JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2022, VOL. 31, NO. 2, 512–528
https://doi.org/10.1080/10618600.2021.2000419

Taylor & Francis
Taylor & Francis Group

Rapid Bayesian Inference for Expensive Stochastic Models

David J. Warne [a], Ruth E. Baker[b], and Matthew J. Simpson[a]

[a]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia; [b]Mathematical Institute, University of Oxford, Oxford, UK

ABSTRACT

Almost all fields of science rely upon statistical inference to estimate unknown parameters in theoretical and computational models. While the performance of modern computer hardware continues to grow, the computational requirements for the simulation of models are growing even faster. This is largely due to the increase in model complexity, often including stochastic dynamics, that is necessary to describe and characterize phenomena observed using modern, high resolution, experimental techniques. Such models are rarely analytically tractable, meaning that extremely large numbers of stochastic simulations are required for parameter inference. In such cases, parameter inference can be practically impossible. In this work, we present new computational Bayesian techniques that accelerate inference for expensive stochastic models by using computationally inexpensive approximations to inform feasible regions in parameter space, and through learning transforms that adjust the biased approximate inferences to closer represent the correct inferences under the expensive stochastic model. Using topical examples from ecology and cell biology, we demonstrate a speed improvement of an order of magnitude without any loss in accuracy. This represents a substantial improvement over current state-of-the-art methods for Bayesian computations when appropriate model approximations are available. Supplementary files for this article are available online.

ARTICLE HISTORY
Received January 2020
Revised July 2021

KEYWORDS
Approximate Bayesian computation; Bayesian inference; Continuum-limit approximation; Parameter estimation; Sequential Monte Carlo

1. Introduction

Modern experimental techniques allow us to observe the natural world in unprecedented detail and resolution (Chen and Zhang 2014). Advances in machine learning and artificial intelligence provide many new techniques for pattern recognition and prediction, however, in almost all scientific inquiry there is a need for detailed mathematical models to provide mechanistic insight into the phenomena observed (Coveney, Dougherty, and Highfield 2016; Baker et al. 2018). This is particularly true in the biological and ecological sciences, where detailed stochastic models are routinely applied to develop and validate theory as well as interpret and analyze data (Wilkinson 2009; Black and McKane 2012; Drawert et al. 2017).

Two distinct computational challenges arise when stochastic models are considered, they are: (i) the *forwards problem*; and (ii) the *inverse problem*, sometimes called the *backwards problem* (Warne, Baker, and Simpson 2019a). While the computational generation of a single sample path, that is the *forwards problem*, may be feasible, generating hundreds or thousands or more such sample paths may be required to gain insight into the range of possible model predictions and to conduct parameter sensitivity analysis (Gunawan et al. 2005; Marino et al. 2008; Lester, Yates, and Baker 2017). The problem is further compounded if the models must be calibrated using experimental data, that is the *inverse problem* of parameter estimation, since millions of sample paths may be necessary.

In many cases, the forwards problem can be sufficiently computationally expensive to render both parameter sensitivity analysis and the inverse problem completely intractable, despite recent advances in computational inference (Sisson, Fan, and Beaumont 2018). This has prompted recent interest in the use of mathematical approximations to circumvent the computational burden, both in the context of the forwards and inverse problems. For example, linear approximations are applied to the forwards problem of chemical reaction networks with bimolecular and higher-order reactions (Cao and Grima 2018), and various approximations, including surrogate models (Rynn et al. 2019; Bon, Lee, and Drovandi 2020), emulators (Buzbas and Rosenberg 2015) and transport maps (Parno and Marzouk 2018), are applied to inverse problems with expensive forwards models, for example, in the study of climate science (Holden et al. 2018). Furthermore, a number of developments, such as multilevel Monte Carlo methods (Giles 2015), have demonstrated that families of approximations can be combined to improve computational performance without sacrificing accuracy.

In the recent years, the Bayesian approach to the inverse problem of model calibration and parameter inference has been particularly successful in many fields of science including, astronomy (EHT Collaboration et al. 2019), anthropology and archaeology (King et al. 2014; Malaspinas et al. 2016), paleontology and evolution (Tavaré et al. 1997; Pritchard et al. 1999; O'Dea et al. 2016), epidemiology (Liu et al. 2018; Warne et al. 2020b), biology (Guindani et al. 2014; Vo et al. 2015; Woods

CONTACT David J. Warne *david.warne@qut.edu.au* School of Mathematical Sciences, Queensland University of Technology - QUT, Level 6 O Block, Gardens Point Campus, Brisbane 4001, Australia.

Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JCGS*.

and Barnes 2016; Lawson et al. 2018), and ecology (Ellison 2004; Stumpf 2014). For complex stochastic models, parameterized by $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, computing the likelihood of observing data $\mathcal{D} \in \mathbb{D}$ is almost always impossible (Browning et al. 2018; Vankov, Guindani, and Ensor 2019). Thus, approximate Bayesian computation (ABC) methods (Sisson, Fan, and Beaumont 2018) are essential. ABC methods replace likelihood evaluation with an approximation based on stochastic simulations of the proposed model, this is captured directly in *ABC rejection sampling* (Tavaré et al. 1997; Pritchard et al. 1999; Beaumont, Zhang, and Balding 2002) (Section 2) where samples are generated from an approximate posterior using stochastic simulations of the forwards problem as a replacement for the likelihood.

Unfortunately, ABC rejection sampling can be computationally expensive or even completely prohibitive, especially for high-dimensional parameter spaces, since a very large number of stochastic simulations are required to generate enough samples from the approximate Bayesian posterior distribution (Sisson, Fan, and Beaumont 2018; Warne, Baker, and Simpson 2020a). This is further compounded when the forwards problem is computationally expensive. In contrast, an appropriately chosen approximate model may yield a tractable likelihood that removes the need for ABC methods (Browning, Haridas, and Simpson 2019; Warne, Baker, and Simpson 2017, 2019b). This highlights a key advantage of such approximations because no ABC sampling is required. However, approximations can perform poorly in terms of their predictive capability, and inference based on such models will always be biased, with the extent of the bias dependent on the level of accuracy.

We consider ABC-based inference algorithms for the challenging problem of parameter inference for computationally expensive stochastic models when an appropriate approximation is available to inform the search in parameter space. Under our approach, the approximate model need not be quantitatively accurate in terms of the forwards problem, but must qualitatively respond to changes in parameter values in a similar way to the stochastic model. In particular, we extend the sequential Monte Carlo ABC sampler (SMC-ABC) of Sisson, Fan, and Tanaka (2007) (Section 2) to exploit the approximate model in two ways: (i) to generate an intermediate proposal distribution, that we call a *preconditioner*, to improve ABC acceptance rates for the stochastic model; and (ii) to construct a biased ABC posterior, then reduce this bias using a *moment-matching* transform. We describe both methods and then present relevant examples from ecology and cell biology. Example calculations demonstrate that our methods generate ABC posteriors with a significant reduction in the number of required expensive stochastic simulations, leading to as much as a tenfold computational speedup. The methods we demonstrate here enable substantial acceleration of accurate statistical inference for a broad range of applications, since many areas of science utilize model approximations out of necessity despite potential inference inaccuracies.

As a motivating case study for this work, we focus on stochastic models that can replicate many spatiotemporal patterns that naturally arise in biological and ecological systems. Stochastic discrete random walk models (Section 3), henceforth called *discrete models*, can accurately characterize the microscale interactions of individual agents, such as animals, plants, microorganisms, and cells (von Hardenberg et al. 2001; Law, Murrell, and Dieckmann 2003; Taylor and Hastings 2005; Codling, Plank, and Benhamou 2008; Agnew et al. 2014; Vincenot et al. 2016). Mathematical modeling of populations as complex systems of agents can enhance our understanding of real biological and ecological populations with applications in cancer treatment (Böttger et al. 2015), wound healing (Callaghan et al. 2006), wildlife conservation (McLane et al. 2011; DeAngelis and Grimm 2014), and the management of invasive species (Taylor and Hastings 2005; Chkrebtii et al. 2015).

For example, the discrete model formulation can replicate many realistic spatiotemporal patterns observed in cell biology. Figure 1(A),(B) demonstrates typical microscopy images obtained from *in vitro* cell culture assays; ubiquitous and important experimental techniques used in the study of cell motility, cell proliferation and drug design. Various patterns are observed: prostate cancer cells (PC-3 line) tend to be highly motile, and spread uniformly to invade vacant spaces (Figure 1(A)); in contrast breast cancer cells (MBA-MD-231 line) tend be relatively stationary with proliferation events driving the formation of aggregations (Figure 1(B)). These phenomena may be captured using a lattice-based discrete model framework by varying the ratio P_p/P_m where $P_p \in [0, 1]$ and $P_m \in [0, 1]$ are, respectively, the probabilities that an agent attempts to proliferate and attempts to move during a time interval of duration $\tau > 0$ (See Section 3.2). For $P_p/P_m \ll 1$, behavior akin to PC-3 cells is recovered (Figure 1(C)–(F)) (Jin et al. 2016). Setting $P_p/P_m \gg 1$, as in Figure 1(H)–(K), leads to clusters of occupied lattices sites that are similar to the aggregates of MBA-MD-231 cells (Agnew et al. 2014; Simpson et al. 2013).

It is common practice to derive approximate continuum-limit differential equation descriptions of discrete models (Callaghan et al. 2006; Simpson, Landman, and Hughes 2010; Jin et al. 2016) (supplementary material). Such approximations provide a means of performing analysis with significantly reduced computational requirements, since evaluating an exact analytical solution, if available, or otherwise numerically solving a differential equation is typically several orders of magnitude faster than generating a single realization of the discrete model, of which hundreds or thousands may be required for reliable ABC sampling (Browning et al. 2018). However, such approximations are generally only valid within certain parameter regimes, for example here when $P_p/P_m \ll 1$ (Callaghan et al. 2006; Simpson, Landman, and Hughes 2010). Consider Figure 1(G), the population density growth curve from the continuum-limit logistic growth model is superimposed with stochastic data for four realizations of a discrete model with $P_p/P_m \ll 1$ and $P_p/P_m \gg 1$, under initial conditions simulating a proliferation assay, where each lattice site is randomly populated with constant probability, such that there are no macroscopic gradients present at $t = 0$. The continuum-limit logistic growth model is an excellent match for the $P_p/P_m \ll 1$ case (Figure 1(C)–(F)), but severely overestimates the population density when $P_p/P_m \gg 1$ since the mean-field assumptions underpinning the continuum-limit model are violated by the presence of clustering (Figure 1(H)–(K)) (Simpson et al. 2013; Agnew et al. 2014).

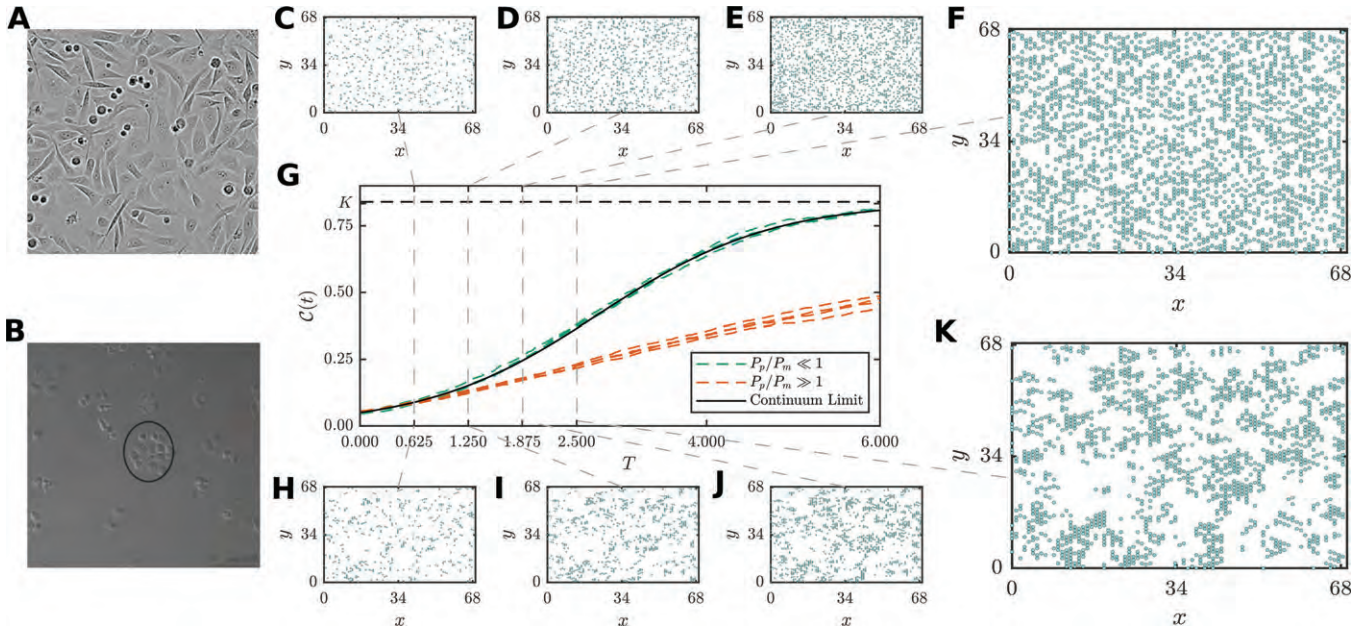

Figure 1. Discrete random walk models can replicate observed spatial patterns in cell culture: (A) PC-3 prostate cancer cells (reprinted from Jin et al. (2017) with permission); and (B) MBA-MD-231 breast cancer cells (reprinted from Simpson et al. (2013) with permission). (C)–(F) Discrete simulations with $P_p/P_m \ll 1$ replicate the uniform distribution of (A) PC-3 cells. (H)–(K) Discrete simulations with $P_p/P_m \gg 1$ replicate spatial clustering of (B) MBA-MD-231 cells. (G) Averaged population density profiles $\mathcal{C}(t)$ for the discrete model with highly motile agents, $P_m = 1$ (dashed green), and near stationary agents, $P_m = 5 \times 10^{-4}$ (dashed orange), compared with the logistic growth continuum limit (solid black), time is nondimensionalized with $T = P_p t/\tau$.

As we demonstrate in Section 3, our methods generate accurate ABC posteriors for inference on the discrete problem for a range of biologically relevant parameter regimes, including those where the continuum-limit approximation is poor. In this respect we demonstrate a novel use of approximations that qualitatively respond to changes in parameters in a similar way to the full exact stochastic model.

2. Methods

In this section, we present details of two new algorithms for the acceleration of ABC inference for expensive stochastic models when an appropriate approximation is available. First, we present essential background in ABC inference and sequential Monte Carlo (SMC) samplers for ABC (Sisson, Fan, and Tanaka 2007; Toni et al. 2009). We then describe our extensions to SMC samplers for ABC and provide numerical examples of our approaches using topical examples from ecology and cell biology.

2.1. Sequential Monte Carlo for ABC

Bayesian analysis techniques are powerful tools for the quantification of uncertainty in parameters, models and predictions (Gelman et al. 2014). Unfortunately, for many stochastic models of practical interest, the likelihood function is intractable. ABC methods replace likelihood evaluation with an approximation based on stochastic simulations of the proposed model, this is captured directly in *ABC rejection sampling* (Tavaré et al. 1997; Pritchard et al. 1999) where $\mathcal{M}$ samples are generated from an approximate posterior, denoted by $p(\boldsymbol{\theta}|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon)$. Here $\mathcal{D}_s \sim s(\mathcal{D}|\boldsymbol{\theta})$ is a data generation process based on simulation of the model, $\rho(\mathcal{D}, \mathcal{D}_s)$ is a discrepancy metric, and ϵ is the discrepancy threshold. The resulting accepted parameter samples are distributed according to $p(\boldsymbol{\theta}|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon) \to p(\boldsymbol{\theta}|\mathcal{D})$ as $\epsilon \to 0$.

