Multifidelity multilevel Monte Carlo to accelerate approximate Bayesian parameter inference for partially observed stochastic processes

David J. Warne [a,b,*], Thomas P. Prescott [c,d], Ruth E. Baker [d], Matthew J. Simpson [a,b]

[a] *School of Mathematical Sciences, Faculty of Science, Queensland University of Technology (QUT), Brisbane, Queensland 4001, Australia*
[b] *Centre for Data Science, QUT, Brisbane, Queensland 4001, Australia*
[c] *The Alan Turing Institute, London, NW1 2DB, United Kingdom*
[d] *Mathematical Institute, University of Oxford, Oxford, OX2 6GG, United Kingdom*

ARTICLE INFO

ABSTRACT

Models of stochastic processes are widely used in almost all fields of science. Theory validation, parameter estimation, and prediction all require model calibration and statistical inference using data. However, data are almost always incomplete observations of reality. This leads to a great challenge for statistical inference because the likelihood function will be intractable for almost all partially observed stochastic processes. This renders many statistical methods, especially within a Bayesian framework, impossible to implement. Therefore, computationally expensive likelihood-free approaches are applied that replace likelihood evaluations with realisations of the model and observation process. For accurate inference, however, likelihood-free techniques may require millions of expensive stochastic simulations. To address this challenge, we develop a new method based on recent advances in multilevel and multifidelity methods for parameter inference using partially observed Markov processes. Our novel approach combines the multilevel Monte Carlo telescoping summation, applied to a sequence of approximate Bayesian posterior targets, with a multifidelity rejection sampler that learns from computationally inexpensive model approximations to minimise the number of computationally expensive exact simulations required for accurate inference. We present the derivation of our new algorithm for likelihood-free Bayesian inference, discuss practical implementation details, and demonstrate substantial performance improvements. Using examples from systems biology, we demonstrate improvements of more than two orders of magnitude over standard rejection sampling techniques. Our approach is generally applicable to accelerate other sampling schemes, such as sequential Monte Carlo, to enable feasible Bayesian analysis for realistic practical applications in physics, chemistry, biology, epidemiology, ecology and economics. We provide source code implementations of our methods and demonstrations (available at https://github.com/davidwarne/MLMCandMultifidelityForABC).

* Corresponding author at: School of Mathematical Sciences, Faculty of Science, Queensland University of Technology (QUT), Brisbane, Queensland 4001, Australia.
E-mail address: david.warne@qut.edu.au (D.J. Warne).

1. Introduction

Stochastic processes are used to model complex systems in almost all fields of science and engineering. Partially observed stochastic processes result in some of the most computationally challenging problems for Bayesian inference [1–3]. Given the ubiquity of stochastic processes for real applications it is essential that efficient methods are developed to enable the analysis of modern high resolution data sets without sacrificing accuracy.

An important application of stochastic processes occurs in the study of cellular processes [4,5]. Here, stochastic models of biochemical reaction networks often provide a more accurate description of system dynamics than deterministic models [6]. This is largely due to intrinsic noise in the system dynamics of many biochemical processes that are significantly influenced by relatively low populations of certain chemical species [7]. For example, in eukaryotic cells, molecules that regulate gene expression occur in relatively low numbers; as a result, stochastic fluctuations have a direct effect on the production rates of proteins [4,8,9]. In addition, there are other interesting phenomena that occur in biological systems that can only be captured by stochastic models, for example, self-induced stochastic resonance [10–12], stochastic focusing [13], and stochastic bi-stability [14,15].

To quantify uncertainty in parameters or predictions, it is typical to consider an expectation of a function of an unknown vector of model parameters, $\theta \in \Theta$, conditional on some observational data, $\mathcal{D}$,

$$\mathbb{E}\left[f(\theta)\right] = \int_{\Theta} f(\theta) p(\theta \mid \mathcal{D}) \, \mathrm{d}\theta, \tag{1}$$

where $p(\theta \mid \mathcal{D})$ is the Bayesian posterior probability density [16],

$$p(\theta \mid \mathcal{D}) = \frac{\mathcal{L}(\theta; \mathcal{D}) p(\theta)}{p(\mathcal{D})}. \tag{2}$$

Here, $p(\theta)$ is the *a priori* probability density of unknown parameters, $\mathcal{L}(\theta; \mathcal{D})$ is the likelihood function that maps a parameter vector θ to the probability of the observations $\mathcal{D}$, and $p(\mathcal{D})$ is a normalising constant often referred to as the evidence. The function $f(\cdot)$ can be any scalar function of the parameters θ. For example, Equation (1) could represent a posterior moment, probability density or cumulative probability. Similarly $f(\cdot)$ could include a model prediction to enable equivalent expectation of the posterior predictive distribution.

Computational challenges arise due to the evidence term, given by

$$p(\mathcal{D}) = \int_{\Theta} \mathcal{L}(\theta; \mathcal{D}) p(\theta) \, \mathrm{d}\theta. \tag{3}$$

This term is rarely tractable and the expectation in Equation (1) must be numerically estimated using sampling techniques, such as Markov chain Monte Carlo (MCMC) [17,18] and sequential Monte Carlo (SMC) [19], that only require point-wise evaluation of the likelihood function, $\mathcal{L}(\theta; \mathcal{D})$.

For partially observed Markov processes, point-wise evaluation of the likelihood requires the solution to the forward Kolmogorov equation, which must be computed approximately. Therefore, standard Bayesian tools cannot be applied and likelihood-free methods are needed, such as approximate Bayesian computation (ABC) [20–22], pseudo-marginal methods [23–25], and Bayesian synthetic likelihood (BSL) [26–28]. In the machine learning literature, there are a multitude of likelihood-free, or simulation-based, approaches available [31], for example the use of deep neural networks to learn a surrogate model of the posterior [29] or the likelihood [30]. Regardless of the likelihood-free approach, realisations of the stochastic model are generated by stochastic simulation in place of likelihood function evaluation. Thus, the computational cost of evaluating Equation (1) depends on the efficiency of the stochastic simulation algorithm and the posterior sampler used as a basis for likelihood-free inference. For example, ABC can be implemented using either rejection sampling [32–34], MCMC [35] or SMC approaches [36–38]. However, if the model is even moderately expensive then such techniques are still infeasible.

Recently, there has been substantial research activity in the application of approximate model simulations and posterior samplers in combination with bias correction adjustments to accelerate likelihood-free applications that would be impractical otherwise. Techniques such as transport maps [44] and moment-matching transforms [45] aim to transform a set of approximate posterior samples, using a surrogate or reduced model, into posterior samples under an expensive exact model. Other approaches such as preconditioning utilise approximate models to inform efficient proposal mechanisms [45]. Various early-rejection and delayed-acceptance schemes [39–43] probabilistically simulate the accurate model based on the rejection/acceptance status of an approximation; this family of methods is generalised for ABC schemes under the multifidelity framework [46–48].

There has also been substantial research in the last decade on the use of control variates to improve the rate of convergence in *mean-square* of Monte Carlo estimates to expectations. In particular, the multilevel Monte Carlo (MLMC) method [49,50] expands an expectation as a telescoping summation of bias corrections and achieves variance reductions by exploiting path-wise convergence properties of numerical schemes for solving stochastic differential equations (SDEs) [51]

or discrete-state Markov processes [52–54]. Recently, the MLMC telescoping summation approach has also been successfully applied to Bayesian inference [55–59], as per Equation (1), including ABC-based samplers [60–62].

Many of the above approaches use approximate simulations in ways that are not always mutually exclusive. However, it is an open question as to how these various schemes could be combined to obtain compounding effects in performance.

1.1. Contribution

In this work, we derive and demonstrate an effective method for approximate Bayesian inference that combines the benefits of multifidelity ABC sampling with that of variance reduction with MLMC applied to the expectations with respect to an ABC posterior density. Acknowledging the wide range of variations in these algorithm classes, we specifically consider a natural application of these approaches to rejection samplers and extend the methods of Prescott and Baker [46], and Warne et al. [62]. Specifically, we apply the MLMC telescoping summation of Warne et al. [62] to the multifidelity ABC sampler of Prescott and Baker [46] including practical implementation of an adaptive tuning scheme for the multifidelity approach [63]. These methods combined achieve multiple orders of magnitude improvement over direct ABC rejection sampling. In addition, for situations where rejection sampling is practically intractable, our methodology is equally extendable to other sampling approaches, such as SMC, using developments by Prescott and Baker [47] and Jasra et al. [61]. Using biochemical reaction networks as a characteristic application area, we present key computational features of ABC methods and introduce the fundamental algorithms developed by Prescott and Baker [46] and Warne et al. [62]. We then highlight the distinct features of the inference problem exploited by each approach to accelerate ABC rejection sampling, and derive our novel approach by leveraging the complementary nature of these features. We then explore various performance results using stochastic models of Michaelis–Menten enzyme kinetics and the repressilator gene regulatory network. These examples enable us to establish requirements for practical application of our approach and provide rules-of-thumb to simplify the tuning of algorithm parameters. Finally, we demonstrate the efficacy of our multifidelity MLMC approach using a challenging model of a two–step MAPK cascade reaction, which forms the basis for many real biological functions such as cell-to-cell signalling through the epithelial growth factor receptor (EGFR) [64,65]. Our results demonstrate the potential of multilevel and multifidelity methods to accelerate ABC sampling by several orders of magnitude.

2. Methods

In this section we describe standard numerical methods for simulation and inference of stochastic biochemical network models [22,66,67]. Then we present recent developments in MLMC [22,62] and multifidelity [46] methods for inference, and develop novel extensions that achieve the benefits from both the MLMC and multifidelity approaches. Finally, we highlight practical challenges and provide guidelines for the application and tuning of our new method.

2.1. Stochastic models of biochemical reaction networks

We consider biochemical reaction network models that involve a well-mixed population of $\mathcal{N}$ chemical species, $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_\mathcal{N}$, that react via a network of $\mathcal{M}$ chemical reactions,

$$\sum_{i=1}^{\mathcal{N}} \nu_{i,j}^- \mathcal{X}_i \to \sum_{i=1}^{\mathcal{N}} \nu_{i,j}^+ \mathcal{X}_i, \quad j = 1, 2, \ldots, \mathcal{M},$$

where $\nu_{i,j}^-$ and $\nu_{i,j}^+$ are, respectively, the reactant and product stoichiometries for species $\mathcal{X}_i$ in reaction j. Given the state of the system at time t, $\mathbf{X}_t = [X_{1,t}, X_{2,t}, \ldots, X_{\mathcal{N},t}]^{\mathrm{T}}$ with $X_{i,t}$ denoting the copy number of species $\mathcal{X}_i$ at time t, it can be shown that, for a sufficiently small time interval $[t, t + \Delta t)$, the probability of reaction j occurring within this interval is proportional to $a_j(\mathbf{X}_t)\Delta t$ where $a_j(\mathbf{X}_t)$ is the propensity function of reaction j. Generally, the propensity function will take the form

$$a_j(\mathbf{X}_t, \boldsymbol{\theta}) = k_j \prod_{i=1}^{\mathcal{N}} \nu_{i,j}^-! \binom{X_{i,t}}{\nu_{i,j}^-},$$

where k_j is the non-dimensionalised kinetic rate parameter for reaction j. However, other more complex forms are also possible that may depend on a general parameter vector such as $\boldsymbol{\theta} = [k_1, k_2, \ldots, k_\mathcal{M}, \boldsymbol{\lambda}]$ where $\boldsymbol{\lambda}$ could include other parameters like hill constants, observation error, or domain sizes.

