Statistica Sinica Preprint No: SS-2023-0332	
Title	Likelihood-free Gibbs Sequential Monte Carlo Sampling
Manuscript ID	SS-2023-0332
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202023.0332
Complete List of Authors	Weixuan Zhu, Wei Li and Weining Shen
Corresponding Authors	Weining Shen
E-mails	weinings@uci.edu
Notice: Accepted author version.	

Statistica Sinica

Likelihood-free Gibbs Sequential Monte Carlo Sampling

Weixuan Zhu[1,2], Wei Li[2] and Weining Shen[3]

[1] *Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen, China*

[2] *Department of Statistics and Data Science, School of Economics, Xiamen University, Xiamen, China*

[3] *Department of Statistics, University of California, Irvine*

Abstract: Approximate Bayesian computation (ABC) has become a standard tool to conduct Bayesian inference for models with intractable likelihoods. However, most existing ABC methods suffer from the curse of dimensionality when the number of parameters is large. To solve this problem, we introduce a Gibbs Sequential Monte Carlo (SMC) method that utilizes a Gibbs kernel to update parameters within the SMC framework and approximate the conditional distribution of the parameters using a variety of regression adjustment methods. We discuss the computational advantage of our method over existing approaches and establish the theoretical property of the Gibbs kernel. We further demonstrate the superior numerical performance of our method using simulation studies and an application to cell motility example.

Key words and phrases: Approximate Bayesian computation, Cell motility, Markov chain Monte Carlo, Regression adjustment, Random forest.

1. Introduction

In Bayesian statistics, Approximate Bayesian Computation (ABC) is a powerful method designed to tackle the challenges posed by intractable likelihood functions. The first ABC-related ideas date back to the 1980s (Rubin, 1984). The main idea of the ABC is to simulate data from the model with different sets of parameters and retain the parameter values if their corresponding simulated data are sufficiently close to the observed data (e.g., within a tolerance level). Since its first appearance in genetics (Tavaré et al., 1997; Pritchard et al., 1999), ABC has received a lot of attention in many applications such as ecology, epidemiology, and material science (François et al., 2008; Marin et al., 2012; Blum and Tran, 2010; Ravandi and Hajizadeh, 2022). Numerous extensions to the standard ABC methodology have been proposed in the literature. Some studies focus on discrepancy measurement between simulated and observed data without relying on summary statistics (Bernton et al., 2019; Zhu et al., 2023). Others explore the integration of MCMC techniques within the ABC framework to enhance parameter space exploration efficiency (Marjoram et al., 2003). Additionally, there are approaches employing regression adjustment methods to mitigate discrepancies arising from mismatches between observed and simulated summaries (Beaumont et al., 2002; Blum and François, 2010). Among these extensions, Sequential Monte Carlo ABC (SMC-ABC) (Toni et al., 2009) emerges as a useful method that improves the sampling efficiency and parameter estimation accuracy over the original ABC. This

approach draws inspiration from sequential Monte Carlo methods (Liu and Chen, 1995). The key idea is to draw samples with a sequence of decreasing tolerance levels and for each tolerance level, using importance sampling techniques to obtain samples that provide a better fit to the observed data as the tolerance level shrinks.

Despite their wide successes, several challenges still remain not fully addressed for standard ABC and SMC-ABC. One such challenge arises in high-dimensional parameter spaces, where designing an appropriate transition kernel for SMC-ABC while maintaining a high acceptance ratio for new proposals becomes difficult, e.g., a random-walk type of proposal will suffer from low sampling efficiency. Various strategies have been developed to broaden the scope of ABC methods for accommodating higher-dimensional models. These approaches include techniques such as regression adjustment (Beaumont et al., 2002; Blum and François, 2010), marginal adjustment (Nott et al., 2014), and the Gaussian copula approach (Li et al., 2017). Nevertheless, it is essential to note that these methods fall under the category of post-processing techniques, meaning they are primarily designed to enhance existing ABC methodologies rather than serving as principled approaches for inherently extending ABC methods to higher dimensions. Moreover, in practical scenarios, there is often an abundance of available summary statistics to summarize the raw data, e.g., in our data example presented in Section 4, we have a total of 145 summary statistics. The selection of informative summary statistics poses a challenge for most

ABC methods, including SMC-ABC. Kousathanas et al. (2016) proposed an ABC-MCMC algorithm which only updates one parameter per iteration. The new candidate is accepted or rejected based on a small subset of the summary statistics, which are informative for that particular parameter. Consequently, this method necessitates the identification of conditionally sufficient statistics for each parameter.

In this paper, we aim to tackle these challenges by introducing a Gibbs-SMC method within the ABC framework. The key idea is to utilize a Gibbs kernel to update each parameter at a time and approximate the conditional distribution of the parameters using regression adjustment methods. The resulting Gibbs kernel essentially corresponds to an invariant distribution and we demonstrate its optimality by showing that it minimizes the variance of the incremental weights. For the conditional distribution approximation, our method offers flexibility by allowing integration with various regression techniques such as GLM, Lasso, random forest, and neural network (Beaumont et al., 2002; Blum and François, 2010; Bi et al., 2022). This flexibility empowers us to capture potential nonlinear relationships between summary statistics and parameters while bypassing the need to pre-select from available summary statistics.

Our method is closely related to but different with the recent work on likelihood-free Gibbs-type MCMC sampler (Rodrigues et al., 2020) for two reasons. First, in their work, a separate regression model is needed for each parameter at each MCMC iteration, which results in a high computational

cost. Our method, in contrast, only requires fitting a regression every SMC iteration, hence is computationally more efficient. Secondly, our method utilizes an importance sampling mechanism which effectively helps avoiding regions of low posterior density (Sisson et al., 2007); and this is not considered in Rodrigues et al. (2020).

The rest of the paper is organized as follows. We discuss our proposed method, including its computational implementation and theoretical property in Section 2. In Section 3, we evaluate the sampling efficiency and parameter estimation accuracy of our method using simulation studies. We apply our method to a cell movement and proliferation example in Section 4 and discuss several future work directions in Section 5.

2. Likelihood-free Gibbs SMC

In this section, we first provide a summary of the two critical types of ABC algorithms, namely SMC-ABC (Toni et al., 2009) and Gibbs-ABC (Clarté et al., 2021). The former utilizes sequential sampling while the latter employs Gibbs sampling in the ABC settings. Then we discuss the computational and theoretical properties of our proposed method, ABC using the likelihood-free Gibbs SMC (ABC-GSMC),

2.1 SMC-ABC

The motivation behind the SMC-ABC algorithm is to approximate the posterior distribution of the model parameters by generating samples from a

sequence of intermediate distributions. This process is known as sequential Monte Carlo or particle filtering.

More specifically, the SMC-ABC algorithm samples from the sequence of target distributions $\{\pi_{\epsilon_t}(\theta|y)\}$, such that the thresholds $\epsilon_1 > \epsilon_2 > \cdots > \epsilon_T > 0$. In iteration t, the algorithm generates a new set of particles $\theta_t^{(1)}, ..., \theta_t^{(N)}$ (i.e., parameter values) by perturbing the previous set of particles $\theta_{t-1}^{(1)}, ..., \theta_{t-1}^{(N)}$ according to a transition kernel distribution K_t. These particles are accepted or rejected as regular ABC with threshold ϵ_t. The rejected ones will be re-perturbed until accepted. The accepted ones are then weighted with the importance weights $\tilde{w}_t^{(i)} = \pi\left(\theta_t^{(i)}\right) / \sum_{j=1}^{N} w_{t-1}^{(j)} K_t\left(\theta_t^{(i)} \mid \theta_{t-1}^{(j)}\right)$, and resampled (with replacement) with the normalised weights $w_t^{(i)}$. Obviously, at each iteration t, it is an importance sampling algorithm with the proposal distribution $q_t(\cdot) = \sum_{j=1}^{N} w_{t-1}^{(j)} K_t(\cdot|\theta_{t-1}^{(j)})$, which is the denominator in the weights $\tilde{w}_t^{(i)}$. The output of SMC-ABC are the weighted samples from the last iteration, i.e. $(\theta_T^{(1)}, w_T^{(1)}), ..., (\theta_T^{(N)}, w_T^{(N)})$. See Algorithm 1 for a detailed description.

SMC-ABC improves the efficiency of the usual ABC algorithms by taking advantage of the sequential importance sampling scheme. Only at the initial iteration, the proposed parameter values are obtained from the prior distribution. Subsequently, adaptive proposal distributions, $q_t(\cdot)$, which are progressively closer to the true posterior, are employed for drawing the proposed parameter values.

An interesting area of research involves developing the transition kernel

Algorithm 1 SMC-ABC

Initialize $\epsilon_1 > \epsilon_2 > ... > \epsilon_T$, set population indicator $t = 1$
for $i = 1 : N$ **do**
  repeat sample $\theta^* \sim \pi(\theta)$; generate a dataset $y^i \sim f(y|\theta^*)$
until $||(S(y^i), S_{obs}|| < \epsilon_t$.