The average acceptance probability of a proposed parameter sample $\boldsymbol{\theta}^*$ is $\mathcal{O}(\epsilon^d)$ (Fearnhead and Prangle 2012), where d is the dimensionality of the data space, $\mathbb{D}$. This renders rejection sampling computationally expensive or even completely prohibitive, especially for high-dimensional parameter spaces (Marjoram et al. 2003; Sisson, Fan, and Tanaka 2007). Summary statistics can reduce the data dimensionality, however, they will often incur information loss (Barnes et al. 2012; Blum et al. 2013; Fearnhead and Prangle 2012). However, strategies including regression adjustment and marginal adjustment strategies can improve the accuracy of dimension reductions (Beaumont, Zhang, and Balding 2002; Nott et al. 2014).

In the SMC-ABC method, importance resampling is applied to a sequence of R ABC posteriors with discrepancy thresholds $\epsilon_1 > \cdots > \epsilon_R$, with ϵ_R indicating the target ABC posterior. Given $\mathcal{M}$ weighted samples $\{(\boldsymbol{\theta}_r^i, w_r^i)\}_{i=1}^{\mathcal{M}}$, called particles, initially from the prior $p(\boldsymbol{\theta})$ (with $r = 0$ corresponding to $\epsilon_0 \to \infty$), particles are filtered through each ABC posterior, $p(\boldsymbol{\theta}|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)$ for $r = 1, \ldots, R$, using three main steps in each iteration: (i) generate a new set of $\mathcal{M}$ particles using ABC rejection sampling with discrepancy threshold, ϵ_r, and proposals drawn from $\eta_{r-1}(\boldsymbol{\theta})$ given by

$$\eta_{r-1}(\boldsymbol{\theta}) = \sum_{i=1}^{\mathcal{M}} q_r(\boldsymbol{\theta}|\boldsymbol{\theta}_{r-1}^i) w_{r-1}^i, \tag{1}$$

where $q_r(\boldsymbol{\theta}|\boldsymbol{\theta}^i)$ is a proposal kernel; (ii) compute new particle importance weights; then (iii) resample particles to avoid particle degeneracy. For reference, the SMC-ABC algorithm as initially developed by Sisson, Fan, and Tanaka (2007) and

Toni et al. (2009) is given in Algorithm 1. Here, lines (4)–(8) correspond the rejection sampling (step (i)) with lines (5)–(6) describing how to draw a sample from the proposal $\boldsymbol{\theta}^{**} \sim \eta_{r-1}(\boldsymbol{\theta})$ as defined in Equation (1), line (10) corresponds to the computation of importance weights (step (ii)) using optimal backward kernels (see Sisson, Fan, and Tanaka 2007), and lines (12)–(13) perform the resampling (step (iii)). The number of particles, $\mathcal{M}$, and the number of intermediate distributions, R, influence the accuracy and performance, respectively, of the sampler. Setting $\mathcal{M}$ too small can lead to large estimator variability and particle degeneracy, and setting R too small leads to large divergence between successive distributions that can result in high rejection rates. Note that throughout all algorithms used

Algorithm 1 SMC-ABC

1: Initialize $\boldsymbol{\theta}_0^i \sim p(\boldsymbol{\theta})$ and $w_0^i = 1/\mathcal{M}$, for $i = 1, \ldots, \mathcal{M}$;
2: **for** $r = 1, \ldots, R$ **do**
3: **for** $i = 1, \ldots, \mathcal{M}$ **do**
4: **repeat**
5: Draw $\boldsymbol{\theta}^*$ from $\{\boldsymbol{\theta}_{r-1}^j\}_{j=1}^{\mathcal{M}}$ according to the probability mass function

$$\mathbb{P}(\boldsymbol{\theta}^* = \boldsymbol{\theta}_{r-1}^j) = \frac{w_{r-1}^j}{\sum_{k=1}^{\mathcal{M}} w_{r-1}^k}, \quad \text{for } j = 1, \ldots, \mathcal{M};$$

6: Sample transition kernel, $\boldsymbol{\theta}^{**} \sim q_r(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$;
7: Generate data, $\mathcal{D}_s \sim s(\mathcal{D}|\boldsymbol{\theta}^{**})$;
8: **until** $\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r$
9: Set $\boldsymbol{\theta}_r^i \leftarrow \boldsymbol{\theta}^{**}$;
10: Set $w_r^i \leftarrow p(\boldsymbol{\theta}_r^i) / \left[\sum_{j=1}^{\mathcal{M}} w_{r-1}^j q_r(\boldsymbol{\theta}_r^i|\boldsymbol{\theta}_{r-1}^j)\right]$;
11: **end for**
12: Resample weighted particles, $\{(\boldsymbol{\theta}_r^i, w_r^i)\}_{i=1}^{\mathcal{M}}$, with replacement;
13: Set $w_r^i \leftarrow 1/\mathcal{M}$ for all $i = 1, \ldots, \mathcal{M}$;
14: **end for**

in this manuscript, we assume that the initial set of weighted particles, $\{\boldsymbol{\theta}_0^i, w_0^i\}_{i=1}^{\mathcal{M}}$, are independent, identically distributed samples from the prior, $p(\boldsymbol{\theta})$, and therefore have uniform weight, $w_0^i = 1/\mathcal{M}$, for all $i = 1, \ldots, \mathcal{M}$. However, the methods are general enough to deal with prior distributions that require importance sampling to draw truly weighted particles.

For a fixed choice of R, efficient use of SMC-ABC depends critically on the selection of appropriate proposal kernels and threshold sequences. Following Filippi et al. (2013), we explore the choice of proposal kernel by considering the density $\xi_r(\boldsymbol{\theta}_r|\mathcal{D})$ that represents the process of sampling at the target threshold, ϵ_r, given the weights of the previous threshold, ϵ_{r-1}, extended as a function $w_{r-1}(\boldsymbol{\theta}_{r-1})$ for any $\boldsymbol{\theta}_{r-1} \in \boldsymbol{\Theta}$. This density is given by

$$\xi_r(\boldsymbol{\theta}_r|\mathcal{D}) = \frac{1}{a_{r,r-1}} \int_{\boldsymbol{\Theta}} \int_{\mathbb{D}} \mathbb{1}_{B(\mathcal{D},\epsilon_r)}(\mathcal{D}_s)\, s(\mathcal{D}_s|\boldsymbol{\theta}_r) \times q_r(\boldsymbol{\theta}_r|\boldsymbol{\theta}_{r-1}) w_{r-1}(\boldsymbol{\theta}_{r-1})\, \mathrm{d}\mathcal{D}_s \mathrm{d}\boldsymbol{\theta}_{r-1}, \quad (2)$$

where the data space, $\mathbb{D}$, has dimensionality d, $B(\mathcal{D}, \epsilon_r)$ is a d-dimensional ball centred on the data with radius ϵ_r, $\mathbb{1}_A(x)$ denotes the indicator function with $\mathbb{1}_A(x) = 1$ if $x \in A$, otherwise $\mathbb{1}_A(x) = 0$. The normalization constant, $a_{r,r-1}$, can be interpreted as the average acceptance probability across all particles. We see this by noting that Equation (2) can be reduced to

$$\xi_r(\boldsymbol{\theta}_r|\mathcal{D}) = \frac{\eta_{r-1}(\boldsymbol{\theta}_r)\mathbb{P}(\mathcal{D}_s \in B(\mathcal{D}, \epsilon_r)|\boldsymbol{\theta}_r)}{a_{r,r-1}}. \quad (3)$$

Here, the distribution $\eta_{r-1}(\boldsymbol{\theta}_r)$ represents the equivalent of the proposal mechanism in Equation (1)

$$\eta_{r-1}(\boldsymbol{\theta}_r) = \int_{\boldsymbol{\Theta}} q_r(\boldsymbol{\theta}_r|\boldsymbol{\theta}_{r-1}) w_{r-1}(\boldsymbol{\theta}_{r-1}) \mathrm{d}\boldsymbol{\theta}_{r-1}, \quad (4)$$

and $\mathbb{P}(\mathcal{D}_s \in B(\mathcal{D}, \epsilon_r)|\boldsymbol{\theta}_r)$ is the probability that simulated data is within ϵ_r of the data $\mathcal{D}$ given a parameter value $\boldsymbol{\theta}_r$. Therefore, the normalizing constant is

$$a_{r,r-1} = \mathbb{E}\left[\mathbb{P}(\mathcal{D}_s \in B(\mathcal{D}, \epsilon_r)|\boldsymbol{\theta}_r)\right], \quad (5)$$

that is, $a_{r,r-1}$ is the average acceptance probability.

It is important to note that, Equation (2) is not, in general, equal to the ABC posterior for arbitrary proposal kernels. From a computational perspective, the goal is to choose $\eta_{r-1}(\boldsymbol{\theta}_r)$ to maximize $a_{r,r-1}$. However, this would not necessarily result in a $\xi_r(\boldsymbol{\theta}_r|\mathcal{D})$ that is an accurate approximation to the true target ABC posterior $p(\boldsymbol{\theta}_r|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)$. To achieve this goal, we require $\eta_{r-1}(\boldsymbol{\theta}_r)$ such that the Kullback–Leibler divergence (Kullback and Leibler 1951), $D_{\mathrm{KL}}\left(\xi_r(\cdot|\mathcal{D}); p(\cdot|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)\right)$, is minimized. Beaumont et al. (2009) and Filippi et al. (2013) demonstrated how the latter goal provides insight into how to optimally choose $\eta_{r-1}(\boldsymbol{\theta}_r)$. The key is to note that $D_{\mathrm{KL}}\left(\xi_r(\cdot|\mathcal{D}); p(\cdot|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)\right)$ can be decomposed as follows,

$$D_{\mathrm{KL}}\left(\xi_r(\cdot|\mathcal{D}); p(\cdot|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)\right) = D_{\mathrm{KL}}\left(\eta_{r-1}(\cdot); p(\cdot|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)\right) - E(\boldsymbol{\theta}_r) + \log_e a_{r,r-1}, \quad (6)$$

where $E(\boldsymbol{\theta}_r) = \mathbb{E}\left[\log_e\left(\int_{\mathbb{D}} \mathbb{1}_{B(\mathcal{D},\epsilon_r)}(\mathcal{D}_s)\, s(\mathcal{D}_s|\boldsymbol{\theta}_r)\, \mathrm{d}\mathcal{D}_s\right)\right]$ is independent of $\eta_{r-1}(\boldsymbol{\theta}_r)$. By rearranging Equation (6), we obtain

$$D_{\mathrm{KL}}\left(\eta_{r-1}(\cdot); p(\cdot|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)\right) = D_{\mathrm{KL}}\left(\xi_r(\cdot|\mathcal{D}); p(\cdot|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)\right) + E(\boldsymbol{\theta}_r) - \log_e a_{r,r-1}.$$

That is, minimizing $D_{\mathrm{KL}}\left(\eta_{r-1}(\cdot); p(\cdot|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)\right)$ is equivalent to minimizing $D_{\mathrm{KL}}\left(\xi_r(\cdot|\mathcal{D}); p(\cdot|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)\right)$ and maximizing $a_{r,r-1}$ simultaneously. Therefore, any proposal mechanism that is closer, in the Kullback-Leibler sense, to $p(\cdot|\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r)$ is more efficient.

For some families of proposal kernels Equation (6) can be minimized analytically. We apply the optimal adaptive scheme of Beaumont et al. (2009) and Filippi et al. (2013) for multivariate Gaussian proposals. That is, we set

$$q_r(\boldsymbol{\theta}_r|\boldsymbol{\theta}_{r-1}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \times \exp\left(-(\boldsymbol{\theta}_r - \boldsymbol{\theta}_{r-1})^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{\theta}_r - \boldsymbol{\theta}_{r-1})/2\right),$$

where n is the dimensionality of parameter space, $\boldsymbol{\Theta}$, and

$$\Sigma = \frac{2}{\mathcal{M}-1} \sum_{i=1}^{\mathcal{M}} (\boldsymbol{\theta}_{r-1}^i - \mu_{r-1})(\boldsymbol{\theta}_{r-1}^i - \mu_{r-1})^{\mathrm{T}} \quad \text{with}$$

$$\mu_{r-1} = \frac{1}{\mathcal{M}} \sum_{i=1}^{\mathcal{M}} \boldsymbol{\theta}_{r-1}^i.$$

2.2. Preconditioning SMC-ABC

Consider a fixed sequence of ABC posteriors for the stochastic model inference problem, $\{p(\boldsymbol{\theta}|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)\}_{r=1}^R$. We want to apply SMC-ABC (Algorithm 1) to efficiently sample from this sequence with adaptive proposal kernels, $\{q_r(\boldsymbol{\theta}^*|\boldsymbol{\theta})\}_{r=1}^R$ (Beaumont et al. 2009; Filippi et al. 2013). Our method exploits an approximate model to further improve the average acceptance probability.

2.2.1. Algorithm Development

Say we have a set of weighted particles that represent the ABC posterior at threshold ϵ_{r-1} using the stochastic model, that is, $\{(\boldsymbol{\theta}_{r-1}^i, w_{r-1}^i)\}_{i=1}^{\mathcal{M}} \approx p(\boldsymbol{\theta}_{r-1}|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_{r-1})$. Now, consider applying the next importance resampling step using an approximate data-generation step, $\tilde{\mathcal{D}}_s \sim \tilde{s}(\tilde{\mathcal{D}}_s|\boldsymbol{\theta})$, where $\tilde{s}(\tilde{\mathcal{D}}_s|\boldsymbol{\theta})$ is the simulation process of an approximate model[1]. Furthermore, assume the computational cost of simulating the approximate model, $\text{Cost}(\tilde{\mathcal{D}}_s)$, is significantly less than the computational cost of the exact model, $\text{Cost}(\mathcal{D}_s)$, that is, $\text{Cost}(\tilde{\mathcal{D}}_s)/\text{Cost}(\mathcal{D}_s) \ll 1$. The result will be a new set of particles that represent the ABC posterior at threshold ϵ_r using this approximate model, denoted $\{(\tilde{\boldsymbol{\theta}}_r^i, \tilde{w}_r^i)\}_{i=1}^{\mathcal{M}} \approx \tilde{p}(\tilde{\boldsymbol{\theta}}_r|\rho(\mathcal{D},\tilde{\mathcal{D}}_s)\leq\epsilon_r)$. As noted in the examples in Section 1, approximate models are not always valid. This implies that $\tilde{p}(\tilde{\boldsymbol{\theta}}_r|\rho(\mathcal{D}\tilde{\mathcal{D}}_s)\leq\epsilon_r)$ is always biased and will not in general converge to $p(\boldsymbol{\theta}|\mathcal{D})$ as $\epsilon_r \to 0$. However, since $\text{Cost}(\tilde{\mathcal{D}}_s)/\text{Cost}(\mathcal{D}_s) \ll 1$, it is computationally inexpensive to compute the distribution

$$\tilde{\eta}_r(\boldsymbol{\theta}_r) = \sum_{j=1}^{\mathcal{M}} \tilde{q}_r(\boldsymbol{\theta}_r|\tilde{\boldsymbol{\theta}}_r^j)\tilde{w}_r^j, \tag{7}$$

in comparison to computing $\eta_{r-1}(\boldsymbol{\theta}_r)$ (Equation (1)). In Equation (7), the proposal kernel $\tilde{q}_r(\boldsymbol{\theta}_r|\tilde{\boldsymbol{\theta}}_r^j)$ is possibly distinct from the $q_r(\boldsymbol{\theta}_r|\boldsymbol{\theta}_{r-1}^j)$ used in $\eta_{r-1}(\boldsymbol{\theta}_r)$ (Equation (1)). To improve the efficiency of the sampling process we require

$$\begin{aligned} &D_{\text{KL}}(\eta_{r-1}(\cdot); p(\cdot|\rho(\mathcal{D}\mathcal{D}_s)\leq\epsilon_r)) \\ &\quad > D_{\text{KL}}(\tilde{\eta}_r(\cdot); p(\cdot|\rho(\mathcal{D},\tilde{\mathcal{D}}_s)\leq\epsilon_r)), \end{aligned} \tag{8}$$

for $\tilde{\eta}_r(\boldsymbol{\theta}_r)$ (Equation (7)) to be more efficient as a proposal mechanism compared with $\eta_{r-1}(\boldsymbol{\theta}_r)$ (Equation (1)). Provided the condition $\text{Cost}(\tilde{\mathcal{D}}_s)/\text{Cost}(\mathcal{D}_s) \ll 1$ holds, any improvements in sampling efficiency will translate directly into computational performance improvements. That is, it does not matter that $\tilde{p}(\tilde{\boldsymbol{\theta}}_r|\rho(\mathcal{D},\tilde{\mathcal{D}}_s)\leq\epsilon_r)$ is biased, it just needs to be less biased than $p(\boldsymbol{\theta}_{r-1}|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_{r-1})$ and computationally inexpensive.

