\theta}_g\}$. However, since the latent variables exist in the state space models, it is necessary to perform marginalization with respect to the latent variables in order to obtain the marginal posterior distribution $p(\Theta \mid \mathbf{y}_{1:N})$. Because it is often difficult to calculate the marginal posterior distribution $p(\Theta \mid \mathbf{y}_{1:N})$ analytically, we propose a new method for estimating the marginal posterior distribution $p(\Theta \mid \mathbf{y}_{1:N})$ based on the PMMH method in this paper.

2.2. Particle Marginal Metropolis–Hastings Method

The PMMH method combines the sequential Monte Carlo (SMC) method [2,4,8–12,14,16–20,23–25,27] with the Metropolis–Hastings (MH) algorithm [8,25,31,32]. The PMMH method was proposed to estimate the marginal posterior distribution of parameters $p(\Theta \mid \mathbf{y}_{1:N})$ for time-series observations $\mathbf{y}_{1:N}$ represented as a state space model [12].

In the PMMH method, the marginal likelihood $p(\mathbf{y}_{1:N} \mid \Theta)$ is used to evaluate the appropriateness of parameters Θ . Here, the SMC method is used to calculate the marginal likelihood $p(\mathbf{y}_{1:N} \mid \Theta)$ of the parameters Θ obtained by marginalization over the distribution of latent variables $\mathbf{z}_{1:N}$. A new sample candidate of parameters $\Theta^* = \{\boldsymbol{\theta}_f^*, \boldsymbol{\theta}_g^*\}$ is proposed from an arbitrary proposal distribution $q(\Theta \mid \Theta[k - 1])$ given the sample one

step before $\Theta[k - 1]$, where k is the sample number. Moreover, whether to accept or reject the sample candidate Θ^* is determined based on the following acceptance probability:

$$p_{\text{accept}} = \min\left(1, \frac{p(\mathbf{y}_{1:N} | \Theta^*)p(\Theta^*)}{p(\mathbf{y}_{1:N} | \Theta[k - 1])p(\Theta[k - 1])} \frac{q(\Theta[k - 1] | \Theta^*)}{q(\Theta^* | \Theta[k - 1])}\right), \tag{3}$$

where $p(\Theta)$ represents the prior distribution of parameters. $p(\mathbf{y}_{1:N} | \Theta)$ is the marginal likelihood obtained by marginalization over the distributions of latent variables $\mathbf{z}_{1:N}$ as follows:

$$\begin{aligned} p(\mathbf{y}_{1:N} | \Theta) &= \int p(\mathbf{y}_{1:N}, \mathbf{z}_{1:N} | \Theta) d\mathbf{z}_{1:N} \\ &= \int g(\mathbf{y}_1 | \mathbf{z}_1, \theta_g) p(\mathbf{z}_1) \prod_{n=2}^N g(\mathbf{y}_n | \mathbf{z}_n, \theta_g) f(\mathbf{z}_n | \mathbf{z}_{n-1}, \theta_f) d\mathbf{z}_{1:N}, \end{aligned} \tag{4}$$

where $p(\mathbf{z}_1)$ is the distribution of latent variables \mathbf{z}_1 at time step 1. Since it is difficult to obtain the marginal likelihood $p(\mathbf{y}_{1:N} | \Theta)$ analytically, the SMC method is used in the PMMH method to calculate the marginal likelihood $p(\mathbf{y}_{1:N} | \Theta)$ numerically.

The SMC method estimates the distribution of latent variables by approximating the distribution with the density of the particles $\{\mathbf{z}_{1:N}^{(1)}, \mathbf{z}_{1:N}^{(2)}, \dots, \mathbf{z}_{1:N}^{(M)}\}$ as follows:

$$p(\mathbf{z}_{1:N} | \mathbf{y}_{1:N}, \Theta) \simeq \frac{1}{M} \sum_{m=1}^M \delta(\mathbf{z}_{1:N} - \mathbf{z}_{1:N}^{(m)}), \tag{5}$$

where $\mathbf{z}_{1:N}^{(m)}$ is the m -th particle and M is the number of particles. $\delta(\mathbf{z}_{1:N})$ is the Dirac delta distribution.

To obtain particles $\{\mathbf{z}_n^{(1)}, \mathbf{z}_n^{(2)}, \dots, \mathbf{z}_n^{(M)}\}$ at a time step n , we sample the m -th particle $\mathbf{z}_n^{(m)}$ at the time step n from the m -th particle $\mathbf{z}_{n-1}^{(m)}$ at the previous time step $n - 1$ for each $m \in \{1, 2, \dots, M\}$ with the system model as follows:

$$\mathbf{z}_n^{(m)} \sim f(\mathbf{z}_n | \mathbf{z}_{n-1}^{(m)}, \theta_f). \tag{6}$$

Moreover, the obtained particles $\{\mathbf{z}_n^{(1)}, \mathbf{z}_n^{(2)}, \dots, \mathbf{z}_n^{(M)}\}$ are resampled based on the normalized weights $\{W_n^{(1)}, W_n^{(2)}, \dots, W_n^{(M)}\}$ obtained as follows:

$$W_n^{(m)} = \frac{w_n^{(m)}}{\sum_{l=1}^M w_n^{(l)}}, \tag{7}$$

$$w_n^{(m)} = g(\mathbf{y}_n | \mathbf{z}_n^{(m)}, \theta_g). \tag{8}$$

By iterating the above flow for time step $n \in \{1, 2, \dots, N\}$, particles that approximate the distribution of latent variables $\mathbf{z}_{1:N}$ can be obtained. Here, the marginal likelihood $p(\mathbf{y}_{1:N} | \Theta)$ can be calculated approximately as follows:

$$p(\mathbf{y}_{1:N} | \Theta) = \prod_{n=1}^N p(\mathbf{y}_n | \mathbf{y}_{1:n-1}, \Theta) \simeq \frac{1}{M} \prod_{n=1}^N \sum_{m=1}^M w_n^{(m)}. \tag{9}$$

By calculating the acceptance probability p_{accept} in Equation (3) with the marginal likelihood $p(\mathbf{y}_{1:N} | \Theta^*)$ for the sample candidate Θ^* obtained by Equation (9), it is determined whether to accept or reject the proposed sample candidate Θ^* . We show the flow of the PMMH method described above in Algorithm 1.

Algorithm 1 Particle Marginal Metropolis–Hastings (PMMH) Method.

```

1: initialize the parameters  $\Theta[0]$ 
2: for  $k = 1, \dots, K$  ( $K$  is the number of samples) do
3:   draw the sample candidate of parameters  $\Theta^* \sim q(\Theta^* | \Theta[k-1])$ 
4:   draw the initial particles  $z_1^{(m)} \sim p(z_1)$  for  $m = 1, \dots, M$  ( $m$  is the particle number of
   the particle that is the source of resampling)
5:   calculate the weights of particles  $\{w_1^{(1)}, w_1^{(2)}, \dots, w_1^{(M)}\}$  with Equation (8)
6:   normalize the weights of particles  $\{W_1^{(1)}, W_1^{(2)}, \dots, W_1^{(M)}\}$  with Equation (7)
7:   resample the particles  $\{z_1^{(1)}, z_1^{(2)}, \dots, z_1^{(M)}\}$  according to the normalized weights
    $\{W_1^{(1)}, W_1^{(2)}, \dots, W_1^{(M)}\}$ 
8:   for  $n = 2, \dots, N$  do
9:     draw the particles  $\{z_n^{(1)}, z_n^{(2)}, \dots, z_n^{(M)}\}$  at time step  $n$  with Equation (6)
10:    calculate the weights of particles  $\{w_n^{(1)}, w_n^{(2)}, \dots, w_n^{(M)}\}$  with Equation (8)
11:    normalize the weights of particles  $\{W_n^{(1)}, W_n^{(2)}, \dots, W_n^{(M)}\}$  with Equation (7)
12:    resample the particles  $\{z_n^{(1)}, z_n^{(2)}, \dots, z_n^{(M)}\}$  according to the normalized weights
    $\{W_n^{(1)}, W_n^{(2)}, \dots, W_n^{(M)}\}$ 
13:   end for
14:   calculate the marginal likelihood  $p(\mathbf{y}_{1:N} | \Theta^*)$  with Equation (9)
15:   calculate the acceptance probability  $p_{\text{accept}}$  with Equation (3)
16:   draw a uniform random number  $\alpha \sim \mathcal{U}(0, 1)$  ( $\mathcal{U}(a, b)$  is a uniform distribution with
   range  $[a, b]$ )
17:   if  $\alpha \leq p_{\text{accept}}$  then
18:     set the sample of parameters  $\Theta[k] \leftarrow \Theta^*$ 
19:   else
20:     set the sample of parameters  $\Theta[k] \leftarrow \Theta[k-1]$ 
21:   end if
22: end for
23: return  $\{\Theta[k]\}_{k=1}^K$ 

```

2.3. Proposed Method

In our study, we propose the REPMMH method, which combines the PMMH method with the replica exchange method [24,38–40] to improve the problem of initial value dependence in the PMMH method. By employing the REPMMH method, we estimate the marginal posterior distribution of parameters from the time-series observations.