Mathematically, the stochastic dynamics of a biochemical network is governed by a discrete-state, continuous-time Markov process [67–70]. The transitional density function of this process, $p(\mathbf{X}_t \mid \mathbf{X}_s, \boldsymbol{\theta})$, describes the probability of the system state at time t given the state at a previous time s. To obtain $p(\mathbf{X}_t \mid \mathbf{X}_s, \boldsymbol{\theta})$ one needs to solve the forward Kolmogorov equation, also referred to as the chemical master equation (CME) [67,71],

$$\frac{\mathrm{d}p(\mathbf{X}_t \mid \mathbf{X}_s), \boldsymbol{\theta}}{\mathrm{d}t} = \sum_{j=1}^{\mathcal{M}} a_j(\mathbf{X}_t - \boldsymbol{\nu}_j; \boldsymbol{\theta}) p(\mathbf{X}_t - \boldsymbol{\nu}_j \mid \mathbf{X}_s, \boldsymbol{\theta}) - a_j(\mathbf{X}_t; \boldsymbol{\theta}) p(\mathbf{X}_t \mid \mathbf{X}_s, \boldsymbol{\theta}), \tag{4}$$

where $\boldsymbol{v}_j$ is the state change corresponding to the occurrence of reaction j, that is $\boldsymbol{v}_j = [v_{1,j}^+ - v_{1,j}^-, v_{2,j}^+ - v_{2,j}^-, \dots, v_{\mathcal{N},j}^+ - v_{\mathcal{N},j}^-]^\mathsf{T}$. While we focus on Markov processes in this work, equivalent concepts exist for non-Markovian systems; though with more complexity [9]. Our new methods may be generalised to non-Markovian models without substantial modification.

2.1.1. Stochastic simulation methods

The CME is intractable for all but the simplest of models [22,66,67]. As a result, exact stochastic simulation [68,72,73] is required to study the system behaviour without introducing potentially substantial bias [45]. Stochastic simulation schemes generate realisations of the network state, $\mathbf{X}_t$, from some initial time t_0 up to a termination time T. Exact stochastic simulation schemes, such as Gillespie's direct method [68] (Algorithm 1), simulate every reaction event, and are computationally prohibitive for systems with very large copy numbers or very high reaction rates.

Algorithm 1 Gillespie's direct method for stochastic simulation.

1: Initialise $t = t_0$ and $\mathbf{X} = \mathbf{x}_0$;
2: **loop**
3:     Set $a_0 \leftarrow \sum_{j=1}^{\mathcal{M}} a_j(\mathbf{X})$;
4:     Sample next reaction time $\Delta t \sim \mathrm{Exp}(a_0)$;
5:     **if** $t + \Delta t > T$ **then**
6:         **return**
7:     **else**
8:         Select reaction $j \in [1, 2, \dots, \mathcal{M}]$ with probabilities

$$\mathbb{P}(j = 1) = a_1(\mathbf{X})/a_0, \quad \mathbb{P}(j = 2) = a_2(\mathbf{X})/a_0, \quad \dots, \quad \mathbb{P}(j = \mathcal{M}) = a_{\mathcal{M}}(\mathbf{X})/a_0;$$

9:         Set $\mathbf{X} \leftarrow \mathbf{X} + \boldsymbol{v}_j$ and $t \leftarrow t + \Delta t$.
10:    **end if**
11: **end loop**

Various approximate stochastic simulation schemes, such as the tau-leaping method [74] (Algorithm 2), can be applied [67,74–77] to improve the computational performance, but there will be bias incurred due to the simplifying approximations [22,78]. For example, assume the propensities do not change substantially over a time interval of length τ. The resulting sample path $\mathbf{Z}_t$ will be a discrete-time Markov chain approximation to a true path $\mathbf{X}_t$ from the full continuous-time process.

Algorithm 2 The tau-leaping method for approximate stochastic simulation.

1: Initialise $t = t_0$, and $\mathbf{Z} = \mathbf{x}_0$;
2: **while** $t + \tau > T$ **do**
3:     Generate event counts, $Y_j \sim \mathrm{Po}(a_j(\mathbf{Z})\tau)$, for $j = 1, 2, \dots, \mathcal{M}$;
4:     Set $\mathbf{Z} \leftarrow \mathbf{Z} + \sum_{j=1}^{\mathcal{M}} Y_j \boldsymbol{v}_j$, and $t \leftarrow t + \tau$.
5: **end while**

2.1.2. Acceleration using multilevel Monte Carlo

Often the goal of stochastic simulation is to estimate the expectation, $\mathbb{E}[f(\mathbf{X}_T)]$, where $f(\cdot)$ is a function of the process state at time $t = T > 0$. Note that this expectation can, through the specification of $f(\cdot)$, resolve to any raw or central moment, it may also be used to estimate the full probability density or cumulative distribution of the system state $\mathbf{X}_T$ [62, 79,80]. MLMC provides a mechanism to exploit approximations that computationally accelerate the estimation [22,51,53]. Assume we have a stochastic process $\{\mathbf{X}_t\}_{t \geq 0}$, such as biochemical reaction network model, that is computationally expensive to simulate with the Gillespie direct method (Algorithm 1) or equivalent. Now consider a sequence of L stochastic processes, $\{\{\mathbf{Z}_{t,\ell}\}_{t \geq 0}\}_{\ell=1}^{\ell=L}$, that approximate $\{\mathbf{X}_t\}_{t \geq 0}$. This sequence is constructed such that the bias decreases and computational cost increases with ℓ. For example, $\{\mathbf{Z}_{t,\ell}\}_{t \geq 0}$ could be a tau-leaping approximation with time step $\tau_\ell = c2^{-\ell}$ and c constant for all $\ell = 1, 2, \dots, L$. The insight of Giles [51] was to expand the desired expectation as a telescoping summation by exploiting linearity of expectation,

$$\mathbb{E}[f(\mathbf{X}_T)] = \mathbb{E}[f(\mathbf{Z}_{T,1})] + \mathbb{E}[f(\mathbf{X}_T) - f(\mathbf{Z}_{T,L})] + \sum_{\ell=2}^{L} \mathbb{E}[f(\mathbf{Z}_{T,\ell}) - f(\mathbf{Z}_{T,\ell-1})], \tag{5}$$

where each of the difference terms acts to correct for the bias of the initial biased approximation, $\mathbb{E}[f(\mathbf{Z}_{T,1})]$. Giles [51] demonstrated that substantial computational improvements can be obtained if positive correlations can be induced between the terms in the bias corrections, that is between $\mathbf{Z}_{T,\ell}$ and $\mathbf{Z}_{T,\ell-1}$ for $\ell = 2, \dots, L$, and between $\mathbf{X}_T$ and $\mathbf{Z}_{T,L}$. For SDEs and discrete-state continuous-time Markov processes, coupling schemes that induce sufficiently strong positive correlations have been well studied [22,50,52,54] and performance improvements of many orders of magnitude can be obtained without any loss in accuracy in terms of mean-square error.

2.2. Approximate Bayesian computation

For the purposes of inference, the likelihood function for sequences of n observations, denoted by $\mathbf{Y}_{\text{obs}} = [\mathbf{y}_{\text{obs}}(t_1), \mathbf{y}_{\text{obs}}(t_2), \ldots, \mathbf{y}_{\text{obs}}(t_n)]$, is given by

$$
p(\mathbf{Y}_{\text{obs}} \mid \boldsymbol{\theta}) = \int_{\mathbb{X}^{n+1}} p(\mathbf{Y}_{\text{obs}}, \mathbf{X}_{t_0}, \ldots, \mathbf{X}_{t_n} \mid \boldsymbol{\theta}) \prod_{i=0}^{n} \mathrm{d}\mathbf{x}_{t_i}
$$

$$
= \int_{\mathbb{X}^{n+1}} p(\mathbf{X}_{t_0}) \mathrm{d}\mathbf{X}_{t_0} \prod_{i=1}^{n} g(\mathbf{y}_{\text{obs}}(t_i) \mid \mathbf{X}_{t_i}, \boldsymbol{\theta}) p(\mathbf{X}_{t_i} \mid \mathbf{X}_{t_{i-1}}, \boldsymbol{\theta}) \, \mathrm{d}\mathbf{X}_{t_i}, \tag{6}
$$

where $g(\mathbf{y}_{\text{obs}}(t_i) \mid \mathbf{X}_{t_i}, \boldsymbol{\theta})$ is the probability density of noisy observation $\mathbf{y}_{\text{obs}}(t_i)$, given the true state, $\mathbf{X}_{t_i} \in \mathbb{X} \subseteq \mathbb{N}^{\mathcal{N}}$, at time t_i, and $p(\mathbf{X}_{t_i} \mid \mathbf{X}_{t_{i-1}}, \boldsymbol{\theta})$ is the solution to the CME for the state transition over the time interval $(t_{i-1}, t_i]$. The intractability of the CME immediately implies the intractability of the Bayesian inference problem. However, depending on the structure of the observation process, Equation (6) may still be intractable even when Equation (4) has an analytical solution. Therefore, almost any practical application of partially observed continuous-time Markov processes will require some form of likelihood-free inference.

ABC, pseudo-marginal methods and BSL are all possible approaches to avoid the dependence of Bayesian inference on the likelihood (Equation (6)) by generating simulated data from the likelihood using exact stochastic simulation along with the observation noise process. Here, we focus on the ABC approach that is based on the posterior approximation,

$$
p(\boldsymbol{\theta} \mid \mathbf{Y}_{\text{obs}}) \approx p(\boldsymbol{\theta} \mid \rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) \leq \epsilon) \propto \mathbb{P}(\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) \leq \epsilon \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}), \tag{7}
$$

where $\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s)$ is a discrepancy metric between noisy observations, $\mathbf{Y}_{\text{obs}}$, and simulated noisy observations, $\mathbf{Y}_s$, and ϵ is a sufficiently small discrepancy threshold. The most direct method to implement ABC is to use rejection sampling (Algorithm 3). Here, independent identically distributed samples of rate parameters are drawn from the prior, and the

---

Algorithm 3 ABC rejection sampling to generate N approximated posteriors samples.

1: **for** $i \in [1, 2, \ldots, N]$ **do**
2:     **repeat**
3:         Sample the prior $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta})$;
4:         Generate simulated data $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta}^*)$;
5:     **until** $\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) \leq \epsilon$
6:     $\boldsymbol{\theta}^i \leftarrow \boldsymbol{\theta}^*$.
7: **end for**

---

sample is accepted if a resulting stochastic simulation of the model is within ϵ of the observations under the discrepancy metric. In practice, rejection sampling may not be computationally feasible due to the prohibitively high rejection rates when the data dimensionality is high. There are many techniques that can be applied to improve the computational efficiency of ABC methods that we do not discuss here due to the wealth of available literature [20,21,35,36,38,81]. Instead we specifically investigate recent developments in MLMC and multifidelity methods [46,62], as they provide a new avenue to explore computational improvements for ABC inference. While we focus on these methods as applied directly to ABC rejection sampling (Algorithm 3), we note that our work is also applicable to other schemes such as SMC [47,61].

2.3. Multilevel Monte Carlo and multifidelity methods for ABC inference

In this section, we describe two recent methods for acceleration of ABC inference that are foundational to the main contribution of this work. These methods are presented by Warne et al. [62] and Prescott and Baker [46], respectively. Both methods build upon the MLMC work of Giles [51] and Rhee and Glynn [82] for efficient computation of expectations with respect to a stochastic process [51,52]. We refer the reader to Warne et al. [22], Lester et al. [53], Schnoerr et al. [66], and Peherstorfer et al. [83], for accessible introductions to simulation and inference methods including MLMC and multifidelity methods.