This idea yields an intuitive new algorithm for SMC-ABC that proceeds through the sequential sampling of $\{p(\boldsymbol{\theta}|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)\}_{r=1}^R$ by applying two resampling steps for each iteration. The first moves the particles from acceptance threshold ϵ_{r-1} to ϵ_r using the computationally inexpensive approximate model, and the second corrects for the bias between $\tilde{p}(\tilde{\boldsymbol{\theta}}_r|\rho(\mathcal{D},\tilde{\mathcal{D}}_s)\leq\epsilon_r)$ and $p(\boldsymbol{\theta}_r|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)$ using the expensive stochastic model, but at an improved acceptance rate. Since the intermediate distribution acts on the proposal mechanism to accelerate the convergence time of SMC-ABC, we denote the sequence $\{\tilde{p}(\tilde{\boldsymbol{\theta}}_r|\rho(\mathcal{D},\tilde{\mathcal{D}}_s)\leq\epsilon_r)\}_{r=1}^R$ as the preconditioner distribution sequence. The algorithm, called *preconditioned SMC-ABC* (PC-SMC-ABC), is given in Algorithm 2. We note that similar notions of preconditioning with approximation informed proposals have been applied in the context of Markov chain Monte Carlo samplers (Parno and Marzouk 2018). However, to the best of our knowledge, our approach represents the first application of preconditioning ideas to SMC-ABC. One particular advantage of the PC-SMC-ABC method is that it is asymptotically unbiased (supplementary material). Effectively, one can consider PC-SMC-ABC as standard SMC-ABC method with a specialized proposal mechanism based on the preconditioner distribution. This means that PC-SMC-ABC is completely general, as discussed in Section 4, and is independent of the specific stochastic models that we consider here. This property of unbiasedness holds even for cases where the approximate model is a poor approximation of the forward

Algorithm 2 Preconditioned SMC-ABC

Initialize $\boldsymbol{\theta}_0^i \sim p(\boldsymbol{\theta})$ and $w_0^i = 1/\mathcal{M}$, for $i = 1,\ldots,\mathcal{M}$;
for $r = 1,\ldots,R$ **do**
  for $i = 1,\ldots,\mathcal{M}$ **do**
    repeat
      Draw $\boldsymbol{\theta}^*$ from $\{\boldsymbol{\theta}_{r-1}^j\}_{j=1}^{\mathcal{M}}$ according to the probability mass function

$$\mathbb{P}(\boldsymbol{\theta}^* = \boldsymbol{\theta}_{r-1}^j) = \frac{w_{r-1}^j}{\sum_{k=1}^{\mathcal{M}} w_{r-1}^k}, \quad \text{for } j = 1,\ldots,\mathcal{M};$$

      Sample transition kernel, $\tilde{\boldsymbol{\theta}}^{**} \sim q_r(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}^*)$;
      Generate data, $\tilde{\mathcal{D}}_s \sim \tilde{s}(\mathcal{D}|\tilde{\boldsymbol{\theta}}^{**})$;
    until $\rho(\mathcal{D},\tilde{\mathcal{D}}_s)\leq\epsilon_r$
    Set $\tilde{\boldsymbol{\theta}}_r^i \leftarrow \tilde{\boldsymbol{\theta}}^{**}$;
    Set $\tilde{w}_r^i \leftarrow p(\tilde{\boldsymbol{\theta}}_r^i)/\left[\sum_{j=1}^{\mathcal{M}} w_{r-1}^j q_r(\tilde{\boldsymbol{\theta}}_r^i|\boldsymbol{\theta}_{r-1}^j)\right]$;
  end for
  for $i = 1,\ldots,\mathcal{M}$ **do**
    repeat
      Draw $\tilde{\boldsymbol{\theta}}^*$ from $\{\tilde{\boldsymbol{\theta}}_{r-1}^j\}_{j=1}^{\mathcal{M}}$ according to the probability mass function

$$\mathbb{P}(\tilde{\boldsymbol{\theta}}^* = \tilde{\boldsymbol{\theta}}_{r-1}^j) = \frac{\tilde{w}_{r-1}^j}{\sum_{k=1}^{\mathcal{M}} \tilde{w}_{r-1}^k}, \quad \text{for } j = 1,\ldots,\mathcal{M};$$

      Sample transition kernel, $\boldsymbol{\theta}^{**} \sim \tilde{q}_r(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}}^*)$;
      Generate data, $\mathcal{D}_s \sim s(\mathcal{D}|\boldsymbol{\theta}^{**})$;
    until $\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r$
    Set $\boldsymbol{\theta}_r^i \leftarrow \boldsymbol{\theta}^{**}$;
    Set $w_r^i \leftarrow p(\boldsymbol{\theta}_r^i)/\left[\sum_{j=1}^{\mathcal{M}} \tilde{w}_r^j \tilde{q}_r(\boldsymbol{\theta}_r^i|\tilde{\boldsymbol{\theta}}_r^j)\right]$;
  end for
  Resample weighted particles, $\{(\boldsymbol{\theta}_r^i, w_r^i)\}_{i=1}^{\mathcal{M}}$, with replacement;
  Set $w_r^i \leftarrow 1/\mathcal{M}$ for all $i = 1,\ldots,\mathcal{M}$;
end for

[1] Throughout, the overbar tilde notation, for example, $\tilde{x}$, is used to refer to the ABC entities related to the approximate model, whereas quantities without the overbar tilde notation, for example, x, are used to refer fo the ABC entities related to the exact model.

dynamics of the model. However, the closer that $\tilde{\eta}_r(\boldsymbol{\theta}_r)$ is to $p(\boldsymbol{\theta}_r|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)$ the better the performance improvement will be, as we demonstrate in Section 3.

2.3. Moment-Matching SMC-ABC

The PC-SMC-ABC method is a promising modification to SMC-ABC that can accelerate inference for expensive stochastic models without introducing bias. However, other approaches can be used to obtain further computational improvements. Here, we consider an alternate approach to using approximate models that aims to get the most out of a small sample of expensive stochastic simulations. Unlike PC-SMC-ABC, this method is generally biased, but it has the advantage of yielding a small and fixed computational budget. Specifically, we define a parameter $\alpha\in[0,1]$, such that $1/\alpha$ is the target computational speedup, for example, $\alpha=1/10$ should result in approximately 10 times speedup. We apply the SMC-ABC method using $\tilde{\mathcal{M}}=\lfloor(1-\alpha)\mathcal{M}\rfloor$ particles based on the approximate model, and then use $\hat{\mathcal{M}}=\lceil\alpha\mathcal{M}\rceil$ particles based on the stochastic model to construct a hybrid population of $\mathcal{M}=\hat{\mathcal{M}}+\tilde{\mathcal{M}}$ particles that will represent the final inference on the stochastic model. The key idea is that we use the $\hat{\mathcal{M}}$ particles of the expensive stochastic model to inform a transformation on the $\tilde{\mathcal{M}}$ particles of the approximation such that they the emulate particles of expensive stochastic model. Here, $\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ are, respectively, the floor and ceiling functions.

2.3.1. Algorithm Development

Assume that we have applied SMC-ABC to sequentially sample $\tilde{\mathcal{M}}$ particles through the ABC posteriors from the approximate model, $\{\tilde{p}(\tilde{\boldsymbol{\theta}}_r|\rho(\mathcal{D},\tilde{\mathcal{D}}_s)\leq\epsilon_r)\}_{r=1}^R$, with $\epsilon_R=\epsilon$. For the sake of the derivation, say that for all $r\in[1,R]$ we have available the mean vector, $\boldsymbol{\mu}_r$, and the covariance matrix, $\boldsymbol{\Sigma}_r$, of the ABC posterior $p(\boldsymbol{\theta}_r|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)$ under the stochastic model. In this case, we use particles $\tilde{\boldsymbol{\theta}}_r^1,\ldots,\tilde{\boldsymbol{\theta}}_r^{\tilde{\mathcal{M}}}\sim\tilde{p}(\tilde{\boldsymbol{\theta}}_r|\rho(\mathcal{D},\tilde{\mathcal{D}}_s)\leq\epsilon_r)$ to emulate particles $\boldsymbol{\theta}_r^1,\ldots,\boldsymbol{\theta}_r^{\hat{\mathcal{M}}}\sim p(\boldsymbol{\theta}_r|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)$ by using the moment matching transform (Lei and Bickel 2011; Sun, Feng, and Saenko 2016)

$$\boldsymbol{\theta}_r^i=\mathbf{L}_r\left[\tilde{\mathbf{L}}_r^{-1}\left(\tilde{\boldsymbol{\theta}}_r^i-\tilde{\boldsymbol{\mu}}_r\right)\right]+\boldsymbol{\mu}_r,\quad i=1,2,\ldots,\tilde{\mathcal{M}},\tag{9}$$

where $\tilde{\boldsymbol{\mu}}_r$ and $\tilde{\boldsymbol{\Sigma}}_r$ are the empirical mean vector and convariance matrix of particles $\tilde{\boldsymbol{\theta}}_r^1,\ldots,\tilde{\boldsymbol{\theta}}_r^{\tilde{\mathcal{M}}}$, $\mathbf{L}_r$ and $\tilde{\mathbf{L}}_r$ are lower triangular matrices obtained through the Cholesky factorization (Press et al. 1997) of $\boldsymbol{\Sigma}_r$ and $\tilde{\boldsymbol{\Sigma}}_r$, respectively, and $\tilde{\mathbf{L}}_r^{-1}$ is the matrix inverse of $\tilde{\mathbf{L}}_r$. This transform will produce a collection of particles that has a sample mean vector of $\boldsymbol{\mu}_r$ and covariance matrix $\boldsymbol{\Sigma}_r$. That is, the transformed sample matches the ABC posterior under the stochastic model up to the first two moments. In Section 3, we demonstrate that matching two moments is sufficient for the problems we investigate here, however, in principle we could extend this matching to higher order moments if required. For discussion on the advantages and disadvantages of matching higher moments, see Section 4.

In practice, it would be rare that $\boldsymbol{\mu}_r$ and $\boldsymbol{\Sigma}_r$ are known. If $p(\boldsymbol{\theta}_{r-1}|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_{r-1})$ is available, then we can use importance resampling to obtain $\hat{\mathcal{M}}$ particles, $\boldsymbol{\theta}_r^1,\ldots,\boldsymbol{\theta}_r^{\hat{\mathcal{M}}}$, from $p(\boldsymbol{\theta}_r|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)$, that is, we perform a step from SMC-ABC using the expensive stochastic model. We can then use the unbiased estimators

$$\hat{\boldsymbol{\mu}}_r=\frac{1}{\hat{\mathcal{M}}}\sum_{i=1}^{\hat{\mathcal{M}}}\boldsymbol{\theta}_r^i\quad\text{and}\quad\hat{\boldsymbol{\Sigma}}_r=\frac{1}{\hat{\mathcal{M}}-1}\sum_{i=1}^{\hat{\mathcal{M}}}(\boldsymbol{\theta}_r^i-\hat{\boldsymbol{\mu}}_r)(\boldsymbol{\theta}_r^i-\hat{\boldsymbol{\mu}}_r)^{\mathrm{T}},\tag{10}$$

to obtain estimates of $\boldsymbol{\mu}_r$ and $\boldsymbol{\Sigma}_r$. Substituting Equation (10) into Equation (9) gives an approximate transform

$$\hat{\boldsymbol{\theta}}_r^i=\hat{\mathbf{L}}_r\left[\tilde{\mathbf{L}}_r^{-1}\left(\tilde{\boldsymbol{\theta}}_r^i-\tilde{\boldsymbol{\mu}}_r\right)\right]+\hat{\boldsymbol{\mu}}_r,\quad i=1,2,\ldots,\tilde{\mathcal{M}},\tag{11}$$

where $\hat{\boldsymbol{\Sigma}}_r=\hat{\mathbf{L}}_r\hat{\mathbf{L}}_r^{\mathrm{T}}$. This enables us to construct an estimate of $p(\boldsymbol{\theta}_r|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)$ by applying the moment-matching transform (Equation (11)) to the particles $\tilde{\boldsymbol{\theta}}_r^1,\ldots,\tilde{\boldsymbol{\theta}}_r^{\tilde{\mathcal{M}}}$ then pooling the transformed particles $\hat{\boldsymbol{\theta}}_r^1,\ldots,\hat{\boldsymbol{\theta}}_r^{\tilde{\mathcal{M}}}$ with the particles $\boldsymbol{\theta}_r^1,\ldots,\boldsymbol{\theta}_r^{\hat{\mathcal{M}}}$ that were used in the estimates $\hat{\boldsymbol{\mu}}_r$ and $\hat{\boldsymbol{\Sigma}}_r$. The goal of the approximate transform application is for the transforms particles $\hat{\boldsymbol{\theta}}_r^1,\ldots,\hat{\boldsymbol{\theta}}_r^{\tilde{\mathcal{M}}}$ to be more accurate in higher moments due, despite only matching the first two moments (see Section 3.6 for numerical justification of this property for some specific examples). This results in an approximation of $p(\boldsymbol{\theta}_r|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)$ using a set of $\mathcal{M}$ particles $\boldsymbol{\theta}_r^1,\ldots,\boldsymbol{\theta}_r^{\mathcal{M}}$ with $\boldsymbol{\theta}_r^{i+\hat{\mathcal{M}}}=\hat{\boldsymbol{\theta}}_r^i$ where $1\leq i\leq\tilde{\mathcal{M}}$.

This leads to our *moment-matching* SMC-ABC (MM-SMC-ABC) method. First, SMC-ABC inference is applied using the approximate model with $\tilde{\mathcal{M}}$ particles. Then, given $\mathcal{M}$ samples from the prior, $p(\boldsymbol{\theta})$, we can sequentially approximate $\{p(\boldsymbol{\theta}_r|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)\}_{r=1}^R$. At each iteration the following steps are performed: (i) generate a small number of particles from $p(\boldsymbol{\theta}_r|\rho(\mathcal{D},\mathcal{D}_s)\leq\epsilon_r)$ using importance resampling and stochastic model simulations; (ii) compute $\hat{\boldsymbol{\mu}}_r$ and $\hat{\boldsymbol{\Sigma}}_r=\hat{\mathbf{L}}_r\hat{\mathbf{L}}_r^{\mathrm{T}}$; (iii) apply the transform from Equation (11) to the particles at ϵ_r from the approximate model; (iv) pool the resulting particles with the stochastic model samples; and (v) reweight particles and resample. The final MM-SMC-ABC algorithm is provided in Algorithm 3.

The performance of this method depends on the choice of α. Note that in Algorithm 3, standard SMC-ABC for the expensive stochastic model is recovered as $\alpha\to1$ (no speedup, inference unbiased), and standard SMC-ABC using the approximate model is recovered as $\alpha\to0$ (maximum speedup, but inference biased). Therefore, we expect there is a choice of $\alpha\in(0,1)$ that provides an optimal tradeoff between computational improvement and accuracy. Clearly, the expected speed improvement is proportional to $1/\alpha$, however, if α is chosen to be too small, then the statistical error in the estimates in Equation (10) will be too high. We explore this tradeoff in detail in Section 3.6 and find that $0.05\leq\alpha\leq0.2$ seems to give a reasonable result.