  Set $\theta_t^{(i)} = \theta^*$ and $\tilde{w_t}^{(i)} = 1/N$.
end for
for $t = 2 : T$ **do**
repeat sample θ^{**} from the weighted set $\{\theta_{t-1}^j, w_{t-1}^j\}_{j=1}^N$; perturb $\theta^* \sim K_t(\cdot|\theta^{**})$ according to a transition kernel; generate a dataset $y^i \sim f(y|\theta^*)$
until $||(S(y^i), S_{obs}|| < \epsilon_t$.

  Set $\theta_t^{(i)} = \theta^*$ and $\tilde{w}_t^{(i)} = \pi\left(\theta_t^{(i)}\right) / \sum_{j=1}^N w_{t-1}^{(j)} K_t\left(\theta_t^{(i)} \mid \theta_{t-1}^{(j)}\right)$.

  Normalise the weights $w_t^{(i)} = \tilde{w}_t^{(i)} / \sum_{j=1}^N \tilde{w}_t^{(j)}$.
end for

K_t. Filippi et al. (2013) suggest a d-dimensional Gaussian $K_t = N(\theta^{**}, 2\Sigma)$, with Σ being the empirical weighted covariance matrix from the particles accepted at previous iteration. They also propose to use another covariance matrix that is specific to θ^{**}, which, however, increases non-negligible computational burden. Lee (2012) proposes an adaptive r-hit kernels that have more robust properties as the threshold $\epsilon \to 0$. However, these kernels are not specifically designed for high-dimensional parameter space. As the dimension of θ increases, the acceptance rate of proposed θ' by these kernels decreases rapidly, resulting in a significant drop in algorithmic efficiency. A very recent paper (Picchini and Tamborrino, 2022) proposes to construct K_t that are conditional on observed summaries, by using a conditional Gaussian distribution or a copula-based method. The so-called

"guided" sampler increases the efficiency by guiding the proposed parameters to rapidly reach regions of the posterior that are compatible with the observed data.

In the SMC literature that does not involve intractable likelihood, Del Moral et al. (2006) suggests to use an MCMC kernel K_t of invariant distribution π_{ϵ_t}. In particular, this approach is justified if K_t is fast mixing, so that we can expect the importance proposal distribution q_t to be reasonably close to the target intermediate distribution π_{ϵ_t}. When faced with the challenge of sampling from the complete conditional distributions necessary for a Gibbs kernel with an invariant distribution π_{ϵ_t}, such as in models with intractable likelihoods, one can turn to approximations of these distributions to construct K_t. Several approaches are proposed in Doucet et al. (2000) to approximate these conditional distributions. It is important to note that these approximations are primarily tailored for state-space models and may not be directly applicable in the context of ABC. In the following subsection, we introduce a SMC algorithm specifically designed to address high-dimensional parameter spaces within the ABC framework (ABC-GSMC).

2.2 Likelihood-free Gibbs-SMC (ABC-GSMC)

When dealing with high-dimensional parameters, traditional ABC and SMC-ABC methods encounter difficulties in generating simulated summary data that can closely resemble the observed summary statistics. This chal-

lenge is commonly referred to as the "curse of dimensionality", which affects most standard ABC techniques. In parameter spaces with a large dimension, finding an effective proposal distribution q_t, or a suitable kernel K_t for SMC-ABC becomes a non-trivial task.

To address this issue, we propose the utilization of a Gibbs kernel that updates one parameter at a time from its conditional distribution. Suppose that $\boldsymbol{\theta} = (\theta_1, ..., \theta_D)^\intercal$ is a D-dimensional parameter vector, and $\boldsymbol{\theta_t^{(i)}} = (\theta_{t,1}^{(i)}, ..., \theta_{t,D}^{(i)})^\intercal$ is the i-th particle at iteration t. We update the d-th parameter $\theta_{t,d}^{(i)}$ from $\pi_{\epsilon_t}(\theta_d|\boldsymbol{\theta_{-d}})$, where $\boldsymbol{\theta_{-d}}$ is the vector $\boldsymbol{\theta}$ excluding θ_d. By subsequently updating each parameter from its conditional distribution, we implicitly construct the Markov kernel with an invariant distribution π_{ϵ_t}. Unlike the "random-walk" type of moves in Filippi et al. (2013); Picchini and Tamborrino (2022), this approach uses the information of π_{ϵ_t} to guide the move of particles and allows us to select the optimal backward kernel which minimizes the variance of the incremental weights (more details are provided in section 2.3).

Note that the conditional distribution $\pi_{\epsilon_t}(\theta_d|\boldsymbol{\theta_{-d}})$ is intractable. Therefore, at iteration t, we approximate the conditional distribution of each parameter employing regression adjustment methods (Beaumont et al., 2002; Blum and François, 2010; Bi et al., 2022). Furthermore, this strategy offers additional advantages. In the context of ABC, the selection of informative summary statistics poses another significant challenge. By employing specific regression techniques, we can identify from a wide range of summary

statistics the most informative ones. Regression methods such as regularized regression or random forest regression (Bi et al., 2022) are particularly robust in handling such scenarios. Thus, our proposed approach not only addresses the challenges associated with high-dimensional parameter spaces but also provides a solution to the selection of informative summary statistics.

Regression adjustment post-processing methods assume that the relation between parameter θ_d and summary statistics $\boldsymbol{S}$ can be fitted with regression models of the form $\theta_d|\boldsymbol{S} \sim f_d(\theta_d|\boldsymbol{S})$ for $d = 1, ..., D$ using samples $(\boldsymbol{\theta^{(i)}}, \boldsymbol{s^{(i)}})_{i=1}^N$. To construct the Gibbs kernel in SMC-ABC, we similarly build regression models of the form $\theta_d|(\boldsymbol{S}, \boldsymbol{\theta_{-d}}) \sim f_d(\theta_d|\boldsymbol{S}, \boldsymbol{\theta_{-d}})$, where $\boldsymbol{\theta_{-d}}$ is the vector $\boldsymbol{\theta}$ excluding θ_d, such that $f_d(\theta_d|\boldsymbol{S}_{obs}, \boldsymbol{\theta_{-d}})$ locally approximate the true conditional distribution $\pi(\theta_d|\boldsymbol{S}_{obs}, \boldsymbol{\theta_{-d}})$. The choice of relevant dependent variables is evidently dependent on the dimension d. When the dependent variables (including summary statistics specific for θ_d) associated with θ_d are easy to identify, they tend to possess a lower dimensionality. However, in cases where the identification of these variables becomes more challenging, employing techniques such as regularized regression or random forest regression can alleviate the need for a careful selection process. The approximate Gibbs update will proceed by iterating through each of the conditional distributions in a cyclical manner. It will draw $\theta_d \sim f_d(\theta_d|\boldsymbol{S}_{obs}, \boldsymbol{\theta_{-d}})$ for $d = 1, ..., D$, while conditioning on the updated values of the other parameters. The resulting

Gibbs kernel K_t can then be approximated as a product of these conditional distributions, i.e., $K_t = \prod_{d=1}^{D} f_d^t(\theta_{t,d}|\boldsymbol{S}_{obs}, \boldsymbol{\theta_{t,-d}})$, where $\boldsymbol{\theta_{t,-d}} = (\theta_{t,1}, ..., \theta_{t,d-1}, ..., \theta_{t-1,d+1}, ..., \theta_{t-1,D})^\intercal$. Though the Gibbs kernel is constructed with invariant distribution π_{ϵ_t}, updating the particle $\boldsymbol{\theta}$ in one cycle does not guarantee its convergence. This can be addressed by a follow-up importance weighting process. Selecting K_t as an MCMC kernel leads to a weight update

$$w_t^{(i)} \propto \frac{\pi(\boldsymbol{\theta}_t^{(i)})}{\sum_{j=1}^{N} w_{t-1}^{(j)} \prod_{d=1}^{D} f_d^t(\theta_{t,d}^{(i)}|\boldsymbol{S}_{obs}, \boldsymbol{\theta}_{\boldsymbol{t,-d}}^{\boldsymbol{(j)}})}. \tag{2.1}$$

The resulting likelihood-free Gibbs-SMC algorithm is in Algorithm 2.

The algorithm initially follows a similar procedure to rejection ABC, where a first pool of samples $(\boldsymbol{\theta^{(i)}}, \boldsymbol{S^{(i)}})_{i=1}^{N}$ is obtained using a relatively large threshold ϵ_1. In t-iteration, D regression models $f_d^t(\theta_d|\boldsymbol{S}, \boldsymbol{\theta_{-d}})$ are fitted using the set of samples from the previous iteration. The particles are then resampled based on their importance weights, and each particle $\boldsymbol{\theta^{(i)}}$ is subsequently updated in a Gibbs manner from the distribution $f_d^t(\theta_d|\boldsymbol{S}_{obs}, \boldsymbol{\theta_{-d}})$. The weights are computed for each particle before transitioning to the next iteration.

