Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

Bayesian Synthetic Likelihood

L. F. Price, C. C. Drovandi, A. Lee & D. J. Nott

To cite this article: L. F. Price, C. C. Drovandi, A. Lee & D. J. Nott (2018) Bayesian Synthetic Likelihood, Journal of Computational and Graphical Statistics, 27:1, 1-11, DOI: 10.1080/10618600.2017.1302882

To link to this article: https://doi.org/10.1080/10618600.2017.1302882

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2018, VOL. 27, NO. 1, 1–11
https://doi.org/10.1080/10618600.2017.1302882

Bayesian Synthetic Likelihood

L. F. Price [a], C. C. Drovandi [a], A. Lee [b], and D. J. Nott[c]

[a]School of Mathematical Sciences, Queensland University of Technology, Australia and Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS); [b]Department of Statistics, University of Warwick, Coventry, UK; [c]Department of Statistics and Applied Probability, National University of Singapore, Singapore

ABSTRACT

Having the ability to work with complex models can be highly beneficial. However, complex models often have intractable likelihoods, so methods that involve evaluation of the likelihood function are infeasible. In these situations, the benefits of working with likelihood-free methods become apparent. Likelihood-free methods, such as parametric Bayesian indirect likelihood that uses the likelihood of an alternative parametric auxiliary model, have been explored throughout the literature as a viable alternative when the model of interest is complex. One of these methods is called the synthetic likelihood (SL), which uses a multivariate normal approximation of the distribution of a set of summary statistics. This article explores the accuracy and computational efficiency of the Bayesian version of the synthetic likelihood (BSL) approach in comparison to a competitor known as approximate Bayesian computation (ABC) and its sensitivity to its tuning parameters and assumptions. We relate BSL to pseudo-marginal methods and propose to use an alternative SL that uses an unbiased estimator of the SL, when the summary statistics have a multivariate normal distribution. Several applications of varying complexity are considered to illustrate the findings of this article. Supplemental materials are available online. Computer code for implementing the methods on all examples is available at *https://github.com/cdrovandi/Bayesian-Synthetic-Likelihood*.

ARTICLE HISTORY
Received March 2016
Accepted January 2017

KEYWORDS
Approximate Bayesian computation; Bayesian indirect likelihood; Indirect inference; Pseudo-marginal methods; Synthetic likelihood

1. Introduction

Statisticians and applied practitioners often desire the ability to work with complex statistical models. Such models can lead to a more complete understanding of the process believed to generate the observed data, in comparison to simpler models that are easy to fit computationally. One computational issue with complex models is that the likelihood function can be very difficult or impossible to compute, precluding the use of standard, likelihood-based approaches to inference. In such settings, likelihood-free methods facilitate inference by approximating the likelihood function in particular ways.

There are some likelihood-free methods that work on the full data level, but it is common practice to reduce the data to a summary statistic for computational or practical purposes (see Blum et al. 2013 for an outline of data reduction techniques). When summary statistics are used, a range of likelihood-free methods including approximate Bayesian computation (ABC, see Sisson and Fan 2011, e.g.) and the synthetic likelihood (SL) method of Wood (2010), which uses a multivariate normal approximation of the distribution of a set of summary statistics, are applicable. Even though Wood (2010) incorporated the SL within a Markov chain Monte Carlo (MCMC) algorithm, the focus of Wood (2010) is to determine the maximum SL estimator, that is, a classical approach. It is trivial to consider a Bayesian version of this, by assigning a prior distribution on the parameter. Then the output of the MCMC algorithm of Wood (2010) would be a sample from an approximated probability distribution of the parameter conditional on the observed summary statistic. We refer to this approach as Bayesian synthetic likelihood (BSL), which is the focus of this article.

In contrast, ABC uses a nonparametric auxiliary likelihood as a replacement to the intractable likelihood (Blum 2010; Drovandi, Pettitt, and Lee 2015) and is currently regarded as the state-of-the-art method of approximation when data have been reduced to a summary statistic. Assume that there is interest in the parameter $\boldsymbol{\theta} \in \Theta$ of a stochastic process and the observed data $\boldsymbol{y} \in \mathsf{Y}$ have been generated from this model. In ABC, data $\boldsymbol{x}$ are simulated from the model based on a proposed $\boldsymbol{\theta}$ and the observed, $\boldsymbol{s_y} \in \mathsf{S}$, and simulated, $\boldsymbol{s_x} \in \mathsf{S}$, summary statistics are compared. Denote n independent simulated datasets as $\boldsymbol{x}_{1:n} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ where each $\boldsymbol{x}_i \in \mathsf{Y}$ is simulated from the model $p(\cdot|\boldsymbol{\theta})$. The summary statistic associated with $\boldsymbol{x}_i$ is denoted $\boldsymbol{s}_i$ for notational convenience. ABC estimates the intractable summary statistic likelihood, $p(\boldsymbol{s_y}|\boldsymbol{\theta})$, nonparametrically:

$$p_{\epsilon,n}(\boldsymbol{s_y}|\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} K_\epsilon(\rho(\boldsymbol{s_y}, \boldsymbol{s}_i)), \tag{1}$$

where $\rho(\boldsymbol{s_y}, \boldsymbol{s_x})$ measures the distance between the observed and simulated summary statistics and $K_\epsilon(\cdot)$ is a kernel weighting function with bandwidth ϵ (known as the ABC tolerance) designed to give larger weight to smaller ρ. Despite the extensive research performed on ABC, there still remains no standard way to select these tuning parameters. ABC can be viewed as a pseudo-marginal method (Andrieu and Roberts 2009), since

CONTACT L. F. Price *leah.south@hdr.qut.edu.au* School of Mathematical Sciences, Queensland University of Technology, Brisbane City, QLD 4000, Australia; and Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS).
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/JCGS*.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JCGS*.

$p_{\epsilon,n}(\boldsymbol{s}_y|\boldsymbol{\theta})$ in (1) is an unbiased estimate of the ABC likelihood, $p_\epsilon(\boldsymbol{s}_y|\boldsymbol{\theta})$, which is given by

$$p_\epsilon(\boldsymbol{s}_y|\boldsymbol{\theta}) = \int_{\mathsf{Y}} p(\boldsymbol{x}|\boldsymbol{\theta}) K_\epsilon(\rho(\boldsymbol{s}_y, \boldsymbol{s}_x)) d\boldsymbol{x}.$$

Using the theory of Andrieu and Roberts (2009), the ABC target is not influenced by the choice of n when the ABC likelihood (1) is incorporated in an MCMC algorithm. However, ABC is known to be highly sensitive to the choice of ϵ, the discrepancy function and to a lesser extent the kernel weighting function (Marin et al. 2012).

Since BSL by definition uses a multivariate normal approximation, it depends only on a single tuning parameter, n, which is the number of replicated simulations of the model used in estimating the mean and covariance matrix of the multivariate normal auxiliary model, denoted by $\boldsymbol{\mu}_n$ and $\boldsymbol{\Sigma}_n$, respectively. BSL uses the following estimate:

$$p_{A,n}(\boldsymbol{s}_y|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}_n(\boldsymbol{\theta}), \boldsymbol{\Sigma}_n(\boldsymbol{\theta})), \tag{2}$$

in place of the intractable likelihood of the summary statistic. The BSL target depends on n as $\mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}_n(\boldsymbol{\theta}), \boldsymbol{\Sigma}_n(\boldsymbol{\theta}))$ is not an unbiased estimate of $\mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$. Due to this dependence on n, there is interest in investigating the sensitivity of BSL to this tuning parameter. The subscript A denotes that the target also depends on the choice of auxiliary model, which in the case of BSL is multivariate normal. Alternative parametric auxiliary models can be considered as part of a wider framework referred to as parametric Bayesian indirect likelihood on the summary statistic level (psBIL) by Drovandi, Pettitt, and Lee (2015). This general framework is described in Appendix A of the supplementary materials, though the focus of this article is on BSL, which is a natural and convenient choice in many applications.

The main aim of this article is to explore the use of the SL within a Bayesian framework. This involves exploring the sensitivity of BSL to its tuning parameter n, the multivariate normal assumption, and assessing the computational efficiency of the approach. We find empirically that the target distribution provided by BSL is remarkably insensitive to n. Due to this, we are able to choose n to maximize computational efficiency and consider guidance from the literature on pseudo-marginal methods (Andrieu and Roberts 2009). Furthermore, we propose to use an estimator of the normal likelihood that is exactly unbiased when the summary statistics are multivariate normal, based on a density estimator due to Ghurye and Olkin (1969). Using this exactly unbiased estimator creates a novel pseudo-marginal BSL method that is theoretically unaffected by n under the multivariate normality assumption for the summary statistic. In a toy application, we show that BSL becomes increasingly more computationally efficient than ABC as the dimension of the summary statistic rises beyond 2. BSL also seems to be more efficient than ABC as the dimension of the summary statistic is increased in more realistic examples. Finally, we consider scenarios where BSL does not perform well.