3. Results

In this section, we provide numerical examples to demonstrate the accuracy and performance of the PC-SMC-ABC and MM-SMC-ABC methods. First we apply PC-SMC-ABC to a tractable

Algorithm 3 Moment-matching SMC-ABC

Given $\alpha \in [0,1]$, initialize $\hat{\mathcal{M}} = \lceil \alpha \mathcal{M} \rceil$ and $\tilde{\mathcal{M}} = \lfloor (1-\alpha)\mathcal{M} \rfloor$;
Initialize $\tilde{\boldsymbol{\theta}}_0^i \sim p(\boldsymbol{\theta})$ and $\tilde{w}_0^i = 1/\tilde{\mathcal{M}}$, for $i = 1, \ldots, \tilde{\mathcal{M}}$;
Initialize $\boldsymbol{\theta}_0^i \sim p(\boldsymbol{\theta})$ and $w_0^i = 1/\mathcal{M}$, for $i = 1, \ldots, \mathcal{M}$;
Apply SMC-ABC to generate the sequence of approximate particles $\{(\tilde{\boldsymbol{\theta}}_1^i, \tilde{w}_1^i)\}_{i=1}^{\tilde{\mathcal{M}}}, \{(\tilde{\boldsymbol{\theta}}_2^i, \tilde{w}_2^i)\}_{i=1}^{\tilde{\mathcal{M}}}, \ldots, \{(\tilde{\boldsymbol{\theta}}_R^i, \tilde{w}_R^i)\}_{i=1}^{\tilde{\mathcal{M}}}$
for $r = 1, \ldots, R$ **do**
  for $i = 1, \ldots, \hat{\mathcal{M}}$ **do**
    repeat
      Draw $\boldsymbol{\theta}^*$ from $\{\boldsymbol{\theta}_{r-1}^j\}_{j=1}^{\mathcal{M}}$ according to the probability mass function

$$\mathbb{P}(\boldsymbol{\theta}^* = \boldsymbol{\theta}_{r-1}^j) = \frac{w_{r-1}^j}{\sum_{k=1}^{\mathcal{M}} w_{r-1}^k}, \quad \text{for } j = 1, \ldots, \mathcal{M};$$

      Sample transition kernel, $\boldsymbol{\theta}^{**} \sim q_r(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$;
      Generate data, $\mathcal{D}_s \sim s(\mathcal{D}|\boldsymbol{\theta}^{**})$;
    until $\rho(\mathcal{D}, \mathcal{D}_s) \leq \epsilon_r$
    Set $\boldsymbol{\theta}_r^i \leftarrow \boldsymbol{\theta}^{**}$;
    Set $w_r^i \leftarrow p(\boldsymbol{\theta}_r^i) / \left[\sum_{j=1}^{\mathcal{M}} w_{r-1}^j q_r(\boldsymbol{\theta}_r^i mid \boldsymbol{\theta}_{r-1}^j)\right]$;
  end for
  Estimate means and covariances $\tilde{\boldsymbol{\mu}}_r$, $\tilde{\boldsymbol{\Sigma}}_r$, $\hat{\boldsymbol{\mu}}_r$, and $\hat{\boldsymbol{\Sigma}}_r$;
  Compute Cholesky decompositions $\tilde{\boldsymbol{\Sigma}}_r = \tilde{\mathbf{L}}_r \tilde{\mathbf{L}}_r^{\mathrm{T}}$ and $\hat{\boldsymbol{\Sigma}}_r = \hat{\mathbf{L}}_r \hat{\mathbf{L}}_r^{\mathrm{T}}$
  for $i = 1, \ldots, \tilde{\mathcal{M}}$ **do**
    Set $\boldsymbol{\theta}_r^{i+\hat{\mathcal{M}}} \leftarrow \hat{\mathbf{L}}_r[\tilde{\mathbf{L}}_r^{-1}(\tilde{\boldsymbol{\theta}}_r^i - \tilde{\boldsymbol{\mu}}_r)] + \hat{\boldsymbol{\mu}}_r$ and $w_r^{i+\hat{\mathcal{M}}} \leftarrow \tilde{w}_r^i$;
  end for
  Resample weighted particles $\{(\boldsymbol{\theta}_r^i, w_r^i)\}_{r=1}^{\mathcal{M}}$ with replacement;
  Set $w_r^i \leftarrow 1/\mathcal{M}$ for all $i = 1, \ldots, \mathcal{M}$;
end for

example to demonstrate the mechanisms of the method and provide insight into effective choices of approximate model. The tractable example considered here is inference for an Ornstein-Uhlenbeck process (Uhlenbeck and Ornstein 1930). We then consider two intractable problems based on expensive discrete models. For our first example, we consider the analysis of spatially averaged population growth data. The discrete model used in this instance is relevant in the ecological sciences as it describes population growth subject to a weak Allee effect (Taylor and Hastings 2005). We then analyze data that is typical of *in vitro* cell culture scratch assays in experimental cell biology using a discrete model that leads to the well-studied Fisher-KPP model (Murray 2002; Edelstein-Keshet 2005). In both examples, we present the discrete model and its continuum limit, then compute the full Bayesian posterior for the model parameters using the PC-SMC-ABC (Algorithm 2) and MM-SMC-ABC (Algorithm 3) methods, and compare the results with the SMC-ABC (Algorithm 1) using either the discrete model or continuum limit alone. We also provide numerical experiments to evaluate the effect of the tuning parameter α on the accuracy and performance of the MM-SMC-ABC method.

It is important to clarify that when we refer to the accuracy of our methods, we refer to their ability to sample from the target ABC posterior under the expensive stochastic model. The evaluation of this accuracy requires sampling from the target ABC posterior under the expensive stochastic model using SMC-ABC. As a result, the target acceptance thresholds are chosen to ensure this is computationally feasible.

3.1. A Tractable Example: Ornstein–Uhlenbeck Process

The Ornstein–Uhlenbeck process is a mean reverting stochastic process with many applications in finance, biology and physics (Uhlenbeck and Ornstein 1930). The evolution of the continuous state X_t is given by an Itō stochastic differential equation (SDE) of the form

$$\begin{aligned} &\mathrm{d}X_t = \gamma(\mu - X_t)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad \text{if } t > 0, \\ &X_0 = x_0, \quad \text{if } t = 0, \end{aligned} \tag{12}$$

where μ is the long-term mean, γ is the rate of mean reversion, σ is the process volatility, W_t is a Wiener process and x_0 is a constant initial condition. Example realizations are shown in Figure 2(A).

We consider data consisting of N independent realizations of the Ornstein–Uhlenbeck processs at time $T < \infty$, that is, $\mathcal{D} = [X_T^1, X_T^2, \ldots, X_T^N]$. This inference problem is analytically tractable since the Fokker–Planck equation can be solved to obtain a Gaussian distribution for the data

$$X_T \sim \mathcal{N}\left(\mu + (x_0 + \mu)e^{-\gamma t}, 1 - \frac{\sigma^2}{2\gamma}e^{-2\gamma t}\right). \tag{13}$$

For demonstration purposes, we will assume a solution for the full Fokker–Planck equation is unavailable and perform ABC inference to estimate the volatility parameter, σ, using stochastic simulation with the Euler–Maruyama discretisation (Maruyama 1955)

$$X_{t+\Delta t} = X_t + \gamma(\mu - X_t)\Delta t + \sigma\sqrt{\Delta t}\xi_t, \tag{14}$$

where ξ_t is a standard normal variate and Δt is a small time step. For the approximate model, we take the stationary distribution of the Ornstein–Uhlenbeck process obtained by taking $T \to \infty$ and solving the steady state of the Fokker–Planck equation,

$$X_\infty \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{2\gamma}\right). \tag{15}$$

This kind of approximation will often be possible since the steady state Fokker–Planck equation is more likely to be tractable than the transient solution for most SDE models. As shown in Figure 2(B), the stationary solution is a better approximation for the true variance than for the true mean. Therefore, for small T this approximation would be more appropriate as a preconditioner for inference of the volatility parameter, σ, in Equation (12) than for the long-time mean, μ. In general, the performance expected from preconditioning will increase as T increases since the stationary approximation will become more accurate.

Figure 3 demonstrates the results of applying PC-ABC-SMC (Algorithm 2) with $\mathcal{M} = 1000$ particles, showing intermediate distributions, for inferring $D = \sigma^2/2$ given data at $T = 1$, using

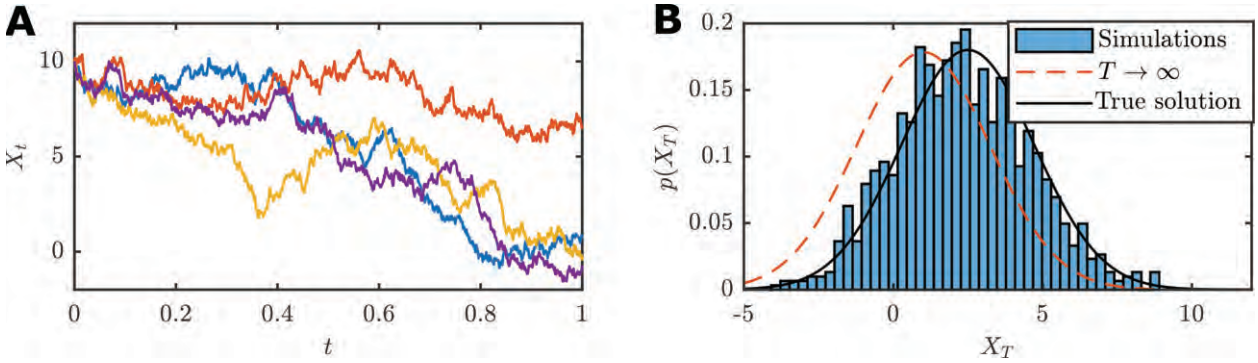

Figure 2. (A) Four example realizations of the Ornstein–Uhlenbeck process with parameters $\mu = 1, \gamma = 2, \sigma = 2\sqrt{5}$ and initial condition $x_0 = 10$. (B) Empirical distribution of X_T for $T = 1$ compared with the exact Fokker–Planck solution and the stationary solution as $T \to \infty$. Simulations are performed using the Euler–Maruyama discretization with time step $\Delta t = 0.001$.

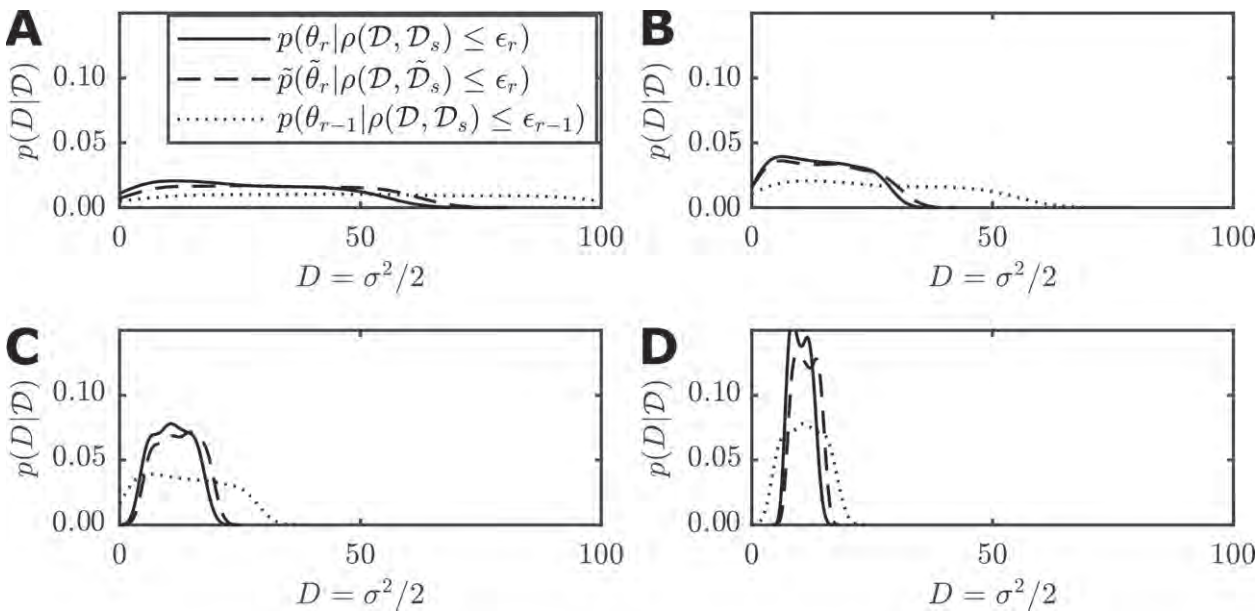

Figure 3. PC-SMC-ABC intermediate steps for the Ornstein–Uhlenbeck SDE example. Each panel demonstrates the transition from threshold ϵ_{r-1} (dotted line) to ϵ_r (solid line) via the preconditioner (dashed line). (A) ϵ_0 to ϵ_1; (B) ϵ_1 to ϵ_2; (C) ϵ_2 to ϵ_3; and (D) ϵ_3 to ϵ_4. The threshold sequence is $\epsilon_r = \epsilon_{r-1}/2$ for $r = 1, 2, 3, 4$ with $\epsilon_0 = 6.4$.

stochastic simulation for the exact model (Equation (14)) with $\Delta t = 0.01$, and the stationary distribution (Equation (15)) for the approximate model used in the preconditioning step.

In this example, all other parameters are treated as known with $x_0 = 10, \mu = 1$ and $\gamma = 2$. Data is generated using $D = 10$ as in Figure 2. In each step, the preconditioner distribution for threshold ϵ_r (dashed line) is a better proposal distribution for the target (solid line) than that of the exact model at threshold ϵ_{r-1} (dotted line). The overall speedup factor is approximately 1.5 for this example, and it continues to improve as T increases since Equation (15) becomes an increasingly better approximation to Equation (13). For truly intractable problems, such as the lattice-based random walk models presented in Sections 3.4 and 3.5, we obtain superior performance gains of up to a factor of four.

3.2. Lattice-Based Stochastic Discrete Random Walk Model

The stochastic discrete model we consider is a lattice-based random walk model that is often used to describe populations of motile cells (Jin et al. 2016). The model involves initially placing a population of N agents of size δ on a lattice, L (Callaghan et al. 2006; Simpson, Landman, and Hughes 2010), for example an $I \times J$ hexagonal lattice (Jin et al. 2016). This hexagonal lattice is defined by a set of indices $L = \{(i,j) : i \in [0,1,\ldots,I-1], j \in [0,1,\ldots,J-1]\}$, and a neighborhood function,

$$\mathcal{N}(i,j) = \begin{cases} \{(i-1,j-1),(i,j-1),(i+1,j-1), \\ \quad (i+1,j),(i,j+1),(i-1,j)\} & \text{if } i \text{ is even}, \\ \{(i-1,j),(i,j-1),(i+1,j), \\ \quad (i+1,j+1),(i,j+1),(i-1,j+1)\} & \text{if } i \text{ is odd}. \end{cases}$$

Lattice indices are mapped to Cartesian coordinates using

$$(x_i, y_j) = \begin{cases} \left(i\frac{\sqrt{3}}{2}\delta, j\delta\right) & \text{if } i \text{ is even}, \\ \left(i\frac{\sqrt{3}}{2}\delta, \left(j+\frac{1}{2}\right)\delta\right) & \text{if } i \text{ is odd}. \end{cases} \quad (16)$$

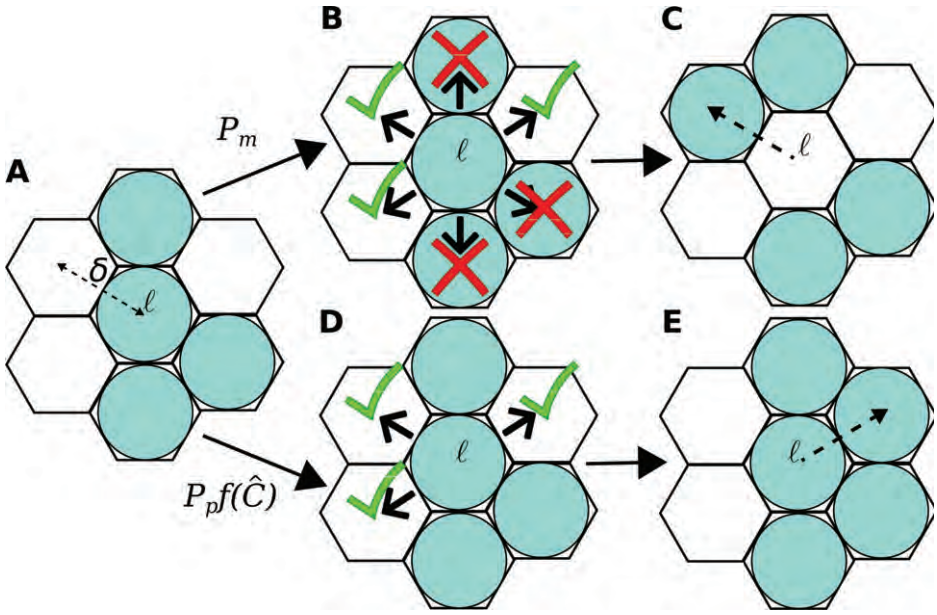

Figure 4. Example of movement and proliferation events in a lattice-based random walk model, using a hexagonal lattice with lattice spacing, δ. (A) An example hexagonal lattice neighborhood $\mathcal{N}(\ell)$. An agent at site ℓ attempts a motility event (A)–(C) with probability P_m. (B) Motility events are aborted when the randomly selected neighbor site is occupied. (C) The agent moves to the selected site, if unoccupied. An agent at site ℓ attempts a proliferation event (A),(D)–(E) with probability P_p. (D) Proliferation events are successful with probability $f(\hat{C}(\ell, t))$, resulting in an unoccupied site being selected. (E) The daughter agent is placed at the selected site and the number of agents in the populations is increased by one.