It is worthy to note that a variety of suitable regression techniques can be employed to construct the models $f_d(\theta_d|\boldsymbol{S}, \boldsymbol{\theta_{-d}})$. These techniques include but are not limited to generalized linear models (GLMs), nonparametric models, Lasso, neural networks, random forest, and others. Following

Algorithm 2 Likelihood-free Gibbs-SMC

Initialize $\epsilon_1 > \epsilon_2 > ... > \epsilon_T$, set population indicator $t = 1$
for $i = 1 : N$ **do**
  repeat sample $\boldsymbol{\theta}^* \sim \pi(\boldsymbol{\theta})$;
  generate a dataset $y^i \sim f(y|\boldsymbol{\theta}^*)$ **until** $||(\boldsymbol{S}(y^i), \boldsymbol{S}_{obs}|| < \epsilon_t$.
  Set $\boldsymbol{\theta}_t^{(i)} = \boldsymbol{\theta}^*$ and $w_t^{(i)} = 1/N$.
end for
for $t = 2 : T$ **do**
  for $d = 1 : D$ **do**
    Fit a suitable regression model $\theta_d|\boldsymbol{S}, \boldsymbol{\theta_{-d}} \sim f_d^t(\theta_d|\boldsymbol{S}, \boldsymbol{\theta_{-d}})$ using the samples $(\boldsymbol{\theta}_{t-1}^{(i)}, \boldsymbol{S}(y_{t-1}^i))_{i=1}^N$ from the previous iteration $t-1$.
  end for
  for $i = 1 : N$ **do**
    Sample $\boldsymbol{\theta_t^{(i)**}}$ from the weighted set $\{\boldsymbol{\theta}_{t-1}^j, w_{t-1}^j\}_{j=1}^N$.
    for $d = 1 : D$ **do**
      repeat sample $\theta_{t,d}^{(i)*} \sim f_d^t(\theta_d|\boldsymbol{S}_{obs}, \boldsymbol{\theta_{-d}})$;
      generate a dataset y_t^i with updated $\theta_{t,d}^{(i)*}$ **until** $||(\boldsymbol{S}(y_t^i), \boldsymbol{S}_{obs}|| < \epsilon_t$.
      Set $\boldsymbol{\theta_t^{(i)*}} = (\theta_{t,1}^{(i)*}, ..., \theta_{t,d}^{(i)*}, \theta_{t-1,d+1}^{(i)}, ..., \theta_{t-1,D}^{(i)})$.
    end for
    Set $\boldsymbol{\theta_t^{(i)}} = \boldsymbol{\theta_t^{(i)*}}$ and $\tilde{w}_t^{(i)} = \frac{\pi(\boldsymbol{\theta}_t^{(i)})}{\sum_{j=1}^N w_{t-1}^{(j)} \prod_{d=1}^D f_d^t(\theta_{t,d}^{(i)}|\boldsymbol{S}_{obs}, \boldsymbol{\theta_{t,-d}^{(j)}})}$.
  end for
  Normalise the weights $w_t^{(i)} = \tilde{w}_t^{(i)} / \sum_{j=1}^N \tilde{w}_t^{(j)}$.
end for

Rodrigues et al. (2020), we propose two ways to draw samples from each conditional regression model. The first is to assume a parametric error distribution, so that a new sample can be drawn directly from the fitted distribution. For example, a normal error assumption leads to a normal distribution for θ_d. The second approach deals with the nonparametric regression case. Specifically, at iteration t of SMC, we obtain the empirical distribution $\hat{F}_{t,r}$ of the residuals $r_{t,d}^{(i)} = \theta_{t,d}^{(i)} - \hat{\theta}_{t,d}^{(i)}$, where $\hat{\theta}_{t,d}^{(i)}$ is the fitted value of $\theta_{t,d}^{(i)}$. A new value of $\theta_{t,d}^{(i)*}$ is then given by $\theta_{t,d}^{(i)*} = \hat{\mu} + r$, where $r \sim \hat{F}_{t,r}$ and $\hat{\mu} = E(\theta_{t,d}^{(i)} | \boldsymbol{S}_{obs}, \boldsymbol{\theta}_{\boldsymbol{t,-d}}^{\boldsymbol{(i)}})$.

The concept of utilizing regression adjustment models to approximate conditional distributions was also adopted in the work of Rodrigues et al. (2020), where they employed this approximation to construct a Gibbs-type Markov chain Monte Carlo (MCMC) sampler. However, in our SMC framework, the idea is more naturally suited for several reasons. Firstly, fitting a separate regression model for each parameter θ_d in every iteration of the Gibbs sampler incurs a significant computational burden, particularly when considering a typically long number of MCMC iterations. Although the authors suggest fitting the models only once prior to the Gibbs sampler to mitigate this issue, this would result in a substantial decrease in the algorithm's accuracy. While in the SMC framework, the regression models are fitted once per SMC iteration, resulting in reduced computational requirements compared to the MCMC approach. Additionally, it is common for the number of SMC iterations to be considerably smaller than the number

of MCMC iterations. Secondly, ABC-MCMC may encounter difficulties in navigating regions of low posterior density since it is unlikely to generate a dataset that closely resembles the observed dataset. In contrast, ABC-SMC mitigates this issue through the utilization of an importance sampling mechanism, which helps avoid getting stuck in regions of low posterior density (Sisson et al., 2007).

In the next section, we show that our Gibbs kernel minimizes the variance of the unnormalized importance weights in Eq (2.1).

2.3 Properties of the Gibbs kernel

The Gibbs kernel K_t proposed in the previous section serves as the forward kernel within the SMC framework. As a forward kernel, it generates new particles based on the previous set of particles. Conversely, the backward kernel $L_{t-1}(\theta_{t-1}|\theta_t)$ is utilized to assign importance weights to the particles, determining their relative significance during the subsequent resampling step. The weights in Eq (2.1) are obtained using the optimal choice for L_{t-1} to minimize the variance of the incremental weight $\Delta w_t^{(i)}$, defined as follows:

$$w_t^{(i)} \propto w_{t-1}^{(i)} \Delta w_t^{(i)}. \tag{2.2}$$

According to Del Moral et al. (2006), the optimal backward kernel takes the form:

$$L_{t-1}(\theta|\theta') = \frac{\pi_{t-1}(\theta')K_t(\theta'|\theta)}{\int \pi_{t-1}(u)K_t(u|\theta)du}, \tag{2.3}$$

which leads to the weights update as follows:

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{\pi_t(\theta_t^{(i)})}{\int \pi_{t-1}(\theta_{t-1}) K_t(\theta_{t-1}|\theta_t^{(i)}) d\theta_{t-1}}. \tag{2.4}$$

However, in the ABC context, computing the integration in the denominator is infeasible. As a result, Del Moral et al. (2006); Sisson et al. (2007) propose to approximate $\pi_{t-1}(d\theta)$ with an SMC point-wise approximation $\hat{\pi}_{t-1}(d\theta) = \sum_{i=1}^{N} w_{t-1}^{(i)} \delta_{\theta_{t-1}^{(i)}}(d\theta)$. This approximation leads to the following weights update:

$$\tilde{w}_t^{(i)} = \pi\left(\theta_t^{(i)}\right) / \sum_{j=1}^{N} w_{t-1}^{(j)} K_t\left(\theta_t^{(i)} \mid \theta_{t-1}^{(j)}\right).$$

Consider the case where $\boldsymbol{\theta_t} = (\theta_{t,1}, ..., \theta_{t,D})$ and we only want to update the d-th component $\theta_{t,d}$. It is established in Del Moral et al. (2006) that the proposal distribution minimizing the variance of the incremental weights is a Gibbs update, i.e., $K_t(\boldsymbol{\theta_{t-1}}|\boldsymbol{\theta_t}) = \delta_{\boldsymbol{\theta_{t-1,-d}}}(\boldsymbol{\theta_{t,-d}}) \pi_t(\theta_{t,d}|\boldsymbol{\theta_{t,-d}})$, where $\boldsymbol{\theta_{t,-d}} = (\theta_{t,1}, ..., \theta_{t,d-1}, \theta_{t,d+1}, ..., \theta_{t,D})$. Now in this paper we propose to update $\boldsymbol{\theta_t}$ by updating each d-th component $\theta_{t,d}$ in a Gibbs style. It is straightforward to establish that the proposal distribution is:

$$K_t(\boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}}) = \prod_{d=1}^{D} \pi_t(\theta_{t,d}|(\theta_{t,1}, ..., \theta_{t,d-1}, \theta_{t-1,d+1}, ..., \theta_{t-1,D})). \tag{2.5}$$

The following theorem establishes that equation (2.5) is the optimal Markov kernel.

Theorem 1. *The sequence of kernels* $\{L_k\}(k = t, ..., T)$ *minimizing the variance of the unnormalized importance weights* $w_t^{(i)}$ *is given by*

$$L_t(\boldsymbol{\theta_{t-1}}|\boldsymbol{\theta_t}) = \prod_{d=1}^{D} \pi_{t-1}(\theta_{t-1,d}|(\theta_{t-1,1}, ..., \theta_{t-1,d-1}, \theta_{t,d+1}, ..., \theta_{t,D})). \tag{2.6}$$

The proof of Theorem 1 is a straightforward adaptation of that in Del Moral et al. (2006).