Since its inception, the SL has shown promising results in substantive applications (see, e.g., Brown et al. 2014). The SL has also been considered within a Bayesian framework. Hartig et al. (2014) applied BSL to the FORMIND forest model. Fasiolo, Pya, and Wood (2016) employed BSL and compared it with particle MCMC (Andrieu, Doucet, and Holenstein 2010) for applications in ecology and epidemiology (see also Fasiolo and Wood 2016). Everitt et al. (2017) used BSL for models that have an intractable normalizing constant. Meeds and Welling (2014), Wilkinson (2014), Moores et al. (2015), and Gutmann and Corander (2015) used emulation to speed up the calculations involved in BSL. Meeds and Welling (2014) considered Gaussian process surrogate modeling of each marginal summary statistic, ignoring potential correlation between summaries. Wilkinson (2014) considered Gaussian process surrogate modeling of the log SL. Gutmann and Corander (2015) developed a similar approach to Wilkinson (2014) but use Bayesian optimization (e.g., Jones 2001) to train the GP rather than the sequential history matching approach (e.g., Craig et al. 1997) used by Wilkinson (2014). Moores et al. (2015) used nonparametric modeling of the relationship between the mean and variance parameters of the SL to a single parameter in the hidden Potts model. Despite this research, no attention has been given to how the value of n affects BSL and no comparison to its natural competitor, ABC. Furthermore, as mentioned earlier, we create a new BSL method.

In Section 2 of this article, the BSL method is described in further detail, including an outline of the notation used throughout this article. Section 3 compares the computational efficiency of ABC and BSL on a toy example. The empirical results are provided in Section 4 on models and datasets with varying complexity. A discussion is provided in Section 5, which also points to further research directions.

2. Bayesian Synthetic Likelihood

2.1. Framework

In a Bayesian framework, the objective is to determine an approximation to the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$, where $\boldsymbol{y} \in \mathsf{Y}$ is data of dimension N coming from the assumed stochastic model. The model is considered to be so complex that the likelihood function, $p(\boldsymbol{y}|\boldsymbol{\theta})$, is not computationally tractable. In the SL method and in psBIL methods more generally, a summary statistic of this data, $\boldsymbol{s}_y \in \mathsf{S} \subseteq \mathbb{R}^d$ where d is the number of statistics, is all that is required. The posterior distribution and likelihood become $p(\boldsymbol{\theta}|\boldsymbol{s}_y)$ and $p(\boldsymbol{s}_y|\boldsymbol{\theta})$, respectively. If $\boldsymbol{s}_y$ carries most of the information contained in the observed data, $\boldsymbol{y}$, then $p(\boldsymbol{\theta}|\boldsymbol{s}_y)$ can be close to $p(\boldsymbol{\theta}|\boldsymbol{y})$. We note also that the aim is not always to target the full posterior as discarding some information can be useful to improve the behavior of the likelihood, for example, in the near-chaotic systems of Wood (2010), or to make the inference more robust to model misspecifications. Since the full data likelihood is intractable, it is likely that $p(\boldsymbol{s}_y|\boldsymbol{\theta})$ is also intractable. In Appendix A of the supplementary materials, a general framework for applying parametric approximations of this intractable likelihood is given.

Due to its suitability under many circumstances and its computational convenience, Wood (2010) used an auxiliary likelihood based on a multivariate normal approximation, referred to as the synthetic likelihood. Here, we have auxiliary parameters $\boldsymbol{\mu}(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\theta})$, where $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\Sigma}$ ($d \times d$ symmetric positive definite matrix) denote the mean and covariance matrix of the multivariate normal distribution. In general, $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are

unknown but can be estimated by simulating n iid datasets of size N from the model based on $\boldsymbol{\theta}$ and fitting the auxiliary likelihood to the summary statistics, $\boldsymbol{s}_{1:n}$. The estimated auxiliary likelihood is $\mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}_n(\boldsymbol{\theta}), \boldsymbol{\Sigma}_n(\boldsymbol{\theta}))$, with the following analytic expression for the auxiliary parameter estimates:

$$\begin{aligned}\boldsymbol{\mu}_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{s}_i,\\ \boldsymbol{\Sigma}_n(\boldsymbol{\theta}) &= \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{s}_i-\boldsymbol{\mu}_n(\boldsymbol{\theta}))(\boldsymbol{s}_i-\boldsymbol{\mu}_n(\boldsymbol{\theta}))^{\top}.\end{aligned} \tag{3}$$

BSL arises when this auxiliary likelihood is combined with a prior distribution on the parameter. Following Drovandi, Pettitt, and Lee (2015), BSL samples from the following target:

$$p_{A,n}(\boldsymbol{\theta}|\boldsymbol{s}_y) \propto p_{A,n}(\boldsymbol{s}_y|\boldsymbol{\theta})p(\boldsymbol{\theta}), \tag{4}$$

where

$$p_{A,n}(\boldsymbol{s}_y|\boldsymbol{\theta}) = \int_{\mathsf{S}^n}\mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}_n(\boldsymbol{\theta}), \boldsymbol{\Sigma}_n(\boldsymbol{\theta}))\prod_{i=1}^{n}p(\boldsymbol{s}_i|\boldsymbol{\theta})d\boldsymbol{s}_{1:n}. \tag{5}$$

By taking a single draw of $\boldsymbol{s}_{1:n} \overset{\text{iid}}{\sim} p(\cdot|\boldsymbol{\theta})$ and computing $\mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}_n(\boldsymbol{\theta}), \boldsymbol{\Sigma}_n(\boldsymbol{\theta}))$, we obtain an unbiased estimate of $p_{A,n}(\boldsymbol{s}_y|\boldsymbol{\theta})$. Under the mild condition that (4) is integrable as a function of $\boldsymbol{\theta}$, using the stochastic and nonnegative estimator $\mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}_n(\boldsymbol{\theta}), \boldsymbol{\Sigma}_n(\boldsymbol{\theta}))$ in an MCMC algorithm (see later) produces a pseudo-marginal algorithm (Andrieu and Roberts 2009) that targets (4). It is important to note that $p_{A,n}(\boldsymbol{\theta}|\boldsymbol{s}_y)$ is not equal to the "ideal" BSL target, $p_A(\boldsymbol{\theta}|\boldsymbol{s}_y) \propto \mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))p(\boldsymbol{\theta})$. Nevertheless, Drovandi, Pettitt, and Lee (2015) did show under mild conditions that the ideal BSL target will be achieved as $n \to \infty$. Later in this section, we develop an alternative BSL method that does target $p_A(\boldsymbol{\theta}|\boldsymbol{s}_y)$, under the assumption that the summary statistic is indeed multivariate normal.

As demonstrated above, the target distribution of BSL depends on the multivariate normal approximation and on the selection of n (since $p_A(\boldsymbol{s}_y|\boldsymbol{\mu}_n(\boldsymbol{\theta}), \boldsymbol{\Sigma}_n(\boldsymbol{\theta}))$ is not an unbiased estimator of $p_A(\boldsymbol{s}_y|\boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$). There is interest, then, on the sensitivity of BSL to the choice of n and its robustness toward departures from normality of the summary statistic, which we investigate empirically in Section 4. We use an MCMC algorithm with T iterations to sample from $p_{A,n}(\boldsymbol{\theta}|\boldsymbol{s}_y)$, which is shown in Appendix B of the supplementary materials. The approach is similar to a standard MCMC algorithm but includes, at each iteration, a simulation step to obtain $\boldsymbol{\mu}_n(\boldsymbol{\theta})$ and $\boldsymbol{\Sigma}_n(\boldsymbol{\theta})$ of the SL. Wood (2010) adopted the same approach but uses the output to maximize the SL, rather than using the samples to construct a posterior distribution. We refer to this algorithm as MCMC BSL.