2.3.1. Brief Summary of Our Proposed Method

We show the schematic diagram of the REPMMH method in Figure 2. In our proposed REPMMH method, we introduce multiple different replicas of parameters $\{\Theta\} = \{\Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(r)}, \dots, \Theta^{(R)}\}$ at temperatures $T = [T^{(1)}, T^{(2)}, \dots, T^{(r)}, \dots, T^{(R)}]$ into the PMMH method. As shown in the middle part of Figure 2, we employ the PMMH method in parallel at each temperature. In the PMMH method at each temperature $T^{(r)}$, we obtain the respective marginal likelihood $p(\mathbf{y}_{1:N} | \Theta^{(r)*})_{T^{(r)}}^{\frac{1}{T^{(r)}}$ by employing the SMC method (Figure 2c) with the respective sample candidate $\Theta^{(r)*}$ proposed in the MH algorithm (Figure 2b).

For each temperature $T^{(r)}$, the SMC method and the MH algorithm are conducted as follows. In the SMC method (Figure 2c), the marginal likelihood $p(\mathbf{y}_{1:N} | \Theta^{(r)*})_{T^{(r)}}^{\frac{1}{T^{(r)}}$ is obtained by iterative procedures of predictions, likelihood calculations and resampling; the latent variables z_n of the current time step n are predicted and the likelihood is calculated

for each particle, and resampling is performed according to the calculated likelihoods of particles at each time step. In the MH algorithm (Figure 2b), the sample candidate $\Theta^{(r)*}$ is determined to be accepted or rejected with the marginal likelihood $p(\mathbf{y}_{1:N} | \Theta^{(r)*})^{\frac{1}{T^{(r)}}}$. At this time, the target distribution becomes smooth as the temperature becomes high. As a result, it becomes easier to obtain samples from a wide range. Furthermore, exchanges between the samples at different temperatures are conducted in order to realize the transitions that are difficult depending on the initial values by the conventional PMMH method.

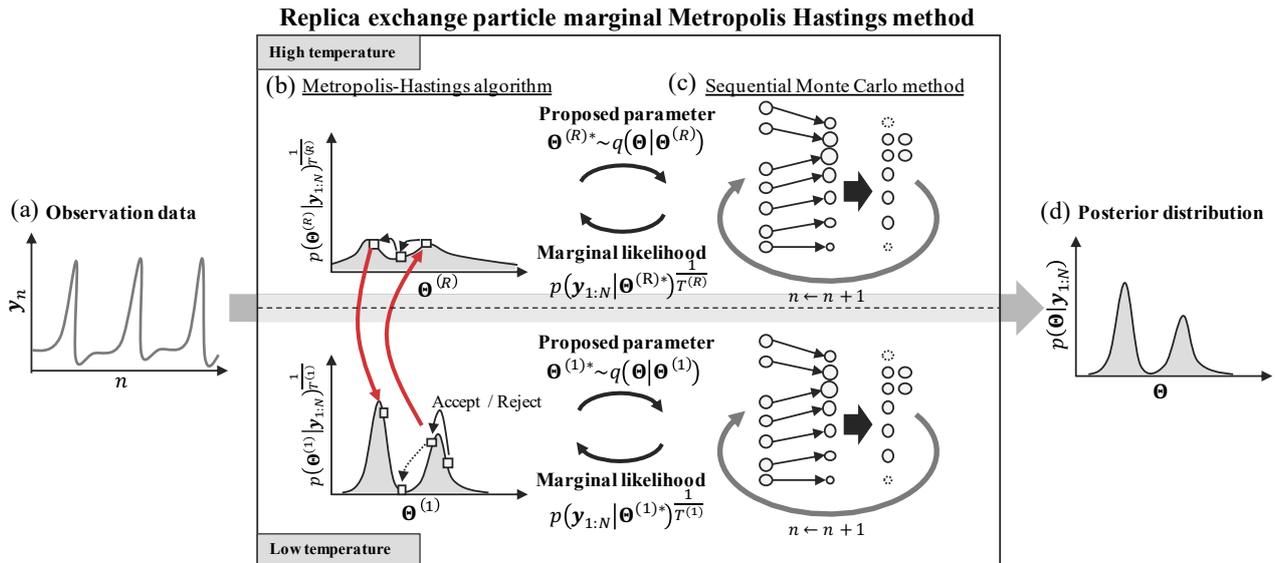


Figure 2. Schematic diagrams of the proposed replica exchange particle marginal Metropolis–Hastings (REPMMH) method. (a) The time-series observations $\mathbf{y}_{1:N}$ as inputs. (b,c) The REPMMH method consisting of (b) the Metropolis–Hastings (MH) algorithms and (c) the sequential Monte Carlo (SMC) methods parallelly conducted at multiple temperatures. In the SMC method, the sample candidate $\Theta^{(r)*}$ proposed by the MH algorithm is used to obtain the marginal likelihood $p(\mathbf{y}_{1:N} | \Theta^{(r)*})^{\frac{1}{T^{(r)}}}$. By the SMC method, the marginalization over time-series of latent variables $\mathbf{z}_{1:N}$ is conducted iteratively for time steps $n = 1, 2, \dots, N$. In the MH algorithm, the marginal likelihood $p(\mathbf{y}_{1:N} | \Theta^{(r)*})^{\frac{1}{T^{(r)}}}$ is used to determine whether to accept or reject the sample candidate. In the REPMMH method, exchanges between samples at different temperatures are considered in order to achieve the transitions that are difficult to achieve with the particle marginal Metropolis–Hastings (PMMH) method. The transitions can be realized by passing through a high temperature due to exchange between temperatures, as shown by the red arrows in the MH algorithm. (d) The estimated posterior distributions of parameters Θ as the output.

2.3.2. Introducing the Replica Exchange Method into the PMMH Method

Here, we propose the REPMMH method to accurately estimate the distribution of parameters from observed data $\mathbf{y}_{1:N}$. In our proposed method, we introduce replicas of parameters $\{\Theta\} = \{\Theta^{(1)}, \Theta^{(2)}, \dots, \Theta^{(r)}, \dots, \Theta^{(R)}\}$ at different temperatures $T = [T^{(1)}, T^{(2)}, \dots, T^{(r)}, \dots, T^{(R)}]$ and consider the extended joint marginal posterior distribution as follows:

$$\pi_{\text{EX}}(\{\Theta\} | \mathbf{y}_{1:N}) = \prod_{r=1}^R \pi_{T^{(r)}}(\Theta^{(r)} | \mathbf{y}_{1:N}), \quad (10)$$

where $\pi_{T^{(r)}}(\Theta^{(r)} \mid \mathbf{y}_{1:N})$ expresses the marginal posterior distribution at temperature $T^{(r)}$, which is expressed by using the original marginal posterior distribution $p(\Theta^{(r)} \mid \mathbf{y}_{1:N})$ of the parameter $\Theta^{(r)}$ at the temperature $T^{(1)} = 1.0$ as follows:

$$\pi_{T^{(r)}}(\Theta^{(r)} \mid \mathbf{y}_{1:N}) = \frac{1}{z(T^{(r)})} p(\Theta^{(r)} \mid \mathbf{y}_{1:N})^{\frac{1}{T^{(r)}}} \quad (r = 1, 2, \dots, R), \tag{11}$$

where $z(T^{(r)})$ is a partition function. Note that, as expressed in Equation (11), at sufficiently high temperatures, the distribution of parameters becomes closer to a uniform distribution, independent of the values of $\mathbf{y}_{1:N}$. The distribution with $T^{(1)} = 1.0$ corresponds to the original marginal posterior distribution $p(\Theta \mid \mathbf{y}_{1:N})$ to be investigated.

The marginal posterior distribution at each temperature $p(\Theta^{(r)} \mid \mathbf{y}_{1:N})$ is obtained using Bayes' theorem as follows:

$$p(\Theta^{(r)} \mid \mathbf{y}_{1:N}) = \frac{p(\mathbf{y}_{1:N} \mid \Theta^{(r)}) p(\Theta^{(r)})}{p(\mathbf{y}_{1:N})} \quad (r = 1, 2, \dots, R). \tag{12}$$

Namely, the marginal posterior distribution $p(\Theta^{(r)} \mid \mathbf{y}_{1:N})$ is proportional to a product of a marginal likelihood $p(\mathbf{y}_{1:N} \mid \Theta^{(r)})$ and a prior distribution $p(\Theta^{(r)})$ of parameters $\Theta^{(r)}$. To obtain the marginal likelihood $p(\mathbf{y}_{1:N} \mid \Theta^{(r)})$, marginalization of the joint distribution at each temperature should be conducted as follows:

$$p(\mathbf{y}_{1:N} \mid \Theta^{(r)}) = \int p(\mathbf{y}_{1:N}, \mathbf{z}_{1:N}^{(r)} \mid \Theta^{(r)}) d\mathbf{z}_{1:N}^{(r)} \quad (r = 1, 2, \dots, R), \tag{13}$$

where $\mathbf{z}_{1:N}^{(r)}$ are the latent variables at the temperature $T^{(r)}$. By performing the SMC method for all the time steps at each temperature, the marginalization is conducted numerically.

As shown in Figure 2b,c, in the proposed method, the SMC method and the MH algorithm are conducted for each temperature. In the SMC method, the marginal likelihood of parameters $p(\mathbf{y}_{1:N} \mid \Theta^{(r)*})$ is determined by the numerical marginalization using the candidate of parameters $\Theta^{(r)*}$ proposed in the MH algorithms.