Remark 1. In our approach, we utilize a Gibbs kernel with an invariant distribution π_t, enabling the use of an approximation for the backward Markov kernel. This approximation is based on the reversal kernel suggested in Del Moral et al. (2012) and takes the form:

$$L_{t-1}(\boldsymbol{\theta_{t-1}}|\boldsymbol{\theta_t}) = \frac{\pi_t(\boldsymbol{\theta_t})K_t(\boldsymbol{\theta_t}|\boldsymbol{\theta_{t-1}})}{\pi_t(\boldsymbol{\theta_{t-1}})}. \tag{2.7}$$

By adopting this backward kernel approximation, we obtain a weight update that is more computationally efficient, although it may come at the cost of increased variance in the weights.

In this work, we adopt the backward kernel in Equation (2.6) and the associated weights update as it provides the minimum variance of the incremental weights.

3. Simulations

3.1 General linear model

We consider a general linear model (GLM) that has been studied by Kousathanas et al. (2016). The model has m statistics $\boldsymbol{S}$ that is a linear function of parameters $\boldsymbol{\theta}$, $\boldsymbol{S} = \boldsymbol{C}\boldsymbol{\theta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \boldsymbol{I})$, where $\boldsymbol{C}$ is a square design matrix, and the error vector $\boldsymbol{\epsilon}$ follows a multivariate normal distribution. Assuming noninformative priors for the parameters $\boldsymbol{\theta}$, their posterior distribution is a multivariate normal, $\boldsymbol{\theta}|\boldsymbol{S} \sim N\left((\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}'\boldsymbol{s}, (\boldsymbol{C}'\boldsymbol{C})^{-1}\right)$. Following Kousathanas et al. (2016), we construct the design matrices $\boldsymbol{C}$ cyclically. This configuration ensures that all parameters receives input from every statistic, but with varying contributions from each statistic.

For the estimation, we set $\boldsymbol{\theta} = \mathbf{0}$ and use uniform priors $U(-10, 10)$ for all the parameters. We then evaluate the performance of ABC-GSMC and ABC-Gibbs across increasing parameter dimensions. This simulation setting inherently favors ABC-Gibbs, because the conditional distribution of θ_d given $\boldsymbol{S}$ and other parameters is a normal distribution with mean being a linear combination of $\boldsymbol{S}$ and other parameters, which can be easily learned with a linear regression model without including any interaction terms. Therefore, for both ABC-Gibbs and ABC-GSMC, we learn each of the conditional distributions of $\boldsymbol{\theta}$ by a single linear regression. For ABC-GSMC, as dimensionality increases, the tolerance is adjusted accordingly. Hence, we set the threshold ϵ as $(5, 4, 3, 2, 1) \times p/4$, where p represents the

dimension of $\boldsymbol{\theta}$. Ultimately, we obtain $N = 1000$ samples. For ABC-Gibbs, we conduct 20,000 iterations. We assess the performance of ABC-Gibbs and ABC-GSMC by calculating the total variation distance (L_1) between the inferred and the true posterior distribution, and then averaging over each dimension.

Figure 1 shows the mean L_1 distance between the true and estimated marginal posterior distributions obtained by both ABC-Gibbs and ABC-GSMC as the number of parameters increases from 3 to 10. Interestingly, we observe that both methods exhibit similar performance in terms of estimation accuracy. Neither method's performance is significantly affected by dimensionality. This finding suggests that ABC-GSMC adeptly manages varying-dimensional parameters, demonstrating its flexibility in such contexts.

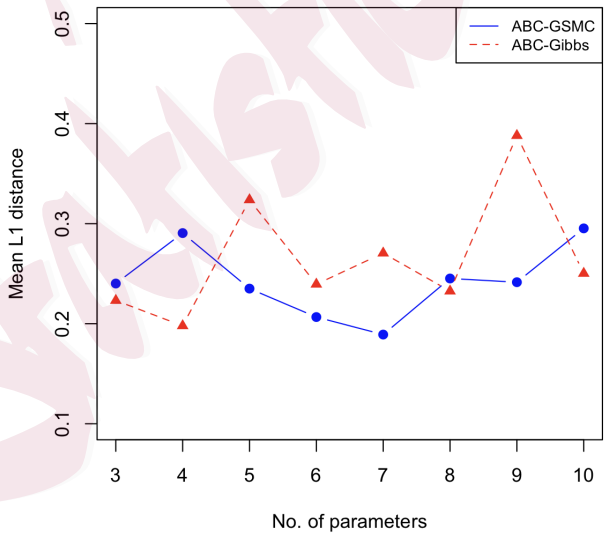

Figure 1: The mean L_1 distance between the true and estimated marginal posterior distributions for increasing numbers of parameters.

3.2 Gaussian mixture example

We conduct simulation studies to evaluate the sampling efficiency and accuracy of our method and compare to other popular ABC approaches. We first consider a Gaussian mixture model example following Nott et al. (2014). The likelihood function of this model has an explicit form, which provides an easy way to assess the performance of ABC methods. Consider $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p) \in \mathrm{R}^p$, $\boldsymbol{b} = (b_1, \ldots, b_p)$, where each b_i follows a Bernoulli distribution with $P(b_i = 0) = w \in (0, 1)$, and Σ is a p by p covariance matrix with $\Sigma_{ii} = 1$ and $\Sigma_{ij} = \rho$ for $1 \leq i \neq j \leq p$. Let $\boldsymbol{\mu(b, \theta)} = ((1 - 2b_1)\theta_1, \ldots, (1 - 2b_p)\theta_p)$, i.e., the i-th component of $\boldsymbol{\mu(b, \theta)}$ is θ_i with probability of w, and $-\theta_i$ with probability of $1 - w$. Given $\boldsymbol{\theta}$ and $\boldsymbol{b}$, $\boldsymbol{S}$ is generated from a p-dimensional normal distribution, i.e., $\boldsymbol{S} \mid \boldsymbol{\theta, b} \sim N(\boldsymbol{\mu(b, \theta)}, \Sigma)$. In other words, each component of $\boldsymbol{S}$ follows a two-component Gaussian mixture distribution, and in total $\boldsymbol{S}$ has 2^p mixture components.

We first consider $p = 2$, and set $w = 0.3$, $\rho = 0.7$. In our implementation, we choose observed summary statistic $\boldsymbol{S_{obs}} = (5/2, 5/2)$, tolerance $\epsilon \in \{5, 3, 1\}$, and $N = 1000$. We use the Euclidean distance $\|\boldsymbol{S} - \boldsymbol{S_{obs}}\|$ to measure the discrepancy between the simulated summary statistic $\boldsymbol{S}$ and the observed one $\boldsymbol{S_{obs}}$. We use independent uniform$(-80, 80)$ as the prior for each component of $\boldsymbol{\theta}$. For our proposed method (ABC-GSMC), we first use linear regression adjustment to approximate the conditional distribution of each element of $\boldsymbol{\theta}$. To facilitate a fair comparison, we use the same

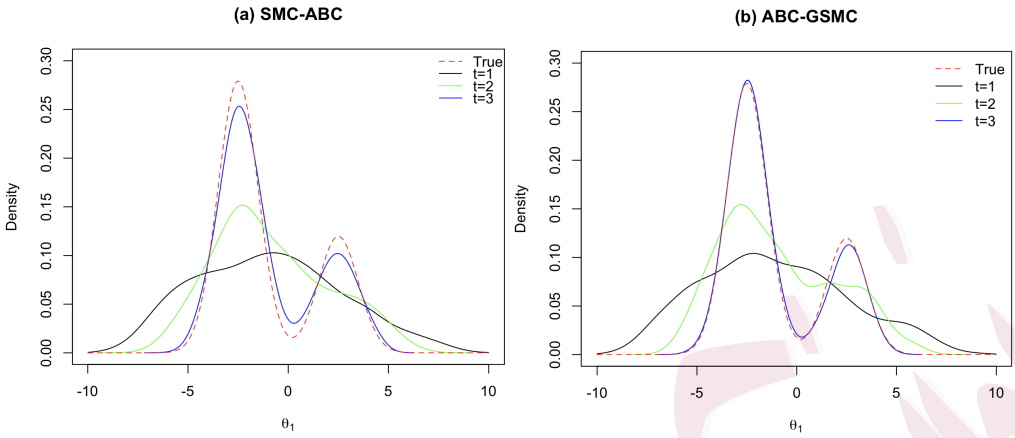

Figure 2: Posterior density of θ_1 obtained by SMC-ABC and ABC-GSMC.

setting when implementing SMC-ABC. We choose the transition kernel as $N(\boldsymbol{\theta}, \Sigma_1)$ where Σ_1 is a two-dimensional identity matrix.

Figure 2 shows the posterior density of θ_1 obtained by SMC-ABC and our proposed ABC-GSMC methods. By comparing the posterior from three updates ($t = 1, 2, 3$) with the truth (dashed line), we find a significant improvement. In the first update, the fit is not satisfactory as expected, since the parameters are sampled from the prior and the tolerance value is large (e.g., 5). For the third update, both methods manage to approximate the true posterior very well.