Under the assumption that the summary statistic is normally distributed and the observed summary statistic $\boldsymbol{s}_y$ is fixed, we point out that there is an exactly unbiased, nonnegative estimator of a normal density function due to Ghurye and Olkin (1969, sec. 3.4), and which results in a valid pseudo-marginal algorithm targeting $p_A(\boldsymbol{\theta}|\boldsymbol{s}_y)$ for any $n > d+3$ rather than $p_{A,n}(\boldsymbol{\theta}|\boldsymbol{s}_y)$, which is only an approximation of $p_A(\boldsymbol{\theta}|\boldsymbol{s}_y)$. We refer to the unbiased estimator of the SL as uSL and the resulting Bayesian procedure as uBSL, where "u" denotes unbiased. Using the notation of Ghurye and Olkin (1969), let

$$c(k,\upsilon) = \frac{2^{-k\upsilon/2}\pi^{-k(k-1)/4}}{\prod_{i=1}^{k}\Gamma\left(\frac{1}{2}(\upsilon-i+1)\right)},$$

and for a square matrix $\boldsymbol{A}$ write $\psi(\boldsymbol{A}) = |\boldsymbol{A}|$ if $\boldsymbol{A} > 0$ and $\psi(\boldsymbol{A}) = 0$ otherwise, where $|\boldsymbol{A}|$ is the determinant of $\boldsymbol{A}$ and $\boldsymbol{A} > 0$ means that $\boldsymbol{A}$ is positive definite. The result of Ghurye and Olkin (1969) shows that an exactly unbiased estimator of $\mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}(\boldsymbol{\theta}), \boldsymbol{\Sigma}(\boldsymbol{\theta}))$ is (in the case where the summary statistics are normal and $n > d+3$)

$$\begin{aligned}\hat{p}_A(\boldsymbol{s}_y|\boldsymbol{\theta}) = (2\pi)^{-d/2}\frac{c(d,n-2)}{c(d,n-1)(1-1/n)^{d/2}}|\boldsymbol{M}_n(\boldsymbol{\theta})|^{-(n-d-2)/2}\\ \times\psi\left(\boldsymbol{M}_n(\boldsymbol{\theta})-(\boldsymbol{s}_y-\boldsymbol{\mu}_n(\boldsymbol{\theta}))(\boldsymbol{s}_y-\boldsymbol{\mu}_n(\boldsymbol{\theta}))^{\top}/(1-1/n)\right)^{(n-d-3)/2},\end{aligned}$$

where $\boldsymbol{M}_n(\boldsymbol{\theta}) = (n-1)\boldsymbol{\Sigma}_n(\boldsymbol{\theta})$. The estimated likelihood, $\hat{p}_A(\boldsymbol{s}_y|\boldsymbol{\theta})$, replaces $\mathcal{N}(\boldsymbol{s}_y; \boldsymbol{\mu}_n(\boldsymbol{\theta}), \boldsymbol{\Sigma}_n(\boldsymbol{\theta}))$ in MCMC BSL to create the novel MCMC uBSL algorithm. We stress that this algorithm targets $p_A(\boldsymbol{\theta}|\boldsymbol{s}_y)$ (not $p(\boldsymbol{\theta}|\boldsymbol{s}_y)$) under the multivariate normality assumption of the summary statistic, which can be made approximately valid through appropriate choice or transformation of the summary statistic (Wood 2010). The appeal of this approach is that the approximate posterior it induces may have less dependence on n compared to the standard implementation of MCMC BSL. We discuss this further in the next section and investigate it empirically in Section 4.

2.2. Choice of n

MCMC BSL is similar to the grouped independence Metropolis–Hastings (GIMH) algorithm of Beaumont (2003) in that a stochastic estimator replaces an intractable likelihood and the likelihood estimate for the current $\boldsymbol{\theta}$ is carried over to the next iteration (it is not re-estimated). Specifically, the GIMH algorithm is a pseudo-marginal method (Andrieu and Roberts 2009) where a nonnegative and unbiased likelihood estimator is used in place of the intractable likelihood. Andrieu and Roberts (2009) showed that the GIMH method has as its limiting distribution the desired posterior distribution. In our case, the auxiliary likelihood estimator based on n is not an unbiased estimator of the auxiliary likelihood obtained if it were possible to take $n \to \infty$, despite using unbiased estimators of the mean vector, $\boldsymbol{\mu}(\boldsymbol{\theta})$, and covariance matrix, $\boldsymbol{\Sigma}(\boldsymbol{\theta})$. However, we demonstrate with strong empirical evidence in Section 4 that under the assumptions that the distribution of the summary statistic is not highly irregular and the model is able to recover $\boldsymbol{s}_y$, the BSL target shows little sensitivity to n. Therefore, we conjecture that, under these assumptions, the bias in the SL estimator decreases fairly rapidly so that small to moderate n is sufficient.

Given the apparent insensitivity of the BSL posterior to n, we suggest to choose n to maximize the overall computational efficiency. A small value of n reduces the computation time per iteration, but will result in highly variable estimates of the SL. It is well known that the GIMH algorithm can become stuck if the likelihood for some $\boldsymbol{\theta}$ is grossly overestimated. On the other hand, if n is set large, the SL is estimated precisely, but the computation time per iteration is high. Borrowing the theoretical

result for the GIMH method outlined in Doucet et al. (2015), the value of n should be chosen such that the log SL at some $\boldsymbol{\theta}$ with high (BSL) posterior support should be estimated with a standard deviation of roughly 1. We investigate this recommendation in Section 4.

The MCMC uBSL target is theoretically unaffected by n if the multivariate normality assumption of the summary statistic holds. However, given that this assumption is unlikely to hold in practice, the sensitivity of MCMC uBSL to the value of n is of interest, which we explore empirically in Section 4. It is important to note that the random variable describing the uSL is a mixture of a discrete and a continuous random variable; it may be identically 0 if the argument of $\psi(\cdot)$ is not positive definite, implying that the log uSL is $-\infty$ in such cases. Hence, the standard deviation of the log uSL is infinite generally, meaning that we cannot consider the guidance of Doucet et al. (2015) for pseudo-marginal methods. In Section 4, we compare the performance of uBSL against BSL.

In Section 3, we demonstrate on a toy example that BSL becomes increasingly more computationally efficient relative to ABC with an increase in the dimension of the summary statistic. This is an expected result as BSL uses a parametric approximation to the summary statistic likelihood as opposed to the nonparametric one used in ABC. However, we find empirically in Section 4 that for BSL the optimal n in terms of computational efficiency increases with the dimension of the summary statistic. Thus, BSL is unable to completely escape the curse of dimensionality issue. Therefore, as in the case of ABC, the choice of summary statistic in BSL is important to keep the computation to a manageable level.

2.3. Normality Assumption

In many applications, the central limit theorem may justify the use of the multivariate normal approximation of the distribution of the summary statistic (Wood 2010). In cases where the normality assumption does not hold, Wood (2010) suggested to apply an appropriate transformation. However, in complex problems with a high-dimensional summary statistic, it may not be feasible to investigate such transformations in great detail. Hence, of interest is the robustness of the BSL method to departures from normality.

A possible online diagnostic tool for BSL would involve performing a hypothesis test for multivariate normality (or at least a test for normality on each component of the summary statistic) at every iteration of MCMC BSL. Unfortunately, to achieve sufficient power in these tests a large value of n is required, which is not computationally efficient. This is consistent with Wood (2010) in the context of SL. Despite this, we adopt this approach to investigate the robustness of BSL when there is evidence against the normality assumption.

3. Efficiency Comparison to ABC

ABC approximates the intractable summary statistic likelihood, $p(\boldsymbol{s}_{\boldsymbol{y}}|\boldsymbol{\theta})$, with the following approximate likelihood:

$$p_\epsilon(\boldsymbol{s}_{\boldsymbol{y}}|\boldsymbol{\theta}) = \int_{\mathcal{Y}} p(\boldsymbol{x}|\boldsymbol{\theta}) K_\epsilon(\rho(\boldsymbol{s}_{\boldsymbol{y}}, \boldsymbol{s}_{\boldsymbol{x}}))d\boldsymbol{x}. \tag{6}$$

A common choice for the kernel weighting function $K_\epsilon(\cdot)$ is the indicator function $K_\epsilon(\rho(\boldsymbol{s}_{\boldsymbol{y}}, \boldsymbol{s}_{\boldsymbol{x}})) \propto \mathbb{I}(\rho(\boldsymbol{s}_{\boldsymbol{y}}, \boldsymbol{s}_{\boldsymbol{x}}) \leq \epsilon)$ while in this article we consider the Gaussian kernel, $K_\epsilon(\rho(\boldsymbol{s}_{\boldsymbol{y}}, \boldsymbol{s}_{\boldsymbol{x}})) \propto \exp(-\frac{\rho(\boldsymbol{s}_{\boldsymbol{y}}, \boldsymbol{s}_{\boldsymbol{x}})^2}{2\epsilon})$. The integral to compute the ABC likelihood in (6) is analytically intractable but it can be unbiasedly estimated by simulation, which is sufficient to generate an MCMC kernel that targets $p_\epsilon(\boldsymbol{\theta}|\boldsymbol{s}_{\boldsymbol{y}}) \propto p_\epsilon(\boldsymbol{s}_{\boldsymbol{y}}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ (Andrieu and Roberts 2009). The estimated ABC likelihood is given in (1). The results of Bornn et al. (2017) suggested that, in a serial computing environment, $n = 1$ is close to optimal. Below we compare the computational efficiency of ABC and BSL on a toy example.

It is well known that ABC suffers a curse of dimensionality with respect to the size of the summary statistic (Blum et al. 2013). The multivariate normality assumption in BSL may deteriorate with the size of the summary statistic, so that BSL still suffers from a curse of dimensionality but in a different manner. However, intuitively one might expect BSL to be more computationally efficient than ABC due to the use of a parametric auxiliary model. The following toy example provides some insight. We will assume the unknown likelihood is $p(\boldsymbol{y}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is some fixed and known covariance matrix. The auxiliary model is $p_A(\boldsymbol{y}|\boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{\phi}, \boldsymbol{\Sigma})$, and one is interested in knowing which of the two methods (standard ABC vs. BSL) is more effective. Such a comparison is aided by the fact that there is a correspondence between $p_{A,n}(\boldsymbol{y}|\boldsymbol{\theta})$ and $p_\epsilon(\boldsymbol{y}|\boldsymbol{\theta})$ in this particular case, with $K_\epsilon(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{x}, \epsilon\boldsymbol{\Sigma})$ and $\boldsymbol{\phi}_n(\boldsymbol{\theta}) = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i$. Indeed, elementary calculations provide that when $\epsilon = n^{-1}$, $p_{A,n}(\boldsymbol{y}|\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\Sigma}(1+\epsilon)) = p_\epsilon(\boldsymbol{y}|\boldsymbol{\theta})$.