2.3.1. Multilevel rejection sampling

For ABC inference, Warne et al. [22,62] consider the direct application of the telescoping summation (Equation (5)) to a sequence of L ABC rejection samplers (Algorithm 3). That is,

$$
\mathbb{E}[f(\boldsymbol{\theta}_L)] = \mathbb{E}[f(\boldsymbol{\theta}_1)] + \sum_{\ell=2}^{L} \mathbb{E}[f(\boldsymbol{\theta}_\ell) - f(\boldsymbol{\theta}_{\ell-1})], \tag{8}
$$

with $\boldsymbol{\theta}_\ell \sim p(\cdot \mid \rho(\mathbf{Y}_{\mathrm{obs}}, \mathbf{Y}_s) \le \epsilon_\ell)$ for threshold $\epsilon_\ell = \epsilon_0 m^{-\ell}$ for $\ell = 1, \ldots, L$, where ϵ_0 is a large discrepancy threshold leading to a high acceptance rate (typically close to the prior) and $m > 1$. The greatest challenge in application of MLMC for inference is the construction of a coupling to generate positively correlated sample pairs $(\boldsymbol{\theta}_\ell, \boldsymbol{\theta}_{\ell-1})$ without introducing additional bias that would violate the telescoping summation. For this Warne et al. [62] apply a novel construction using the marginal empirical distribution at level ℓ and the inverse marginal distributions obtained from the bias corrections up to level $\ell - 1$. That is, given N_ℓ i.i.d. samples $\boldsymbol{\theta}_\ell^1, \ldots, \boldsymbol{\theta}_\ell^{N_\ell}$, one generates

$$\tilde{\boldsymbol{\theta}}_{\ell-1}^i = [\hat{F}_{\ell-1,1}^{-1}(\bar{F}_{\ell,1}(\theta_{\ell,1})), \ldots, \hat{F}_{\ell-1,k}^{-1}(\bar{F}_{\ell,k}(\theta_{\ell,k}))], \quad \text{for } i = 1, \ldots, N_\ell, \tag{9}$$

where $\bar{F}_{\ell,j}(\cdot)$ is the marginal empirical distribution of the jth dimension of $\boldsymbol{\theta}_\ell$, denoted by $\theta_{\ell,j}$, and $\hat{F}_{\ell-1,j}^{-1}(\cdot)$ is an estimate of the marginal distribution inverse for the jth dimension of $\boldsymbol{\theta}_{\ell-1}$. The result is a sampling procedure for $(\boldsymbol{\theta}_\ell, \tilde{\boldsymbol{\theta}}_\ell)$ with a positive correlation induced between pairs. Warne et al. [62] note that this approach only strictly satisfies the telescoping summation in each marginal rather than the full distribution of $\boldsymbol{\theta}_L$. As a result, the sequence of discrepancies, $\epsilon_1, \ldots, \epsilon_L$, needs to be chosen so that the correlation structures between $p(\cdot \mid \rho(\mathbf{Y}_{\mathrm{obs}}, \mathbf{Y}_s) \le \epsilon_\ell)$ and $p(\cdot \mid \rho(\mathbf{Y}_{\mathrm{obs}}, \mathbf{Y}_s) \le \epsilon_{\ell-1})$ are similar. This requires the bias correction terms in Equation (8) to be computed in order from $\ell = 2 \to L$ due to the dependence on the previous level for the estimation of the marginal distribution inverses at level $\ell - 1$ (Equation (9)). The complete process is given in Algorithm 4. Given an appropriate discrepancy sequence, the sample size sequence, $N_1, \ldots, N_L$, can be optimised to achieve improved convergence rates in mean squared error (see Section 6.5.1 in Warne [84], Theorem 3.1 in Giles [51], and Section 3.2 in Lester et al. [53]). However, the target accuracy needs to be sufficiently small for this improved convergence rate to take effect due to an overhead computational cost in generating the trial simulations needed to optimise the sequence $N_1, \ldots, N_L$.

Algorithm 4 Multilevel Monte Carlo for ABC rejection sampling (MLMC-ABC).

1: Initialise $\epsilon_1, \ldots, \epsilon_L$, $N_1, \ldots, N_L$ and prior $p(\boldsymbol{\theta})$.
2: **for** $\ell = 1, \ldots, L$ **do**
3:     Sample $\boldsymbol{\theta}_\ell^1, \ldots, \boldsymbol{\theta}_\ell^{N_\ell} \sim p(\cdot \mid \rho(\mathbf{Y}_{\mathrm{obs}}, \mathbf{Y}_s) \le \epsilon_\ell)$ using ABC rejection sampling;
4:     Set $\bar{F}_{\ell,j}(s) \leftarrow \sum_{i=1}^{N_\ell} \mathbb{1}_{(-\infty,s]}(\theta_{\ell,j})/N_\ell$ for $j = 1, \ldots, k$;
5:     **if** $\ell = 1$ **then**
6:         $\hat{f}_\ell \leftarrow \sum_{i=1}^{N_\ell} f(\boldsymbol{\theta}_\ell^i)/N_\ell$;
7:     **else**
8:         **for** $i = 1, \ldots, N_\ell$ **do**
9:             Set $\tilde{\boldsymbol{\theta}}_{\ell-1}^i \leftarrow \left[\hat{F}_{\ell-1,1}^{-1}\left(\bar{F}_{\ell,1}^{N_\ell}\left(\theta_{\ell,1}^i\right)\right), \ldots, \hat{F}_{\ell-1,k}^{-1}\left(\bar{F}_{\ell,k}^{N_\ell}\left(\theta_{\ell,k}^i\right)\right)\right]$;
10:         **end for**
11:         Set $\hat{F}_{\ell,j}(s) \leftarrow \hat{F}_{\ell-1,j}(s) + \sum_{i=1}^{N_\ell} \left[\mathbb{1}_{(-\infty,s]}\left(\theta_{\ell,j}^i\right) - \mathbb{1}_{(-\infty,s]}\left(\tilde{\theta}_{\ell-1,j}^i\right)\right]/N_\ell$, $j = 1, \ldots, k$;
12:         Set $\hat{f}_\ell \leftarrow \hat{f}_{\ell-1} + \sum_{i=1}^{N_\ell} \left[f(\boldsymbol{\theta}_\ell^i) - f(\tilde{\boldsymbol{\theta}}_{\ell-1}^i)\right]/N_\ell$.
13:     **end if**
14: **end for**

2.3.2. Multifidelity rejection sampling

An alternative approach developed by Prescott and Baker [46] utilises the telescoping summation in a probabilistic manner akin to the de-biasing approach of Rhee and Glynn [82]. Instead of considering a sequence of ABC samplers defined in terms of acceptance thresholds, they consider ABC rejection samples with different simulator fidelities. That is, a high fidelity simulator $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta})$ that is computationally expensive, such as the Gillespie direct method, and a computationally cheaper low fidelity simulator $\tilde{\mathbf{Y}}_s \sim \tilde{s}(\cdot \mid \boldsymbol{\theta})$, such as tau-leaping approximation with time step τ. The idea is to perform ABC rejection sampling with the low fidelity simulator with discrepancy $\tilde{\rho}(\mathbf{Y}_{\mathrm{obs}}, \tilde{\mathbf{Y}}_s)$ and acceptance threshold $\tilde{\epsilon}$, and then perform a probabilistic bias correction that requires ABC rejection using a high fidelity simulation with discrepancy $\rho(\mathbf{Y}_{\mathrm{obs}}, \mathbf{Y}_s)$ and acceptance threshold ϵ. The resulting estimator is given by

$$\mathbb{E}[f(\boldsymbol{\theta})] \approx \frac{\sum_{i=1}^N w(\boldsymbol{\theta}^i) f(\boldsymbol{\theta}^i)}{\sum_{i=1}^N w(\boldsymbol{\theta}^i)}, \tag{10}$$

where $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^N$ are samples from the prior and the weight function is

$$w(\boldsymbol{\theta}) = \mathbb{1}_{(0,\tilde{\epsilon}]}\left(\tilde{\rho}(\mathbf{Y}_{\mathrm{obs}}, \tilde{\mathbf{Y}}_s)\right) + \xi\left[\mathbb{1}_{(0,\epsilon]}(\rho(\mathbf{Y}_{\mathrm{obs}}, \mathbf{Y}_s)) - \mathbb{1}_{(0,\tilde{\epsilon}]}\left(\tilde{\rho}(\mathbf{Y}_{\mathrm{obs}}, \tilde{\mathbf{Y}}_s)\right)\right]. \tag{11}$$

Here, ξ is a random variable given by

$$\xi = \frac{\mathbb{1}_{(0,\eta(\tilde{\mathbf{Y}}_s)]}(U)}{\eta(\tilde{\mathbf{Y}}_s)}, \tag{12}$$

where $U \sim \mathcal{U}(0,1)$ and $\eta(\tilde{\mathbf{Y}}_s)$ is the probability of generating a high-fidelity simulation given a realisation from the low-fidelity simulation. This so-called continuation probability can take many forms, however, the method originally proposed by Prescott and Baker [46] is

$$\eta(\tilde{\mathbf{Y}}_s) = \eta_1 \mathbb{1}_{(0,\tilde{\epsilon}]} \left(\tilde{\rho}(\mathbf{Y}_{\text{obs}}, \tilde{\mathbf{Y}}_s) \right) + \eta_2 \mathbb{1}_{(\tilde{\epsilon},\infty)} \left(\tilde{\rho}(\mathbf{Y}_{\text{obs}}, \tilde{\mathbf{Y}}_s) \right), \tag{13}$$

where η_1 and η_2 are, respectively, the continuation probabilities when the low-fidelity simulation, $\tilde{\mathbf{Y}}_s \sim \tilde{s}(\cdot \mid \boldsymbol{\theta})$, is accepted and rejected. The sampler proceeds according to Algorithm 5.

Algorithm 5 Multifidelity ABC rejection sampling (MF-ABC).

1: Initialise $\eta_1, \eta_2, N, \epsilon, \tilde{\epsilon}, \rho(\mathbf{Y}_{\text{obs}}, \cdot), \tilde{\rho}(\mathbf{Y}_{\text{obs}}, \cdot)$ and prior $p(\boldsymbol{\theta})$;
2: **for** $i = 1, 2, \ldots, N$ **do**
3:      Sample the prior $\boldsymbol{\theta}^i \sim p(\boldsymbol{\theta})$;
4:      Simulate the low-fidelity model $\tilde{\mathbf{Y}}_s \sim \tilde{s}(\cdot \mid \boldsymbol{\theta}_i)$;
5:      Set $\tilde{w} \leftarrow \mathbb{1}_{(0,\tilde{\epsilon}]} \left(\tilde{\rho}(\mathbf{Y}_{\text{obs}}, \tilde{\mathbf{Y}}_s) \right)$ and $\eta \leftarrow \eta_1 \tilde{w} + \eta_2 (1 - \tilde{w})$;
6:      **if** $U < \eta$ where $U \sim \mathcal{U}(0,1)$ **then**
7:          Simulate the high-fidelity model $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta}_i)$;
8:          Set $w_i \leftarrow \tilde{w} + (\mathbb{1}_{(0,\epsilon]} (\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s)) - \tilde{w})/\eta$;
9:      **else**
10:         Set $w_i \leftarrow \tilde{w}$;
11:      **end if**
12: **end for**
13: Set $\hat{f} \leftarrow \sum_{i=1}^{N} w_i f(\boldsymbol{\theta}^i) / \sum_{i=1}^{N} w_i$.

Prescott and Baker [46] prove that when $\eta_1 > 0$ and $\eta_2 > 0$ the multifidelity estimator (Equation (10)) is asymptotically unbiased. The multifidelity ABC estimator may be viewed as a form of importance sampling and thereby has a bias of order $\mathcal{O}(1/N)$ (Supplementary Material). Unlike standard rejection and importance samplers, the multifidelity weights can be negative. Despite this fact, the multifidelity ABC estimator is both asymptotically unbiased and consistent [63]. Under certain conditions η_1, η_2 may be optimised such that the multifidelity ABC rejection sampler (Algorithm 5) is more computationally efficient than direct rejection sampling with the high fidelity simulator (Algorithm 3). This improvement represents a decrease in the average computational cost for a given target mean-square error, but it does not improve the convergence rate. However, the multifidelity approach will not incur a significant overhead for selecting the continuation probabilities since adaptive schemes can be applied [63] (Supplementary Material).

2.4. Multifidelity MLMC for ABC inference

The MLMC and multifidelity approaches to ABC inference, MLMC-ABC and MF-ABC, obtain computational improvements in distinct ways that are also complementary. MLMC-ABC (Algorithm 4) combines a sequence of ABC rejection samplers with different discrepancy thresholds [62] while the stochastic simulation scheme is fixed and must be unbiased. Conversely, MF-ABC (Algorithm 5) combines two stochastic simulation schemes [46] with the other elements of the ABC sampler remaining largely unchanged. Computationally, MLMC-ABC empirically improves the convergence rate of the mean-square error as $\epsilon \to 0$ at the expense of a tuning step that incurs an additional cost [62]. This is consistent with theoretical and empirical results from other MLMC applications [57,60,61]. In addition, MF-ABC reduces the average simulation cost without improvements in the convergence rate [46,63]. Our novel contribution in this work is to show how these methods can be combined to exploit the computational advantages of both.