Next we let the parameter dimension p take values in $\{2, \ldots, 5\}$, and evaluate the sampling efficiency and posterior approximation accuracy for both methods. We consider four iterations with the threshold $\epsilon \in \{20, 5, 3, 1\}$. For ABC-GSMC, we still use the linear regression adjustment to approximate the conditional distribution of each element of $\boldsymbol{\theta}$. For SMC-ABC,

we use the a Gaussian transition kernel $N(\boldsymbol{\theta}, \Sigma_1)$ with $\Sigma_1 = I_2$. For the first iteration, note that both SMC-ABC and ABC-GSMC are essentially drawing samples from the prior of $\boldsymbol{\theta}$. Therefore we only compare their performance starting from the second iteration. For each method, we calculate the acceptance rate, defined as $R = N_{\text{accept}}/N_{\text{sample}}$, where N_{accept} is the number of accepted particles and N_{sample} is the number of total particles being sampled. We repeat this process for ten times and report the average acceptance rate for both methods in Figure 3 panel (a) while varying the parameter dimension p from 2 to 5. We find that both methods have an acceptance rate that decreases as p increases, which is as expected because it is harder to satisfy $\|\boldsymbol{S} - \boldsymbol{S}_{\text{obs}}\| < \epsilon$ as p increases. Meanwhile our proposed ABC-GSMC method always has a higher acceptance rate than SMC-ABC because SMC-ABC only uses a random perturbation for the particle while ABC-GSMC is based on a more accurate regression-based conditional distribution approximation.

We further compare the acceptance rates for both SMC-ABC and ABC-GSMC over first several iterations (t) and parameter dimension p in Table 1. The first iteration is the same for both methods because the particles are sampled from the prior. For the next two iterations ($t = 2, 3$), the acceptance rate increases significantly because a better proposal distribution is used. As the acceptance threshold continues to decrease, the acceptance rate starts to decrease for $t = 4$. Under all scenarios, ABC-GSMC has a higher acceptance rate, and the gain is substantial when p is large. Figure 3

and Table 1 confirm the excellent performance of our proposed ABC-GSMC in terms of estimation accuracy and sampling efficiency.

Table 1: Acceptance rate during each iteration t.

Dimension p	2	3	4	5
SMC-ABC($t=1$)	0.05063	0.00829	0.00123	0.00016
ABC-GSMC($t=1$)	0.05063	0.00829	0.00123	0.00016
SMC-ABC($t=2$)	0.05666	0.01585	0.00432	0.00113
ABC-GSMC($t=2$)	0.10440	0.03532	0.01334	0.00533
SMC-ABC($t=3$)	0.15765	0.04278	0.01287	0.00346
ABC-GSMC($t=3$)	0.22381	0.09552	0.04029	0.12526
SMC-ABC($t=4$)	0.01958	0.00222	0.00022	0.00002
ABC-GSMC($t=4$)	0.03931	0.00703	0.00114	0.00009

We further examine the posterior density of θ_1 obtained by ABC rejection sampling (ABC-REJ), SMC-ABC, and our proposed ABC-GSMC when $p = 5$. In order to obtain 1,000 posterior samples for ABC-REJ, we generate a total of 60,000,000 samples and then accept the best 1,000 particles. The posterior densities of θ_1 are given in panel (b) of Figure 3. Both SMC-ABC and ABC-GSMC have a significantly better performance than ABC-REJ, and ABC-GSMC is doing slightly better than SMC-ABC, especially when θ_1 is around 0.

Sensitivity to regression models: We also evaluate the sensitivity of our method to difference choices of the regression model. We set $p = 5$ and consider a Gaussian mixture model, with threshold $\epsilon \in \{30, 20, 10, 5, 3, 2, 1\}$ for iteration $t = 1$ to 7. We firstly consider three regression models (linear, quadratic, and cubic) to approximate the conditional of $\boldsymbol{\theta}$. The posterior distribution of θ_1 (at the $7th$ iteration) is given in Figure 4 and the num-

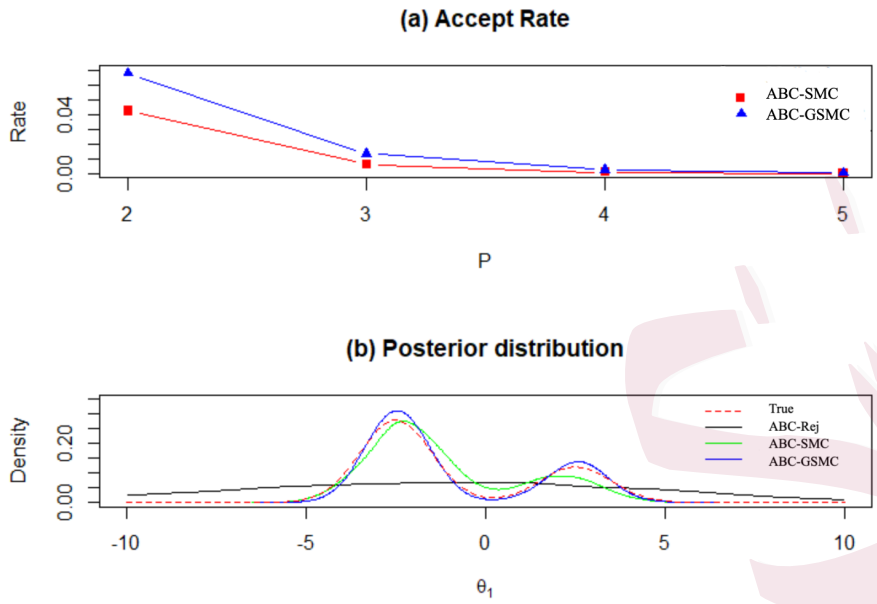

Figure 3: (a) Acceptance rate and parameter dimension p; and (b) posterior distribution of θ_1 obtained by several methods.

ber of particles required to produce 1000 samples is given in Table 2. By using the same threshold, all three regression models yield a similar level of estimation accuracy. Interestingly, the numbers of required particles are about the same when the threshold is large (e.g., $t \leq 5$). As the threshold decreases down to 1, both quadratic and cubic regressions need a larger number of sampled particles (hence more computational time) because the model does not provide a good fit to approximate the conditional distribution.

An alternative approach is to select a flexible nonparametric regression model, particularly when the true relationship between the parameter θ_d and the summary statistics is non-linear. However, in this specific exam-

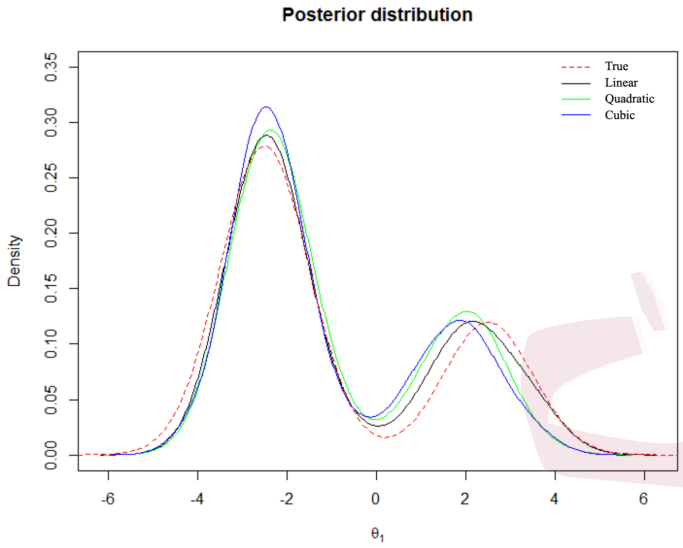

Figure 4: Posterior distribution of θ_1 obtained by different regression models (linear, quadratic, and cubic)

Table 2: Number of particles needed for each iteration t

Regression model	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$
Linear	5557	9240	15266	19142	24327	122527
Quadratic	6898	10314	16282	22056	30817	311736
Cubic	6077	9606	15111	19613	31753	512213

ple, we did not observe significant improvements when using random forest or neural network regressions. Our conjecture is that while nonparametric regression models may globally fit the relationship between θ_d and the summary statistics more accurately, linear regression is sufficiently effective in approximating $\pi(\theta_d|\boldsymbol{S_{obs}}, \boldsymbol{\theta_{-d}})$ when $\boldsymbol{\theta_{-d}}$ is fixed at a specific value. This aspect is crucial for our algorithm, as it uses $\pi(\theta_d|\boldsymbol{S_{obs}}, \boldsymbol{\theta_{-d}})$ to guide the movement of the particles, which is then corrected by the re-weighting step in the subsequent stage. Consequently, even a coarse approximation

performs well, making our method robust to the choice of regression models.