Since the two different auxiliary models define the same approximate likelihood, they induce the same posterior distribution for a given prior distribution. The main difference is how the likelihood $\mathcal{N}(\boldsymbol{y}; \boldsymbol{\theta}, \boldsymbol{\Sigma}(1+\epsilon))$ is estimated within a Monte Carlo method. In ABC, the approximation is $Z_1(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{x}_1, \epsilon\boldsymbol{\Sigma}), \boldsymbol{x}_1 \sim \mathcal{N}(\cdot; \boldsymbol{\theta}, \boldsymbol{\Sigma})$, whereas in BSL the approximation is $Z_2(\boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{x}_2, \boldsymbol{\Sigma}), \boldsymbol{x}_2 \sim \mathcal{N}(\cdot; \boldsymbol{\theta}, n^{-1}\boldsymbol{\Sigma})$.

The full details on the efficiency comparison for this toy example can be found in Appendix C of the supplementary materials. Here, we provide the main results. In a rejection sampling framework, we find that the ratio of the ABC acceptance rate against the BSL acceptance rate is $\epsilon^{d/2}$. ABC requires on average $n^{d/2}$ more samples of $\boldsymbol{x}_1$ than BSL requires of $\boldsymbol{x}_2$. Taking into account the fact that $\boldsymbol{x}_2$ is an average of n $\mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ random variables, and recalling that $\epsilon = n^{-1}$, we obtain that an ABC rejection sampler is more computationally efficient when $d = 1$, equally computationally efficient when $d = 2$, and becomes significantly less efficient as d increases beyond 2.

For other Monte Carlo methods, it is of some interest to consider the relative variance of the estimates $Z_1(\boldsymbol{\theta})$ and $Z_2(\boldsymbol{\theta})$. We find that $\text{var}(Z_2(\boldsymbol{\theta})) < \text{var}(Z_1(\boldsymbol{\theta}))$ for any $(\boldsymbol{y}, \boldsymbol{\theta}, \boldsymbol{\Sigma})$, $\epsilon \in (0, 1)$, and $d \geq 1$. The relative variance of $Z_2(\boldsymbol{\theta})$ converges to 0 as $\epsilon \to 0$. In fact, we have that the relative variance of $Z_1(\boldsymbol{\theta})$ is in $\mathcal{O}(\epsilon^{-d/2})$ as $\epsilon \to 0$. It follows that if we define $Z_{1,M}(\boldsymbol{\theta})$ to be an average of M iid replicates of $Z_1(\boldsymbol{\theta})$, then one would need to increase M at least by $n^{d/2}$ to control the relative variance. Once again, by taking into account the fact that simulating $Z_2(\boldsymbol{\theta})$ involves averaging n $\mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ random variables, this implies that as d increases beyond 2 BSL becomes significantly more computationally efficient.

Since the above result is limited to the toy application considered and that BSL estimates $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ rather than using a fixed $\boldsymbol{\Sigma}$, we attempt to compare the computational efficiency of BSL and ABC on the examples in the next section.

4. Examples

Here, results for BSL and uBSL are obtained for several simulation studies. These involve investigating the sensitivity of the BSL approaches to n, robustness against the normality assumption, and comparisons to ABC in terms of computational efficiency. We note that it is generally difficult to fairly compare the efficiency of different methods, especially when they result in different target densities. In addition, ABC has a number of tuning parameters, which further complicates the comparison. For some of the examples the targets produced are similar so in these cases we use the same proposal distribution in the BSL and ABC MCMC algorithms. To simplify the comparisons, we use efficient starting values and proposal distributions that are informed by pilot runs as described in Appendix D of the supplementary materials. We assume that model simulation consumes the majority of computing time in the simulation-based methods we consider in this article. Thus to compare efficiency, we compute the effective sample size (ESS) for each parameter obtained from the CODA package in R (Plummer et al. 2006) and standardize it by the number of model simulations used. We then multiply this number by a large constant scalar to increase the magnitude to facilitate easy comparison. We refer to this as the normalized ESS value.

We attempt to give an advantage to the ABC approach by assigning what might be considered an efficient discrepancy function. Using the true parameter value, which is obviously unavailable in practice, a number of simulations are generated from the model and the summary statistics calculated. Then the sample covariance matrix of the summary statistic is computed and subsequently used in a Mahalanobis discrepancy function for ABC. Devising an efficient discrepancy function in ABC is a nontrivial task and may involve some pilot runs. BSL avoids this extra tuning and computation. A Gaussian function with variance ϵ is used as the kernel weighting function. We choose the ϵ value for ABC so that it produces similar normalized ESS values to the BSL run that maximized these values. Then we perform the regression adjustment approach of Beaumont, Zhang, and Balding (2002), which attempts to predict the ABC posterior marginals for $\epsilon = 0$. If the regression adjustment has an impact, we suggest that the accuracy of ABC can be improved by further lowering ϵ. In this case, we suggest that BSL is more computationally efficient than ABC.

An additional example based on the four parameter g-and-k quantile distribution (e.g., Rayner and MacGillivray 2002) was investigated, and full details are provided in Appendix E of the supplementary materials. The findings from this example support the claim that the BSL posteriors are insensitive to n and that BSL appears to be more efficient than ABC for this example.

4.1. Difference With True Posterior—Toy Example

This example allows for the comparison of ABC and BSL to the true posterior conditional on the full data, which is known analytically. As previously explained, the aim is not always to compare to the full posterior, but we are using a sufficient statistic in this application so a comparison is sensible.

In this toy example, the data are drawn from a Poisson distribution, $\mathcal{P}(\lambda)$, with mean λ and the prior for λ is a gamma distribution, $\mathcal{G}(\alpha, \beta)$, with mean α/β. The summary statistic here is the sample mean, which is sufficient for λ. We compare the results of the BSL approaches with the true posterior, $\lambda|\boldsymbol{y} \sim \mathcal{G}(\sum_{i=1}^{N} y_i + \alpha, \beta + N)$.

4.1.1. Sensitivity to n

For this simple example, the MCMC (u)BSL algorithm is run with $\lambda = 30$, $\alpha = \beta = 0.001$, $N = 100$, and $T = 100{,}000$. By setting $\lambda = 30$, the choice of a normal distribution for the auxiliary likelihood may be approximately valid. The range of n values investigated here are shown in Table 1.

In Figure 1(a) and 1(b), the estimated posterior distributions implied by BSL and uBSL for a range of n values are shown, respectively. For ease of presentation, the figure only shows results for some values of n. It is evident from the graph that the BSL posteriors are surprisingly insensitive to n, with small departures in the tails for $n = 2$. We find that the BSL posteriors are very similar for $n \geq 5$. The posterior results for uBSL are unsurprisingly insensitive to n given that the distribution of the sample sum at the true value of λ is $\mathcal{P}(3000)$, which can be well approximated by a normal distribution.

From Table 1, the efficiency of BSL and uBSL is very similar. The normalized ESS suggests that n values of 5–7 give efficient results. However, we find that the posterior results for $n = 2$ are slightly away from the other values of n for BSL.

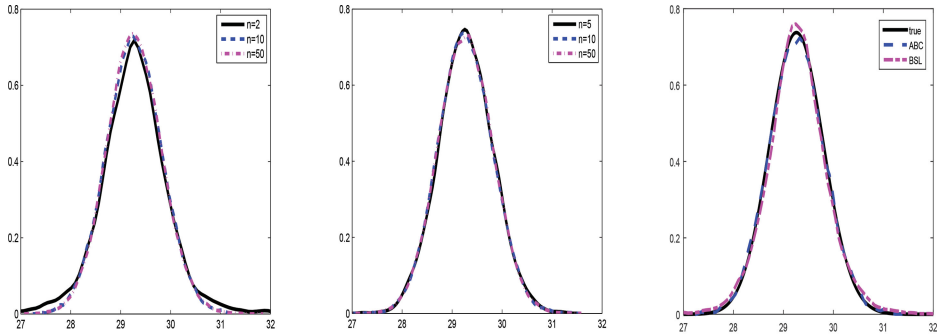
At the true value of λ with $n = 2$, the log SL has a very heavy tail and thus a large standard deviation. Despite this, BSL shows relatively high efficiency. For $n = 5$, the tail is less heavy but the standard deviation is well above 1. The standard deviation for $n = 10$ gives a value closer to that recommended of pseudo-marginal methods. However, $n = 10$ produces less efficient results than smaller values of n. The value of $n = 20$ produces accurate synthetic likelihoods but requires too much computation to be useful.