We now derive our new method, called *multifidelity MLMC for ABC* (MF-MLMC-ABC). Assume we have two sequences of L ABC acceptance thresholds, $\{\epsilon_\ell\}_{\ell=1}^{\ell=L}$ and $\{\tilde{\epsilon}_\ell\}_{\ell=1}^{\ell=L}$, with $\epsilon_\ell > \epsilon_{\ell+1}$ and $\tilde{\epsilon}_\ell > \tilde{\epsilon}_{\ell+1}$ for all $\ell = 1, \ldots, L-1$, and a sequence of L time-step lengths, $\{\tau_\ell\}_{\ell=1}^{\ell=L}$. Note that there are no constraints on the relation between the two acceptance thresholds sequences, nor any requirement that the sequence of time-step lengths is strictly decreasing or even monotonic. Given a set of model parameters, $\boldsymbol{\theta} \in \Theta$, let $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta})$ and $\mathbf{Y}_s^\tau \sim s^\tau(\cdot \mid \boldsymbol{\theta})$ denote, respectively, exact stochastic simulation (i.e., using the Gillespie direct method [68]) and approximate stochastic simulation with time-step τ (i.e., using the tau-leaping method [74]). Finally, let $\rho(\mathbf{Y}_{\text{obs}}, \cdot)$ and $\rho^\tau(\mathbf{Y}_{\text{obs}}, \cdot)$ be discrepancy metrics used to compare observed data $\mathbf{Y}_{\text{obs}}$ with, respectively, exact simulation output, $\mathbf{Y}_s$, or approximate simulation output, $\mathbf{Y}_s^\tau$.

Given the above notation, for any $\ell = 1, \ldots, L$ we can apply MF-ABC sampling using weights,

$$w^{\tau_\ell}(\boldsymbol{\theta}_\ell) = \mathbb{1}_{(0,\tilde{\epsilon}_\ell]} \left(\rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell}) \right) + \frac{\mathbb{1}_{(0,\eta^{\tau_\ell}(\mathbf{Y}_s^{\tau_\ell})]}(U)}{\eta^{\tau_\ell}(\mathbf{Y}_s^{\tau_\ell})} \left[\mathbb{1}_{(0,\epsilon_\ell]} (\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s)) - \mathbb{1}_{(0,\tilde{\epsilon}_\ell]} \left(\rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell}) \right) \right], \tag{14}$$

with continuation probability function

$$\eta^{\tau_\ell}(\mathbf{Y}_s^{\tau_\ell}) = \eta_{\ell,1} \mathbb{1}_{(0,\tilde{\epsilon}_\ell]} \left(\rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell}) \right) + \eta_{\ell,2} \mathbb{1}_{(\tilde{\epsilon}_\ell,\infty)} \left(\rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell}) \right), \tag{15}$$

Algorithm 6 Multifidelity multilevel Monte Carlo for ABC rejection sampling (MF-MLMC-ABC).

1: Initialise $\{\epsilon_\ell\}_{\ell=1}^{\ell=L}$, $\{\tilde{\epsilon}_\ell\}_{\ell=1}^{\ell=L}$, $\{\tau_\ell\}_{\ell=1}^{\ell=L}$, $\{N_\ell\}_{\ell=1}^{\ell=L}$, $\{(\eta_{\ell,1}, \eta_{\ell,2})\}_{\ell=1}^{\ell=L}$, $\rho(\mathbf{Y}_{\text{obs}}, \cdot)$, $\rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \cdot)$ and prior $p(\boldsymbol{\theta})$;
2: **for** $\ell = 1, \ldots, L$ **do**
3:     **for** $i = 1, 2, \ldots, N_\ell$ **do**
4:         Sample the prior $\boldsymbol{\theta}_\ell^i \sim p(\boldsymbol{\theta})$;
5:         Simulate the low-fidelity model with $\tau = \tau_\ell$, $\mathbf{Y}_s^{\tau_\ell} \sim s^{\tau_\ell}(\cdot \mid \boldsymbol{\theta}_\ell^i)$;
6:         Set $\tilde{w} \leftarrow \mathbb{1}_{(0,\tilde{\epsilon}_\ell]} \left(\rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell}) \right)$ and $\eta^{\tau_\ell} \leftarrow \eta_{\ell,1}\tilde{w} + \eta_{\ell,2}(1-\tilde{w})$;
7:         **if** $U < \eta^{\tau_\ell}$ where $U \sim \mathcal{U}(0,1)$ **then**
8:             Simulate the high-fidelity model $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta}_\ell^i)$;
9:             Set $w_\ell^i \leftarrow \tilde{w} + (\mathbb{1}_{(0,\epsilon]}(\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s)) - \tilde{w})/\eta^{\tau_\ell}$;
10:         **else**
11:             Set $w_\ell^i \leftarrow \tilde{w}$;
12:         **end if**
13:     **end for**
14:     Set $W_\ell = 1 / \sum_{i=1}^{N_\ell} w_\ell^i$;
15:     Set $\bar{F}_{\ell,j}(s) \leftarrow W_\ell \sum_{i=1}^{N_\ell} w_\ell^i \mathbb{1}_{(-\infty,s]}(\theta_{\ell,j})$ for $j = 1, \ldots, k$;
16:     **if** $\ell = 1$ **then**
17:         $\hat{f}_\ell \leftarrow W_\ell \sum_{i=1}^{N_\ell} w_\ell^i f(\boldsymbol{\theta}_\ell^i)$;
18:     **else**
19:         **for** $i = 1, \ldots, N_\ell$ **do**
20:             Set $\tilde{\boldsymbol{\theta}}_{\ell-1}^i \leftarrow \left[\hat{F}_{\ell-1,1}^{-1} \left(\bar{F}_{\ell,1}^{N_\ell}\left(\theta_{\ell,1}^i\right) \right), \ldots, \hat{F}_{\ell-1,k}^{-1} \left(\bar{F}_{\ell,k}^{N_\ell}\left(\theta_{\ell,k}^i\right) \right) \right]$;
21:         **end for**
22:         Set $\hat{F}_{\ell,j}(s) \leftarrow \hat{F}_{\ell-1,j}(s) + W_\ell \sum_{i=1}^{N_\ell} w_\ell^i \left[\mathbb{1}_{(-\infty,s]}\left(\theta_{\ell,j}^i\right) - \mathbb{1}_{(-\infty,s]}\left(\tilde{\theta}_{\ell-1,j}^i\right) \right]$, $j = 1, \ldots, k$;
23:         Set $\hat{f}_\ell \leftarrow \hat{f}_{\ell-1} + W_\ell \sum_{i=1}^{N_\ell} w_\ell^i \left[f(\boldsymbol{\theta}_\ell^i) - f(\tilde{\boldsymbol{\theta}}_{\ell-1}^i) \right]$.
24:     **end if**
25: **end for**

---

for constants $\eta_{\ell,1} > 0$, $\eta_{\ell,2} > 0$ for each $\ell = 1, \ldots, L$ and $U \sim \mathcal{U}(0,1)$. Based on the results of Prescott and Baker [46], an expectation estimated this way would be aysmptotically unbiased with respect to the ABC posterior under the exact stochastic simulation and discrepancy measure, that is, $\boldsymbol{\theta}_\ell \sim p(\cdot \mid \rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) \leq \epsilon_\ell)$. This provides a connection to the MLMC-ABC telescoping summation in Equation (8). Therefore, we can apply the MF-ABC estimator (Equation (10)) using weights defined by Equation (14) to each of the L terms in the MLMC-ABC telescoping summation (Equation (4)) and thereby arrive at the MF-MLMC-ABC estimator,

$$\mathbb{E}[f(\boldsymbol{\theta}_L)] \approx \hat{f} = \sum_{\ell=1}^{L} \sum_{i=1}^{N_\ell} \frac{w^{\tau_\ell}(\boldsymbol{\theta}_\ell^i) g_\ell(\boldsymbol{\theta}_\ell^i)}{\sum_{j=1}^{N_\ell} w^{\tau_\ell}(\boldsymbol{\theta}_\ell^j)}, \tag{16}$$

where

$$g_\ell(\boldsymbol{\theta}_\ell^i) = \begin{cases} f(\boldsymbol{\theta}_\ell^i) & \text{if } \ell = 1 \\ f(\boldsymbol{\theta}_\ell^i) - f(\tilde{\boldsymbol{\theta}}_{\ell-1}^i) & \text{if } \ell > 1 \end{cases}, \tag{17}$$

and $\tilde{\boldsymbol{\theta}}_{\ell-1}^i$ is constructed from $\boldsymbol{\theta}_\ell^i$ and estimated marginal distribution functions obtained from the previous $\ell - 1$ terms, as given in Equation (9), to implement an approximate coupling between levels [22,62]. Due to the properties of MF-ABC, Equation (16) is an asymptotically unbiased estimator of Equation (8) [46] and therefore an asymptotically unbiased estimator of $\mathbb{E}[f(\boldsymbol{\theta}_L)]$ up to the approximate coupling scheme [62].

We therefore arrive at the MF-MLMC-ABC method presented in Algorithm 6. Note that the proposed approach, just as with MLMC-ABC and MF-ABC, requires a number of algorithmic hyperparameters to be selected appropriately to ensure efficient sampling. In the next section we discuss theoretical results that guide how these parameters should be selected.

2.5. Optimal algorithm configuration

There are several important algorithmic hyperparameters that must be appropriately chosen to practically apply the MF-MLMC-ABC method. Each of these has various influences on the accuracy and performance of the MF-MLMC-ABC method. In this section, we will step through each of these algorithmic hyperparameters and provide theoretical results to optimally configure the method. The algorithmic hyperparameters that require optimisation are the number of levels, L, the form of the sequences $\{\epsilon_\ell\}_{\ell=1}^{\ell=L}$ and $\{\tau_\ell\}_{\ell=1}^{\ell=L}$, the sequence of samples to draw from each level, $\{N_\ell\}_{\ell=1}^{\ell=L}$, and the sequence of continuation probabilities, $\{(\eta_{\ell,1}, \eta_{\ell,2})\}_{\ell=1}^{\ell=L}$. Of these, L and $\{\tau_\ell\}_{\ell=1}^{\ell=L}$ need to be selected heuristically, however, for any given L and $\{\tau_\ell\}_{\ell=1}^{\ell=L}$, the sequences $\{N_\ell\}_{\ell=1}^{\ell=L}$ and $\{(\eta_{\ell,1}, \eta_{\ell,2})\}_{\ell=1}^{\ell=L}$ may be optimised.