We define an occupancy function such that $C(\ell, t) = 1$ if site ℓ is occupied by an agent at time $t \geq 0$, otherwise $C(\ell, t) = 0$. This means that in our discrete model each lattice site can be occupied by, at most, one agent.

During each discrete time step of duration τ, agents attempt to move with probability $P_m \in [0, 1]$ and attempt to proliferate with probability $P_p \in [0, 1]$. If an agent at site ℓ attempts a motility event, then a neighboring site will be selected uniformly at random. The motility event is aborted if the selected site is occupied, otherwise the agent will move to the selected site (Figure 4(A)–(C)). For proliferation events, the local neighborhood average occupancy,

$$\hat{C}(\ell, t) = \frac{1}{6} \sum_{\ell' \in N(\ell)} C(\ell', t),$$

is calculated and a uniform random number $u \sim \mathcal{U}(0, 1)$ is drawn. If $u > f(\hat{C}(\ell, t))$, where $f(\hat{C}(\ell, t)) \in [0, 1]$ is called the *crowding function* (Jin et al. 2016; Browning, McCue, and Simpson 2017), then the proliferation event is aborted due to local crowding effects and contact inhibition. If $u \leq f(\hat{C}(\ell, t))$, then proliferation is successful and a daughter agent is placed at a randomly chosen unoccupied lattice site in $\mathcal{N}(\ell)$ (Figure 4(A),(D)–(E)). The evolution of the model is generated through repeating this process though M time steps, $t_1 = \tau$, $t_2 = 2\tau, \ldots$, $t_M = M\tau$. This approach, based on the work by Jin et al. (2016), supports a generic proliferation mechanism since $f(\hat{C}(\ell, t))$ is an arbitrary smooth function satisfying $f(0) = 1$ and $f(K) = 0$, where $K > 0$ is the carrying capacity density. However, in the literature there are also examples that include other mechanisms such as cell-cell adhesion (Johnston, Simpson, and Plank 2013), directed motility (Binny et al. 2016), and Allee effects (Böttger et al. 2015).

3.3. Approximate Continuum-Limit Descriptions

Discrete models do not generally lend themselves to analytical methods, consequently, their application is intrinsically tied to computationally intensive stochastic simulations and Monte Carlo methods (Jin et al. 2016). As a result, it is common practice to approximate mean behavior using differential equations by invoking mean-field assumptions, that is, to treat the occupancy status of lattice sites as independent (Callaghan et al. 2006; Simpson, Landman, and Hughes 2010). The resulting approximate continuum-limit descriptions (supplementary material) are partial differential equations (PDEs) of the form

$$\frac{\partial \mathcal{C}(x, y, t)}{\partial t} = D\nabla^2 \mathcal{C}(x, y, t) + \lambda \mathcal{C}(x, y, t) f(\mathcal{C}(x, y, t)), \quad (17)$$

where $\mathcal{C}(x, y, t) = \mathbb{E}[C(\ell, t)]$, $D = \lim_{\delta \to 0, \tau \to 0} P_m \delta^2 / (4\tau)$ is the diffusivity, $\lambda = \lim_{\tau \to 0} P_p / \tau$ is the proliferation rate with $P_p = \mathcal{O}(\tau)$, and $f(\cdot)$ is the crowding function that is related to the proliferation mechanism implemented in the discrete model (Jin et al. 2016; Browning, McCue, and Simpson 2017). For spatially uniform populations there will be no macroscopic spatial gradients on average, that is $\nabla \mathcal{C}(x, y, t) = \mathbf{0}$. Thus, $\mathcal{C}(x, y, t)$ is just a function of time, $\mathcal{C}(t)$, and the continuum limit reduces to an ordinary differential equation (ODE) describing the net population growth,

$$\frac{\mathrm{d}\mathcal{C}(t)}{\mathrm{d}t} = \lambda \mathcal{C}(t) f(\mathcal{C}(t)). \quad (18)$$

For many standard discrete models, the crowding function is implicitly $f(\mathcal{C}) = 1 - \mathcal{C}$ (Callaghan et al. 2006). That is, the continuum limits in Equation (17) and Equation (18) yield the Fisher-KPP model (Edelstein-Keshet 2005; Murray 2002) and the logistic growth model (Tsoularis and Wallace 2002; Warne,

Baker, and Simpson 2017), respectively. However, nonlogistic growth models , for example, $f(\mathcal{C}) = (1-\mathcal{C})^n$ for $n > 1$, have also been considered (Tsoularis and Wallace 2002; Simpson, Landman, and Hughes 2010; Jin et al. 2016).

3.4. Temporal Example: A Weak Allee Model

The Allee effect refers to the reduction in growth rate of a population at low densities. This is particularly well studied in ecology where there are many mechanisms that give rise to this phenomenon (Taylor and Hastings 2005; Johnston et al. 2017). We introduce an Allee effect into our discrete model by choosing a crowding function of the form

$$f(\hat{C}(\ell,t)) = \left(1 - \frac{\hat{C}(\ell,t)}{K}\right)\left(\frac{A + \hat{C}(\ell,t)}{K}\right),$$

where $\hat{C}(\ell,t) \in [0,1]$ is the local density at the lattice site $\ell \in L$, at time t, $K > 0$ is the carrying capacity density, and A is the Allee parameter which yields a weak Allee effect for $A \geq 0$ (Wang, Shi, and Wang 2019). Note that smaller values of A entail a more pronounced Allee effect with $A < 0$ leading to a strong Allee effect that can lead to species extinction (Wang, Shi, and Wang 2019). For simplicity, we only consider the weak Allee effect here, but our methods are general enough to consider any sufficiently smooth $f(\cdot)$.

Studies in ecology often involve population counts of a particular species over time (Taylor and Hastings 2005). In the discrete model, the initial occupancy of each lattice site is independent, and hence there are no macroscopic spatial gradients on average. It is reasonable to summarize simulations of the discrete model at time t by the average occupancy over the entire lattice, $\bar{C}(t) = (1/IJ)\sum_{\ell \in L} C(\ell,t)$. Therefore, the continuum limit for this case is given by (Wang, Shi, and Wang 2019)

$$\frac{d\mathcal{C}(t)}{dt} = \lambda \mathcal{C}(t)\left(1 - \frac{\mathcal{C}(t)}{K}\right)\left(\frac{A + \mathcal{C}(t)}{K}\right), \qquad (19)$$

with $\mathcal{C}(t) = \mathbb{E}\left[\bar{C}(t)\right]$, $\lambda = \lim_{\tau \to 0} P_p/\tau$, and $\mathcal{C}(0) = \mathbb{E}\left[\bar{C}(0)\right]$.

We generate synthetic time-series ecological data using the discrete model, with observations made at times $t_1 = \tau \times 10^3, t_2 = 2\tau \times 10^3, \ldots, t_{10} = \tau \times 10^4$, resulting in data $\mathcal{D} = [C_{\text{obs}}(t_1), C_{\text{obs}}(t_2), \ldots, C_{\text{obs}}(t_{10})]$ with $C_{\text{obs}}(t) = \bar{C}(t)$ where $\bar{C}(t)$ is the average occupancy at time t for a single realization of the discrete model (supplementary material). For this example, we consider an $I \times J$ hexagonal lattice with $I = 80$, $J = 68$, and parameters $P_p = 1/1000$, $P_m = 0$, $\delta = \tau = 1$, $K = 5/6$, and $A = 1/10$. Reflecting boundary conditions are applied at all boundaries and a uniform initial condition is applied, specifically, each site is occupied with probability $\mathbb{P}(C(\ell,0) = 1) = 1/4$ for all $\ell \in L$, giving $\mathcal{C}(0) = 1/4$. This combination of parameters is selected since it is known that the continuum limit (Equation (19)) will not accurately predict the population growth dynamics of the discrete model in this regime since $P_p/P_m \gg 1$ (supplementary material).

For the inference problem we assume P_m is known, and we seek to compute $p(\boldsymbol{\theta}|\mathcal{D})$ under the discrete model with $\boldsymbol{\theta} = [\lambda, A, K]$ and $\lambda = P_p/\tau$. We use uninformative priors, $P_p \sim \mathcal{U}(0, 0.005)$, $K \sim \mathcal{U}(0,1)$ and $A \sim \mathcal{U}(0,1)$ with the additional constraint that $A \leq K$, that is, A and K are not independent in the prior. The discrepancy metric used is the Euclidean distance. For the discrete model, this is

$$\rho(\mathcal{D}, \mathcal{D}_s) = \left[\sum_{k=1}^{10}\left(C_{\text{obs}}(t_k) - \bar{C}(t_k)\right)^2\right]^{1/2},$$

where $\bar{C}(t_k)$ is the average occupancy at time t_k of a realization of the discrete model given $\boldsymbol{\theta}$. Similarly, for the continuum limit we have

$$\rho(\mathcal{D}, \tilde{\mathcal{D}}_s) = \left[\sum_{j=1}^{10}\left(C_{\text{obs}}(t_k) - \mathcal{C}(t_k)\right)^2\right]^{1/2},$$

where $\mathcal{C}(t_k)$ is the solution to the continuum limit (Equation (19)), computed numerically (Fehlberg 1969; Iserles 2008) (supplementary material). We compute the posterior using our PC-SMC-ABC and MM-SMC-ABC methods to compare with SMC-ABC under the continuum limit and SMC-ABC under the discrete model. In each instance, $\mathcal{M} = 1000$ particles are used to approach the target threshold $\epsilon = 0.125$ using the sequence $\epsilon_1, \epsilon_2, \ldots, \epsilon_5$ with $\epsilon_r = \epsilon_{r-1}/2$. In the case of MM-SMC-ABC the tuning parameter is $\alpha = 0.1$. The Gaussian proposal kernels, $q_r(\boldsymbol{\theta}_r|\boldsymbol{\theta}_{r-1})$ and $\tilde{q}_r(\boldsymbol{\theta}_r|\tilde{\boldsymbol{\theta}}_r)$, are selected adaptively (Beaumont et al. 2009; Filippi et al. 2013).

Figure 5 and Table 1 present the results. SMC-ABC using the continuum-limit model is a poor approximation for SMC-ABC using the discrete model, especially for the proliferation rate parameter, λ (Figure 5(a)), which is expected because $P_m = 0$. However, the posteriors estimated using PC-SMC-ABC are an excellent match to the target posteriors estimated using SMC-ABC with the expensive discrete model, yet the PC-SMC-ABC

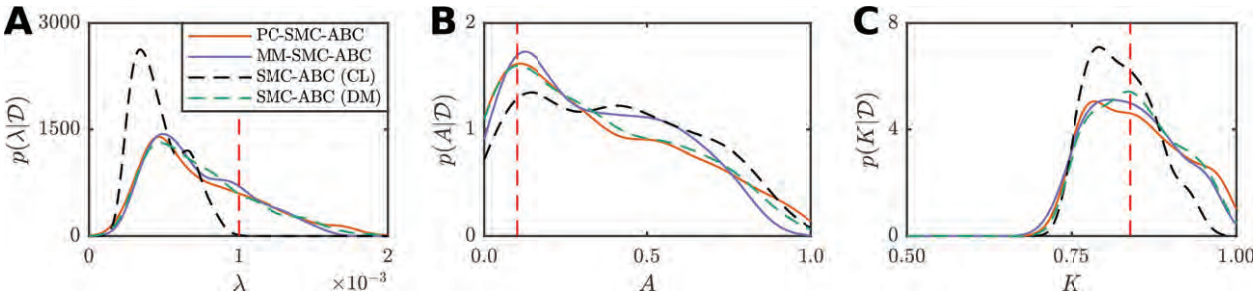

Figure 5. Comparison of estimated posterior marginal densities for the weak Allee model. There is a distinct bias in the SMC-ABC density estimate using the continuum limit (CL) (black dashed) compared with SMC-ABC with the discrete model (DM) (green dashed). However, the density estimates computed using the PC-SMC-ABC (orange solid) and MM-SMC-ABC (purple solid) methods match well with a significantly reduced computational overhead.

Table 1. Computational performance comparison of the SMC-ABC, PC-SMC-ABC, and MM-SMC-ABC methods for the weak Allee model inference problem.

Method	Stochastic samples	Continuum samples	Run time (hours)	Speedup
SMC-ABC	28,588	0	47.1	1×
PC-SMC-ABC	13,799	58,752	21.1	2×
MM-SMC-ABC	3,342	36,908	5.6	8×

NOTE: Computations are performed using an Intel® Xeon™ E5-2680v3 CPU (2.5 GHz).

method requires only half the number of stochastic simulations (Table 1). The MM-SMC-ABC method is not quite as accurate as the PC-SMC-ABC method, however, the number of expensive stochastic simulations is reduced by more than a factor of eight (Table 1) leading to considerable increase in computational efficiency.

3.5. Spatiotemporal Example: A Scratch Assay

We now look to a discrete model commonly used in studies of cell motility and proliferation, and use spatially extended data that is typical of *in vitro* cell culture experiments, specifically scratch assays (Liang, Park, and Guan 2007).

In this case we use a crowding function of the form $f(\hat{C}(\ell,t)) = 1 - \hat{C}(\ell,t)/K$, where $K > 0$ is the carrying capacity density, since it will lead to a logistic growth source term in Equation (17) which characterizes the growth dynamics of many cell types (Simpson, Landman, and Hughes 2010; Jin et al. 2017). The discrete model is initialized such that initial density is independent of y. Therefore, we summarize the discrete simulation by computing the average occupancy for each x coordinate, that is, we average over the y-axis in the hexagonal lattice (Jin et al. 2016), that is, $\bar{C}(x,t) = (1/J)\sum_{(x,y)\in L} C((x,y),t)$. Thus, one arrives at the Fisher-KPP model (Murray 2002; Edelstein-Keshet 2005) for the continuum limit,

$$\frac{\partial \mathcal{C}(x,t)}{\partial t} = D\frac{\partial^2 \mathcal{C}(x,t)}{\partial x^2} + \lambda \mathcal{C}(x,t)\left(1 - \frac{\mathcal{C}(x,t)}{K}\right), \qquad (20)$$

where $\mathcal{C}(x,t) = \mathbb{E}\left[\bar{C}(x,t)\right]$, $D = \lim_{\delta\to 0,\tau\to 0} P_m\delta^2/(4\tau)$, and $\lambda = \lim_{\tau\to 0} P_p/\tau$.