4.1.2. Comparison to ABC

Using the squared difference between summary statistics as the discrepancy function, we find that $\epsilon = 0.001$ results in an ABC posterior close to the true posterior (see Figure 1(c)). The MCMC ABC acceptance probability is 17.2%. The normalized ESS for ABC is 25, indicating that ABC is more efficient than BSL and uBSL for this one parameter and summary statistic example.

Table 1. Sensitivity of BSL/uBSL to n for the simple example with regards to MCMC acceptance rate and normalized ESS. Shown also is the estimated standard deviation of the log SL at the true parameter value $\lambda = 30$. A "–" indicates that a result is not available for uBSL as the value of n is too small.

n	acc. rate (%)	ESS	sd(log $p_{A,n}(\boldsymbol{s}_{\boldsymbol{y}}\|\lambda = 30)$)
2	33.6/–	13/–	large
5	55.9/54.2	14/19	5
6	58.4/57.2	17/19	1.6
7	60.2/59.8	19/22	1.4
10	62.8/63.0	10/11	0.8
20	66.6/66.6	5/6	0.4



(a) BSL (b) uBSL (c) ABC and BSL

Figure 1. Results for the toy example. (a) BSL posterior estimates for different values of n. (b) uBSL posterior estimates for different values of n. (c) Comparison of the true posterior (solid) with the BSL posterior (with $n = 5$, dot-dash) and ABC posterior (with ϵ chosen so that the acceptance probability is roughly 17.2%, dash). The online figure is in color.

4.1.3. Normality

Here, we investigate how the normality assumption of the BSL approaches affects the accuracy of the results. The sum of N iid $\mathcal{P}(\lambda)$ variables is distributed as $\mathcal{P}(N\lambda)$. A general rule of thumb is that a Poisson distribution is approximately normally distributed for mean greater than or equal to 30. To investigate the effects of the normality assumption, $N = 100$ with $\lambda = 30$ is chosen for an example where the normality assumption is appropriate and $N = 10$ with $\lambda = 1$ is chosen for an example where the normality assumption is violated. For this investigation, we use $n = 50$ and $T = 100{,}000$ so that any error can be mostly attributed to the lack of normality of the summary statistic.

An Anderson–Darling test is performed using the summary statistic sample at each MCMC iteration. Some graphs are displayed in Appendix F of the supplementary materials, showing the accuracy of the estimated posterior along with some histograms of the p-values. These graphs are shown for examples with $\lambda = 1$ and $\lambda = 30$. As expected, the p-values do not appear to be uniformly distributed between zero and one in the example where $\lambda = 1$. When $\lambda = 30$ the assumption of normality appears much more reasonable. The BSL approaches appear to have very accurate estimates of the posterior distribution in both cases, which is remarkable given the strong departure from normality when $\lambda = 1$.

4.2. Departure from Normality—Ricker Model

The summary statistics in this example are non-Gaussian so it is interesting to investigate the sensitivity of BSL and uBSL to n and to compare their output with ABC, which does not suffer from the multivariate normality assumption.

We consider the Ricker model presented in Wood (2010). Here, a population of size N_t at time t evolves according to $N_{t+1} = rN_t e^{-N_t+e_t}$ where $e_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2)$. However, a more realistic scenario is where the N_t are unobserved and what is observed are the random variables, Y_t, such that $Y_t \sim \mathcal{P}(\phi N_t)$.

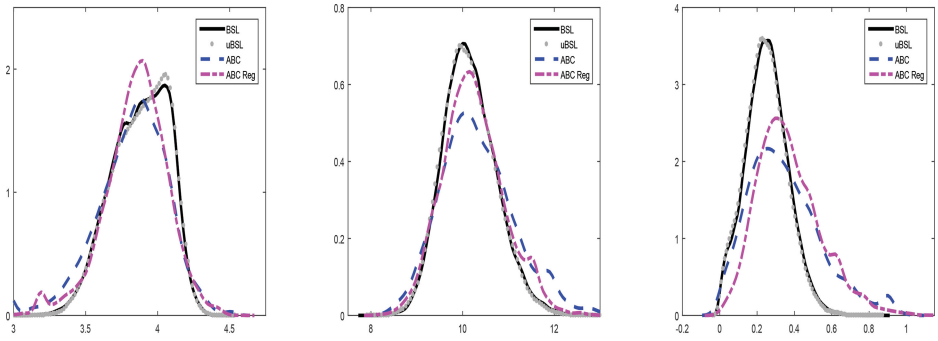
We set the model parameter as $\boldsymbol{\theta} = (\log r, \phi, \sigma_e)$. The observed data of size $N = 100$ is generated from the Ricker model with parameter $\boldsymbol{\theta} = (3.8, 10, 0.3)^\top$ and $N_0 = 1$. Here, we use the same summary statistics to that used in Wood (2010). These include the average observation, the number of zeros, the autocovariances up to lag 5 (including the variance, lag 0), the parameter estimates of β_0 and β_1 based on the regression, $y_t^{0.3} = \beta_0 y_{t-1}^{0.3} + \beta_1 y_{t-1}^{0.6} + \eta_t$ where $\eta_t \sim \mathcal{N}(0, \sigma_\eta^2)$, and the coefficients of a cubic regression of the ordered differences on their observed values (see Wood 2010 for more details). This constitutes a total of 13 summary statistics. The prior distributions on the parameters are independent, uniform, and improper (with positive σ_e and ϕ).

To determine if the normality assumption of the summary statistic is reasonable, we use the Anderson–Darling test for normality on each of the (marginal) summary statistics when $n = 100$ for all values of $\boldsymbol{\theta}$ proposed in MCMC BSL. We find that the components of the summary statistic do not follow a normal distribution; there are departures away from the uniform distribution for the p-values, significantly for many components (results not shown). However, it does not appear that the distributions of the summary statistics are highly irregular for $\boldsymbol{\theta}$ with high posterior support, with some statistics showing skewness without very heavy tails.

For this example, it is possible to use an unbiased particle filtering estimate of the likelihood in particle MCMC (Andrieu, Doucet, and Holenstein 2010) to target the posterior conditional on the full data. Given that ABC and BSL work on the summary statistic level, a comparison has not been included. See Fasiolo, Pya, and Wood (2016) who provided a comparison of particle MCMC and BSL.

4.2.1. Sensitivity to n

To investigate the sensitivity of the BSL posteriors to n, we run the algorithm for a variety of values of n. The results are shown in a table and in figures presented in Appendix G of the supplementary materials. These figures demonstrate that the BSL target distributions are again remarkably insensitive to n, given the lack of normality of the summary statistic. Owing to the insensitivity of the target to n, the optimal n may be considered as the one that maximizes the normalized ESS. The efficiency of BSL and uBSL are again quite similar. The optimal value of n out of the values tested appears to be 50 (with normalized ESS values of 30, 35, and 45 for the three parameters), but values of n in the range 30–100 seem to provide relatively efficient results. For this



(a) $\log r$ (b) ϕ (c) σ_e

Figure 2. Posterior estimates for $\log r$, ϕ, and σ_e of the Ricker model when using ABC (dash, based on 0.26% acceptance rate), ABC with regression adjustment (dot-dash), BSL (solid) and uBSL (dot). The BSL approaches use $n = 50$. The online figure is in color.

range of n, values of the standard deviation of the estimated log SL at the true parameter are roughly 1–4.

4.2.2. Comparison to ABC

We run ABC for 25 million iterations with an ABC tolerance that produces an acceptance rate of roughly 0.26%. Our chosen ABC tolerance leads to normalized ESS values of 19, 20, and 26, which suggests that we have allocated a small tolerance relative to the BSL results with efficient n values. This indicates that further reductions in ϵ may have an effect on the ABC posterior, suggesting that BSL may be more efficient for this example. There is some discrepancy between the (u)BSL and ABC posteriors for σ_e but overall the BSL approaches produce an approximate posterior in the vicinity of the ABC approximation despite the mild departure from normality of the summary statistics.

4.3. High-Dimensional Summary Statistic—Cell Biology Model

In this realistic example, the summary statistic is of dimension 145. ABC suffers a curse of dimensionality with respect to the size of the summary statistic, and intuitively one might suspect that BSL will also suffer from a curse of dimensionality as the multivariate normality assumption deteriorates in higher dimensions. The performance of ABC and BSL in this application reveals how the methods scale with increasing dimension of the summary statistic.

Cell motility and proliferation are important parts of many biological processes. Cell motility causes random movement, which, together with proliferation or reproduction, can cause tumors to spread (Swanson et al. 2003) or wounds to heal (Dale, Maini, and Sherratt 1994; Zahm et al. 1997). The main function of many medical treatments is to influence the rates of these processes. To measure the efficacy of such treatments, it is important that a measure of cell motility and proliferation can be accurately obtained. Unfortunately, stochastic models for collective cell spreading do not possess a tractable likelihood function. Several articles have adopted an ABC approach to estimate the parameters (e.g., Johnston et al. 2014; Vo et al. 2015b). One difficulty with these cell biology applications is that the observed data are typically available as sequences of images and therefore it is not trivial to reduce the dimension of the summary statistic to a suitable level for ABC while simultaneously retaining relevant information contained in the images.