For a given level ℓ and assuming τ_ℓ is selected, the optimal $\eta_{\ell,1}$ and $\eta_{\ell,2}$ can be determined through optimising the limiting efficiency as the number of samples $N_\ell \to \infty$. Prescott and Baker [46] show that this corresponds to minimising the function

$$\phi(\eta_{\ell,1}, \eta_{\ell,2}; g_\ell) = \mathbb{E}\left[w^{\tau_\ell}(\boldsymbol{\theta}_\ell)^2(g_\ell(\boldsymbol{\theta}_\ell) - \mathbb{E}\left[g_\ell(\boldsymbol{\theta}_\ell)\right])^2\right]\mathbb{E}\left[C_\ell(\boldsymbol{\theta}_\ell)\right], \tag{18}$$

where $C_\ell(\boldsymbol{\theta}_\ell)$ is the cost of computing the weight $w^{\tau_\ell}(\boldsymbol{\theta}_\ell)$ (Equation (14)). If the acceptance state of an approximate simulation is interpreted as a classifier for the predicted acceptance state for an exact simulation, and the true positive rate exceeds the false positive rate, then Equation (18) can be minimised for $(\eta_{\ell,1}, \eta_{\ell,2}) \in (0, 1]^2$ (See Lemmas 4.2 and 4.3 in Prescott and Baker [46]). The optimal continuation probabilities are given by (See Corollary 4.4 in Prescott and Baker [46])

$$(\eta_{\ell,1}^*, \eta_{\ell,2}^*) = \begin{cases} \left(\sqrt{\dfrac{R_p^\ell}{R_0^\ell}}, \sqrt{\dfrac{R_n^\ell}{R_0^\ell}}\right), & \text{if } \max\{R_p^\ell, R_n^\ell\} \le R_0^\ell, \\ (1, \bar{\eta}_{\ell,2}), & \text{if } \max\{R_p^\ell, R_n^\ell\} > R_0^\ell \text{ and } \phi(1, \bar{\eta}_{\ell,2}) \le \phi(\bar{\eta}_{\ell,1}, 1), \\ (\bar{\eta}_{\ell,1}, 1), & \text{otherwise,} \end{cases} \tag{19}$$

where

$$R_p^\ell = \frac{p_{fp}^\ell \mathbb{E}\left[c^{\tau_\ell}(\boldsymbol{\theta}_\ell)\right]}{c_p^\ell}, \quad R_n^\ell = \frac{p_{fn}^\ell \mathbb{E}\left[c^{\tau_\ell}(\boldsymbol{\theta}_\ell)\right]}{c_n^\ell}, \quad R_0^\ell = p_{tp}^\ell - p_{fp}^\ell,$$

$$\bar{\eta}_{\ell,1} = \min\left\{1, \sqrt{\frac{R_p^\ell + p_{fp}^\ell c_n^\ell/c_p^\ell}{R_0^\ell + p_{fn}^\ell}}\right\}, \quad \bar{\eta}_{\ell,2} = \min\left\{1, \sqrt{\frac{R_n^\ell + p_{fn}^\ell c_p^\ell/c_n^\ell}{R_0^\ell + p_{fp}^\ell}}\right\}, \tag{20}$$

and

$$p_{tp}^\ell = \mathbb{E}\left[\mathbb{1}_{(0,\epsilon_\ell]}\left(\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s)\right) \mathbb{1}_{(0,\tilde{\epsilon}_\ell]}\left(\rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell})\right)(g_\ell(\boldsymbol{\theta}_\ell) - \mathbb{E}\left[g_\ell(\boldsymbol{\theta}_\ell)\right])^2\right],$$

$$p_{fp}^\ell = \mathbb{E}\left[\mathbb{1}_{(\epsilon_\ell,\infty]}\left(\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s)\right) \mathbb{1}_{(0,\tilde{\epsilon}_\ell]}\left(\rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell})\right)(g_\ell(\boldsymbol{\theta}_\ell) - \mathbb{E}\left[g_\ell(\boldsymbol{\theta}_\ell)\right])^2\right],$$

$$p_{fn}^\ell = \mathbb{E}\left[\mathbb{1}_{(0,\epsilon_\ell]}\left(\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s)\right) \mathbb{1}_{(\tilde{\epsilon}_\ell,\infty]}\left(\rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell})\right)(g_\ell(\boldsymbol{\theta}_\ell) - \mathbb{E}\left[g_\ell(\boldsymbol{\theta}_\ell)\right])^2\right], \tag{21}$$

$$c_p^\ell = \mathbb{E}\left[c(\boldsymbol{\theta}_\ell) \mid \rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell}) \le \tilde{\epsilon}_\ell\right],$$

$$c_n^\ell = \mathbb{E}\left[c(\boldsymbol{\theta}_\ell) \mid \rho^{\tau_\ell}(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^{\tau_\ell}) > \tilde{\epsilon}_\ell\right].$$

In Equations (20) and (21), $c(\boldsymbol{\theta}_\ell)$ and $c^{\tau_\ell}(\boldsymbol{\theta}_\ell)$ denote, respectively, the cost of generating an exact realisation, $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta}_\ell)$, and an approximate realisation, $\mathbf{Y}_s^{\tau_\ell} \sim s^{\tau_\ell}(\cdot \mid \boldsymbol{\theta}_\ell)$.

To optimise the sequence of samples $\{N_\ell\}_{\ell=1}^{\ell=L}$, we aim to minimise the total expected computational cost of computing the MF-MLMC-ABC estimator, that is, $\mathbb{E}\left[C(\hat{f})\right] = \sum_{\ell=1}^L N_\ell \mathbb{E}\left[C_\ell(\boldsymbol{\theta}_\ell)\right]$, subject to the constraint $\text{Var}\left[\hat{f}\right] \propto h^2$ where h^2 is the target variance. Using a Lagrange multiplier, it can be shown (See Giles [51], Lester et al. [53], and Warne et al. [62]) that the following scaling is optimal,

$$N_\ell \propto h^{-2} \sqrt{\frac{v_\ell}{c_\ell}} \sum_{m=1}^L \sqrt{v_m c_m}, \quad \text{for } \ell = 1, 2, \ldots, L, \tag{22}$$

where $c_\ell = \mathbb{E}\left[C_\ell(\boldsymbol{\theta}_\ell)\right]$ and $v_\ell = \mathbb{E}\left[w^{\tau_\ell}(\boldsymbol{\theta}_\ell)^2(g_\ell(\boldsymbol{\theta}_\ell) - \mathbb{E}\left[g_\ell(\boldsymbol{\theta}_\ell)\right])^2\right]/\mathbb{E}\left[w^{\tau_\ell}(\boldsymbol{\theta}_\ell)\right]^2$.

Relative to ABC rejection sampling, the MF-MLMC-ABC method is asymptotically unbiased, with the bias at level ℓ being of the order importance sampling $\mathcal{O}(1/N_\ell)$ (Supplementary Material). The effect of this bias to be considered along with the optimal sample size scaling in Equation (23), especially for the terminal level L. That is, we require the bias due to MF-ABC sampling to be small compared with the bias incurred from the ABC-based approximations, which are $\mathcal{O}(\epsilon)$. This is difficult to test in practice, however, it is a common feature of any ABC method based on importance sampling.

2.6. Practical algorithm tuning

A practical choice for the selection of most components of MF-MLMC-ABC is immediately available from the target ABC-based inference problem, for example, the exact stochastic simulation process $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta})$, the prior probability density, $p(\boldsymbol{\theta})$, the discrepancy metric $\rho(\mathbf{Y}_{\text{obs}}, \cdot)$, and the target acceptance threshold ϵ_L. Other choices are easily motivated. The largest acceptance threshold, ϵ_1, can be chosen so that $p(\boldsymbol{\theta} \mid \rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) \le \epsilon_1) \approx p(\boldsymbol{\theta})$. The approximate stochastic simulation scheme, $\mathbf{Y}_s^\tau \sim s^\tau(\cdot \mid \boldsymbol{\theta})$, could be chosen from a range of candidates, but a first-order method such as tau-leaping will be an appropriate default choice in many cases. Given a tau-leaping scheme for the approximate model, it will often be appropriate to take $\rho^\tau(\mathbf{Y}_{\text{obs}}, \cdot) = \rho(\mathbf{Y}_{\text{obs}}, \cdot)$, and therefore $\tilde{\epsilon}_\ell = \epsilon_\ell$ for all $\ell = 1, \ldots, L$ is applicable. Of course, there is freedom and flexibility in all of these choices, and the discussion section highlights some potentially useful alternative strategies.

Next the number of levels, L, needs to be determined. Unfortunately, there is no general theory for this choice. However one common approach from the MLMC literature is to consider a geometric sequence $\epsilon_\ell = \epsilon_1 m^{-\ell+1}$ then set $L = 1 -$

$\log_m(\epsilon_L/\epsilon_1)$. Some heuristics do exist to determine the appropriate scale factor $m > 1$. For MLMC with SDEs Giles [51] demonstrated $m = 4$ is optimal, however, due to the approximate coupling scheme for MLMC-ABC, Warne et al., [22,62] propose $m \in [1.5, 2.5]$ as a practical choice for inference.

Choice of the sequence of time-steps, $\{\tau_\ell\}_{\ell=1}^{\ell=L}$, can be guided by Equations (19)–(21). Firstly, we wish τ_ℓ to be small enough to exhibit low false positive and false negative rates; this will result in smaller optimal continuation probabilities (Equation (20)) and reduce the number of times the high-fidelity model is simulated [46]. However, the speed-up factor for low-fidelity simulations over high-fidelity simulations needs to be sufficiently high, otherwise there is not enough of a computational benefit [63]. Therefore, τ_ℓ cannot be arbitrary small. To tune τ_ℓ, some experimentation is required to compare the differences computation time and the acceptance state between pairs of exact and approximate simulations. Fortunately, it is possible to identify poor choices of τ_ℓ, since the adaptive tuning scheme for the optimal continuation probabilities will be unable to improve upon standard ABC rejection sampling. It is also important to note that $\{\tau_\ell\}_{\ell=1}^{\ell=L}$ need not be a strictly decreasing or even monotonic sequence, and we demonstrate in the results section that $\tau_1 = \tau_2 = \cdots = \tau_L$ works quite well for many applications. Furthermore, introducing a coupling scheme between the exact and approximate simulations [46,52,53] can reduce mis-classification rates of larger values of τ_ℓ and lead to improved performance.

In practice, Equation (19) is solved for optimal continuation probabilities through the generation of initial trial samples [46]. In this work, we extend this through adaptive updates to $(\eta_{\ell,1}, \eta_{\ell,2})$ while generating samples at level ℓ. For the trial samples, we generate M samples using Algorithm 5 with $\eta_{\ell,1} = \eta_{\ell,2} = 1$, then initial estimates of the expectations in Equation (20) are produced through direct Monte Carlo estimates. Next, gradient descent is applied to iteratively update $\eta_{\ell,1}$ and $\eta_{\ell,2}$ toward the optimum in Equation (19) while also refining the estimates in Equations (20). While it is possible to iteratively refine the solution to Equation (19) directly as the sampling proceeds [46], we find that this is extremely sensitive to the initial estimates. Therefore we utilise our adaptive gradient descent MF-ABC sampler based on recent developments [63] as a robust alternative to Algorithm 5 (Supplementary Material).

Finally, to apply MF-MLMC-ABC the sequence of sample numbers, $\{N_\ell\}_{\ell=1}^{\ell=L}$, are needed. Fortunately, we can rewrite Equation (22) as

$$N_\ell \propto \frac{1}{h^{-2}} \frac{\sqrt{\phi(\eta_{\ell,1}, \eta_{\ell,2}; g_\ell)}}{\mathbb{E}\left[C_\ell(\boldsymbol{\theta}_\ell)\right] \mathbb{E}\left[w^{\tau_\ell}(\boldsymbol{\theta}_\ell)\right]} \sum_{m=1}^{L} \frac{\sqrt{\phi(\eta_{m,1}, \eta_{m,2}; g_m)}}{\mathbb{E}\left[w^{\tau_m}(\boldsymbol{\theta}_m)\right]}, \tag{23}$$

to highlight that the optimal sequence of sample numbers, $\{N_\ell\}_{\ell=1}^{\ell=L}$, is dependent on the optimal continuation probabilities obtained using the adaptive gradient descent MF-ABC scheme at each level. Therefore, we can estimate the terms required for optimal $\{N_\ell\}_{\ell=1}^{\ell=L}$ directly from the same trial samples used to estimate optimal $\{(\eta_{\ell,1}, \eta_{\ell,2})\}_{\ell=1}^{\ell=L}$.

2.7. Summary

The MF-ABC and MLMC-ABC methods exploit the multilevel telescoping summation, and properties of approximate stochastic simulation and ABC sampling in distinct and complementary ways. In the first instance, MF-ABC can improve on the expected cost of stochastic simulation in ABC rejection sampling using the randomized bias correction term [46,82]. MLMC-ABC assumes exact stochastic simulation for ABC rejection sampling, but applies MLMC techniques to a sequence of correlated samplers with discrepancy thresholds $\epsilon_1 > \epsilon_2 > \cdots > \epsilon_L$ with the effect of improving the convergence rate [61,62]. We develop a new method, MF-MLMC-ABC, that results from the application of MF-ABC sampling for each of the terms in the MLMC-ABC telescoping summation. In the next section we demonstrate how to tune these methods practically and show the computational benefits of both can be exploited to achieve improvements of two orders of magnitude.