Just as with the weak Allee model, here we generate synthetic spatiotemporal cell culture data using the discrete model. Observations are made at times $t_1 = 3\tau \times 10^2$, $t_2 = 6\tau \times 10^2, \ldots, t_{10} = 3\tau \times 10^3$, resulting in data

$$\mathcal{D} = \begin{bmatrix} C_{\text{obs}}(x_1,t_1) & C_{\text{obs}}(x_1,t_2) & \ldots & C_{\text{obs}}(x_1,t_{10}) \\ C_{\text{obs}}(x_2,t_1) & C_{\text{obs}}(x_2,t_2) & \ldots & C_{\text{obs}}(x_2,t_{10}) \\ \vdots & \vdots & \ddots & \vdots \\ C_{\text{obs}}(x_I,t_1) & C_{\text{obs}}(x_I,t_2) & \ldots & C_{\text{obs}}(x_I,t_{10}) \end{bmatrix},$$

with $C_{\text{obs}}(x,t) = \bar{C}(x,t)$ where $\bar{C}(x,t)$ is the average occupancy over sites $(x,y_1), (x,y_2), \ldots, (x,y_J)$ at time t for a single realization of the discrete model. As with the weak Allee model, we consider an $I \times J$ hexagonal lattice with $I = 80$, $J = 68$, and parameters $P_p = 1/1000$, $P_m = 1$, $\delta = \tau = 1$ and $K = 5/6$. We simulate a scratch assay by specifying the center 20 cell columns ($31 \le i \le 50$) to be initially unoccupied, and apply a uniform initial condition outside the scratch area such that $\mathbb{E}[C(\ell,0)] = 1/4$ overall. Reflecting boundary conditions are applied at all boundaries. Note, we have selected a parameter regime with $P_p/P_m \ll 1$ for which the continuum limit is an accurate representation of the discrete model average behavior (supplementary material).

Since we have spatial information for this problem, we assume P_m is also an unknown parameter and perform inference on the discrete model to compute $p(\boldsymbol{\theta}|\mathcal{D})$ with $\boldsymbol{\theta} = [\lambda, D, K]$, $\lambda = P_p/\tau$, and $D = P_m\delta^2/4\tau$. We utilize uninformative priors, $P_p \sim \mathcal{U}(0, 0.008)$, $P_m \sim \mathcal{U}(0,1)$, and $K \sim \mathcal{U}(0,1)$. For the discrepancy metric we use the Frobenius norm; for the discrete model, this is

$$\rho(\mathcal{D},\mathcal{D}_s) = \left[\sum_{k=1}^{10}\sum_{i=1}^{I}\left(C_{\text{obs}}(x_i,t_k) - \bar{C}(x_i,t_k)\right)^2\right]^{1/2},$$

where $\bar{C}(x_i,t_k)$ is the average occupancy at site x_i at time t_k of a realization of the discrete model given parameters $\boldsymbol{\theta}$. Similarly, for the continuum limit we have

$$\rho(\mathcal{D},\tilde{\mathcal{D}}_s) = \left[\sum_{k=1}^{10}\sum_{i=1}^{I}\left(C_{\text{obs}}(x_i,t_k) - \mathcal{C}(x_i,t_k)\right)^2\right]^{1/2},$$

where $\mathcal{C}(x_i,t_k)$ is the solution to the continuum-limit PDE (Equation (20)), computed using a backward-time, centered-space finite difference scheme with fixed-point iteration and adaptive time steps (Sloan and Abbo 1999; Simpson, Landman, and Bhaganagarapu 2007) (supplementary material). We estimate the posterior using our PC-SMC-ABC and MM-SMC-ABC methods to compare with SMC-ABC using the continuum limit and SMC-ABC using the discrete model. In each case, $\mathcal{M} = 1000$ particles are used to approach the target threshold, $\epsilon = 2$, using the sequence $\epsilon_1, \epsilon_2, \ldots, \epsilon_5$ with $\epsilon_r = \epsilon_{r-1}/2$. In the case of MM-SMC-ABC the tuning parameter is $\alpha = 0.1$. Again, Gaussian proposal kernels, $q_r(\boldsymbol{\theta}_r|\boldsymbol{\theta}_{r-1})$ and $\tilde{q}_r(\boldsymbol{\theta}_r|\tilde{\boldsymbol{\theta}}_r)$, are selected adaptively (Beaumont et al. 2009; Filippi et al. 2013).

Results are shown in Figure 6 and Table 2. Despite the continuum limit being a good approximation of the discrete model average behavior, using solely this continuum limit in the inference problem still leads to bias. Just as with the weak Allee model, both PC-SMC-ABC and MM-SMC-ABC methods produce a more accurate estimate of the SMC-ABC posterior density with the discrete model. Overall, PC-SMC-ABC is unbiased, however, MM-SMC-ABC is still very accurate. The main point for our work is that the PC-SMC-ABC and MM-SMC-ABC methods both produce posteriors that are accurate compared with the expensive stochastic inference problem, whereas the approximate model alone does not. From Table 2, both PC-SMC-ABC and MM-SMC-ABC require a reduced number of stochastic simulations of the discrete model compared with direct SMC-ABC. For PC-SMC-ABC, the reduction is almost a factor of four and, for MM-SMC-ABC, the reduction is almost a factor of eleven.

3.6. A Guide to Selection of α for MM-SMC-ABC

The performance of MM-SMC-ABC is dependent on the tuning parameter $\alpha \in [0,1]$. Since MM-SMC-ABC will only propagate

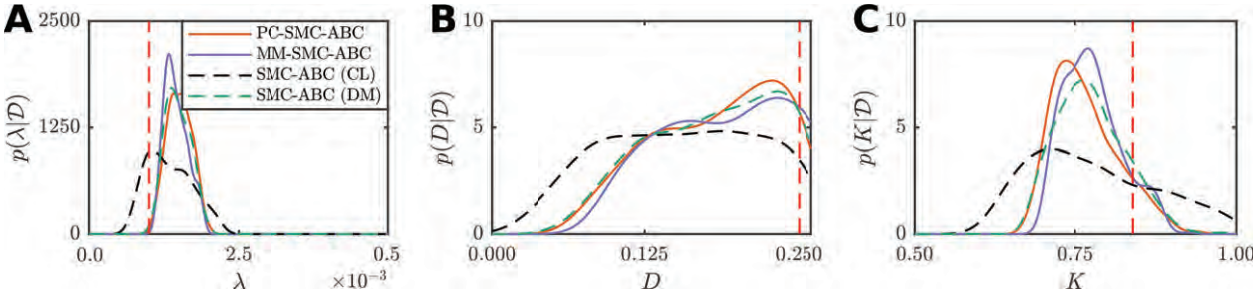

Figure 6. Comparison of estimated posterior marginal densities for the scratch assay model. There is a distinct bias in the SMC-ABC density estimate using the continuum limit (CL) (black dashed) compared with SMC-ABC with the discrete model (DM) (green dashed). However, the density estimates computed using the PC-SMC-ABC (orange solid) and MM-SMC-ABC (purple solid) methods match well with a reduced computational overhead.

Table 2. Computational performance comparison of the SMC-ABC, PC-SMC-ABC, and MM-SMC-ABC methods, using the scratch assay model inference problem.

Method	Stochastic samples	Continuum samples	Run time (hours)	Speedup
SMC-ABC	46,435	0	20.6	1×
PC-SMC-ABC	13,949	13,179	5.6	4×
MM-SMC-ABC	4,457	10,594	1.9	11×

NOTE: Computations are performed using an Intel® Xeon™ E5-2680v3 CPU (2.5 GHz).

$\lceil \alpha \mathcal{M} \rceil$ particles based on the expensive stochastic model, α can be considered as a target computational cost reduction factor with $1/\alpha$ being the target speed up factor. However, intuitively there will be a limit as to how small one can choose α before the statistical error incurred from the estimates of $\boldsymbol{\mu}_r$ and $\boldsymbol{\Sigma}_r$ is large enough to render the approximate moment matching transform inaccurate. It is nontrivial to analyze MM-SMC-ABC to obtain a theoretical guideline for choosing α, therefore we perform a computational benchmark to obtain a heuristic. It should be noted that we use an SMC-ABC sampler with $\mathcal{M} = 1000$ as a benchmark for accuracy and performance. As a result, it may be that repeating the analysis with larger $\mathcal{M}$ could lead to a smaller optimal α.

Here, using different values for α we repeatedly solve the weak Allee model (Section 3.4) and the scratch assay model (Section 3.5). For both inverse problems we applied MM-SMC-ABC under identical conditions as in Sections 3.4 and 3.5 with the exception of the tuning parameter α that takes values from the sequence $\{\alpha_k\}_{k=0}^{5}$ with $\alpha_0 = 0.8$ and $\alpha_k = \alpha_{k-1}/2$ for $k > 0$. For each α_k in the sequence, we consider N independent applications of MM-SMC-ABC. The computational cost for each α_k is denoted by $\text{Cost}(\alpha_k)$ and represents the run time in seconds for an application of MM-SMC-ABC with tuning parameter α_k. We also calculate an error metric,

$$\text{Error}(\alpha_k) = \mathcal{E}\left(\boldsymbol{\Theta}_R, \boldsymbol{\Theta}_R(\alpha_k), P\right),$$

where $\boldsymbol{\Theta}_R = \left\{\boldsymbol{\theta}_R^i\right\}_{i=1}^{\mathcal{M}}$ is a set of particles from an application of SMC-ABC using the expensive stochastic model, and $\boldsymbol{\Theta}_R(\alpha_k) = \left(\left\{\boldsymbol{\theta}_R^i\right\}_{i=1}^{\lceil \alpha_k \mathcal{M} \rceil} \cup \left\{\tilde{\boldsymbol{\theta}}_R^i\right\}_{i=1}^{\lfloor (1-\alpha_k)\mathcal{M} \rfloor}\right)$ is the pooled exact and approximate transformed particles from the jth application of MM-SMC-ABC. For $P \in \mathbb{N}$, the function $\mathcal{E}(\cdot, \cdot, P)$ is the P-th order empirical moment-matching distance function (Lillacci and Khammash 2010; Zechner et al. 2012; Liao, Vejchodský, and Erban 2015), given by

$$\mathcal{E}(\mathbf{X}, \mathbf{Y}, P) = \sum_{m=0}^{P} \sum_{\mathbf{b} \in S_m} \frac{1}{|S_m|^m} \left(\frac{\hat{\mu}(\mathbf{X})^{\mathbf{b}} - \hat{\mu}(\mathbf{Y})^{\mathbf{b}}}{\hat{\mu}(\mathbf{X})^{\mathbf{b}}} \right)^2,$$

for two sample sets $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{\mathcal{M}}\}$ and $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{\mathcal{M}}\}$ with $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^n$ for $n \geq 1$, and $S_m = \{\mathbf{b} : \mathbf{b} \in \mathbb{N}^n, ||\mathbf{b}|| = m\}$. For any n-dimensional discrete vector $\mathbf{b} = [b_1, b_2, \ldots, b_n]^{\mathrm{T}} \in \mathbb{N}^n$, then $\hat{\mu}(\mathbf{X})^{\mathbf{b}}$ is the $\mathbf{b}$th empirical raw moment of the sample set $\mathbf{X}$,

$$\hat{\mu}(\mathbf{X})^{\mathbf{b}} = \frac{1}{\mathcal{M}} \sum_{i=1}^{\mathcal{M}} \mathbf{x}_i^{\mathbf{b}},$$

where $\mathbf{x}_i^{\mathbf{b}} = x_{i,1}^{b_1} \times x_{i,2}^{b_2} \times \cdots \times x_{i,n}^{b_n}$. Note that P must be greater than the number of moments that are matched in the approximate transform (Equation (11)) to ensure that MM-SMC-ABC is improving the accuracy in higher moments also.

We estimate the average $\text{Cost}(\alpha_k)$ and $\text{Error}(\alpha_k)$ for each value of α_k for both the weak Allee effect and the scratch assay inverse problems. Figure 7 displays the estimates and standard errors given $P = 6$ and $N = 10$, with the value of α_k shown. We emphasize that $P > 2$, that is our error measure compares the first six moments while only two moments are matched. There is clearly a threshold for α, below which the error becomes highly variable. For both the weak Allee effect model (Figure 7(A),(C)) and the scratch assay model (Figure 7(B),(D)), the optimal choice of α is located between $\alpha_2 = 0.2$ and $\alpha_4 = 0.05$. Therefore, we suggest a heuristic of $\alpha \in [0.1, 0.2]$ to be a reliable choice. If extra performance is needed $\alpha \in [0.05, 0.1)$ may also be acceptable, but if accuracy is of the utmost importance then $\alpha \approx 0.2$ seems to be the most robust choice. This experiment also provides insight in to the consistency and stability of the MM-SMC-ABC method, where $\alpha \geq 0.1$ leads to results that are consistently fast and have low variability in the error metric. While further work is required to assess theoretically the stability and consistency properties of this method, these numerical results are promising. In general, the choice of optimal α is still an open problem and is likely to be impacted by the specific nature of the relationship between the exact model and the approximate model.

3.7. Summary

This section presented numerical examples to demonstrate our new methods, PC-SMC-ABC and MM-SMC-ABC, for

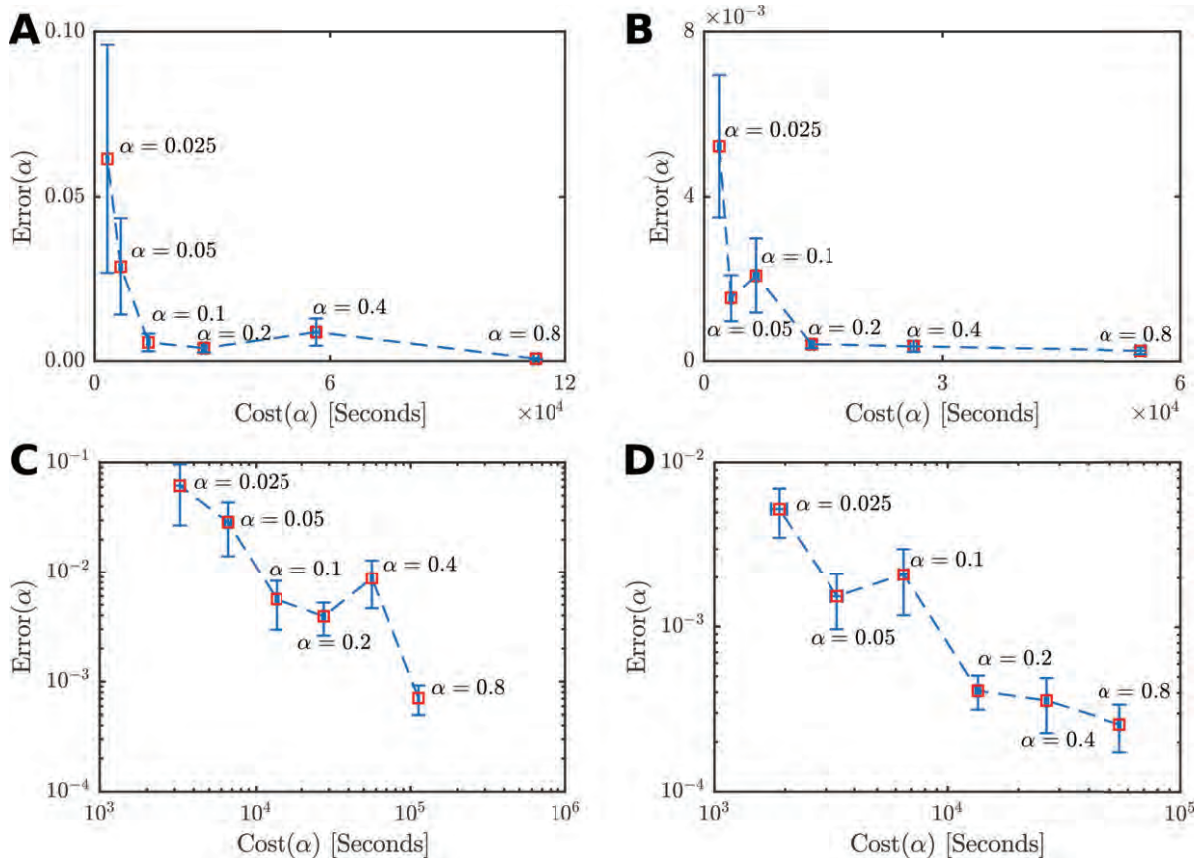

Figure 7. Error versus cost plots for different values of the tuning parameter α. Averages and standard errors are shown for $N = 10$ independent applications of the MM-SMC-ABC method to the (A) weak Allee effect model and (B) the scratch assay model. Results in (A)–(B) are shown in (C)–(D) in a log scale for clarity.