A common method of collecting information about cell diffusivity and proliferation is the scratch assay (e.g., Fronza et al. 2009; Johnston et al. 2014). Scratch assays can be used to measure cell migration in vitro and can be performed with readily available and inexpensive equipment. Once cells have formed a single layer completely covering the assay (i.e., a confluent monolayer), a "scratch" is made, which separates the cells (Liang, Park, and Guan 2007). Images of the cells are taken at regular time intervals until the cells are once again in contact, and often the images are then reduced to summary statistics. In most cases (e.g., Treloar and Simpson 2013; Simpson et al. 2013; Johnston et al. 2014), formal analysis is performed on a small number of images with intervals of at least 1 hr, even when images are taken more frequently. In the experiment of Johnston et al. (2014), images of murine fibroblast cells (3T3 cells) are taken every 5 min for 12 hr and here we consider the possibility of using all 145 images (including the initial image) in the analysis. Johnston et al. (2014) considered three of these images, at 4, 8, and 12 hr. By using 145 images rather than a small subset of this, valuable information about the rates of motility and proliferation could be attained. Here, we investigate the capabilities of BSL to accommodate this high-dimensional summary statistic and compare it with ABC with the same summary statistic and also with the ABC approach of Johnston et al. (2014) who considered only three images. The reader is referred to Johnston et al. (2014) for the summary statistics used in their article.

To create the observed data, the cells can be placed on a two-dimensional discrete lattice using image analysis software and some manual processes. This is a time-consuming process and part of the reason why Johnston et al. (2014) considered only three images (in addition to reducing the dimensionality of the problem). Here, we consider simulated data to determine whether it might be beneficial to manually process more images, in terms of how much additional information is obtained about the parameters. Let $X^t_{x,y} \in \{0, 1\}$ be an indicator that defines whether a cell is present at position (x, y) for $x \in \{1, \ldots, R\}$, $y \in \{1, \ldots, C\}$ at time index $t \in \{0, 1, \ldots, 144\}$. Here R and C are the number of rows and columns in the lattice, respectively. Denote the matrix of indicators at time index t as $\boldsymbol{X}^t$. One

possibly informative summary statistic regarding motility is the Hamming distance between $\boldsymbol{X}^t$ and $\boldsymbol{X}^{t-1}$

$$s_t = \sum_{x=1}^{R}\sum_{y=1}^{C} \left|X_{x,y}^t - X_{x,y}^{t-1}\right|.$$

This summary statistic should be suitable if relatively few motility events take place during the time interval since the Hamming distance does not take into account how far cells might have traveled, only the number of positions in the two matrices that differ. The summary statistic we use to provide information regarding the proliferation is the total number of cells at the end of the experiment, which we denote as K for some simulated dataset. Thus, the simulated summary statistic is given by $\boldsymbol{s} = (s_1, \ldots, s_{144}, K)$ and is of dimension 145.

Random walk models allow for the direct comparison of simulations to their observed counterparts. The random walk used here is a reflection of the cells under consideration. Cells are motile, with the ability to move to a neighboring lattice site (north, east, south, west) during each time period of duration τ, which is fixed and set small enough so it approximates well a stochastic process in continuous time. Assuming that there are a total of $N(t)$ cells present at time t, then during each time step $N(t)$ cells are chosen with replacement and given the opportunity to move (Simpson et al. 2013). Experiments have suggested that the cell movement is random, so cells are equally likely to attempt movement in the x and y directions. If the attempted movement is to a vacant site, then the motility event is successful.

When 3T3 fibroblast cells proliferate, they have a separation distance of one (Simpson, Landman, and Hughes 2010). After all motility events have been attempted during a single time step, $N(t)$ cells are chosen with replacement and given the opportunity to proliferate. The proliferation is successful if the selected neighboring location is empty (Simpson, Landman, and Hughes 2010).

The outcomes of this biological process are determined solely by the cell motility and proliferation so only two parameters are required in the random walk. While the parameters of interest are the diffusivity D and the proliferation rate λ, it is simple to just work with the probabilities of motility and proliferation, $P_m \in [0, 1]$ and $P_p \in [0, 1]$. Conversion back to the biological parameters is done by using the formulas in Johnston et al. (2014). It is also possible to include additional parameters for cell-to-cell adhesion and cell-to-substrate adhesion depending on the type of cell under consideration. For more details on the random walk simulation model, the reader is referred to Johnston et al. (2014).

The data are simulated with $P_m = 0.35$ and $P_p = 0.001$ with $R = 27$ and $C = 36$. Initially, $N(0) = 110$ cells are placed randomly in the rectangle with positions $x \in \{1, 2, \ldots, 13\}$ and $y \in \{1, 2, \ldots, 36\}$.

4.3.1. Sensitivity to n

We run the BSL methods with various values of n, with the table of results and the estimated BSL and uBSL posteriors for different values of n given in Appendix H of the supplementary materials. It is again remarkable how insensitive the (u)BSL posteriors are to n, given the high-dimensional summary statistic. Despite this insensitivity of the target to n, very large values of n are required to estimate the SL precisely and achieve reasonable mixing due to the high-dimensional summary statistic. However, we are able to take advantage of the embarrassingly parallel nature of BSL by performing the n independent model simulations on a computer node with 16 cores. This is trivial to implement, using the *parfor* technology in Matlab, for example.

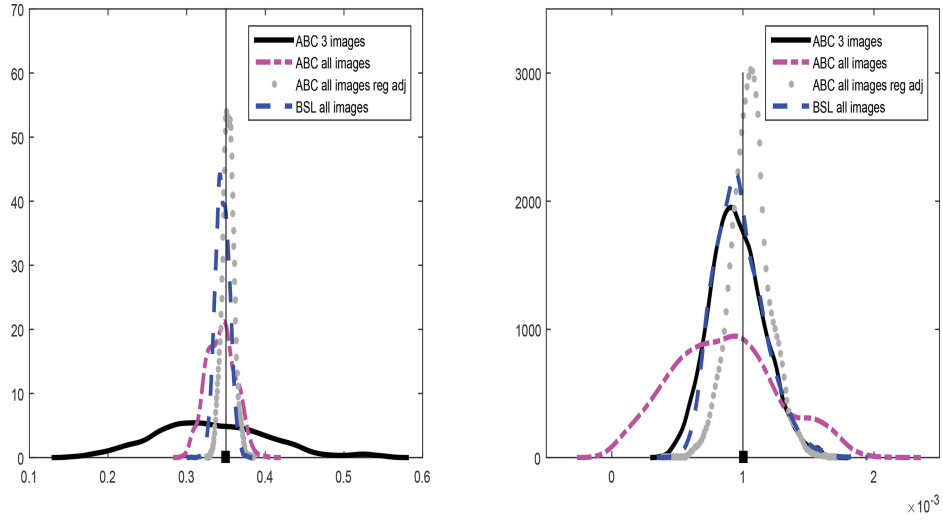
The BSL and uBSL methods have similar efficiency. The optimal value of n appears to be $n = 5000$, which produces an estimated log SL with a standard deviation of 1.4. However, n values between 2500 and 10,000 are also relatively efficient.

4.3.2. Comparison to ABC

We run ABC with the same set of summary statistics. We make use of the 16 processors by taking the average of the kernel weighting function values for 16 independent model simulations for each proposed parameter value. We use 2 million iterations for MCMC ABC. ABC is run with several different tolerance values within $\epsilon = 1000$ and $\epsilon = 1500$. These choices of the tolerances result in acceptance rates between 1% and 8%. The tolerance of $\epsilon = 1000$ produces normalized ESS values closest to the BSL approaches with near optimal n. However, we find that this tolerance leads to poor (nonsmooth) posterior density estimates even after significant thinning. We suggest that the very low acceptance rate is leading to an inaccurate estimate of the ESS. The tolerance of $\epsilon = 1100$ gives smoother posterior estimates. We also perform regression adjustment. In the regressions, we use the first discrepancy value and associated simulated summary statistic produced by the 16 independent simulations at each iteration of MCMC ABC.

The comparison of the posterior results for ABC, BSL ($n = 5000$), and the ABC approach of Johnston et al. (2014) is shown in Figure 3. The results for uBSL and BSL are similar so we only present the BSL results. First, from Figure 3(a), it is evident that a substantial amount of additional information can be obtained about the motility parameter P_m by considering more than three images. Further, the results from BSL are much more precise than that from ABC (cross-validation provides reassurance that we are not over-confident in the parameter values, see Appendix H of the supplementary materials). Remarkably, the ABC results using three images are much more precise than the ABC results for P_p (Figure 3(b)) using all the images. Note that Johnston et al. (2014) considered the number of cells at each of the three time points (including 12 hr). Both BSL and ABC with all the images use the number of cells at 12 hr as a summary statistic. The results for P_p from BSL are very close to the results of the ABC approach of Johnston et al. (2014). It appears that ABC with all the images is being greatly affected by the inability to reduce the ABC tolerance, further demonstrated by the strong impact of the regression adjustment.