3. Results

Using a variety of biologically relevant stochastic biochemical reaction network models, we demonstrate the substantial computational improvements using our new MF-MLMC-ABC method (Algorithm 6). First we consider the properties of the MLMC-ABC (Algorithm 4) and MF-ABC (Algorithm 5) methods for a fundamental biochemical building block, the Michaelis–Menten model, to show practically how to tune these methods. Then we apply these guidelines to tune our new MF-MLMC-ABC method for the repressilator gene regulatory network to show the performance benefits over MF-ABC and MLMC-ABC. Finally, we perform a realistic test on the two-step MAPK cascade network that is of fundamental importance in cell biology. For two computationally challenging networks we show that MF-MLMC-ABC effectively combines the advantages of both MF-ABC and MLMC-ABC to accelerate ABC rejection sampling by two orders of magnitude.

In the sections that follow, we focus on the performance improvements of our new approach for the purpose of estimating posterior means of unknown parameters. However, it is important to note that our approach, arising from Equation (16), handles expectations of an arbitrary function of the unknown parameters, $f(\boldsymbol{\theta})$. This could include posterior probabilities or densities, or even central moments of the posterior predictive distribution, that is the mean and variance of the biochemical network state $\mathbf{X}_T$ given $\boldsymbol{\theta}$. We demonstrate the use of MF-MLMC-ABC for estimation of marginal densities for the two-stage MAPK example.
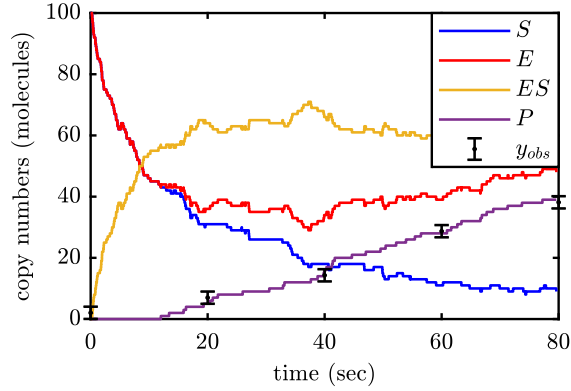
Fig. 1. Example realisation of the Michaelis–Menten model along with noisy observations of the product molecules $y_{\text{obs}}(t) \sim \mathcal{N}(P_t, \sigma^2)$ (error bars indicate $y_{\text{obs}}(t) \pm \sigma$). Here, the initial condition is $E_0 = S_0 = 1000$, the true rate parameters are $k_1 = 0.001$, $k_2 = 0.005$ and $k_3 = 0.01$, and observations are taken at $t_i = 20i$, for $i = 1, 2, 3, 4$ with standard deviation $\sigma = 10$.

3.1. Initial explorations of MF-ABC and MLMC-ABC: Michaelis–Menten kinetics

Using a stochastic network of Michaelis–Menten enzyme kinetics [85,86], we demonstrate the essential requirements and computational benefits of the MF-ABC and MLMC-ABC methods in order to inform the configuration of MF-MLMC-ABC for more challenging networks. The Michaelis–Menten enzyme kinetics model describes the catalytic conversion of a substrate, S, into a product, P, via an enzymatic reaction involving enzyme, E,

$$E + S \xrightarrow{k_1} [ES], \quad [ES] \xrightarrow{k_2} E + S, \quad [ES] \xrightarrow{k_3} E + P, \tag{24}$$

with kinetic rate parameters, k_1, k_2, and k_3. Biologically, this network is of interest since many intracellular processes are built from Michaelis–Menten sub-components. Computationally, the Michaelis–Menten model is a minimal example of a network without a closed-form solution to the CME, however, with only three rate parameters and four chemical species, ABC inference is feasible even with rejection sampling [22].

For our simulated data we consider the realistic scenario where only the product molecules are directly observed (typically via fluorescent tagging of proteins [87–89]), that is,

$$y_{\text{obs}}(t) \sim \mathcal{N}(P_t, \sigma^2), \tag{25}$$

where σ is the standard deviation of the additive Gaussian observation noise. In real applications, especially for low copy numbers, it may be more appropriate to consider multinomial noise or multiplicative Gaussian noise [90,91], however, for the purposes of the numerical experiments we present here, additive Gaussian noise is perfectly reasonable. Fig. 1 shows an example realisation of the Michaelis–Menten model with simulated observations indicated at discrete times, $t_1 = 20, t_2 = 40, t_3 = 60$, and $t_4 = 80$.

Using the Michaelis–Menten model and the noisy partial observations, we explore the effects of varying the parameters L and τ on the performance of MLMC-ABC and MF-ABC, respectively, with a tau-leaping method assumed for the approximate simulation scheme. To this end, we consider the ABC inference problem,

$$\mathbb{E}\left[k_3 \mid \mathbf{Y}_{\text{obs}}\right] \approx \int_{\mathbb{R}^3} k_3 \, p(\boldsymbol{\theta} \mid \rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) \leq \epsilon) \, d\boldsymbol{\theta}, \tag{26}$$

where $\boldsymbol{\theta} = [k_1, k_2, k_3]$ is the vector of unknown rate parameters, $\mathbf{Y}_{\text{obs}} = [y_{\text{obs}}(t_1), \dots, y_{\text{obs}}(t_4)]$ is noisy observations of product copy numbers at discrete times $t_1, \dots, t_4$, $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta})$ is simulated data of the Michaelis–Menten model using the Gillespie direct method and simulating the observation process (Equation (25)), ϵ is the discrepancy threshold, and the discrepancy metric is $\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) = \|\mathbf{Y}_{\text{obs}} - \mathbf{Y}_s\|_2$ where $\|\cdot\|_2$ is the Euclidean norm. Independent uniform priors are used with $k_1 \sim \mathcal{U}(0, 0.003)$, $k_2 \sim \mathcal{U}(0, 0.0015)$, and $k_3 \sim \mathcal{U}(0, 0.05)$.

We first explore MLMC-ABC (Algorithm 4) in terms of the effect of the number of levels, L, using data generated with Equation (25) with $\sigma = 2$ at discrete times $t_1 = 20, t_2 = 40, t_3 = 60$, and $t_4 = 80$ using a single realisation given initial conditions $E_0 = S_0 = 1000$ and $[ES]_0 = P_0 = 0$, and rate parameters $k_1 = 0.001$, $k_2 = 0.005$ and $k_3 = 0.01$. We use $\epsilon_1 = 1600$ and apply MLMC-ABC for different numbers of levels $L \in [2, 3, \dots, 9]$ and different target discrepancy thresholds $\epsilon_L = \epsilon \in [200, 300, \dots, 600]$. Optimal sample numbers $\{N_\ell\}_{\ell=1}^{\ell=L}$ are obtained using a Lagrange multiplier approach [51,53,62] and rescaled so that $N_L = 16$. All stochastic simulations, $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta})$, are exact using the Gillespie direct method (Algorithm 1). Fig. 2 shows how the trade-off between computational cost and estimator variance is affected by the number of levels, L. The trends in terms of L (Fig. 2(A)) are not very meaningful because for a fixed L the scale factor m will be different for each
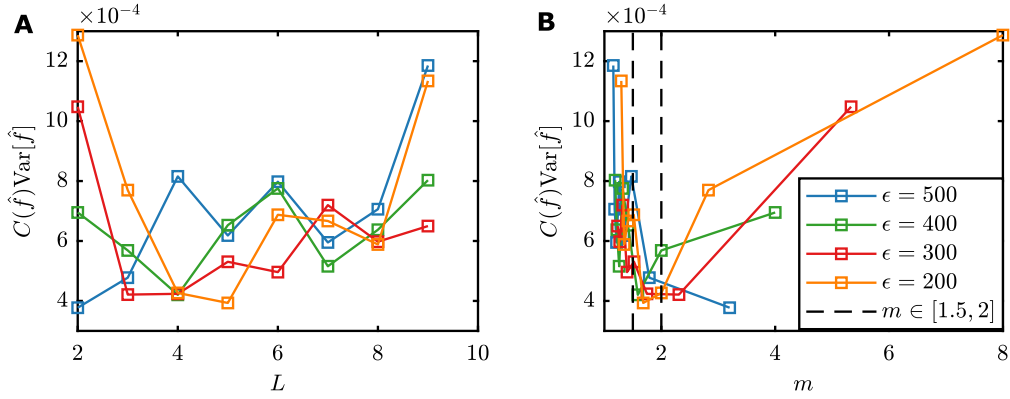
Fig. 2. (A) The effect of the number of levels, L, in MLMC-ABC on the sampling efficiency (as determined by the product of computational cost and the estimator variance). (B) Equivalent plot in terms of $m = (\epsilon_1/\epsilon_L)^{1/(L-1)}$ to highlight that this product is minimised for the scaling $m \in [1.5, 2]$ (black dashed lines) regardless of the target epsilon. Target discrepancy thresholds are indicated by dashed lines and in all cases $\epsilon_1 = 1600$. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)
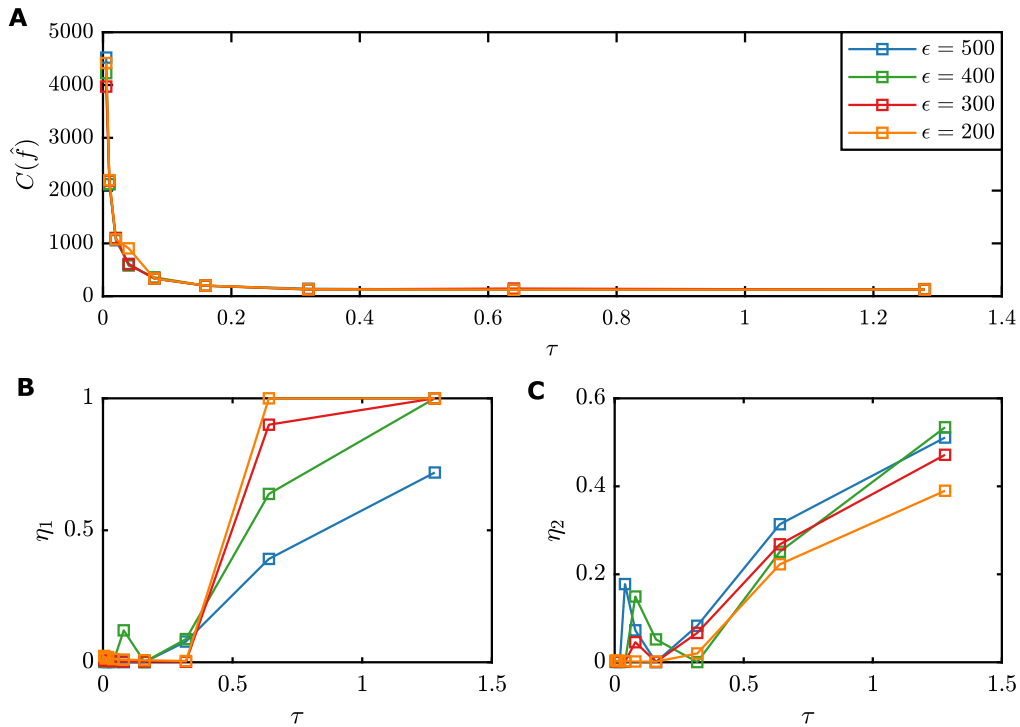
Fig. 3. The relationship between the approximate stochastic simulation time-step, τ and: (A) the cost of generating $N = 10000$ weighted samples; (B) the continuation probability when an approximate stochastic simulation is accepted; and (C) the continuation probability when an approximate stochastic simulation is rejected. Coloured lines indicate the effect of τ for different discrepancy thresholds ϵ. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

target threshold. However, when the same data are presented in terms of the scale factor, m (Fig. 2(B)), the optimal choice is consistently contained within $m \in [1.5, 2]$. This result aligns closely with the previous work on MLMC-ABC [62]. Therefore, we conclude that a good heuristic for a given ϵ_1 and ϵ_L is to choose m within this interval such that $L = 1 - \log_m(\epsilon_L/\epsilon_1)$ is a positive integer.