In the MH algorithm, the candidate of parameters $\Theta^{(r)*}$ is determined to be accepted or rejected at each temperature $T^{(r)}$ with the marginal posterior $\pi_{T^{(r)}}(\Theta^{(r)*} \mid \mathbf{y}_{1:N})$ (Figure 2b). Here, the acceptance probability $p_{\text{accept}}^{(r)}$ at each temperature is calculated as follows:

$$p_{\text{accept}}^{(r)} = \min \left(1, \frac{p(\mathbf{y}_{1:N} \mid \Theta^{(r)*}) p(\Theta^{(r)*})}{p(\mathbf{y}_{1:N} \mid \Theta^{(r)[k-1]}) p(\Theta^{(r)[k-1]})} \frac{q(\Theta^{(r)}[k-1] \mid \Theta^{(r)*})}{q(\Theta^{(r)*} \mid \Theta^{(r)[k-1]})} \right). \tag{14}$$

Moreover, we exchange samples between different temperatures $T^{(r)}$ and $T^{(r+1)}$ according to the exchange probability as follows:

$$p_{\text{EX}}(\{\Theta\}, \{\Theta\}^*) = \min(1, R_{\text{EX}}(\{\Theta\}, \{\Theta\}^*)), \tag{15}$$

$$R_{\text{EX}}(\{\Theta\}, \{\Theta\}^*) = \frac{\pi_{\text{EX}}(\{\Theta\}^* \mid \mathbf{y}_{1:N})}{\pi_{\text{EX}}(\{\Theta\} \mid \mathbf{y}_{1:N})}, \tag{16}$$

where $\{\Theta\}^*$ is expressed as follows:

$$\{\Theta\}^* = \{\Theta^{(1)}, \dots, \Theta^{(r+1)}, \Theta^{(r)}, \dots, \Theta^{(R)}\}. \tag{17}$$

Note that the exchange probability $p_{\text{EX}}(\{\Theta\}, \{\Theta\}^*)$ corresponds to the Metropolis criterion for proposing to exchange the samples between different temperatures $T^{(r)}$ and $T^{(r+1)}$. By deciding whether to accept or reject the proposed samples $\{\Theta\}^*$ with the Metropolis criterion of Equation (15), the transition probability $W(\{\Theta\} \rightarrow \{\Theta\}^*)$ for the exchange process satisfies the detailed balance condition as follows:

$$\begin{aligned} & \pi_{\text{EX}}(\{\Theta\} \mid \mathbf{y}_{1:N})W(\{\Theta\} \rightarrow \{\Theta\}^*) \\ &= \pi_{\text{EX}}(\{\Theta\} \mid \mathbf{y}_{1:N})q(\{\Theta\}^* \mid \{\Theta\})p_{\text{EX}}(\{\Theta\}, \{\Theta\}^*) \\ &= \min(\pi_{\text{EX}}(\{\Theta\} \mid \mathbf{y}_{1:N})q(\{\Theta\}^* \mid \{\Theta\}), \pi_{\text{EX}}(\{\Theta\}^* \mid \mathbf{y}_{1:N})q(\{\Theta\} \mid \{\Theta\}^*)) \\ &= \pi_{\text{EX}}(\{\Theta\}^* \mid \mathbf{y}_{1:N})q(\{\Theta\}^* \mid \{\Theta\}) \min\left(\frac{\pi_{\text{EX}}(\{\Theta\} \mid \mathbf{y}_{1:N})}{\pi_{\text{EX}}(\{\Theta\}^* \mid \mathbf{y}_{1:N})}, 1\right) \\ &= \pi_{\text{EX}}(\{\Theta\}^* \mid \mathbf{y}_{1:N})q(\{\Theta\}^* \mid \{\Theta\})p_{\text{EX}}(\{\Theta\}^*, \{\Theta\}) \\ &= \pi_{\text{EX}}(\{\Theta\}^* \mid \mathbf{y}_{1:N})q(\{\Theta\} \mid \{\Theta\}^*)p_{\text{EX}}(\{\Theta\}^*, \{\Theta\}) \\ &= \pi_{\text{EX}}(\{\Theta\}^* \mid \mathbf{y}_{1:N})W(\{\Theta\}^* \rightarrow \{\Theta\}), \end{aligned}$$

where $q(\{\Theta\}^* \mid \{\Theta\})$ is the proposed probability for $\{\Theta\}^*$ and the proposed probability of the exchange process is symmetric $q(\{\Theta\}^* \mid \{\Theta\}) = q(\{\Theta\} \mid \{\Theta\}^*)$. Thus, since the exchange process in the proposed method satisfies the detailed balance condition, the proposed method can sample from the distribution $\pi_{\text{EX}}(\{\Theta\} \mid \mathbf{y}_{1:N})$.

By this exchange process, the REPMH method makes it possible to improve the problem of initial value dependence in the PMMH method. The sampled distributions of the replica $\pi_{T^{(r)}}(\Theta^{(r)} \mid \mathbf{y}_{1:N})$ at higher temperatures become closer to a uniform distribution ideally as follows:

$$\lim_{T^{(r)} \rightarrow \infty} \pi_{T^{(r)}}(\Theta^{(r)} \mid \mathbf{y}_{1:N}) = \lim_{T^{(r)} \rightarrow \infty} \frac{1}{z(T^{(r)})} p(\Theta^{(r)} \mid \mathbf{y}_{1:N})^{\frac{1}{T^{(r)}}} \propto p(\Theta^{(r)} \mid \mathbf{y}_{1:N})^0 = \text{const.} \quad (18)$$

Therefore, in practice, it becomes possible to escape from local optima at sufficiently high temperatures (Figure 2b). Moreover, the samples may not stay in one local optimum since each replica is exchanged between the high temperature and low temperature repeatedly, and we can sample the parameters efficiently. We show the flow of our REPMH method described above in Algorithm 2.

2.3.3. Relations among Particle Markov Chain Monte Carlo Methods

We briefly summarize the differences among the conventional particle Markov chain Monte Carlo (PMCMC) methods and our proposed REPMH method that can estimate parameters of a state space model in Table 1. The particle Gibbs (PG) method is another PMCMC method, and it samples latent variables and parameters in a state space model alternately by using Gibbs sampling [8,30]. The PMMH method combines the SMC method with the MH algorithm, whereas the PG method combines the SMC method with Gibbs sampling. While the SMC method is employed to calculate the marginal likelihood $p(\mathbf{y}_{1:N} \mid \Theta)$ of parameters Θ in the PMMH method, the SMC method is employed to obtain samples of latent variables $\mathbf{z}_{1:N}$ in the PG method [12]. The PMMH method directly targets the marginal posterior distribution $p(\Theta \mid \mathbf{y}_{1:N})$, whereas the PG method targets the joint posterior distribution $p(\mathbf{z}_{1:N}, \Theta \mid \mathbf{y}_{1:N})$ [12]. Note that the SMC method used in the PG method is called the conditional SMC method and uses the previous sample of latent variables $\mathbf{z}_{1:N}[k-1]$ as a particle in the SMC method [12]. Furthermore, advanced versions of the PG method have been proposed, such as the particle Gibbs with ancestor sampling (PGAS) method for improving sampling efficiency and the replica exchange particle Gibbs with ancestor sampling (REPGAS) method to improve the initial value dependence [16,18,19,24]. Samples obtained by employing the PMMH method also have a problem of initial value dependence, similar to those obtained by employing the PG method, and it is considered that combining the PMMH method with the replica exchange method would be effective.

Algorithm 2 Replica Exchange Particle Marginal Metropolis–Hastings (REPMMH) Method.

```

1: initialize the parameters  $\{\Theta\}[0]$ 
2: for  $k = 1, \dots, K$  do
3:   for  $r = 1, \dots, R$  do
4:     draw the sample candidate of parameters  $\Theta^{(r)*} \sim q(\Theta^{(r)*} | \Theta^{(r)}[k-1])$ 
5:     calculate the marginal likelihood  $p(\mathbf{y}_{1:N} | \Theta^{(r)*})$  by using the SMC method according to Equation (13)
6:     calculate the acceptance probability  $p_{\text{accept}}^{(r)}$  with Equation (14)
7:     draw a uniform random number  $\alpha \sim \mathcal{U}(0, 1)$  ( $\mathcal{U}(a, b)$  is a uniform distribution with range  $[a, b)$ )
8:     if  $\alpha \leq p_{\text{accept}}^{(r)}$  then
9:       set the sample of parameters  $\Theta^{(r)}[k] \leftarrow \Theta^{(r)*}$ 
10:    else
11:      set the sample of parameters  $\Theta^{(r)}[k] \leftarrow \Theta^{(r)}[k-1]$ 
12:    end if
13:  end for
14:  choose the replica number  $r_{\text{EX}} \leftarrow 1$  or  $r_{\text{EX}} \leftarrow 2$  for the exchange
15:  repeat
16:    calculate exchange probability  $p_{\text{EX}}(\{\Theta\}, \{\Theta\}^*)$  with Equation (15) for replica numbers  $r_{\text{EX}}$  and  $r_{\text{EX}} + 1$ 
17:    draw a uniform random number  $\alpha_{\text{EX}} \sim \mathcal{U}(0, 1)$ 
18:    if  $\alpha_{\text{EX}} \leq p_{\text{EX}}(\{\Theta\}, \{\Theta\}^*)$  then
19:      exchange replicas  $(\Theta^{(r_{\text{EX}})}[k], \Theta^{(r_{\text{EX}}+1)}[k]) \leftarrow (\Theta^{(r_{\text{EX}}+1)}[k], \Theta^{(r_{\text{EX}})}[k])$ 
20:    end if
21:    set the replica number  $r_{\text{EX}} \leftarrow r_{\text{EX}} + 2$  for the exchange
22:  until  $r_{\text{EX}} \leq R - 1$ 
23: end for
24: return  $\{\{\Theta\}[k]\}_{k=1}^K$ 

```

Table 1. The PMCMC methods for estimating parameters in a state space model.