ABC inference with expensive stochastic discrete models. The tractable Ornstein–Uhlenbeck process was used to highlight the mechanisms leading to the performance improvements of PC-ABC-SMC. Then two examples based on lattice-based random walks were used to demonstrate the efficacy of both PC-SMC-ABC and MM-SMC-AB. In the weak Allee model example, data were generated using parameters that violate standard continuum-limit assumptions; in the scratch assay model example, the Fisher-KPP continuum limit is known to be a good approximation in the parameter regime of the generated data. In both examples, final inferences are biased when the continuum limit is exclusively relied on in the SMC-ABC sampler. However, the results from our new algorithms, PC-SMC-ABC and MM-SMC-ABC, show significantly more accurate posteriors can be computed at a fraction of the cost of the full SMC-ABC using the discrete model, with speed improvements over an order of magnitude.

As mentioned in Section 2.3, the tuning parameter, α, in the MM-SMC-ABC method effectively determines the tradeoff between the computational speed of the approximate model and the accuracy of the expensive stochastic model. The values $\alpha = 0$ and $\alpha = 1$ correspond to performing inference exclusively with, respectively, the continuum limit and the stochastic discrete model. Based on numerical experimentation, we find that $\alpha \approx 0.1$ is quite reasonable, however, this conclusion will be dependent on the specific model, the parameter space dimensionality, and the number of particles used for the SMC scheme.

4. Discussion

In the life sciences, computationally challenging stochastic discrete models are routinely used to characterize the dynamics of biological populations (Codling, Plank, and Benhamou 2008; Callaghan et al. 2006; Simpson, Landman, and Hughes 2010). In practice, approximations such as the mean-field continuum limit are often derived and used in place of the discrete model for analysis and, more recently, for inference. However, parameter inferences will be biased when the approximate model is solely utilized for inference, even in cases when the approximate model provides an accurate description of the average behavior of the stochastic model.

We provide a new approach to inference for stochastic models that maintains all the methodological benefits of working with discrete mathematical models, while avoiding the computational bottlenecks of relying solely upon repeated expensive stochastic simulations. Our two new algorithms, PC-SMC-ABC and MM-SMC-ABC, utilize samples from the approximate model inference problem in different ways to accelerate SMC-ABC sampling. The PC-SMC-ABC method is asymptotically unbiased, and we demonstrate computational improvements of up to a factor of almost four are possible. While potentially biased, MM-SMC-ABC can provide further improvements. In general, the expected speedup is around $1/\alpha$, and $\alpha \approx 0.1$ is reasonable based on our numerical investigations. For larger values of $\mathcal{M}$ it may be that even smaller values of α could be effective.

There are some assumptions in our approach that could be generalized in future work. First, in PC-SMC-ABC, we assume that the condition in Equation (8) holds for all ϵ_r; this is reasonable for the models we consider since we never observe a decrease in performance. However, it may be possible for the bias in the approximate model to be so extreme for some ϵ_r that the condition in Equation (8) is violated, leading to a decrease in performance at specific generations. Acceptance probabilities could be estimated by performing a small set of trial samples from both $\eta_{r-1}(\boldsymbol{\theta}_{r-1})$ and $\tilde{\eta}_r(\boldsymbol{\theta}_r)$ proposal mechanisms, enabling automatic selection of the optimal proposal mechanism. Second, in the moment matching transform proposed in Equation (9), we use two moments only as this is sufficient for the problems we consider here with numerical examples demonstrating accuracy in the first six moments. However, our methodology is sufficiently flexible that additional moments can be incorporated if necessary. While including higher moments will improve the accuracy of the moment-matching transform, more samples from the exact model will be required to achieve initial estimates of these moments resulting in eroded performance.

While the performance improvements we demonstrate here are significant, it is also possible to obtain improvements of similar order through detailed code optimization techniques applied to standard SMC-ABC. We emphasize that our schemes would also benefit from such optimizations as advanced vectorization and parallelization to further improve their performance (Lee et al. 2010; Warne, Sisson, and Drovandi 2021). Our algorithm extensions are also more direct to implement over advanced high performance computing techniques for acceleration of computational schemes.

There are many extensions to our methods that could be considered. We have based our presentation on a form of an SMC-ABC sampler that uses a fixed sequence of thresholds. However, the ideas of using the preconditioning distribution, as in PC-SMC-ABC, and the moment matching transform, as in MM-SMC-ABC, are applicable to SMC schemes that adaptively select thresholds (Drovandi and Pettitt 2011). Recently, there have been a number of state-of-the-art inference schemes introduced based on multilevel Monte Carlo (MLMC) (Giles 2015; Warne, Baker, and Simpson 2019a). Our new SMC-ABC schemes could exploit MLMC to combine samples from all acceptance thresholds using a coupling scheme and bias correction telescoping summation, such as in the work of Jasra et al. (2019) or Warne, Baker, and Simpson (2018). Early accept/rejection schemes, such as those considered by Prangle (2016), Prescott and Baker (2020), and Lester (2020), could also be introduced for the sampling steps involving the expensive discrete model. Lastly, the PC-SMC-ABC and the MM-SMC-ABC methods could also be applied together and possibly lead to a compounding effect in the performance. Delayed acceptance schemes (Golightly, Henderson, and Sherlock 2015; Banterle et al. 2019; Bon, Lee, and Drovandi 2020; Everitt and Rowińska 2020) are also an alternative approach with similar motivations to the methods we propose in this work. However, these approaches can be highly sensitive to false negatives, that is, cases where a particular value of $\boldsymbol{\theta}$ would be rejected under the approximate model but accepted under the exact model . Our PC-SMC-ABC approach is not affected by false negatives due to the use of the second set of proposal kernels.

We have demonstrated our methods using a two-dimensional lattice-based discrete random-walk model that leads to mean-field continuum-limit approximations with linear diffusion and a source term of the form $\lambda \mathcal{C} f(\mathcal{C})$. However, our methods are more widely applicable. We could further generalize the model to deal with a more general class of reaction-diffusion continuum limits involving nonlinear diffusion (Warne, Baker, and Simpson 2019b; Witelski 1995) and generalized proliferation mechanisms (Simpson et al. 2013; Tsoularis and Wallace 2002). Our framework is also relevant to lattice-free discrete models (Codling, Plank, and Benhamou 2008; Browning et al. 2018) and higher dimensional lattice-based models (Browning, Haridas, and Simpson 2019); we expect the computational improvements will be even more significant in this case. Many other forms of model combinations are also be possible. For example, a sequence of continuum models of increasing complexity could be considered, as in Browning, Haridas, and Simpson (2019). Alternatively, a sequence of numerical approximations of increasing accuracy could be used for inference using a complex target PDE model (Cotter, Dashti, and Stuart 2010). Linear mapping approximations of higher order chemical reaction network models, such as in Cao and Grima (2018), could also exploit our approach. Another relevant and very general application in systems biology is the utilization reaction rate equations or moment closures, that result in deterministic ODEs, as approximations to stochastic chemical kinetics models (Higham 2008; Wilkinson 2009; Ruess and Lygeros 2015; Browning et al. 2020).

Of course, not all approximate models will necessarily provide performance improvements. As demonstrated for the Ornstein–Uhlenbeck example (Section 3.1), the stationary distribution will be more appropriate for inference of σ rather than μ with the approximation improving as the model sample time T increases. However, as shown for lattice-based random walk models (Sections 3.4 and 3.5), even when the assumptions associated with the approximation do not hold, it is still possible to improve sampling with PC-SMC-ABC and MM-ABC-SMC. Therefore, we suggest that approximations that are derived from some limiting, averaged behavior of the exact model will be good initial candidates for our methods. Semi-automated model reduction techniques also are potential approaches to obtain approximations (Transtrum and Qiu 2014) that could be investigated in the future.

In this work, novel methods have been presented for exploiting approximate models to accelerate Bayesian inference for expensive stochastic models. We have shown that, even when the approximation leads to biased parameter inferences, it can still inform the proposal mechanisms for ABC samplers using the stochastic model. Our numerical examples show performance improvements of more than ten-fold. These substantial computational improvements are promising and expands the feasibility of Bayesian analysis for problems involving expensive stochastic models.

Supplementary Material

Appendix: Supplementary sections including extra technical descriptions, numerical results and datasets. (.pdf file) *Software:* Snapshot of GitHub repository including example implementations and examples. (.zip file)

Acknowledgments

Computational resources where provided by the eResearch Office, Queensland University of Technology. The authors thank Wang Jin for helpful discussions. D.J.W. acknowledges continued support from the Centre for Data Science at the Queensland University of Technology. D.J.W. is a member of the Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers. R.E.B. would like to thank the Leverhulme Trust for a Leverhulme Research Fellowship, the Royal Society for a Wolfson Research Merit Award, and the BBSRC for funding via BB/R00816/1.

Funding

This work was supported by the Australian Research Council (DP170100474). M.J.S. appreciates support from the University of Canterbury Erskine Fellowship.

Software Availability

Numerical examples presented in this work are available from GitHub *https://github.com/ProfMJSimpson/Warne_RapidBayesianInference_2019*.

ORCID

David J. Warne http://orcid.org/0000-0002-9225-175X

References

Agnew, D., Green, J., Brown, T., Simpson, M., and Binder, B. (2014), "Distinguishing Between Mechanisms of Cell Aggregation Using Pair-Correlation Functions," *Journal of Theoretical Biology*, *352*, 16–23. [513]

Baker, R. E., . Peña, J.-M., Jayamohan, J., and Jérusalem, A. (2018), "Mechanistic Models Versus Machine Learning, A Fight Worth Fighting for the Biological Community?" *Biology Letters*, 14, 20170660. [512]

Banterle, M., Grazian, C., Lee, A., and Robert, C. P. (2019), "Accelerating Metropolis-Hastings Algorithms by Delayed Acceptance," *Foundations of Data Science*, 1, 103–128. [525]

Barnes, C. P., Filippi, S., Stumpf, M. P. H., and Thorne, T. (2012), "Considerate Approaches to Constructing Summary Statistics for ABC Model Selection," *Statistics and Computing*, 22, 1181–1197. [514]

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009), "Adaptive Approximate Bayesian Computation," *Biometrika*, 96, 983–990. [515,516,521,522]

Beaumont, M. A., Zhang, W., and Balding, D. J. (2002), "Approximate Bayesian Computation in Population Genetics," *Genetics*, 162, 2025–2035. [513,514]

Binny, R. N., Haridas, P., James, A., Law, R., Simpson, M. J., and Plank, M. J. (2016), "Spatial Structure Arising From Neighbour-Dependent Bias in Collective Cell Movement," *PeerJ*, 4, e1689. [520]

Black, A. J., and McKane, A. J. (2012), "Stochastic Formulation of Ecological Models and Their Applications," *Trends in Ecology & Evolution*, 27, 337–345. [512]

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. [514]

Bon, J. J., Lee, A., and Drovandi, C. (2020), "Accelerating Sequential Monte Carlo With Surrogate Likelihoods," arXiv:2009.03699 [stat.CO] [512,525]

Böttger, K., Hatzikirou, H., Voss-Böhme, A., Cavalcanti-Adam, E. A., Herrero, M. A., and Deutsch, A. (2015), "An Emerging Allee Effect is Critical for Tumor Initiation and Persistence," *PLOS Computational Biology*, 11, e1004366. [513,520]

Browning, A. P., Haridas, P., and Simpson, M. J. (2019), "A Bayesian Sequential Learning Framework to Parameterise Continuum Models of Melanoma Invasion Into Human Skin," *Bulletin of Mathematical Biology*, 81, 676–698. [513,525]

Browning, A. P., McCue, S. W., Binny, R. N., Plank, M. J., Shah, E. T., and Simpson, M. J. (2018), "Inferring Parameters for a Lattice-Free Model of Cell Migration and Proliferation Using Experimental Data," *Journal of Theoretical Biology*, 437, 251–260. [513,525]

Browning, A. P., McCue, S. W., and Simpson, M. J. (2017), "A Bayesian Computational Approach to Explore the Optimal Duration of a Cell Proliferation Assay," *Bulletin of Mathematical Biology*, 79, 1888–1906. [520]

Browning, A. P., D. J. Warne, K. Burrage, R. E. Baker, and M. J. Simpson (2020), "Identifiability Analysis for Stochastic Differential Equation Models in Systems Biology," *Journal of the Royal Society Interface*, *17*(173), 20200652. [525]

Buzbas, E. O., and Rosenberg, N. A. (2015), "AABC: Approximate Approximate Bayesian Computation for Inference in Population-Genetic Models," *Theoretical Population Biology*, 99, 31–42. [512]

Callaghan, T., Khain, E., Sander, L. M., and Ziff, R. M. (2006), "A Stochastic Model for Wound Healing," *Journal of Statistical Physics*, 122, 909–924. [513,519,520,524]

Cao, Z., and Grima, R. (2018), "Linear Mapping Approximation of Gene Regulatory Networks With Stochastic Dynamics," *Nature Communications*, 9, 3305. [512,525]

Chen, C. P., and Zhang, C.-Y. (2014), "Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data," *Information Sciences*, 275, 314–347. [512]

Chkrebtii, O. A., Cameron, E. K., Campbell, D. A., and Bayne, E. M. (2015), "Transdimensional Approximate Bayesian Computation for Inference on Invasive Species Models With Latent Variables of Unknown Dimension," *Computational Statistics & Data Analysis*, 86, 97–110. [513]

Codling, E. A., Plank, M. J., and Benhamou, S. (2008), "Random Walk Models in Biology," *Journal of The Royal Society Interface*, 5, 813–834. [513,524,525]

Cotter, S. L., Dashti, M., and Stuart, A. M. (2010), "Approximation of Bayesian Inverse Problems for PDEs," *SIAM Journal on Numerical Analysis*, 48, 322–345. [525]

Coveney, P. V., Dougherty, E. R., and Highfield, R. R. (2016), "Big Data Need Big Theory Too," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374, 20160153. [512]

DeAngelis, D. L., and Grimm, V. (2014), "Individual-Based Models in Ecology After Four Decades," *F1000Prime Reports*, 6, 39. [513]

Del Moral, P., Doucet, A., and Jasra, A. (2006), "Sequential Monte Carlo Samplers," *Journal of the Royal Statistical Society*, Series B, 68, 411–436.