Performance under high dimensionality: To investigate the performance of our method under an increasing dimensionality of parameter space, we compare our method with ABC-Gibbs proposed by Rodrigues et al. (2020). ABC-Gibbs approximates full conditional distributions of posterior with regression-based models using synthetic (simulated) parameter value and summary statistic pairs. Notably, in this scenario, the full conditional distributions of θ and b are analytically tractable, encompassing numerous interaction terms within θ and b. To adequately approximate the conditional distributions, one needs to include all interaction terms in the regression model. However, this endeavor quickly becomes impractical as the dimensionality p expands. For instance, with $p = 5$, each regression model entails 9 regressors, yielding over 6,000 interaction terms of various orders. So we only incorporate first-order interaction terms in ABC-Gibbs for our comparison. We run ABC-Gibbs for 20,000 iterations. For ABC-GSMC, we use the same settings in the previous comparisons. Note that we only use a simple linear regression model without any interaction. We assess the performance of ABC-Gibbs and ABC-GSMC by calculating the total variation distance (L_1) between the inferred and the true posterior distribution across various values of p. The L_1 distance is calculated for each dimension, and the average value is utilized for comparison. Figure 5 illustrates the mean L_1 distance for both methods across varying numbers of parameters. Remarkably, we observe that the performance of both meth-

ods remains consistent regardless of dimensionality, as they are specifically engineered to manage high-dimensional parameters. Notably, ABC-GSMC outperforms ABC-Gibbs consistently by having a lower L_1 estimation error. This advantage primarily arises from the less accurate approximation of conditional distributions by ABC-Gibbs. In contrast, ABC-GSMC, being inherently a sequential Monte Carlo method, relies less on conditional approximation for its efficacy.

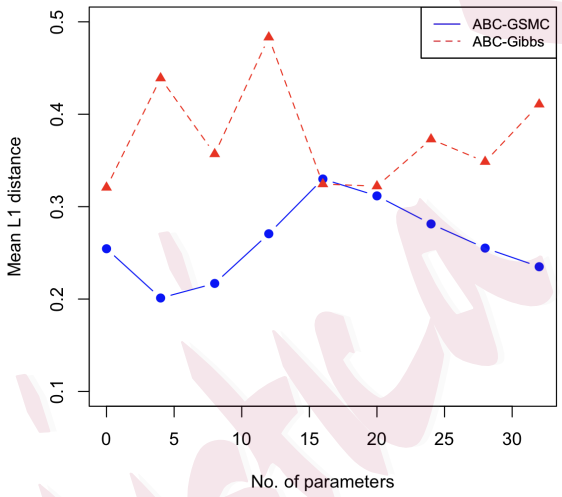

Figure 5: The mean L_1 distance between the true and estimated marginal posterior distributions for increasing numbers of parameters.

3.3 Banana prior model

Next we consider the simulation example used in Section 3.1 of Li et al. (2017), where the name of 'banana prior' comes from the shape of the prior distribution. Let $\boldsymbol{y} = (y_1, \ldots, y_p)$ follow a p-dimensional Gaussian distribution $N(\boldsymbol{\theta}, \Sigma)$ with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$ and $\Sigma = \text{diag}(\sigma_0, \ldots, \sigma_0)$. We

then consider the 'twisted prior' on θ following Haario et al. (1999),

$$\pi(\boldsymbol{\theta}) \propto \exp\left\{-\frac{\theta_1^2}{200} - \frac{(\theta_2 - b\theta_1^2 + 100b)^2}{2} - \sum_{j=3}^{p}\theta_j^2\right\}.$$

This prior can be viewed as a modification of the usual independent Gaussian prior over components of $\boldsymbol{\theta}$ except now θ_1 and θ_2 are dependent where the strength of dependence is decided by the parameter b. We follow the same setting with Li et al. (2017) by choosing $\sigma_0 = 1$ and $b = 0.1$ such that θ_1 and θ_2 are strongly dependent. We set $p = 5$, the observed $y_{\text{obs}} = (10, 0, 0, 0, 0)$, and the summary statistic $\boldsymbol{S} = \boldsymbol{y}$, i.e., no information loss.

To estimate the posterior of $\boldsymbol{\theta}$, we apply both SMC-ABC and ABC-GSMC methods and choose the threshold $\epsilon \in \{10, 8, 5, 3, 1\}$. For each iteration, the number of accepted particles is set as 1,000. We use the Euclidean distance to measure the discrepancy between the simulated and the observed summary statistics. The linear regression is used to approximate the conditional distributions of parameters. For SMC-ABC, we use a Gaussian kernel $N(\boldsymbol{\theta}, \Sigma_1)$ for parameter perturbation with Σ_1 is a covariance matrix with diagonal elements of 1 and off-diagonals of 0.3. For implementation, we draw a total of 275,726 and 68,686 samples for SMC-ABC and ABC-GSMC, respectively. For ABC-REJ, we generate 280,000 and 70,000 parameter samples from the prior to generate simulated data and calculate the summary statistics, and then obtain 1,000 samples for parameter esti-

mation. Contour plots for the estimated posterior $\pi(\theta_1, \theta_2 \mid \boldsymbol{S_{obs}})$ obtained by three methods are shown in Figure 6. The true density is marked as blue dashed line for comparison. For ABC-REJ, there is a considerable improvement as we increase the number of samples from 70,000 to 280,000. Still it misses the dependence structure between θ_1 and θ_2. Both SMC-ABC and ABC-GSMC manage to capture the dependence structure better than ABC-REJ, while ABC-GSMC clearly has the best performance among all three methods in comparison. We further evaluate the Kullback–Leibler (KL) divergence between the estimated and true posterior of (θ_1, θ_2) in Table 3. Our proposed ABC-GSMC performs uniformly better than SMC-ABC at each iteration, and has a clear advantage over ABC-REJ. All these findings confirm the strong performance of our method when there exists a strong dependence between parameters of interest.

Table 3: KL divergence between the estimated posterior and the truth.

Method	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$
SMC-ABC	2.3896	2.0128	1.3911	0.6072	0.2860
ABC-GSMC	2.3991	1.2633	0.7481	0.3316	0.1859
ABC-REJ (70000)	-	-	-	-	0.9503
ABC-REJ (280000)	-	-	-	-	0.5795

4. A cell movement and proliferation example

4.1 Background

Cell motility and proliferation are important components of many biological processes. Cell motility can cause random movement of cells, and

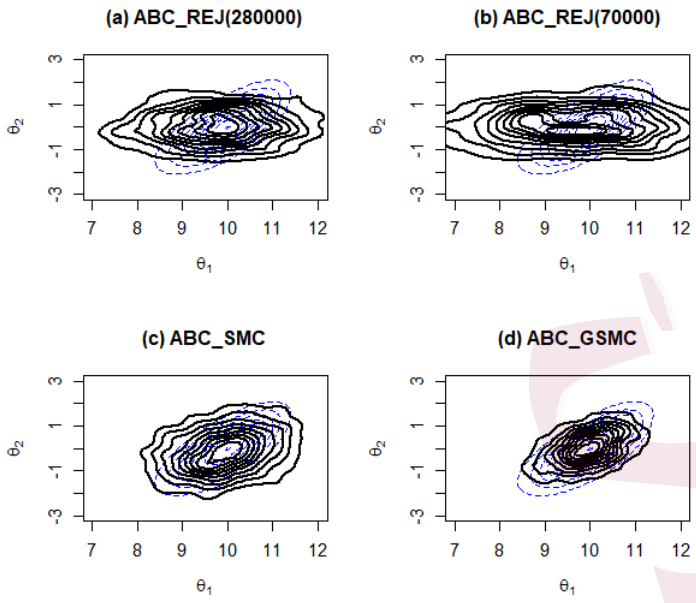

Figure 6: Posterior distribution of (θ_1, θ_2) obtained by several methods.

together with cell proliferation, can lead to tumor metastasis (Swanson et al., 2003) or wound healing (Zahm et al., 1997). Many medical treatments are based on affecting the rate of cell movement or proliferation. In order to assess the effectiveness of medical treatments, it is often necessary to measure the rates of cell movement and proliferation. However, the likelihood function of the stochastic model of cell diffusion is often intractable. A common approach is to use the ABC method to estimate parameters (Johnston et al., 2014; Vo et al., 2015). A main challenge is that the observed data is often in the form of images, and it is difficult to reduce the dimensionality of summary statistics to a level suitable for the use of ABC methods while retaining the information contained in the images.

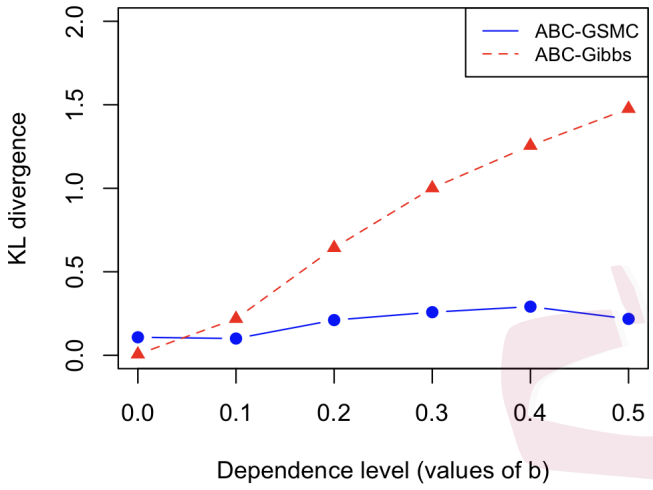

Figure 7: The KL divergence between the true and estimated posterior of (θ_1, θ_2) under different dependence levels.