Method	Target Distribution	Overview
PG	$p(\mathbf{z}_{1:N}, \Theta \mathbf{y}_{1:N})$	Sample parameters Θ and latent variables $\mathbf{z}_{1:N}$ alternately with Gibbs sampling for targeting the joint posterior distribution $p(\mathbf{z}_{1:N}, \Theta \mathbf{y}_{1:N})$. Note that the SMC method is used for sampling latent variables $\mathbf{z}_{1:N}$. The SMC method used in the PG method is called the conditional SMC method and uses the previous sample of latent variables $\mathbf{z}_{1:N}[k-1]$ as a particle in the SMC method [12].
PGAS	$p(\mathbf{z}_{1:N}, \Theta \mathbf{y}_{1:N})$	Sample latent variables $\mathbf{z}_{1:N}$ not only in the forward direction but also in the backward direction in the PG method [16,18,19].
REPGAS	$p(\mathbf{z}_{1:N}, \Theta \mathbf{y}_{1:N})$	Improve the problem of initial value dependence in the PGAS method by combining the replica exchange method and the PGAS method [24].
PMMH	$p(\Theta \mathbf{y}_{1:N})$	Sample parameters Θ with the MH algorithm for targeting directly the marginal posterior distribution $p(\Theta \mathbf{y}_{1:N})$ obtained by marginalization over the distribution of latent variables $\mathbf{z}_{1:N}$. Note that the SMC method is used to calculate the marginal likelihood $p(\mathbf{y}_{1:N} \Theta)$ [12].
REPMMH	$p(\Theta \mathbf{y}_{1:N})$	Improve the problem of initial value dependence in the PMMH method by combining the replica exchange method and the PMMH method.

3. Results

In this section, we show that by employing our proposed replica exchange particle marginal Metropolis–Hastings (REPMMH) method for the Izhikevich neuron model [41,42] and Lévy-driven stochastic volatility model [12,43,44], the marginal posterior distribution of parameters $p(\Theta | y_{1:N})$ can be estimated from observations $y_{1:N}$, and we verify whether the REPMMH method can overcome the problem of initial value dependence in the particle marginal Metropolis–Hastings (PMMH) method. Moreover, we compare the sampling efficiency of the REPMMH method with that of the conventional methods, the PMMH method and the replica exchange particle Gibbs with ancestor sampling (REPGAS) method.

3.1. Izhikevich Neuron Model

To verify the effectiveness of our proposed method, we use the Izhikevich neuron model. The Izhikevich neuron model is a computational model for the membrane potential of a neuron [41,42]:

$$\frac{dv}{dt} = 0.04v^2 + 5v + 140 - u + I_{\text{ext}} + \xi_v(t), \quad (19)$$

$$\frac{du}{dt} = a(bv - u) + \xi_u(t), \quad (20)$$

where v is the membrane potential and u is the membrane recovery variable. I_{ext} is the input current from outside the neuron, and a and b are parameters in the Izhikevich neuron model that represent the characteristic of the neuron. In Equations (19) and (20), we consider additive white Gaussian noise terms $\xi_v(t)$ and $\xi_u(t)$ ($\langle \xi_v(t) \rangle = \langle \xi_u(t) \rangle = 0$, $\langle \xi_v(t)\xi_v(s) \rangle = \sigma_v^2\delta(t-s)$, $\langle \xi_u(t)\xi_u(s) \rangle = \sigma_u^2\delta(t-s)$ and $\langle \xi_v(t)\xi_u(s) \rangle = 0$, where $\delta(t)$ is the Dirac delta function). Here, standard deviations of the membrane potential and membrane recovery variable are expressed by σ_v and σ_u , respectively. If the membrane potential v exceeds the threshold value $v_{\text{th}} = 30$, the membrane potential v and the membrane recovery variable u are reset to c and $u + d$, respectively, as follows:

$$\begin{aligned} v &\leftarrow c, \\ u &\leftarrow u + d, \end{aligned}$$

where c and d are parameters representing the characteristic of the neuron.

Here, we assume that the observations $y_{1:N}$ are the membrane potentials with Gaussian observation noise, and we estimate the parameters $\Theta = \{a, b, c, d\}$ from only the observations $y_{1:N}$. We use the true parameters $\Theta = \{a, b, c, d\} = \{0.02, 0.2, -65, 6\}$ and the number of data $N = 5.0 \times 10^2$ to generate data. In the system model, the means and the variances of the Gaussian noise are $\{\mu_v, \sigma_v^2\} = \{0, 0.25\}$ and $\{\mu_u, \sigma_u^2\} = \{0, 10^{-4}\}$. In the observation model, the mean and the variance of the Gaussian noise are $\{\mu_y, \sigma_y^2\} = \{0, 1\}$. We show the generated data from the Izhikevich neuron model in Figure 3. In Figure 3, complex spike activities with different inter-spike intervals and different peaks are seen in response to external inputs. We assume that only one-dimensional time-series of observed data y_n and external inputs can be used, while the latent dynamics are governed by two-dimensional nonlinear dynamical systems with four parameters $\Theta = \{a, b, c, d\}$. We employ our REPMMH method and the conventional methods, the PMMH method and the REPGAS method, for the generated data in Figure 3 to estimate the posterior distribution of the parameters $\Theta = \{a, b, c, d\}$. In all methods, the initial values of the parameters $\Theta[0]$ are $\{a, b, c, d\} = \{0.025, 0.15, -60, 5.5\}$, the number of samples K is 10^6 , the number of burn-in samples $K_{\text{burn-in}}$ is 10^6 , and the number of particles M is 50. In our REPMMH method and the REPGAS method, the number of temperatures R is 64.

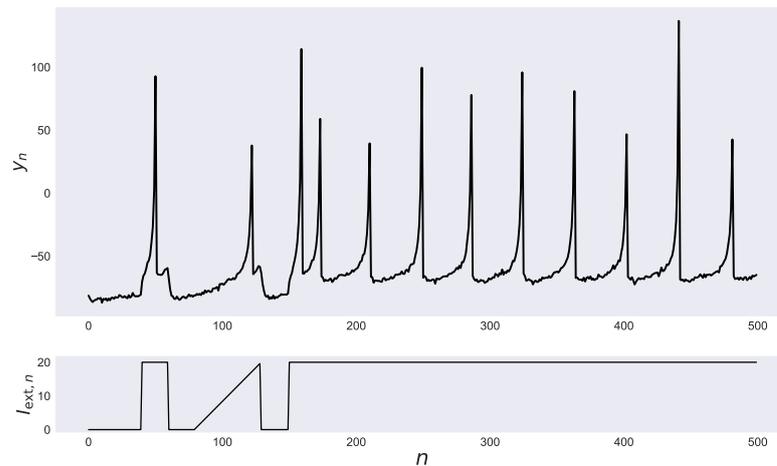


Figure 3. Observations and external inputs used to evaluate the proposed method. Observed membrane potential of the Izhikevich neuron model y_n (top) in response to input current $I_{ext,n}$ (bottom) is shown.

We show in Figure 4 the estimated posterior distribution of parameters $p(\Theta | y_{1:N})$ obtained by employing the PMMH method. In each graph, the vertical axis expresses the value of the probability density function, while the horizontal axis expresses the values of parameters a, b, c and d . Furthermore, the solid lines represent the true values, the dashed lines represent the initial values, and the histograms represent the estimated posterior distributions of the parameters. From Figure 4, we find that a peak of the estimated posterior distribution of parameter d , $p(d | y_{1:N})$, is located around its true value $d = 6.0$. However, the maximum values of the estimated posterior distribution of the other three parameters, a, b and c , remain around their initial values ($a = 0.025, b = 0.15, c = -60$), which are far from their true values ($a = 0.020, b = 0.20, c = -65$). Thus, the joint posterior distribution of four parameters is found to be not adequately estimated. From this result, the samples in the PMMH method are considered to remain in the local optimum since the initial values are far from the true value.