The BSL approaches are able to make use of the multiple cores in a more efficient manner than ABC. Given the much higher acceptance rates of MCMC BSL compared to ABC, fewer iterations are required, which implies fewer calls to the multiple cores. Further, due to the large value of n required to estimate the BSL target precisely, the multiple cores are given a



(a) P_m (b) P_p

Figure 3. Posterior density estimates for (a) P_m and (b) P_p of the collective cell spreading model when using the ABC approach of Johnston et al. (2014) (solid), ABC with all the images without (dot-dash) and with (dot) regression adjustment, and BSL ($n = 5000$) with all the images (dash). Shown in squares with a vertical line are the true values of the parameters used to generate the data. The online figure is in color.

significant amount of work to do. ABC performs on average 67 model simulations per second, whereas BSL is able to produce 820 model simulations per second. It appears that BSL, together with the capabilities of parallel computing, is able to deal with the very high-dimensional summary statistic in this application.

5. Discussion

Our empirical results suggest that the optimal value of the standard deviation of the estimated log SL (at a parameter value with high posterior support) is likely to be above 1, the value recommended generally for pseudo-marginal methods (Doucet et al. 2015). However, in their theoretical results, Doucet et al. (2015) assumed that the log-likelihood estimator has a normal distribution. For BSL, we observe that for moderate values of n the synthetic log-likelihood estimator can have a heavy left tail (i.e., underestimated synthetic log-likelihoods). This might explain the larger optimal standard deviation that we observe for BSL relative to pseudo-marginal methods generally. The odd underestimated log-likelihood does not cause much problem in pseudo-marginal methods as these values are simply rejected. More concerning is when the log-likelihood estimator has a heavy right tail, which can result in the likelihood being grossly overestimated and the MCMC chain becoming stuck at that value for a long period. We do note that the BSL target is remarkably insensitive to n and there appears to be quite a large range of n values that lead to relatively efficient results (an estimated log SL of between 1 and 3, roughly). These results indicate that it may be easier to select a value of n in BSL compared to selecting ϵ in ABC. Even when there are significant departures in normality of the summary statistic, but where the distribution of the summary statistic remains regular, BSL can still produce reasonable approximations. However, when the distribution of the summary statistic is highly irregular as in the example given in Appendix I of the supplementary materials, the output of BSL cannot be trusted, while ABC represents a robust alternative in such cases.

We find that even though uBSL has an estimated log SL with infinite variance, in practice it gives an efficiency that appears similar to BSL. Further, we find in the examples that BSL and uBSL provide similar posterior approximations. Despite the fact that uBSL provides some additional theoretical support in terms of the sensitivity of the posterior approximation to n, we find that the standard BSL posterior is remarkably insensitive to n. Given this, it may be that the standard BSL method is adopted more often in future applications as it is simpler to implement.

In a serial computing environment, some theoretical results suggest that BSL should be more computationally efficient than ABC when the dimension of the summary statistic is greater than 2. ABC is known to suffer from the curse of dimensionality, and this may also be a concern in BSL with the multivariate normal approximation deteriorating in higher dimensions. However, the theoretical results in favor of BSL are supported by empirical results that demonstrate that BSL becomes increasingly efficient relative to ABC as the dimension of the summary statistic increases. Since the optimal value of n in BSL is inherently greater than 1, BSL may benefit more from parallel computing than ABC, where the optimal number of replicated simulations is $n = 1$ in a serial computing environment (Bornn et al. 2017). In the cell biology application, we found an order of magnitude improvement to the computing time when using a computer node with 16 cores.

Meeds and Welling (2014) developed an approach that adaptively chooses the value of n at each iteration to keep the probability of making an incorrect accept/reject decision below a chosen level. When one parameter configuration is clearly preferred over another, only a small value of n is required. Given the insensitivity of the target distribution to n, such an approach may be useful for further improving the efficiency of the BSL approaches.

One aspect of the MCMC BSL approaches we have not investigated is the convergence properties. We found that when the chain is initialized in a negligible point of the posterior support that it can become stuck there for long periods. In these parts of the space, the SL is estimated with very high variability. Lee and Łatuszyński (2014) showed that the ABC MCMC kernel that we use in this article is not geometrically ergodic. It would be interesting to explore the convergence properties of the MCMC BSL approaches. Some studies have investigated the asymptotic properties of various ABC methods (e.g., Li and Fearnhead 2016; Frazier et al. 2016). Further research on the asymptotic properties of BSL would be of interest.

We have done some initial investigations on the performance of BSL when the model is misspecified. Specifically, when the model is unable to recover the observed statistic, $\boldsymbol{s}_y$. We found that a larger value of n than what might be expected for a given d is required to achieve a reasonable acceptance probability in MCMC (u)BSL and that the BSL methods are not necessarily robust to such misspecifications. The reason for the poor efficiency is that $\boldsymbol{s}_y$ is always in the tails of the SL and is thus harder to estimate. It is possible that ABC may be more robust and efficient in such settings, but this requires further investigation. However, these scenarios would typically motivate further model development.

Higher variability in the tails of the SL is a general issue affecting convergence and causing the algorithm to become stuck, especially in cases of model misspecification and highly irregular summary statistics. To help overcome these issues, a parametric auxiliary model other than the multivariate normal can be used, which is equivalent to working in the general psBIL framework. One possible avenue for further research is to investigate the use of a multivariate-t auxiliary model for the summary statistic to improve the robustness of the method. However, the multivariate-t distribution does not have a convenient closed-form expression for its parameter estimates. Alternatively, different marginal distributions of the summary other than normal could be handled using a Gaussian copula model. It would be possible in this framework to consider semiparametric models where the marginals are fitted using kernel density estimates and the dependence between marginals modeled with a Gaussian copula.

The major issue with estimating the SL precisely is in estimating the covariance matrix of the summary statistic accurately. The sample covariance matrix used in this article is an unbiased estimator but it is well known that there are biased but lower variance estimators particularly in the presence of small samples (see Ledoit and Wolf 2004; Friedman, Hastie, and Tibshirani 2008). We plan to investigate such approaches in future research, which may lead to BSL methods that require fewer model simulations.

Following MCMC ABC, a series of sequential Monte Carlo (SMC) ABC approaches (e.g., Sisson, Fan, and Tanaka 2007) have been developed that appear to increase efficiency. The major advantages of the SMC approach over MCMC ABC is that it is straightforward to adapt the parameter proposal in this framework, a population of particles prevents the algorithm from getting stuck and many of the algorithms have natural stopping rules so that ϵ does not need to be explicitly chosen (e.g., Vo et al. 2015a). Further, SMC can facilitate fully Bayesian model comparisons more easily than in MCMC (see, e.g., Drovandi and McCutchan 2016). Everitt et al. (2017) used BSL within an SMC framework to perform model comparisons in models with intractable normalizing constants. There is scope to extend this algorithm, for example, to develop an exact-approximate SMC BSL algorithm and to adaptively select the value of n as the sequence of targets is traversed.

Overall the BSL approaches appear to be useful methods for approximating $p(\boldsymbol{\theta}|\boldsymbol{s}_y)$. The method requires less tuning than ABC, is more computationally efficient than ABC in challenging scenarios, shows some robustness to the normality assumption, and is embarrassingly parallelizable. The clear drawback of the method is the normality assumption, which will be increasingly violated with an increasing dimension of the summary statistic. Although our article suggests that BSL remains an interesting avenue to investigate, further research is required on the theoretical properties of the method and to determine more precisely under what scenarios it may be preferable to ABC.

Supplementary Materials

Additional information to supplement the main article is available in the following files found online.

Appendices: Contains all appendices to the main document. (Appendices.pdf, PDF portable document format)

Code: Contains all of the code required to perform the described methods on the Ricker example from Section 4.2. (Code.zip, compressed (zipped) folder)

Acknowledgments

LFP and CCD are grateful to Mat Simpson for assistance with the cell biology example. CCD thanks NUS and the University of Warwick for supporting a visit where discussions on this research took place. LFP was supported by an Australian Postgraduate Award. CCD was supported by an Australian Research Council's Discovery Early Career Researcher Award funding scheme DE160100741. DJN was supported by a Singapore Ministry of Education Academic Research Fund Tier 2 grant (R-155-000-143-112).