Next we look at the more nuanced problem of tuning the tau-leap time-step, τ, in the context of MF-ABC (Algorithm 5) for performance. Using the same data configuration and ABC problem definition as for MLMC-ABC we apply MF-ABC for different time-steps $\tau \in [0.005, 0.01, 0.02, 0.04, 0.08, 0.16, 0.32, 0.64, 1.28]$ and for the same set of target discrepancy thresholds as for MLMC-ABC. We also take $\tilde{\epsilon} = \epsilon$, and $\rho^\tau(\mathbf{Y}_{obs}, \mathbf{Y}_s^\tau) = \|\mathbf{Y}_{obs} - \mathbf{Y}_s^\tau\|_2$ with $\mathbf{Y}_s^\tau \sim s^\tau(\cdot \mid \boldsymbol{\theta})$ is the approximate stochastic simulation process using the tau-leaping method (Algorithm 2).

Fig. 3 shows the effect of varying τ on the computational cost of generating the N weighted samples, and the optimal continuation probabilities η_1 and η_2, as determined by the adaptive update scheme. Note that some choices of τ result in
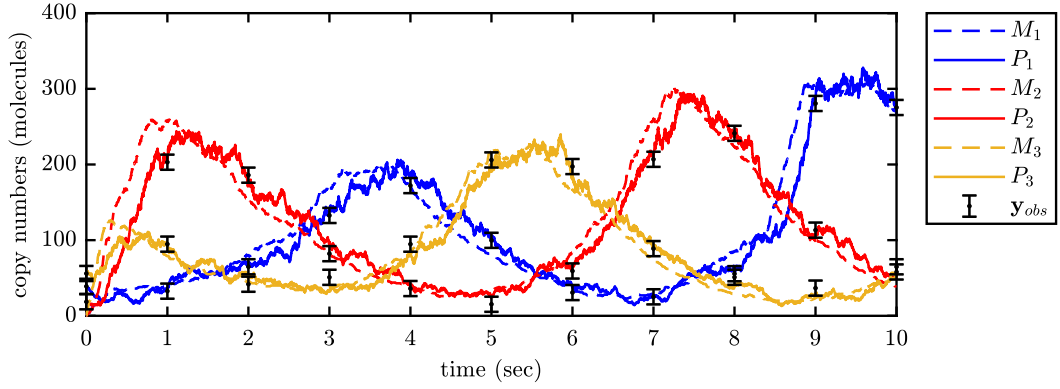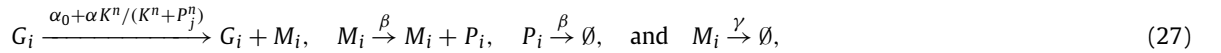
Fig. 4. Example realisation of the repressilator gene regulatory network model along with noisy observations of the three protein molecules, $\mathbf{y}_{\text{obs}}(t) \sim \mathcal{N}([P_{1,t}, P_{2,t}, P_{3,t}]^{\text{T}}, \sigma^2 \mathbf{I})$ (error bars indicate $\mathbf{y}_{\text{obs}}(t) \pm \sigma$). Here, the initial condition is $M_{1,0} = M_{2,0} = M_{3,0} = 0$, $P_{1,0} = 40$, $P_{2,0} = 20$, $P_{3,0} = 60$, the true rate parameters are $\alpha_0 = 1$, $\alpha = 1000$, $K = 20$, $n = 2$, $\beta = 5$ and $\gamma = 1$, and observations are taken at $t_i = i$, for $i = 1, 2, \ldots 10$ with standard deviation $\sigma = 10$.

very small values for both η_1 and η_2 (Fig. 3(B)–(C)). However, this does not translate into a computational improvement, since the total compute time is always larger that the worst case with $\eta_1 = \eta_2 = 1$ (Fig. 3(A)). We therefore conclude that there is no computational advantage in using MF-ABC for the Michaelis–Menten model as specified here. Effectively, the tau-leaping method does not provide enough of a computational benefit over the Gillespie direct method for the Michaelis–Menten model with the given initial conditions.

A key message from this section is that before employing MLMC-ABC or MF-ABC, and by extension our new method MF-MLMC-ABC, some initial exploration should be performed. While MF-ABC provides no benefit for the Michaelis–Menten model, this is largely due to the simplicity of the network. In more complex models that we consider in the next two sections, MF-ABC provides a substantial improvement leading to high efficiency with the MF-MLMC-ABC method.

3.2. Tuning and performance of MF-MLMC-ABC: Repressilator gene regulatory network

We now demonstrate the computational benefits of MF-MLMC-ABC using a stochastic gene regulatory network called the repressilator [92]. The repressilator describes the expression levels of three genes, G_1, G_2, and G_3, in which the expression of G_i inhibits the expression of $G_{(i \bmod 3)+1}$, forming a cycle that results in stochastic oscillations. Each gene, G_i, consists of two reactions that describe gene expression, through the transcription of mRNA, M_i, and translation into protein, P_i, and two reactions that describe the degradation of mRNA and protein molecules. For the ith gene we have

$$G_i \xrightarrow{\alpha_0 + \alpha K^n/(K^n + P_j^n)} G_i + M_i, \quad M_i \xrightarrow{\beta} M_i + P_i, \quad P_i \xrightarrow{\beta} \emptyset, \quad \text{and} \quad M_i \xrightarrow{\gamma} \emptyset, \tag{27}$$

where $j = (i + 1 \bmod 3) + 1$, $\alpha_0 \geq 0$ is the leakage transcription rate (the transcription rate of a maximally inhibited gene), $\alpha + \alpha_0 > 0$ is the free transcription rate (uninhibited transcription rate), $n \geq 0$ is the Hill coefficient that describes the strength of the repressive effect of the inhibitor protein P_j, K is the number of P_j inhibitor proteins required to reduce the transcription rate of G_i by 50% (excluding leakage), $\beta > 0$ is the protein translation and degradation rate, and $\gamma > 0$ is the mRNA degradation rate.

We consider noisy observations of the protein copy numbers since these will be the only observables via fluorescent markers. This yields the observation process

$$\mathbf{y}_{\text{obs}}(t) \sim \mathcal{N}([P_{1,t}, P_{2,t}, P_{3,t}]^{\text{T}}, \sigma^2 \mathbf{I}), \tag{28}$$

where $\mathbf{I}$ the 3×3 identity matrix. Just as with the Michaelis–Menten example, alternate noise models could be utilised. Discrete observations are taken at regular one time unit intervals, $t_i = i$, for $i = 0, 1, \ldots, 10$. Fig. 4 shows an example realisation of the repressilator model along with discrete observations of the protein molecules.

This model is a common choice to benchmark the performance of likelihood-free inference methods since the oscillatory behaviour renders the acceptance probability to be very low [25,46,93]. For our MF-MLMC-ABC method, the repressilator is particularly interesting due to the possibility that coupled pairs of exact simulations and approximate simulations will go out of phase with each other. This means the time-step sequence $\{\tau_\ell\}_{\ell=1}^{\ell=L}$ must be carefully chosen.

In the target ABC inference problem we consider assumes only the parameters of Hill functions to be unknown and evaluates

$$\mathbb{E}[K \mid \mathbf{Y}_{\text{obs}}] \approx \int_{\mathbb{R}^2} K p(\boldsymbol{\theta} \mid \rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) \leq \epsilon) \, \mathrm{d}\boldsymbol{\theta}, \tag{29}$$
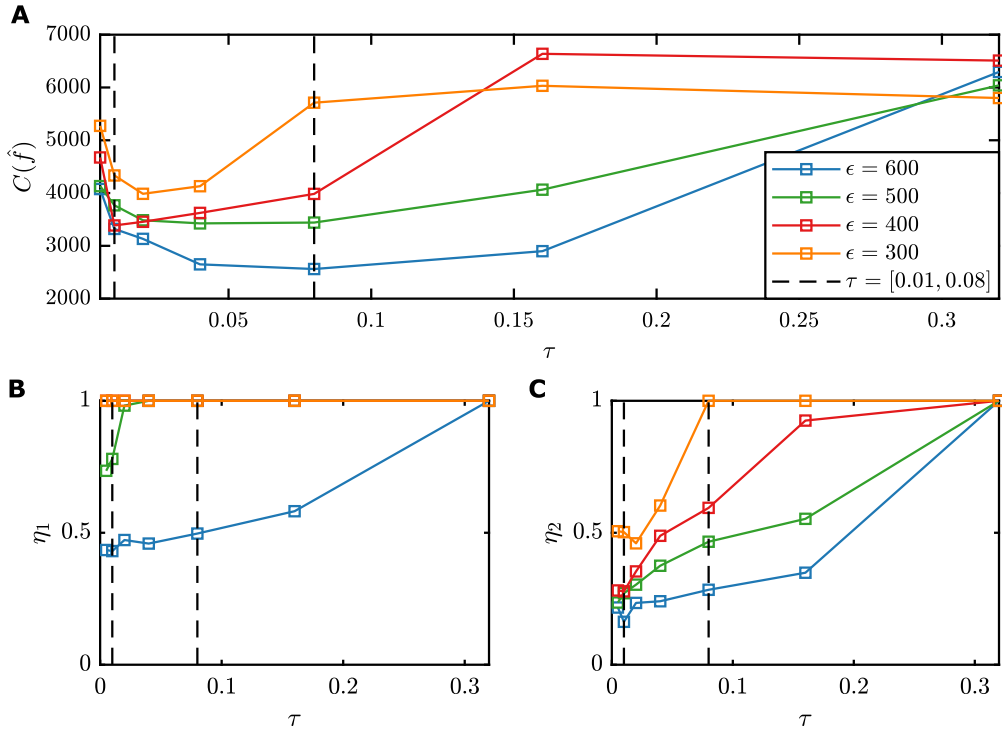
Fig. 5. The relationship between the approximate stochastic simulation time-step, τ, and: (A) the cost of generating $N = 10000$ weighted samples; (B) the continuation probability when an approximate stochastic simulation is accepted; and (C) the continuation probability when an approximate stochastic simulation is rejected. Solid lines indicate the effect of τ for different discrepancy thresholds ϵ and the dashed black lines indicate the range of values for τ that demonstrate performance improvement. (For interpretation of the colours in the figure(s), the reader is referred to the web version of this article.)

where $\boldsymbol{\theta} = [K, n]$ is the vector of unknown Hill function parameters, $\mathbf{Y}_{\text{obs}} = [\mathbf{y}_{\text{obs}}(t_0), \ldots, \mathbf{y}_{\text{obs}}(t_{10})]$ are noisy observations of protein copy numbers at discrete times $t_1, \ldots, t_{10}$ (Equation (28)), $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta})$ is simulated data of the repressilator model generated using the Gillespie direct method and the observation process (Equation (28)), ϵ is the discrepancy threshold, and the discrepancy metric is $\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) = \|\mathbf{Y}_{\text{obs}} - \mathbf{Y}_s\|_2$ where $\|\cdot, \|_2$ is the Euclidean norm. Independent uniform priors are used with $K \sim \mathcal{U}(10, 30)$, $n \sim \mathcal{U}(1, 4)$. We treat the rate parameters as known with $\alpha_0 = 1$, $\alpha = 1000$, $\beta = 5$ and $\gamma = 1$.

To apply MF-MLMC-ABC we require an appropriate sequence of time-steps $\{\tau_\ell\}_{\ell=1}^{\ell=L}$. To tune this sequence, we draw a small number, $N = 10000$, of MF-ABC weighted samples using the adaptive optimisation scheme for continuation probabilities (Supplementary Material) for a range of discrepancy thresholds $\epsilon \in [200, 300, \ldots, 600]$ and time-steps $\tau \in [0.005, 0.01, 0.02, \ldots, 0.64]$. We also take $\tilde{\epsilon} = \epsilon$ and $\rho^\tau(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^\tau) = \|\mathbf{Y}_{\text{obs}} - \mathbf{Y}_s^\tau\|_2$, with $\mathbf{Y}_s^\tau \sim s^\tau(\cdot \mid \boldsymbol{\theta})$ the approximate stochastic simulation process using the tau-leaping method (Algorithm 2). Fig. 5 show the relationship between τ, the computational cost, and continuation probabilities for each of the target discrepancy thresholds. If these discrepancies represented the MF-MLMC-ABC discrepancy sequence $\epsilon_1 = 600, \ldots, \epsilon_L = 300$ with $L = 4$, then Fig. 5(A) can be used to identify the optimal sequence for each τ_ℓ by finding the value of τ_ℓ with the lowest expected cost $C(\hat{f})$. This suggests a sequence of $\tau_1 = 0.08$, $\tau_2 = 0.04$, $\tau_3 = 0.01$, $\tau_4 = 0.02$. If we restrict our choice to a single time-step value to apply to all levels, then $\tau_1 = \tau_2 = \cdots \tau_L = 0.02$ is the best overall as this value for τ_ℓ results in the lowest total cost for N samples from each level.