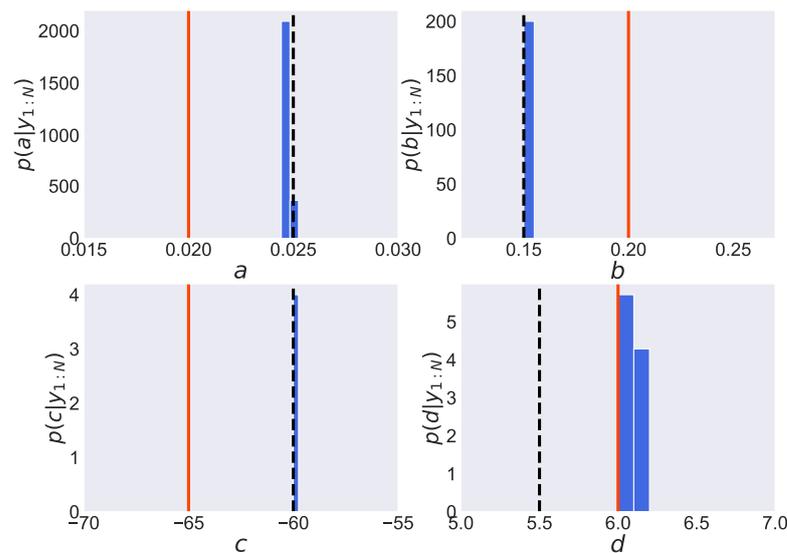


Figure 4. Estimated posterior distributions obtained by employing the PMMH method in the Izhikevich neuron model. In each graph, the estimated probability density function of the parameter (a, b, c and d) is shown by the blue histogram. The red solid and black dashed lines express the true and initial values, respectively.

We show in Figure 5 the estimated posterior distribution of parameters $p(\Theta | y_{1:N})$ obtained by employing the REPMMH method. From Figure 5, we find that the maximum values of the estimated posterior distribution of parameters are located around the true values ($a = 0.020$, $b = 0.20$, $c = -65$, $d = 6.0$), even though the initial values of the parameters ($a = 0.025$, $b = 0.15$, $c = -60$, $d = 5.5$) are set to be far from the true values. This improvement in estimation accuracy would be induced by combination with the replica exchange method. It is easier to obtain samples from a wider range since the replica exchange method allows samples to pass through high temperatures. From these results, we find that the problem of initial value dependence in the PMMH method is improved by employing our proposed method.

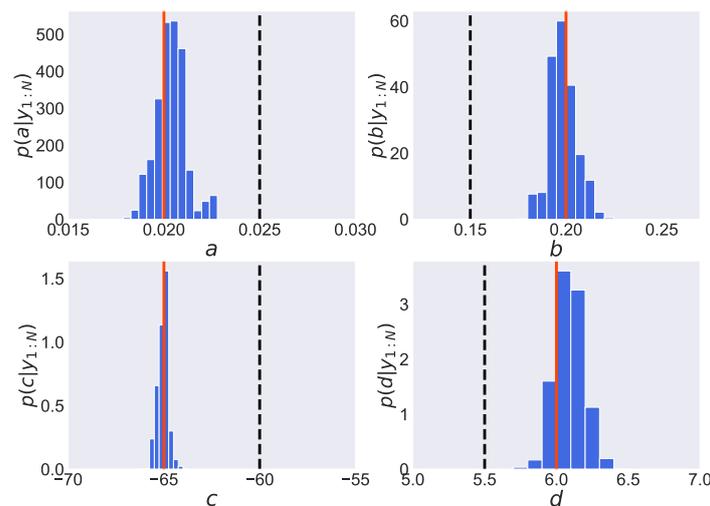


Figure 5. Estimated posterior distributions obtained by employing the REPMMH method in the Izhikevich neuron model. See also the captions of the figure and subfigures for Figure 4.

Moreover, we show in Figure 6 the estimated posterior distribution of parameters $p(\Theta | y_{1:N})$ obtained by employing the REPGAS method. As shown in this figure, the distributions are estimated to be almost the same as those obtained by employing the REPMMH method, which indicates that the true values are estimated properly.

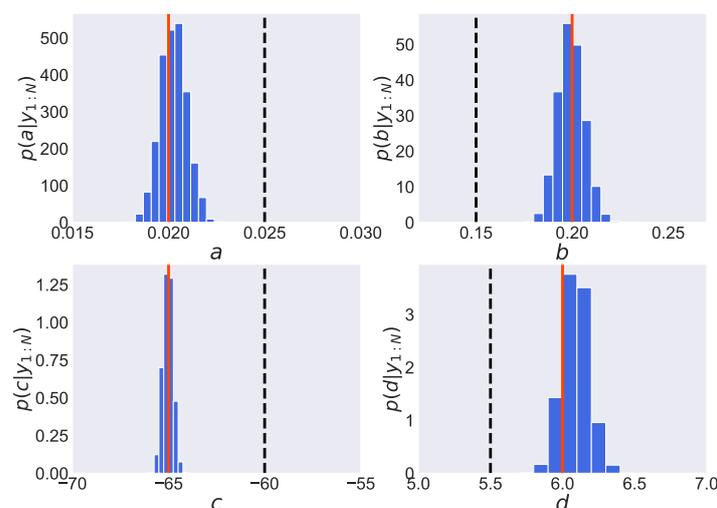


Figure 6. Estimated posterior distributions obtained by employing the replica exchange particle Gibbs with ancestor sampling (REPGAS) method in the Izhikevich neuron model. See also the captions of the figure and subfigures for Figure 4.

In order to investigate the efficiency of sampling parameters in the proposed method and existing methods, we show in Figure 7 the autocorrelation function results calculated using the samples of the PMMH method, the REPGAS method and our proposed REPMMH method. In all the parameters a , b , c and d , the decay of the autocorrelation in the REPMMH samples is faster using the REPMMH samples than those calculated using the PMMH samples and the REPGAS samples. The time constant of the autocorrelation function has a strong influence on the convergence time of the PMCMC method. The time constants of the autocorrelation functions for the REPMMH samples are around 20 for all parameters a , b , c and d , while those of the autocorrelation functions for the PMMH samples are more than 10^5 , as shown in Figure 7. Since the computational cost of the exchange process in the REPMMH method is very small compared to the computational cost of the SMC method, the computational cost of the REPMMH method is approximately $R = 64$ times the computational cost of the PMMH method. Nevertheless, the REPMMH method drastically improves the sampling efficiency compared to the increase in the computational cost; the REPMMH method is $R = 64$ times more computationally expensive than the PMMH method, while the effective sample size of the REPMMH method is much larger (around 10^3 times larger) than that of the PMMH method.

When the same number of temperatures and particles is used, the REPGAS method is more computationally expensive than the REPMMH method since the REPGAS method requires the sampling of the latent variables $z_{1:N}$ and the ancestor sampling, which considers sampling of the latent variables $z_{1:N}$ not only in the forward direction but also in the backward direction in the conditional SMC method. Nevertheless, the REPMMH method has high sampling efficiency compared to the REPGAS method. Thus, we find that the sampling efficiency of our proposed REPMMH method is higher than that of the conventional methods.

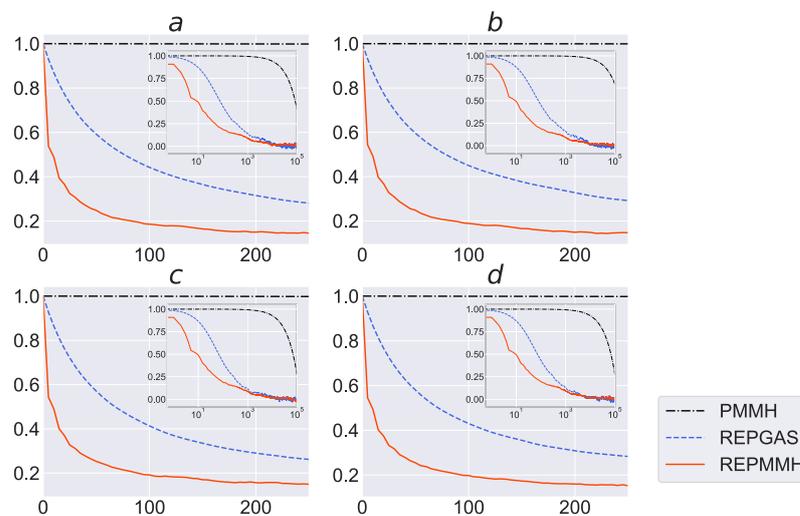


Figure 7. Autocorrelation as a function of the lag length for parameters a , b , c and d in the Izhikevich neuron model. Results for the PMMH method (black dashed-dotted line), the REPGAS method (blue dashed line) and the REPMMH method (red solid line) are shown. Each inset figure represents the result when the horizontal axis is the logarithmic scale. In results obtained by the REPGAS method and the REPMMH method, samples at $T^{(1)} = 1.0$ were used.

Moreover, in order to evaluate the influence of the number of temperatures R and the number of particles M on the estimated results, we compare the estimated results in various settings. We show the estimated results with the numbers of temperatures $R = 1, 4, 16$ and 64 in Table 2. Table 2 shows the mode values of the estimated distributions, the standard deviations (Std) of the estimated distributions and the values of autocorrelation functions (ACF) with the lag length 30 for the numbers of temperatures $R = 1, 4, 16$ and 64 . Note that the numbers of particles M are 50 in all cases and the maximum value of temperature is fixed at $T^{(R)} = 1.1^{63}$ for $R > 1$.

As mentioned above, for the number of temperatures $R = 1$, we can estimate the parameter d around the true value $d = 6.0$, while the other three parameters a, b and c remain at their initial values ($a = 0.025, b = 0.15, c = -60$). We also find that the samples of parameters a, b and c cannot move enough from their initial values since the values of the standard deviations are very small. For the number of temperatures $R = 4$, we find the samples can escape the local optima and we can estimate the true values of all parameters a, b, c and d accurately due to the high temperatures that allow escape from the local optima. However, since the values of the autocorrelation functions are close to 1.0, we need a large number of samples in order to estimate the shape of the distribution $p(\Theta | y_{1:N})$. On the other hand, we find that the values of the autocorrelation functions are smaller for $R = 16$ and 64.