The scratch assay (Fronza et al., 2009) is a commonly used method for collecting information on cell diffusion and proliferation, and can be used to measure cell migration in vitro. It is easy-to-implement and low-cost. Once the cells have formed a single layer completely covering the test substance (i.e., a fused monolayer), a "scratch" is produced by separating the cells. The cells are then imaged at regular time intervals, and these images are converted into summary statistics for further analysis. Due to the curse of dimensionality associated with the high-dimensional summary statistics, previous studies on cell motility and proliferation analyzed the images based on intervals of one hour or more, even when the imaging time interval was less than one hour. In Johnston et al. (2014), cells were imaged every 5 minutes for 12 hours, but in their actual analysis, the authors only considered the cell imaging pictures from the 4th, 8th, and 12th hours,

rather than selecting all the cell imaging pictures for analysis. Our aim is to make use of all available cell imaging pictures, which contain more information on cell motility and proliferation. Here, we use the proposed likelihood-free Gibbs sequential Monte Carlo algorithm to analyze the cell movement and proliferation model and examine how this method performs under high-dimensional summary statistics.

We follow Price et al. (2018) and use a random walk model to simulate the processes of cell movement and proliferation. When setting up the model, we assume that cell imaging is taken every 5 minutes and lasts for 12 hours, so the time label is $t = \{1, 2, ..., 144\}$. Consider a two-dimensional lattice with size of R times C. Let $X^t_{x,y} \in \{0, 1\}$ represent whether a cell is observed at position (x, y) at time t, where $x \in \{1, 2, ..., R\}$ and $y \in \{1, 2, ..., C\}$. Then we use X^t to represent the position information matrix at time t. Here, we follow Price et al. (2018) and calculate the Hamming distance between X^t and X^{t-1} as the summary statistic s_t:

$$s_t = \sum_{x=1}^{R} \sum_{y=1}^{C} |X^t_{x,y} - X^{t-1}_{x,y}|.$$

We also use the total number of cells K observed on the lattice at the end of the experiment as another summary statistic reflecting cell proliferation information. Therefore, we have a 145-dimensional summary statistic $s = (s_1, s_2, \ldots, s_{144}, K)$.

We use a random walk model to simulate the movement and prolifera-

tion process of cells. Suppose that there are $N(t)$ cells on the observation lattice at time t. Then within any given time interval, we can randomly choose $N(t)$ cells repeatedly, with a possibility of choosing the same cells multiple times. The chosen cells are then assigned with a probability of P_m to move. For example, as shown in Figure 8, cells can move in any direction (east, west, south, north) to their adjacent areas, and the probability of moving in each direction is 1/4. If the new position after moving is empty, i.e., there is no cell present at that location, then the movement is successful. If there is a cell on the target position after moving, that move fails. Similarly, when dealing with cell proliferation, we repetitively select $N(t)$ cells and assign them a probability P_p to proliferate. The cells can proliferate in any direction (east, west, south, and north) with an equal probability of 1/4. If the adjacent area selected for proliferation is empty, proliferation will succeed; otherwise, proliferation will fail (Simpson et al., 2013).

This biological process is entirely determined by the movement and proliferation of cells. In practice, scientists are more concerned with the diffusion rate and the growth rate of cells. In the random walk model, these are reflected by the cell movement probability P_m and proliferation probability P_p. For data generation, we set $P_m = 0.3$, $P_p = 0.001$, the observation lattice dimension $R = 27$ and $C = 36$. At $t = 0$, we generate $N(0) = 110$ cells with their locations randomly distributed at the lattice locations $x \in \{1, \ldots, 13\}, y \in \{1, \ldots, 36\}$. The observed data $M^{\text{obs}} =$

$\{M_1^{\text{obs}}, \ldots, M_{144}^{\text{obs}}$ and summary statistics $s^{\text{obs}} = (s_1^{\text{obs}}, \ldots, s_{144}^{\text{obs}}, K^{\text{obs}})$ can then be generated based on the random walk model.

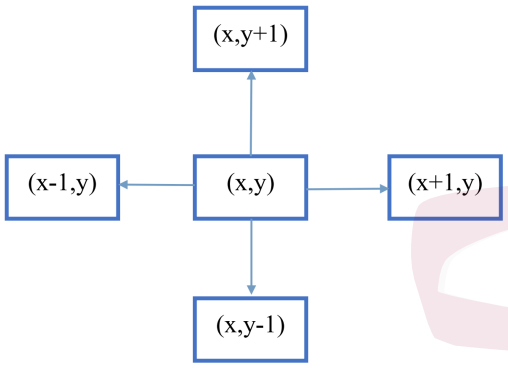

Figure 8: Cell movement and proliferation model demonstration

4.2 Results

We use a uniform distribution on $(0, 1)$ as the prior for cell movement probability P_m, and another uniform distribution on $(0, 0.01)$ for cell proliferation probability P_p. The latter is to reflect the *a priori* knowledge about small values of P_p. When implementing SMC-ABC and ABC-GSMC, we consider four iterations with a threshold of $\epsilon = \{2000, 1500, 1000, 800\}$ and set the number of accepted particles to be $N = 1000$. We use linear regression to approximate the conditional distribution. For SMC-ABC, we use $N(x, \Sigma_2)$ with $\Sigma_2 = \text{diag}(0.1, 0.1)$ as the perturbation kernel. If the parameter falls into $[0, 1]$ after perturbation, we consider it successful (since we are estimating a probability) and use it to generate simulated data and summary statistics. To alleviate a fair comparison, we use the same setting for SMC-ABC and ABC-GSMC.

We summarize the estimated posterior distribution of P_m and P_p as boxplots in Figure 9. Compared to the true parameter value (horizontal dashed line), SMC-ABC tends to overestimate both parameters, while ABC-GSMC manages to rapidly capture the true value after the first two iterations. The gain in efficiency is also significant by our method.

When implementing SMC-ABC, we obtain a total of 13,947 successful samples (parameter values between 0 and 1 after perturbation). For ABC-GSMC, that number is 5,699. We also implement ABC-REJ by sampling 15,000 and 6,000 times from the prior, generating the simulated data and calculating the summary statistics. We then proceed with a threshold of 1/15 and 1/6 to obtain 1,000 parameter values and name these methods as ABC-REJ(6000) and ABC-REJ(15000). We report their posterior summaries together with those obtained by SMC-ABC and ABC-GSMC in Table 4. Despite some minor underestimation issue, our method still enjoys the smallest estimation bias and its estimation efficiency is much better than SMC-ABC and ABC-REJ. When estimating a small probability such as $P_p = .001$, our method achieves a desirable level of accuracy at the second iteration, only requiring a much smaller number of sampled particles. All these findings highlight the excellent utility of our method when dealing with high-dimensional summary statistics.

In the study of cell proliferation and diffusion, most existing research analyzes imaging data from a few time points. Price et al. (2018) suggest that using more imaging data of cells can provide more accurate parameter

estimation. In our study, as we assume the use of all time points of cell imaging, and use the random walk model to approximate the cell diffusion and proliferation process. Therefore, after obtaining parameters, we generate 144 observation matrices accordingly and calculate the corresponding summary statistics. This is undoubtedly a time-consuming process, and each simulation data requires huge computing power. When the quality of the sampled parameters is poor, it will cause a serious computational burden. Compared to the existing methods such as SMC-ABC and ABC rejection sampling, our method provides a valuable alternative.

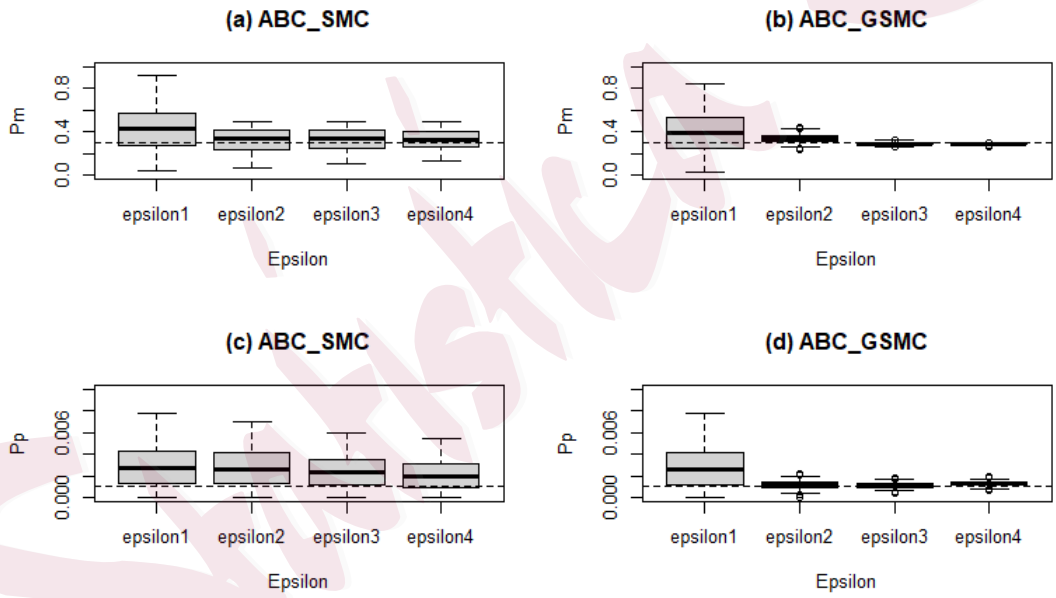

Figure 9: Boxplots for estimated parameters (true value marked as dashed line).