ORCID

L. F. Price http://orcid.org/0000-0002-5646-2963
C. C. Drovandi http://orcid.org/0000-0001-9222-8763
A. Lee http://orcid.org/0000-0001-7765-0616

References

Andrieu, C., Doucet, A., and Holenstein, R. (2010), "Particle Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society*, Series B, 72, 269–342. [2,6]

Andrieu, C., and Roberts, G. O. (2009), "The Pseudo-Marginal Approach for Efficient Monte Carlo Computations," *The Annals of Statistics*, 37, 697–725. [1,2,3,4]

Beaumont, M. A. (2003), "Estimation of Population Growth or Decline in Genetically Monitored Populations," *Genetics*, 164, 1139–1160. [3]

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002), "Approximate Bayesian Computation in Population Genetics," *Genetics*, 162, 2025–2035. [5]

Blum, M. G. B. (2010), "Approximate Bayesian Computation: A Non-Parametric Perspective," *Journal of the American Statistical Association*, 105, 1178–1187. [1]

Blum, M. G. B., Nunes, M. A., Prangle, D., and Sisson, S. A. (2013), "A Comparative Review of Dimension Reduction Methods

in Approximate Bayesian Computation," *Statistical Science*, 28, 189–208. [1,4]

Bornn, L., Pillai, N. S., Smith, A., and Woodard, D. (2017), "The Use of a Single Pseudo-Sample in Approximate Bayesian Computation," *Statistics and Computing*, 27, 583–590. [4,9]

Brown, V. L., Drake, J. M., Barton, H. D., Stallknecht, D. E., Brown, J. D., and Rohani, P. (2014), "Neutrality, Cross-Immunity and Subtype Dominance in Avian Influenza Viruses," *PLOS ONE*, 9, 1–10. [2]

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. (1997), "Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments," in *Case Studies in Bayesian Statistics*, C. Gatsonis, J. S. Hodges, R. E. Kass, R. McCulloch, P. Rossi, and N. D. Singpurwalla, eds., New York: Springer, pp. 37–93. [2]

Dale, P. D., Maini, P. K., and Sherratt, J. A. (1994), "Mathematical Modeling of Corneal Epithelial Wound Healing," *Mathematical Biosciences*, 124, 127–147. [7]

Doucet, A., Pitt, M. K., Deligiannidis, G., and Kohn, R. (2015), "Efficient Implementation of Markov Chain Monte Carlo When Using an Unbiased Likelihood Estimator," *Biometrika*, 102, 295–313. [4,9]

Drovandi, C. C., and McCutchan, R. A. (2016), "Alive SMC^2: Bayesian Model Selection for Low-Count Time Series Models with Intractable Likelihoods," *Biometrics*, 72, 344–353. [10]

Drovandi, C. C., Pettitt, A. N., and Lee, A. (2015), "Bayesian Indirect Inference Using a Parametric Auxiliary Model," *Statistical Science*, 30, 72–95. [1,2,3]

Everitt, R. G., Johansen, A. M., Rowing, E., and Evdemon-Hogan, M. (2017), "Bayesian Model Comparison with Un-Normalised Likelihoods," *Statistics and Computing*, 27, 403–422. [2,10]

Fasiolo, M., Pya, N., and Wood, S. N. (2016), "A Comparison of Inferential Methods for Highly Non-Linear State Space Models in Ecology and Epidemiology," *Statistical Science*, 31, 96–118. [2,6]

Fasiolo, M., and Wood, S. N. (2017), "Approximate Methods for Dynamic Ecological Models," in *Handbook of Approximate Bayesian Computation*, S. Sisson, L. Fan, and M. Beaumont, eds. [2]

Frazier, D. T., Martin, G. M., Robert, C. P., and Rousseau, J. (2016), "Asymptotic Properties of Approximate Bayesian Computation," *https://arxiv.org/pdf/1607.06903.pdf*. [10]

Friedman, J., Hastie, T., and Tibshirani, R. (2008), "Sparse Inverse Covariance Estimation with the Graphical Lasso," *Biostatistics*, 9, 432–441. [10]

Fronza, M., Heinzmann, B., Hamburger, M., Laufer, S., and Merfort, I. (2009), "Determination of the Wound Healing Effect of Calendula Extracts Using the Scratch Assay with 3T3 Fibroblasts," *Journal of Ethnopharmacology*, 126, 463–467. [7]

Ghurye, S. G., and Olkin, I. (1969), "Unbiased Estimation of Some Multivariate Probability Densities and Related Functions," *The Annals of Mathematical Statistics*, 40, 1261–1271. [2,3]

Gutmann, M. U., and Corander, J. (2015), "Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models," *Journal of Machine Learning Research*, 17, 1–47. [2]

Hartig, F., Dislich, C., Wiegand, T., and Huth, A. (2014), "Technical Note: Approximate Bayesian Parameterization of a Process-Based Tropical Forest Model," *Biogeosciences*, 11, 1261–1272. [2]

Johnston, S., Simpson, M. J., McElwain, D. L. S., Binder, B. J., and Ross, J. V. (2014), "Interpreting Scratch Assays Using Pair Density Dynamic and Approximate Bayesian Computation," *Open Biology*, 4, 1–11. [7,8]

Jones, D. R. (2001), "A Taxonomy of Global Optimization Methods Based on Response Surfaces," *Journal of Global Optimization*, 21, 345–383. [2]

Ledoit, O., and Wolf, M. (2004), "A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices," *Journal of Multivariate Analysis*, 88, 365–411. [10]

Lee, A., and Łatuszyński, K. (2014), "Variance Bounding and Geometric Ergodicity of Markov Chain Monte Carlo Kernels for Approximate Bayesian Computation," *Biometrika*, 101, 655–671. [10]

Li, W., and Fearnhead, P. (2016), "Improved Convergence of Regression Adjusted Approximate Bayesian Computation," *https://arxiv.org/pdf/1609.07135.pdf*. [10]

Liang, C.-C., Park, A. Y., and Guan, J.-L. (2007), "In vitro Scratch Assay: A Convenient and Inexpensive Method for Analysis of Cell Migration In Vitro," *Nature Protocols*, 2, 329–333. [7]

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012), "Approximate Bayesian Computation Methods," *Statistics and Computing*, 22, 1167–1180. [2]

Meeds, E., and Welling, M. (2014), "GPS-ABC: Gaussian Process Surrogate Approximate Bayesian Computation," in *Proceedings of the Thirtieth Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-14)*, Corvallis, Oregon, AUAI Press, pp. 593–602. [2,9]

Moores, M. T., Drovandi, C. C., Mengersen, K. L., and Robert, C. P. (2015), "Pre-Processing for Approximate Bayesian Computation in Image Analysis," *Statistics and Computing*, 25, 23–33. [2]

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: Convergence Diagnosis and Output Analysis for MCMC," *R News*, 6, 7–11. [5]

Rayner, G. D., and MacGillivray, H. L. (2002), "Numerical Maximum Likelihood Estimation for the g-and-k and Generalized g-and-h Distribution," *Statistics and Computing*, 12, 57–75. [5]

Simpson, M. J., Landman, K. A., and Hughes, B. D. (2010), "Cell Invasion with Proliferation Mechanisms Motivated by Time-Lapse Data," *Physics A: Statistical Mechanics and its Applications*, 389, 3779–3790. [8]

Simpson, M. J., Treloar, K. K., Binder, B. J., Haridas, P., Manton, K. J., Leavesley, D. I., McElwain, D. L. S., and Baker, R. E. (2013), "Quantifying the Roles of Cell Motility and Cell Proliferation in a Circular Barrier Assay," *Journal of the Royal Society Interface*, 10, 1–11. [7,8]

Sisson, S. A., and Fan, Y. (2011), "Likelihood-Free Markov Chain Monte Carlo," in *MCMC Handbook*, S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, eds., Boca Raton, FL: Chapman & Hall, pp. 313–335. [1]

Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007), "Sequential Monte Carlo Without Likelihoods," *Proceedings of the National Academy of Sciences*, 104, 1760–1765. [10]

Swanson, K. R., Bridge, C., Murray, J. D., and Alvord Jr, E. C. (2003), "Virtual and Real Brain Tumor: Using Mathematical Modeling to Quantify Glioma Growth and Invasion," *Journal of the Neurological Sciences*, 216, 1–10. [7]

Treloar, K. K., and Simpson, M. J. (2013), "Sensitivity of Edge Detection Methods for Quantifying Cell Migration Assays," *PLoS ONE*, 8, e67389. [7]

Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Pettet, G. J. (2015a), "Melanoma Cell Colony Expansion Parameters Revealed by Approximate Bayesian Computation," *PLOS Computational Biology*, 11, e1004635. [10]

Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Simpson, M. J. (2015b), "Quantifying Uncertainty in Parameter Estimates for Stochastic Models of Collective Cell Spreading Using Approximate Bayesian Computation," *Mathematical Biosciences*, 263, 133–142. [7]

Wilkinson, R. (2014), "Accelerating ABC Methods Using Gaussian Processes," *Journal of Machine Learning Research*, 33, 1015–1023. [2]

Wood, S. N. (2010), "Statistical Inference for Noisy Nonlinear Ecological Dynamic Systems," *Nature*, 466, 1102–1107. [1,2,3,4,6]

Zahm, J.-M., Kaplan, H., Herard, A.-L., Doriot, F., Pierrot, D., Somelette, P., and Puchelle, E. (1997), "Cell Migration and Proliferation During the In Vitro Wound Repair of the Respiratory Epithelium," *Cell Motility and the Cytoskeleton*, 37, 33–43. [7]