Table 2. The estimated results with the numbers of temperatures $R = 1, 4, 16$ and 64.

Parameter		$R = 1$	$R = 4$	$R = 16$	$R = 64$
$a = 0.020$	Mode	0.0251	0.0200	0.0205	0.0205
	Std	5.5×10^{-5}	6.9×10^{-4}	6.7×10^{-4}	7.8×10^{-4}
	ACF	0.9999	0.9914	0.5175	0.3074
$b = 0.20$	Mode	0.155	0.200	0.200	0.200
	Std	3.0×10^{-4}	7.2×10^{-3}	7.3×10^{-3}	7.0×10^{-3}
	ACF	0.9999	0.9919	0.5773	0.3082
$c = -65$	Mode	-60.0	-64.75	-65.0	-65.0
	Std	2.1×10^{-3}	2.8×10^{-1}	2.5×10^{-1}	2.6×10^{-1}
	ACF	0.9999	0.9926	0.5359	0.3117
$d = 6.0$	Mode	6.10	6.10	6.05	6.05
	Std	2.0×10^{-2}	8.9×10^{-2}	9.8×10^{-2}	9.8×10^{-2}
	ACF	0.9999	0.9928	0.5222	0.3176

We show the estimated results with the numbers of particles $M = 10, 20, 30, 40$ and 50 in Table 3. Note that the numbers of temperatures R are 64 in all cases. For the numbers of particles $M = 10$ and 20, the estimated values of the parameters a, b, c and d are far from the true values ($a = 0.020, b = 0.20, c = -65, d = 6.0$). We consider that these results are due to the low approximation accuracy of the marginal likelihood $p(y_{1:N} | \Theta)$ in the SMC method with too small numbers of particles. For the numbers of particles $M \geq 30$, we can estimate the true values of parameters a, b, c and d . Since there is no significant difference between the mode values, the standard deviations and the values of the autocorrelation functions for the numbers of particles $M = 30, 40$ and 50, we consider that the number of particles M is sufficient for this problem if it is above 30.

Table 3. The estimated results with the numbers of particles $M = 10, 20, 30, 40$ and 50.

Parameter		$M = 10$	$M = 20$	$M = 30$	$M = 40$	$M = 50$
$a = 0.020$	Mode	0.0220	0.0210	0.0200	0.0200	0.0205
	Std	5.9×10^{-4}	7.5×10^{-4}	7.3×10^{-4}	6.7×10^{-4}	7.8×10^{-4}
	ACF	0.5548	0.2707	0.3072	0.3391	0.3074
$b = 0.20$	Mode	0.195	0.190	0.200	0.195	0.200
	Std	4.9×10^{-3}	8.3×10^{-3}	7.1×10^{-3}	5.9×10^{-3}	7.0×10^{-3}
	ACF	0.2738	0.2611	0.2932	0.3352	0.3082
$c = -65$	Mode	-60.75	-64.50	-65.00	-65.00	-65.00
	Std	2.9×10^{-1}	3.7×10^{-1}	2.7×10^{-1}	2.8×10^{-1}	2.6×10^{-1}
	ACF	0.9223	0.3071	0.2793	0.3701	0.3117
$d = 6.0$	Mode	6.50	6.10	6.10	6.05	6.05
	Std	1.4×10^{-1}	8.2×10^{-2}	1.0×10^{-1}	9.3×10^{-2}	9.8×10^{-2}
	ACF	0.6096	0.2750	0.3220	0.2985	0.3176

3.2. Lévy-Driven Stochastic Volatility Model

Next, we also verify the effectiveness of the proposed method using the Lévy-driven stochastic volatility model [12,43,44]. In this model, the dynamics of the logarithm of asset price $y^*(t)$ are represented by the following differential equation:

$$dy^*(t) = \left\{ \mu + \beta \sigma^2(t) \right\} dt + \sigma(t) dB(t), \tag{21}$$

where μ is the drift parameter and β is the risk premium. $B(t)$ is the Brownian motion and $\sigma^2(t)$ represents the volatility. The dynamics of the volatility $\sigma^2(t)$ are modeled by the following Lévy-driven Ornstein–Uhlenbeck process:

$$d\sigma^2(t) = -\lambda\sigma^2(t)dt + dz(\lambda t), \tag{22}$$

where λ is a positive constant and $z(t)$ is a non-Gaussian Lévy process with positive increments. The observation at the time step n , y_n , in this model is obtained by the following Gaussian distribution:

$$y_n \sim \mathcal{N}(\mu\Delta + \beta\sigma_n^2, \sigma_n^2), \tag{23}$$

where Δ is the length of the time interval.

The stochastic volatility models are numerically investigated by using discretized dynamical models [12,43,44], and the estimation algorithm for the parameters of the stochastic volatility models has been investigated using such the discretized ones [12]. The integrated volatility σ_n^2 at the time step n is calculated as follows:

$$\begin{aligned} \sigma_n^2 &= \int_{(n-1)\Delta}^{n\Delta} \sigma^2(u)du \\ &= \lambda^{-1} \left[z(\lambda n\Delta) - \sigma^2(n\Delta) - z\{\lambda(n-1)\Delta\} + \sigma^2\{(n-1)\Delta\} \right], \end{aligned} \tag{24}$$

where $\sigma^2(n\Delta)$ and $z(\lambda n\Delta)$ are, respectively, represented as follows:

$$\sigma^2(n\Delta) = \exp(-\lambda\Delta)\sigma^2\{(n-1)\Delta\} + \eta_{\sigma,n}, \tag{25}$$

$$z(\lambda n\Delta) = z\{\lambda(n-1)\Delta\} + \eta_{z,n}. \tag{26}$$

Here, we address the case where the volatility $\sigma^2(t)$ follows a tempered stable marginal distribution [44]. Following [2,44], $\eta_{\sigma,n}$ and $\eta_{z,n}$ are obtained as follows:

$$\eta_{\sigma,n} = \sum_{i=1}^{\infty} \min\left(\left(\frac{a_i\kappa}{A\lambda\Delta}\right)^{-1/\kappa}, e_i v_i^{1/\kappa}\right) \exp(-\lambda\Delta r_i) + \sum_{i=1}^{N(\lambda\Delta)} c_i \exp(-\lambda\Delta r_i^*), \tag{27}$$

$$\eta_{z,n} = \sum_{i=1}^{\infty} \min\left(\left(\frac{a_i\kappa}{A\lambda\Delta}\right)^{-1/\kappa}, e_i v_i^{1/\kappa}\right) + \sum_{i=1}^{N(\lambda\Delta)} c_i, \tag{28}$$

where $A = 2^\kappa \delta \kappa^2 / \Gamma(1 - \kappa)$, $a_1 < a_2 < \dots$ are the arrival times of a Poisson process with intensity 1, e_1, e_2, \dots are independent and identically distributed exponential random variables with mean $2\gamma^{-1/\kappa}$, and $v_1, v_2, \dots, r_1, r_2, \dots$ and r_1^*, r_2^*, \dots are standard uniform random variables. c_1, c_2, \dots are obtained from a gamma distribution with the shape parameter $1 - \kappa$ and the scale parameter $2\gamma^{-1/\kappa}$, and $N(\lambda\Delta)$ is obtained from a Poisson distribution with mean $\lambda\Delta\delta\gamma\kappa$. Here, κ, δ, γ and λ are the parameters $\Theta = \{\kappa, \delta, \gamma, \lambda\}$ to be estimated.

In this paper, we employ the proposed method and the PMMH method for the stochastic volatility model. The PMMH-based methods, including the proposed method, can be applied to complex models such as the Lévy-driven stochastic volatility model, as long as the probability density of the observation model can be calculated. On the other hand, the PG-based method is difficult to apply to the stochastic volatility model due to the need to calculate the probability density of the system model [12]. Following [12], we use the true parameters $\Theta = \{\kappa, \delta, \gamma, \lambda\} = \{0.5, 1.41, 2.83, 0.1\}$, the number of data $N = 4.0 \times 10^2$ and the time interval of length $\Delta = 1.0$ to generate data. In order to estimate the parameters Θ , we use the initial values of the parameters $\Theta[0] = \{0.25, 7.41, 9.83, 1.5\}$, the number of samples $K = 1.5 \times 10^5$, the number of burn-in samples $K_{\text{burn-in}} = 10^5$ and the number of particles $M = 200$ in both the proposed REPMMH method and the PMMH method. In the REPMMH method, the number of temperatures R is 64.

We show in Figure 8 the estimated posterior distribution of parameters $p(\Theta | y_{1:N})$ obtained by employing the PMMH method. From Figure 8, we find that a peak of the estimated posterior distribution of parameter λ , $p(\lambda | y_{1:N})$, is located around its true value $\lambda = 0.1$. However, the maximum values of the estimated posterior distribution of the other three parameters, κ , δ and γ , are far from their true values ($\kappa = 0.5$, $\delta = 1.41$, $\gamma = 2.83$). Thus, the joint posterior distribution of the four parameters is found to be not adequately estimated. It is considered that the target distribution is not reached with a small number of samples since the sampling efficiency of the PMMH method is low.