Drawert, B., Griesemer, M., Petzold, L. R., and Briggs, C. J. (2017), "Using Stochastic Epidemiological Models to Evaluate Conservation Strategies For Endangered Amphibians," *Journal of The Royal Society Interface*, 14, 20170480. [512]

Drovandi, C. C., and Pettitt, A. N. (2011), "Estimation of Parameters for Macroparasite Population Evolution Using Approximate Bayesian Computation," *Biometrics*, 67, 225–233. [525]

Edelstein-Keshet, L. (2005), *Mathematical Models in Biology*, Phiadelphia: Society for Industrial and Applied Mathematics. [518,520,522]

Ellison, A. M. (2004), "Bayesian Inference in Ecology," *Ecology Letters*, 7, 509–520. [513]

EHT Collaboration et al. (2019), "First M87 Event Horizon Telescope Results. VI. the Shadow and Mass of the Central Black Hole," *The Astrophysical Journal Letters*, 875, L6. [512]

Everitt, R. G., and Rowińska, P. A. (2020), "Delayed Acceptance ABC-SMC," *Journal of Computational and Graphical Statistics*, 30, 55–66. DOI: 10.1080/10618600.2020.1775617. [525]

Fearnhead, P., and Prangle, D. (2012), "Constructing Summary Statistics for Approximate Bayesian Computation: Semi-Automatic Approximate Bayesian Computation," *Journal of the Royal Statistical Society*, Series B, 74, 419–474. [514]

Fehlberg, E. (1969), "Low-Order Classical Runge–Kutta Formulas With Step Size Control and Their Application to Some Heat Transfer Problems," Technical Report R-315, Washington DC:NASA. [521]

Filippi, S., Barnes, C. P., Cornebise, J., and Stumpf, M. P. H. (2013), "On Optimality of Kernels for Approximate Bayesian Computation Using Sequential Monte Carlo," *Statistical Applications in Genetics and Molecular Biology*, 12, 87–107. [515,516,521,522]

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014), *Bayesian Data Analysis* (3rd ed.). New York: Chapman & Hall/CRC. [514]

Giles, M. B. (2015), "Multilevel Monte Carlo Methods," *Acta Numerica*, 24, 259–328. [512,525]

Golightly, A., Henderson, D. A., and Sherlock, C. (2015), "Delayed Acceptance Particle MCMC for Exact Inference in Stochastic Kinetic Models," *Statistics and Computing*, 25, 1039–1055. [525]

Guindani, M., Sepúlveda, N., Paulino, C. D., and Müller, P. (2014), "A Bayesian Semiparametric Approach for the Differential Analysis of Sequence Counts Data," *Journal of the Royal Statistical Society*, Series C, 63, 385–404. [512]

Gunawan, R., Cao, Y., Petzold, L., and Doyle, F. J. III (2005), "Sensitivity Analysis of Discrete Stochastic Systems," *Biophysical Journal*, 88, 2530–2540. [512]

Higham, D. J. (2008), "Modeling and Simulating Chemical Reactions," *SIAM Review*, 50, 347–368. [525]

Holden, P. B., Edwards, N. R., Ridgwell, A., Wilkinson, R. D., Fraedrich, K., Lunkeit, F., Pollitt, H., Mercure, J.-F., Salas, P., Lam, A., Knobloch, F., Chewpreecha, U., and Viñuales, J. E. (2018), "Climate–Carbon Cycle Uncertainties and the Paris Agreement," *Nature Climate Change*, 8, 609–613. [512]

Iserles, A. (2008), *A First Course in the Numerical Analysis of Differential Equations* (2nd ed.), Cambridge: Cambridge University Press. [521]

Jasra, A., Jo, S., Nott, D., Shoemaker, C., and Tempone, R. (2019), "Multilevel Monte Carlo in Approximate Bayesian Computation," *Stochastic Analysis and Applications*, 37, 346–360. [525]

Jin, W., Penington, C. J., McCue, S. W., and Simpson, M. J. (2016), "Stochastic Simulation Tools and Continuum Models for Describing Two-Dimensional Collective Cell Spreading With Universal Growth Functions," *Physical Biology*, 13, 056003. [513,519,520,521,522]

Jin, W., Shah, E. T., Penington, C. J., McCue, S. W., Maini, P. K., and Simpson, M. J. (2017), "Logistic Proliferation of Cells in Scratch Assays is Delayed," *Bulletin of Mathematical Biology*, 79, 1028–1050. [514,522]

Johnston, S. T., Baker, R. E., McElwain, D. L. S., and Simpson, M. J. (2017), "Co-operation, Competition and Crowding: A Discrete Framework Linking Allee Kinetics, Nonlinear Diffusion, Shocks and Sharp-Fronted Travelling Waves," *Scientific Reports*, 7, 42134. [521]

Johnston, S. T., Simpson, M. J., and Plank, M. J. (2013), "Lattice-Free Descriptions of Collective Motion With Crowding and Adhesion," *Physical Review E*, 88, 062720. [520]

King, T. E., Fortes, G. G., Balaresque, P., Thomas, M. G., Balding, D., Delser, P. M., Neumann, R., Parson, W., Knapp, M., Walsh, S., Tonasso, L., Holt, J., Kayser, M., Appleby, J., Forster, P., Ekserdjian, D., Hofreiter, M., and Schürer, K. (2014), "Identification of the Remains of King Richard III," *Nature Communications*, 5, 5631. [512]

Kullback, S., and Leibler, R. A. (1951), "On Information and Sufficiency," *The Annals of Mathematical Statistics*, 22, 79–86. [515]

Law, R., Murrell, D. J., and Dieckmann, U. (2003), "Population Growth in Space and Time: Spatial Logistic Equations," *Ecology*, 84, 252–262. [513]

Lawson, B. A. J., Drovandi, C. C., Cusimano, N., Burrage, P., Rodriguez, B., and Burrage, K. (2018), "Unlocking Data Sets by Calibrating Populations of Models to Data Density: A Study in Atrial Electrophysiology," *Science Advances*, 4, e1701676. [513]

Lee, A., Yau, C., Giles, M. B., Doucet, A., and Holmes, C. C. (2010), "On the Utility of Graphics Cards to Perform Massively Parallel Simulation of Advanced Monte Carlo Methods," *Journal of Computational and Graphical Statistics*, 19, 769–789. [525]

Lei, J., and Bickel, P. (2011), "A Moment Matching Ensemble Filter for Nonlinear Non-Gaussian Data Assimilation," *Monthly Weather Review*, 139, 3964–3973. [517]

Lester, C. (2020), "Multi-Level Approximate Bayesian Computation," arXiv:1811.08866 [q-bio.QM] [525]

Lester, C., Yates, C. A., and Baker, R. E. (2017), "Efficient Parameter Sensitivity Computation for Spatially Extended Reaction Networks," *The Journal of Chemical Physics*, 146, 044106. [512]

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. [522]

Liao, S., Vejchodský, T., and Erban, R. (2015), "Tensor Methods for Parameter Estimation and Bifurcation Analysis of Stochastic Reaction Networks," *Journal of The Royal Society Interface*, 12, 20150233. [523]

Lillacci, G., and Khammash, M. (2010), "Parameter Estimation and Model Selection in Computational Biology," *PLOS Computational Biology*, 6, e1000696. [523]

Liu, Q.-H., Ajelli, M., Aleta, A., Merler, S., Moreno, Y., and Vespignani, A. (2018), "Measurability of the Epidemic Reproduction Number in Data-Driven Contact Networks," *Proceedings of the National Academy of Sciences*, 115, 12680–12685. [512]

Malaspinas, A.-S., Westaway, M. C., Muller, C., Sousa, V. C., Lao, O., Alves, I., et al. (2016). "A Genomic History of Aboriginal Australia," *Nature*, 538, 207–214. [512]

Marino, S., Hogue, I. B., Ray, C. J., and Kirschner, D. E. (2008), "A Methodology for Performing Global Uncertainty and Sensitivity Analysis in Systems Biology," *Journal of Theoretical Biology*, 254, 178–96. [512]

Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003), "Markov Chain Monte Carlo Without Likelihoods," *Proceedings of the National Academy of Sciences*, 100, 15324–15328. [514]

Maruyama, G. (1955), "Continuous Markov Processes and Stochastic Equations," *Rendiconti del Circolo Matematico di Palermo*, 4, 48–90. [518]

McLane, A. J., Semeniuk, C., McDermid, G. J., and Marceau, D. J. (2011), "The Role of Agent-Based Models in Wildlife Ecology and Management," *Ecological Modelling*, 222, 1544–1556. [513]

Murray, J. D. (2002), *Mathematical Biology: I An Introduction*, New York: Springer. [518,520,522]

Nott, D. J., Fan, Y., Marshall, L., and Sisson, S. A. (2014), "Approximate Bayesian Computation and Bayes' Linear Analysis: Toward High-Dimensional ABC," *Journal of Computational and Graphical Statistics*, 23, 65–86. [514]

O'Dea, A., Lessios, H. A., Coates, A. G., Eytan, R. I., Restrepo-Moreno, S. A., Cione, A. L. de Queiroz, A., Farris, D. W., Norris, R. D., Stallard, R. F., Woodburne, M. O., Aguilera, O., Aubry, M. P., Berggren, W. A., Budd, A. F., Cozzuol, M. A., Coppard, S. E., Duque-Caro, H., Finnegan, S., Gasparini, G. M., Grossman, E. L., Johnson, K. G., Keigwin, L. D., Knowlton, N., Leigh, E. G., Leonard-Pingel, J. S., Marko, P. B., Pyenson, N. D., Rachello-Dolmen, P. G., Soibelzon, E., Soibelzon, L., Todd, J. A., Vermeij, G. J., Jackson, J. B. (2016), " Formation of the Isthmus of Panama," *Science Advances*, 2, e1600883. [512]

Parno, M. D., and Marzouk, Y. M. (2018), "Transport Map Accelerated Markov Chain Monte Carlo," *SIAM/ASA Journal on Uncertainty Quantification*, 6, 645–682. [512,516]

Prangle, D. (2016), "Lazy ABC," *Statistics and Computing*, 26, 171–185. [525]

Prescott, T. P., and Baker, R. E. (2020), "Multifidelity Approximate Bayesian Computation," *SIAM/ASA Journal of Uncertainty Quantification*, 8, 114–138. [525]

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1997), *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge: Cambridge University Press. [517]

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A., and Feldman, M. W. (1999), "Population Growth of Human Y Chromosomes: A Study of Y Chromosome Microsatellites," *Molecular Biology and Evolution*, 16, 1791–1798. [512,513,514]

Ruess, J., and Lygeros, J. (2015), "Moment-Based Methods for Parameter Inference and Experiment Design for Stochastic Biochemical Reaction Networks," *ACM Transactions on Modeling and Computer Simulation*, 25, 8. [525]

Rynn, J. A. J., Cotter, S. L., Powell, C. E., and Wright, L. (2019), "Surrogate Accelerated Bayesian Inversion for the Determination of the Thermal Diffusivity of a Material," *Metrologia*, 56, 015018. [512]

Simpson, M. J., Binder, B. J., Haridas, P., Wood, B. K., Treloar, K. K., McElwain, D. L. S., and Baker, R. E. (2013), "Experimental and Modelling Investigation of Monolayer Development With Clustering," *Bulletin of Mathematical Biology*, 75, 871–889. [513,514,525]

Simpson, M. J., Landman, K. A., and Bhaganagarapu, K. (2007), "Coalescence of Interacting Cell Populations," *Journal of Theoretical Biology*, 247, 525–543. [522]

Simpson, M. J., Landman, K. A., and Hughes, B. D. (2010), "Cell Invasion With Proliferation Mechanisms Motivated by Time-Lapse Data," *Physica A: Statistical Mechanics and its Applications*, 389, 3779–3790. [513,519,520,521,522,524]

Sisson, S. A., Fan, Y., and Beaumont, M. (2018), *Handbook of Approximate Bayesian Computation* (1st ed.). Boca Raton:Chapman & Hall/CRC. [512,513]

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. [513,514,515]

Sloan, S. W., and Abbo, A. J. (1999), "Biot Consolidation Analysis With Automatic Time Stepping and Error Control Part 1: Theory and Implementation," *International Journal for Numerical and Analytical Methods in Geomechanics*, 23, 467–492. [522]

Stumpf, M. P. H. (2014), "Approximate Bayesian Inference for Complex Ecosystems," *F1000Prime Reports 6*, 60. [513]

Sun, B., Feng, J., and Saenko, K. (2016), "Return of Frustratingly Easy Domain Adaptation," in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pp. 2058–2065. AAAI Press. [517]

Tavaré, S., Balding, D. J., Griffiths, R. C., and Donnelly, P. (1997), "Inferring Coalescence Times From DNA Sequence Data," *Genetics*, 145, 505–518. [512,513,514]

Taylor, C. M., and Hastings, A. (2005), "Allee Effects in Biological Invasions," *Ecology Letters*, 8, 895–908. [513,518,521]

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. (2009), "Approximate Bayesian Computation Scheme for Parameter Inference and Model Selection in Dynamical Systems," *Journal of The Royal Society Interface*, 6, 187–202. [514,515]

Transtrum, M. K., and Qiu, P. (2014), "Model Reduction by Manifold Boundaries," *Physical Review Letters*, 113, 098701. [525]

Tsoularis, A., and Wallace, J. (2002), "Analysis of Logistic Growth Models," *Mathematical Biosciences*, 179, 21–55. [520,521,525]

Uhlenbeck, G. E., and Ornstein, L. S. (1930), "On the Theory of Brownian Motion," *Physical Review*, 36, 823–841. [518]

Vankov, E. R., Guindani, M., and Ensor, K. B. (2019), "Filtering and Estimation for a Class of Stochastic Volatility Models With Intractable Likelihoods," *Bayesian Analysis*, 14, 29–52. [513]

Vincenot, C. E., Carteni, F., Mazzoleni, S., Rietkerk, M., and Giannino, F. (2016), "Spatial Self-Organization of Vegetation Subject to Climatic Stress—Insights From a System Dynamics—Individual-Based Hybrid Model," *Frontiers in Plant Science*, 7, 636. [513]

Vo, B. N., Drovandi, C. C., Pettitt, A. N., and Simpson, M. J. (2015), "Quantifying Uncertainty in Parameter Estimates for Stochastic Models of Collective Cell Spreading Using Approximate Bayesian Computation," *Mathematical Biosciences*, 263, 133–142. [512]

von Hardenberg, J., Meron, E., Shachak, M., and Zarmi, Y. (2001), "Diversity of Vegetation Patterns and Desertification," *Physical Review Letters*, 87, 198101. [513]

Wang, Y., Shi, J., and Wang, J. (2019), "Persistence and Extinction of Population in Reaction–Diffusion–Advection Model With Strong Allee Effect Growth," *Journal of Mathematical Biology*, 78, 2093–2140. [521]

Warne, D. J., Baker, R. E., and Simpson M. J. (2017), "Optimal Quantification of Contact Inhibition in Cell Populations," *Biophysical Journal*, 113, 1920—1924. [513,521]

Warne, D. J., Baker, R. E., and Simpson, M. J. (2018), "Multilevel Rejection Sampling for Approximate Bayesian Computation," *Computational Statistics & Data Analysis*, 124, 71–86. [525]

——— (2019a), "Simulation and Inference Algorithms for Stochastic Biochemical Reaction Networks: From Basic Concepts to State-of-the-Art," *Journal of The Royal Society Interface*, 16, 20180943. [512,525]

——— (2019b), "Using Experimental Data and Information Criteria to Guide Model Selection for Reaction–Diffusion Problems in Mathematical Biology," *Bulletin of Mathematical Biology*, 81, 1760–1804. [513,525]

——— (2020a), "A Practical Guide to Pseudo-Marginal Methods for Computational Inference in Systems Biology," *Journal of Theoretical Biology*, 496, 110255. [513]

Warne, D. J., Ebert, A., Drovandi, C., Hu, W., Mira, A., and Mengersen, K. (2020b), "Hindsight is 2020 Vision: A Characterisation of the Global Response to the COVID-19 Pandemic," *BMC Public Health*, 20, 1868. [512]

Warne, D. J., Sisson, S. A., and Drovandi (2021), "Vector Operations for Accelerating Expensive Bayesian Computations – A Tutorial Guide," *Bayesian Analysis*. [525]

Wilkinson, D. J. (2009), "Stochastic Modelling for Quantitative Description of Heterogeneous Biological Systems," *Nature Reviews Genetics*, 10, 122–133. [512,525]

Witelski, T. P. (1995), "Merging Traveling Waves for the Porous-Fisher's Equation," *Applied Mathematics Letters 8*, 57–62. [525]

Woods, M. L., and Barnes, C. P. (2016), "Mechanistic Modelling and Bayesian Inference Elucidates the Variable Dynamics of Double-Strand Break Repair," *PLOS Computational Biology*, 12, e1005131. [513]

Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., and Koeppl, H. (2012), "Moment-Based Inference Predicts Bimodality in Transient Gene Expression," *Proceedings of the National Academy of Sciences*, 109, 8340–8345. [523]