Table 4: Posterior summaries (95% credible interval, posterior mean, and standard deviation) for P_m and P_p estimation.

	P_m (true = .3)			P_p (true=.001)		
Method	95% CI	mean	sd	95% CI	mean	sd
SMC-ABC	(0.3209,0.3323)	0.3266	0.0915	(0.0020,0.0021)	0.0021	0.0013
ABC-GSMC	(0.2991,0.2999)	0.2995	0.0066	(0.0006,0.0007)	0.0006	0.0001
ABC-REJ(6000)	(0.2682 ,0.2769)	0.2726	0.0708	(0.0038,0.0042)	0.0040	0.0028
ABC-REJ(15000)	(0.2833,0.2895)	0.2864	0.0500	(0.0025,0.0028)	0.0027	0.0021

5. Discussion

In this paper, we propose a likelihood-free Gibbs sequential Monte Carlo (ABC-GSMC) method, which uses the accepted parameter values and summary statistics from each iteration to fit a regression model to approximate the conditional distribution. Compared to the original SMC-ABC method that perturbs parameters randomly, our method updates parameters by sampling from approximation of the conditional distribution, guiding the perturbation of parameters more specific. As a result, the number of particles required to achieve the same estimation accuracy as the SMC-ABC method is much smaller, leading to a higher computational efficiency.

Several directions remain open in future research. Firstly, in the implementation of the ABC-GSMC method, we manually set the sequence of threshold values. When the threshold sequence is not well designed, the algorithm may get stuck in a certain iteration, and it may take a long time to proceed to the next iteration. Therefore, we can consider using effective sample size to adaptively determine the threshold required for each iter-

ation. Secondly, we could incorporate wasserstein distance to our model, to avoid the selection of summary statistics. Lastly, other flexible regression models can also be used in the ABC-GSMC algorithm for conditional distribution approximation, such as deep learning.

References

Beaumont, M. A., W. Zhang, and D. J. Balding (2002). Approximate bayesian computation in population genetics. *Genetics 162*(4), 2025–2035.

Bernton, E., P. E. Jacob, M. Gerber, and C. P. Robert (2019). Approximate bayesian computation with the wasserstein distance. *Journal of the Royal Statistical Society Series B: Statistical Methodology 81*(2), 235–269.

Bi, J., W. Shen, and W. Zhu (2022). Random forest adjustment for approximate bayesian computation. *Journal of Computational and Graphical Statistics 31*(1), 64–73.

Blum, M. G. and O. François (2010). Non-linear regression models for approximate bayesian computation. *Statistics and computing 20*, 63–73.

Blum, M. G. and V. C. Tran (2010). Hiv with contact tracing: a case study in approximate bayesian computation. *Biostatistics 11*(4), 644–660.

Clarté, G., C. P. Robert, R. J. Ryder, and J. Stoehr (2021). Componentwise approximate bayesian computation via gibbs-like steps. *Biometrika 108*(3), 591–607.

Del Moral, P., A. Doucet, and A. Jasra (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68*(3), 411–436.

Del Moral, P., A. Doucet, and A. Jasra (2012). An adaptive sequential monte carlo method for approximate bayesian computation. *Statistics and computing 22*, 1009–1020.

REFERENCES

Doucet, A., S. Godsill, and C. Andrieu (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing 10*, 197–208.

Filippi, S., C. P. Barnes, J. Cornebise, and M. P. Stumpf (2013). On optimality of kernels for approximate bayesian computation using sequential monte carlo. *Statistical applications in genetics and molecular biology 12*(1), 87–107.

François, O., M. G. Blum, M. Jakobsson, and N. A. Rosenberg (2008). Demographic history of european populations of arabidopsis thaliana. *PLoS genetics 4*(5), e1000075.

Fronza, M., B. Heinzmann, M. Hamburger, S. Laufer, and I. Merfort (2009). Determination of the wound healing effect of calendula extracts using the scratch assay with 3t3 fibroblasts. *Journal of ethnopharmacology 126*(3), 463–467.

Haario, H., E. Saksman, and J. Tamminen (1999). Adaptive proposal distribution for random walk metropolis algorithm. *Computational statistics 14*, 375–395.

Johnston, S. T., M. J. Simpson, D. S. McElwain, B. J. Binder, and J. V. Ross (2014). Interpreting scratch assays using pair density dynamics and approximate bayesian computation. *Open biology 4*(9), 140097.

Kousathanas, A., C. Leuenberger, J. Helfer, M. Quinodoz, M. Foll, and D. Wegmann (2016). Likelihood-free inference in high-dimensional models. *Genetics 203*(2), 893–904.

Lee, A. (2012). On the choice of mcmc kernels for approximate bayesian computation with smc samplers. In *Proceedings of the 2012 Winter simulation conference (WSC)*, pp. 1–12. IEEE.

Li, J., D. J. Nott, Y. Fan, and S. A. Sisson (2017). Extending approximate bayesian computation methods to high dimensions via a gaussian copula model. *Computational Statistics & Data Analysis 106*, 77–89.

REFERENCES

Liu, J. S. and R. Chen (1995). Blind deconvolution via sequential imputations. *Journal of the american statistical association 90*(430), 567–576.

Marin, J.-M., P. Pudlo, C. P. Robert, and R. J. Ryder (2012). Approximate bayesian computational methods. *Statistics and computing 22*(6), 1167–1180.

Marjoram, P., J. Molitor, V. Plagnol, and S. Tavaré (2003). Markov chain monte carlo without likelihoods. *Proceedings of the National Academy of Sciences 100*(26), 15324–15328.

Nott, D. J., Y. Fan, L. Marshall, and S. Sisson (2014). Approximate bayesian computation and bayes' linear analysis: toward high-dimensional abc. *Journal of Computational and Graphical Statistics 23*(1), 65–86.

Picchini, U. and M. Tamborrino (2022). Guided sequential schemes for intractable Bayesian models. *arXiv preprint arXiv:2206.12235*.

Price, L. F., C. C. Drovandi, A. Lee, and D. J. Nott (2018). Bayesian synthetic likelihood. *Journal of Computational and Graphical Statistics 27*(1), 1–11.

Pritchard, J. K., M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman (1999). Population growth of human y chromosomes: a study of y chromosome microsatellites. *Molecular biology and evolution 16*(12), 1791–1798.

Ravandi, M. and P. Hajizadeh (2022). Application of approximate bayesian computation for estimation of modified weibull distribution parameters for natural fiber strength with high uncertainty. *Journal of Materials Science 57*(4), 2731–2743.

Rodrigues, G., D. J. Nott, and S. A. Sisson (2020). Likelihood-free approximate gibbs sampling. *Statistics and computing 30*, 1057–1073.

Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *The Annals of Statistics*, 1151–1172.

Simpson, M. J., K. K. Treloar, B. J. Binder, P. Haridas, K. J. Manton, D. I. Leavesley, D. S. McElwain, and R. E. Baker (2013). Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. *Journal of the Royal Society Interface 10*(82), 20130007.

Sisson, S. A., Y. Fan, and M. M. Tanaka (2007). Sequential monte carlo without likelihoods. *Proceedings of the National Academy of Sciences 104*(6), 1760–1765.

Swanson, K. R., C. Bridge, J. Murray, and E. C. Alvord Jr (2003). Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion. *Journal of the neurological sciences 216*(1), 1–10.

Tavaré, S., D. J. Balding, R. C. Griffiths, and P. Donnelly (1997). Inferring coalescence times from dna sequence data. *Genetics 145*(2), 505–518.

Toni, T., D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf (2009). Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface 6*(31), 187–202.

Vo, B. N., C. C. Drovandi, A. N. Pettitt, and G. J. Pettet (2015). Melanoma cell colony expansion parameters revealed by approximate bayesian computation. *PLOS computational biology 11*(12), e1004635.

Zahm, J.-M., H. Kaplan, A.-L. Hérard, F. Doriot, D. Pierrot, P. Somelette, and E. Puchelle (1997). Cell migration and proliferation during the in vitro wound repair of the respiratory epithelium. *Cell motility and the cytoskeleton 37*(1), 33–43.

Zhu, W., T. Zuo, and C. Wang (2023). Approximate bayesian computation with semiparametric density ratio model. *Journal of Nonparametric Statistics*, 1–16.