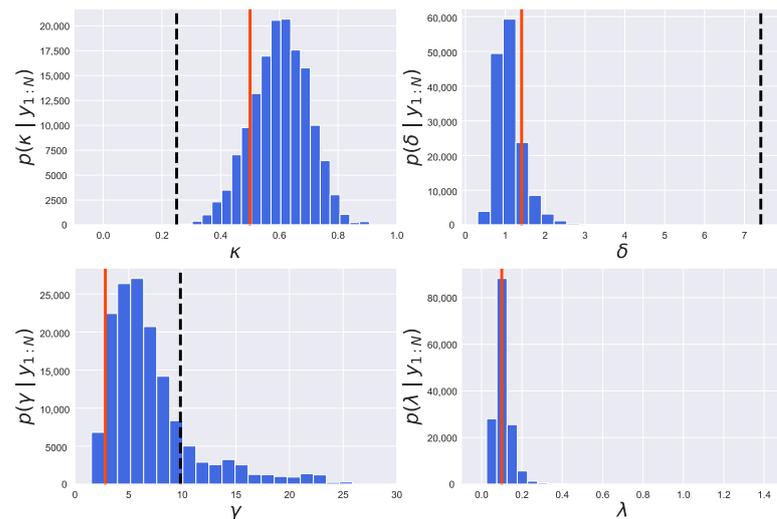


Figure 8. Estimated posterior distributions obtained by employing the PMMH method in the Lévy-driven stochastic volatility model. In each graph, the estimated probability density function of parameters (κ , δ , γ and λ) is shown by the blue histogram. The red solid and black dashed lines express the true and initial values, respectively.

We show in Figure 9 the estimated posterior distributions of parameters $p(\Theta | y_{1:N})$ obtained by employing the REPMMH method. From Figure 9, we find that the true values of parameters $\Theta = \{\kappa, \delta, \gamma, \lambda\} = \{0.5, 1.41, 2.83, 0.1\}$ are estimated appropriately by using the same number of samples and the same initial values $\Theta[0] = \{0.25, 7.41, 9.83, 1.5\}$ as the PMMH method. The results in Figures 8 and 9 show that our REPMMH method has higher sampling efficiency than the PMMH method.

Moreover, we show in Figure 10 the autocorrelation function results calculated using the samples of the PMMH method and the REPMMH method. In all parameters κ , δ , γ and λ , the decay of the autocorrelation is faster using the REPMMH samples than that using the PMMH samples. As shown in Figure 10, the time constants of the autocorrelation functions for the REPMMH samples are less than 15 for all parameters κ , δ , γ and λ , while the time constant of the autocorrelation functions for the PMMH samples for the parameter γ is more than 3.0×10^3 . As mentioned above, since the computational cost of the REPMMH method is approximately $R = 64$ times the computational cost of the PMMH method, the REPMMH method improves the sampling efficiency compared to the increase in the computational cost.

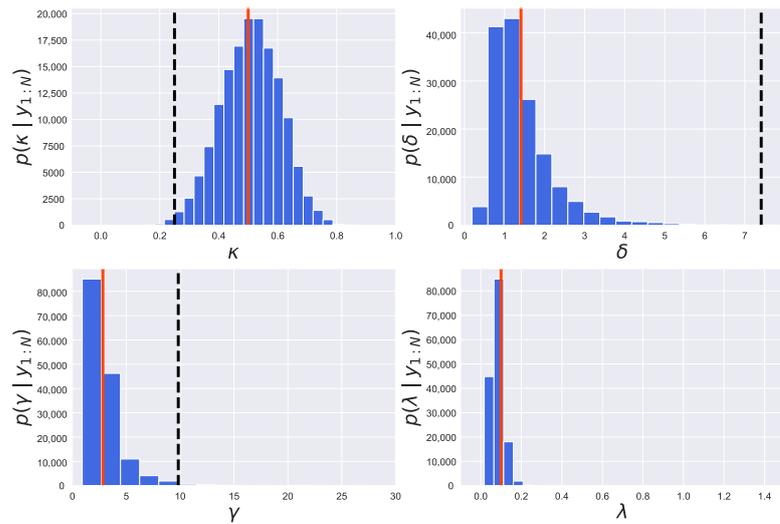


Figure 9. Estimated posterior distributions obtained by employing the REPMMH method in the Lévy-driven stochastic volatility model. See also the captions of the figure and subfigures for Figure 8.

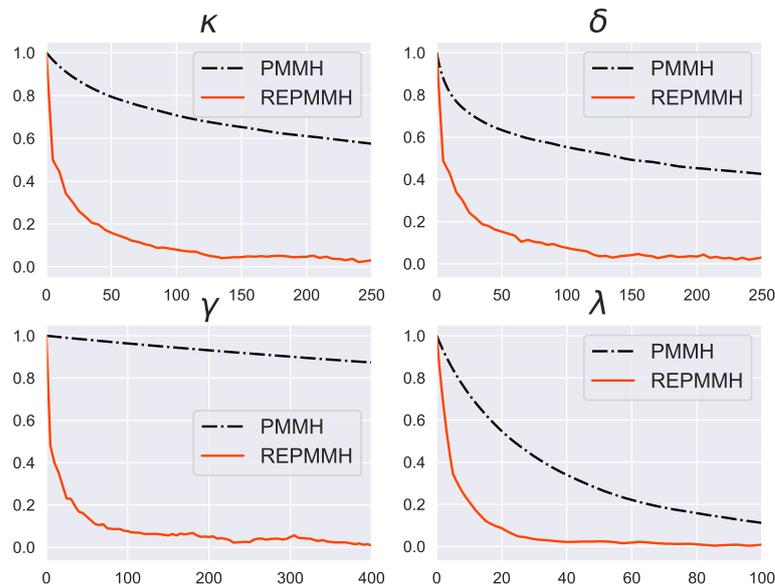


Figure 10. Autocorrelation as a function of the lag length for parameters κ , δ , γ and λ in the Lévy-driven stochastic volatility model. Results for the PMMH method (black dashed-dotted line) and the REPMMH method at $T^{(1)} = 1.0$ (red solid line) are shown.

4. Concluding Remarks

In this paper, we have proposed the replica exchange particle marginal Metropolis–Hastings (REPMMH) method in order to estimate the marginal posterior distribution of parameters $p(\Theta | y_{1:N})$ of the state space model. Our proposed method can be applied to complex models such as the Lévy-driven stochastic volatility model, even if the probability densities of the system models cannot be calculated explicitly. By the proposed method, we introduce the exchange between samples of model parameters Θ at different temperatures and realize an efficient sampling method for model parameters governing the nonlinear dynamical systems.

Using nonlinear dynamical models such as the Izhikevich neuron model and Lévy-driven stochastic volatility model, we show that our proposed REPMMH method can improve the problem of initial value dependence of the particle marginal Metropolis–Hastings (PMMH) method. The results have shown that the proposed REPMMH method

accurately estimates the marginal posterior distribution of parameters. Moreover, by comparing the autocorrelation functions of the obtained samples, it has been also shown that our proposed REPMMH method can sample more efficiently than the conventional methods. In the replica exchange particle Gibbs with ancestor sampling (REPGAS) method, the next sample of latent variables is obtained under the strong influence of the current sample of latent variables. On the other hand, in the REPMMH method, the correlation of the latent variables between the current and next steps is low since the REPMMH method only calculates the marginal likelihood of the next step, regardless of the latent variables obtained in the current step. Therefore, it is considered that the REPMMH method can sample parameters more efficiently than the REPGAS method.

In this paper, although we conducted the experiments by using two specific dynamical models: the Izhikevich neuron model and the Lévy-driven stochastic volatility model, the proposed REPMMH method can be applied to various dynamical systems described by ordinary or partial differential equations. Although the proposed method can be applied to any ordinary or partial differential equations that can be represented as state space models, applications of the proposed method are difficult when the system models for the dynamical systems or the observation models are completely unknown. In such cases, we consider that combining the proposed method with non-parametric Bayesian methods is effective. We leave this for future work.

Author Contributions: Conceptualization, T.O.; Formal analysis, H.I. and T.O.; Investigation, H.I., K.H. and T.O.; Methodology, H.I. and T.O.; Supervision, T.O.; Writing—original draft, H.I.; Writing—review & editing, H.I., K.H. and T.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Grants-in-Aid for Scientific Research for Innovative Areas “Initiative for High-Dimensional Data driven Science through Deepening of Sparse Modeling” (JSPS KAKENHI Grant No. JP25120010) and for Scientific Research (B) (JSPS KAKENHI Grant No. JP17H02923, JP19H04125, JP21H03509), and a Fund for the Promotion of Joint International Research (Fostering Joint International Research) (JSPS KAKENHI Grant No. JP15KK0010) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan, and CREST (Nos. JPMJCR1755, JPMJCR1861, JPMJCR1914), Japan Science and Technology Agency, Japan. This work was partially supported by “Program for Promoting Researches on the Supercomputer Fugaku” (DPMSD, Project ID: JPMXP1020200307) from the MEXT, Japan.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

PMMH	Particle Marginal Metropolis–Hastings
REPMMH	Replica Exchange Particle Marginal Metropolis–Hastings
SMC	Sequential Monte Carlo
EM	Expectation–Maximization
MCMC	Markov Chain Monte Carlo
PMCMC	Particle Markov Chain Monte Carlo
MH	Metropolis–Hastings
PG	Particle Gibbs
PGAS	Particle Gibbs with Ancestor Sampling
REPGAS	Replica Exchange Particle Gibbs with Ancestor Sampling

References

1. Netto, M.A.; Gimeno, L.; Mendes, M.J. A new spline algorithm for non-linear filtering of discrete time systems. *IFAC Proc. Vol.* **1978**, *11*, 2123–2130. [[CrossRef](#)]
2. Doucet, A.; Godsill, S.; Andrieu, C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Stat. Comput.* **2000**, *10*, 197–208. [[CrossRef](#)]
3. Ghahramani, Z.; Hinton, G.E. *The EM Algorithm for Mixtures of Factor Analyzers*; Technical Report CRG-TR-96-1; Department of Computer Science, University of Toronto: Toronto, ON, USA, 1996.
4. Kitagawa, G. A self-organizing state-space model. *J. Am. Stat. Assoc.* **1998**, *93*, 1203–1215. [[CrossRef](#)]
5. Meyer, R.; Christensen, N. Fast Bayesian reconstruction of chaotic dynamical systems via extended Kalman filtering. *Phys. Rev. E* **2001**, *65*, 016206. [[CrossRef](#)]
6. Doucet, A.; De Freitas, N.; Gordon, N. An introduction to sequential Monte Carlo methods. In *Sequential Monte Carlo Methods in Practice*; Doucet, A., De Freitas, N., Gordon, N., Eds.; Springer: New York, NY, USA, 2001; pp. 3–14.
7. Hyndman, R.J.; Koehler, A.B.; Snyder, R.D.; Grose, S. A state space framework for automatic forecasting using exponential smoothing methods. *Int. J. Forecast.* **2002**, *18*, 439–454. [[CrossRef](#)]
8. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
9. Vogelstein, J.T.; Watson, B.O.; Packer, A.M.; Yuste, R.; Jedynek, B.; Paninski, L. Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophys. J.* **2009**, *97*, 636–655. [[CrossRef](#)]
10. Vogelstein, J.T.; Packer, A.M.; Machado, T.A.; Sippy, T.; Babadi, B.; Yuste, R.; Paninski, L. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* **2010**, *104*, 3691–3704. [[CrossRef](#)]
11. Tsunoda, T.; Omori, T.; Miyakawa, H.; Okada, M.; Aonishi, T. Estimation of intracellular calcium ion concentration by nonlinear state space modeling and expectation-maximization algorithm for parameter estimation. *J. Phys. Soc. Jpn.* **2010**, *79*, 124801. [[CrossRef](#)]
12. Andrieu, C.; Doucet, A.; Holenstein, R. Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. B* **2010**, *72*, 269–342. [[CrossRef](#)]
13. Meng, L.; Kramer, M.A.; Eden, U.T. A sequential Monte Carlo approach to estimate biophysical neural models from spikes. *J. Neural Eng.* **2011**, *8*, 065006. [[CrossRef](#)]
14. Paninski, L.; Vidne, M.; DePasquale, B.; Ferreira, D.G. Inferring synaptic inputs given a noisy voltage trace via sequential Monte Carlo methods. *J. Comput. Neurosci.* **2012**, *33*, 1–19. [[CrossRef](#)]
15. Snyder, R.D.; Ord, J.K.; Koehler, A.B.; McLaren, K.R.; Beaumont, A.N. Forecasting compositional time series: A state space approach. *Int. J. Forecast.* **2017**, *33*, 502–512. [[CrossRef](#)]
16. Lindsten, F.; Schön, T.B.; Jordan, M.I. Ancestor sampling for particle Gibbs. In Proceedings of the Advances in Neural Information Processing Systems, Stateline, NV, USA, 3–8 December 2012; pp. 2600–2608.
17. Henriksen, S.; Wills, A.; Schön, T.B.; Ninness, B. Parallel implementation of particle MCMC methods on a GPU. *IFAC Proc. Vol.* **2012**, *45*, 1143–1148. [[CrossRef](#)]
18. Frigola, R.; Lindsten, F.; Schön, T.B.; Rasmussen, C.E. Bayesian inference and learning in Gaussian process state-space models with particle MCMC. In Proceedings of the Advances in Neural Information Processing Systems, Stateline, NV, USA, 5–10 December 2013; pp. 3181–3189.
19. Lindsten, F.; Jordan, M.I.; Schön, T.B. Particle Gibbs with ancestor sampling. *J. Mach. Learn. Res.* **2014**, *15*, 2145–2184.
20. Omori, T.; Kuwatani, T.; Okamoto, A.; Hukushima, K. Bayesian inversion analysis of nonlinear dynamics in surface heterogeneous reactions. *Phys. Rev. E* **2016**, *94*, 033305. [[CrossRef](#)] [[PubMed](#)]
21. Omori, T.; Sekiguchi, T.; Okada, M. Belief propagation for probabilistic slow feature analysis. *J. Phys. Soc. Jpn.* **2017**, *86*, 084802. [[CrossRef](#)]
22. Rangapuram, S.S.; Seeger, M.W.; Gasthaus, J.; Stella, L.; Wang, Y.; Januschowski, T. Deep state space models for time series forecasting. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 2–8 December 2018; pp. 7785–7794.
23. Wang, P.; Yang, M.; Peng, Y.; Zhu, J.; Ju, R.; Yin, Q. Sensor control in anti-submarine warfare—A digital twin and random finite sets based approach. *Entropy* **2019**, *21*, 767. [[CrossRef](#)]
24. Inoue, H.; Hukushima, K.; Omori, T. Replica exchange particle-Gibbs method with ancestor sampling. *J. Phys. Soc. Jpn.* **2020**, *89*, 104801. [[CrossRef](#)]
25. Shapovalova, Y. “Exact” and approximate methods for Bayesian inference: stochastic volatility case study. *Entropy* **2021**, *23*, 466. [[CrossRef](#)]
26. Gregory, P. *Bayesian Logical Data Analysis for the Physical Sciences*; Cambridge University Press: Cambridge, UK, 2005.
27. Andrieu, C.; Doucet, A.; Punskeya, E. Sequential Monte Carlo methods for optimal filtering. In *Sequential Monte Carlo Methods in Practice*; Doucet, A., De Freitas, N., Gordon, N., Eds.; Springer: New York, NY, USA, 2001; pp. 79–95.
28. Wu, C.J. On the convergence properties of the EM algorithm. *Annal. Stat.* **1983**, *11*, 95–103. [[CrossRef](#)]
29. McLachlan, G.J.; Krishnan, T. *The EM Algorithm and Extensions*; Wiley: Hoboken, NY, USA, 1996.
30. Geman, S.; Geman, D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **1984**, *6*, 721–741. [[CrossRef](#)]
31. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [[CrossRef](#)]

32. Hastings, W.K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **1970**, *57*, 97–109. [[CrossRef](#)]
33. Cunningham N.; Griffin J. E.; Wild D. L. ParticleMDI: particle Monte Carlo methods for the cluster analysis of multiple datasets with applications to cancer subtype identification. *Adv. Data Anal. Classif.* **2020**, *14*, 463–484. [[CrossRef](#)]
34. Wang, S.; Wang, L. Particle Gibbs sampling for Bayesian phylogenetic inference. *Bioinformatics* **2021**, *37*, 642–649. [[CrossRef](#)]
35. Jasra, A.; Persing, A.; Beskos, A.; Heine, K.; De Iorio, M. Bayesian inference for duplication–mutation with complementarity network models. *J. Comput. Biol.* **2015**, *22*, 1025–1033. [[CrossRef](#)] [[PubMed](#)]
36. Du D.; Hu,Z.; Du, Y. Model Identification and Physical Exercise Control using Nonlinear Heart Rate Model and Particle Filter. In 2019 IEEE 15th International Conference on Automation Science and Engineering, Vancouver, BC, Canada, 22–26 August 2019; pp. 405–410.
37. Osmundsen, K. K.; Kleppe, T. S.; Liesenfeld, R.; Oglend, A. Estimating the Competitive Storage Model with Stochastic Trends in Commodity Prices. *Econometrics* **2021**, *9*, 40. [[CrossRef](#)]
38. Hukushima, K.; Nemoto, K. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604–1608. [[CrossRef](#)]
39. Urano, R.; Okamoto, Y. Designed-walk replica-exchange method for simulations of complex systems. *Comput. Phys. Commun.* **2015**, *196*, 380–383. [[CrossRef](#)]
40. Motonaka, K.; Miyoshi, S. Connecting PM and MAP in Bayesian spectral deconvolution by extending exchange Monte Carlo method and using multiple data sets. *Neural Netw.* **2019**, *118*, 159–166. [[CrossRef](#)]
41. Izhikevich, E.M. Simple model of spiking neurons. *IEEE Trans. Neural Netw.* **2003**, *14*, 1569–1572. [[CrossRef](#)]
42. Izhikevich, E.M. Which model to use for cortical spiking neurons?. *IEEE Trans. Neural Netw.* **2004**, *15*, 1063–1070. [[CrossRef](#)]
43. Barndorff-Nielsen, O.E.; Shephard, N. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. B* **2001**, *63*, 167–241. [[CrossRef](#)]
44. Barndorff-Nielsen, O.E.; Shephard, N. Normal modified stable processes. *Theor. Probab. Math. Stat.* **2001**, *65*, 1–19.