Using the above heuristics we arrive at the choice of $L = 5$ and $\tau = 0.04$. We apply our new MF-MLMC-ABC method (Algorithm 6) to the ABC inference problem in Equation (29) and compare with MLMC-ABC (Algorithm 4), MF-ABC (Algorithm 5), and ABC rejection sampling (Algorithm 3) for different values for the target discrepancy $\epsilon \in [350, 500]$, and in all cases $\epsilon_1 = 1600$. For each target, we perform optimal tuning steps for MF-MLMC-ABC, MLMC-ABC and MF-ABC to adapt the continuation probabilities $\{(\eta_{\ell,1}, \eta_{\ell,1})\}_{\ell=1}^{\ell=L}$ and sample sizes $\{N_\ell\}_{\ell=1}^{\ell=L}$. We repeat this for different target variances, h^2, then estimate the variance, $\text{Var}\left[\hat{f}\right]$, and the computational cost, $C(\hat{f})$, to obtain an estimate of the convergence rate, γ, by least-squares fitting $\text{Var}\left[\hat{f}\right] \propto C(\hat{f})^{-\gamma}$ to align with theory from the MLMC literature [50,51]. Fig. 6 demonstrates the substantial computational advantage of MF-MLMC-ABC.

While MF-ABC is consistently computationally cheaper than ABC rejection sampling, the convergence rate is $\gamma \approx 1$ which is equivalent to the limiting behaviour of ABC rejection and the theoretical rate under the central limit theorem. MLMC-ABC achieves a higher convergence rate of around $\gamma \approx 1.5$. However, the computational benefit is not realised until the target variances are small due to the overhead of tuning for $\{N_\ell\}_{\ell=1}^{\ell=L}$ along with the exact stochastic simulations performed at every level. MF-MLMC-ABC out-performs the other methods in every target discrepancy and target variance, and succeeds
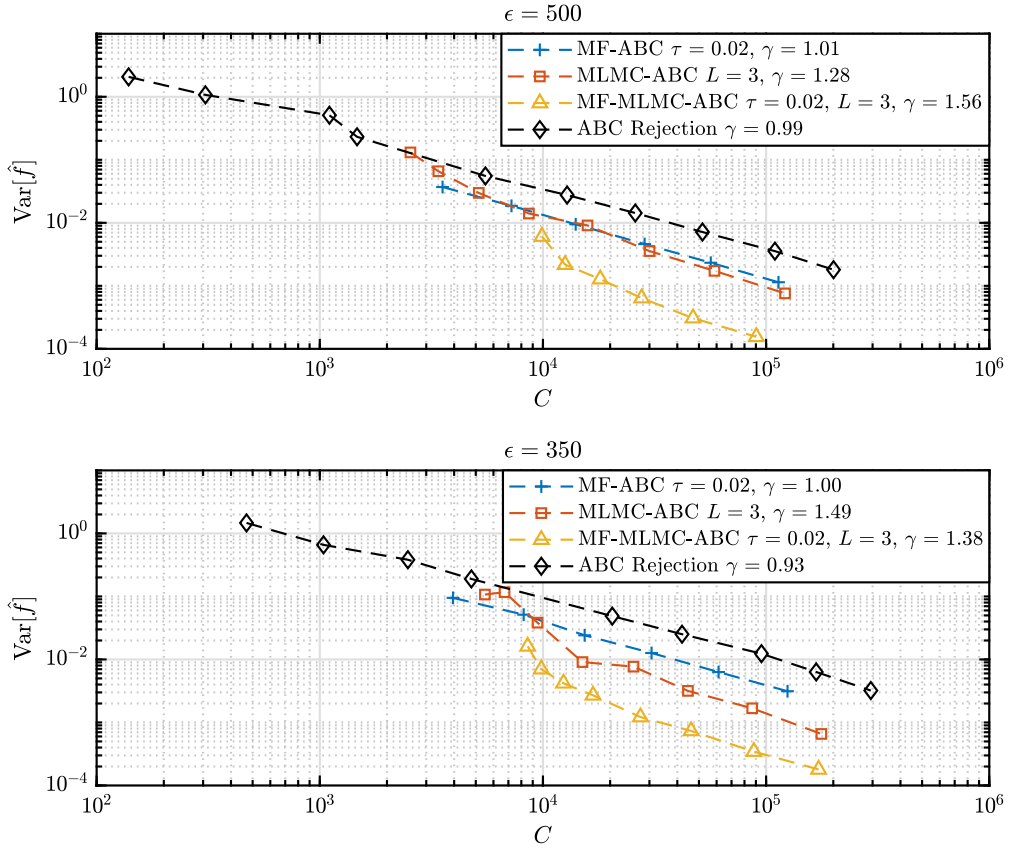
Fig. 6. Comparison of convergence rates for MF-MLMC-ABC (yellow triangles) with MF-ABC (blue crosses), MLMC-ABC (red squares), and ABC rejection sampling (black diamonds) using the repressilator model with thresholds: (A) $\epsilon = 500$; and (B) $\epsilon = 350$. Rates are estimated by fitting $\mathrm{Var}\left[\hat{f}\right] \propto C(\hat{f})^{-\gamma}$ to benchmark data using least squares.

in both improving the convergence rate and substantially reducing the overall computational cost, including the tuning steps. The convergence rate is improved in a similar way to MLMC-ABC with $\gamma \approx 1.5$, however, the overall computational reduction in MF-MLMC-ABC compared with MLMC-ABC is larger than the reduction in MF-ABC vs ABC rejection. This is largely due to the fact that MF-ABC is most effective for larger discrepancies, as noted by the tendency for optimal continuation probabilities to be smaller as the discrepancy threshold increases (Fig. 5(B)–(C)). Furthermore, MLMC will utilise variance reduction via the coupling in the telescoping summation to allocate fewer samples for the smaller discrepancies. As a result, the earlier levels benefit from both the higher acceptance rates that come with large discrepancy thresholds, and smaller continuation probabilities so exact stochastic simulations are rarely executed. The effect extends to the optimisation of $\{N_\ell\}_{\ell=1}^{\ell=L}$ through Equation (23). This results in a substantial reduction in the usual overheads associated with MLMC-ABC and computational benefit is realised for much smaller target variances. Consistently, for equivalent computational cost the MF-MLMC-ABC is between one and two orders of magnitude lower in terms of variance and the improvement increases for larger computation times due to the convergence rate. This example demonstrates practically how MF-MLMC-ABC may be tuned without substantial overhead to provide a very high-performance inference method. We also explore the effect of different choices of L and τ to demonstrate the efficiency of our heuristics (Supplementary Material).

One final aspect of the MF-MLMC-ABC approach that is important to consider is that of the additional bias that can be incurred from either selecting a poor choice of L (that is, m is too large), or from the bias in the MF-ABC scheme for small N_ℓ. To explore this we compare the expectations of the parameter K across algorithms. Fig. 7, demonstrates that this bias, as expected, decays with computational effort. However, it is also important to note that the magnitude of the bias is small compared with that of the ABC approximation as additional experimentation shows that a smaller discrepancy of $\epsilon < 200$ leads to $\mathbb{E}\left[K \mid \mathbf{Y}_{\mathrm{obs}}\right] \approx 18$.

3.3. A challenging problem: Two-step MAPK cascade reaction

The last example we consider is a challenging biological network involving a two-step Mitogen Activated Protein Kinase (MAPK) enzymatic cascade [64]. Such cascade reactions are essential components of cell signalling processes, such as Epidermal Growth Factor Receptor (EGFR) signalling, that regulates cell growth, death, proliferation, and differentiation
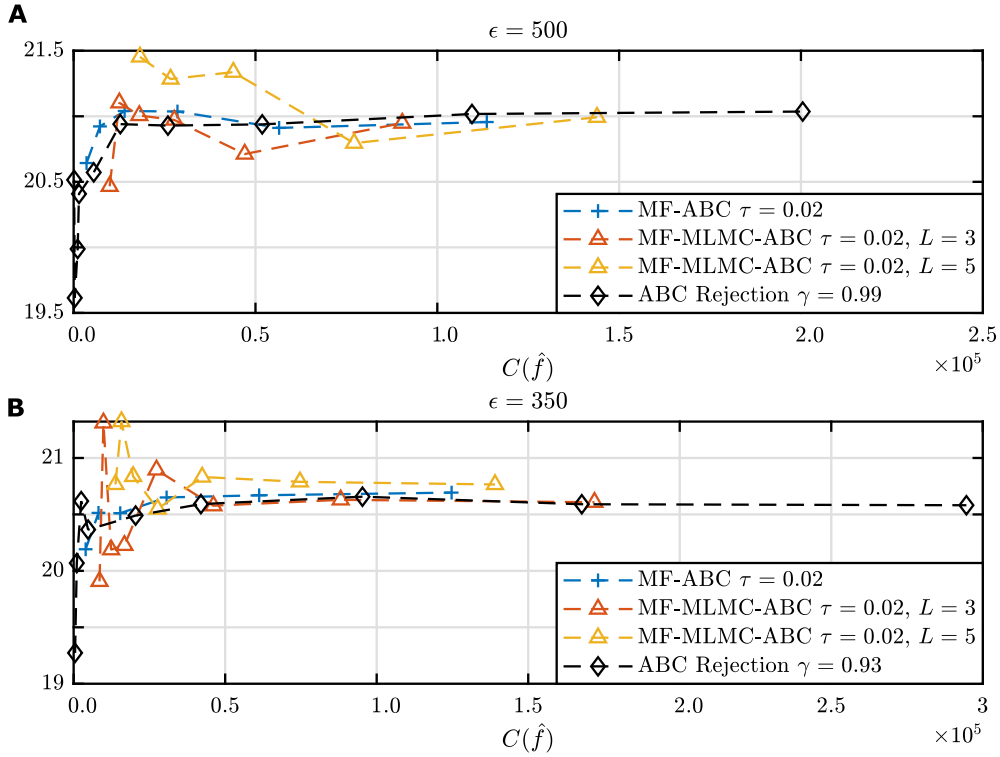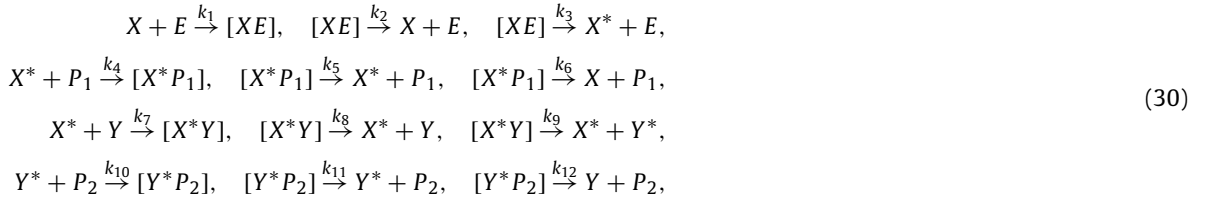
Fig. 7. Comparison of expectations for MF-MLMC-ABC (yellow and red triangles) with MF-ABC (blue crosses), and ABC rejection sampling (black diamonds) using the repressilator model with thresholds: (A) $\epsilon = 500$; and (B) $\epsilon = 350$.

in mammalian cells [65,94]. This two-step MAPK cascade model involves four coupled Michaelis–Menten components that govern the phosphorylation and dephosphorylation of two proteins X and Y,

$$
\begin{aligned}
X + E &\xrightarrow{k_1} [XE], \quad [XE] \xrightarrow{k_2} X + E, \quad [XE] \xrightarrow{k_3} X^* + E, \\
X^* + P_1 &\xrightarrow{k_4} [X^*P_1], \quad [X^*P_1] \xrightarrow{k_5} X^* + P_1, \quad [X^*P_1] \xrightarrow{k_6} X + P_1, \\
X^* + Y &\xrightarrow{k_7} [X^*Y], \quad [X^*Y] \xrightarrow{k_8} X^* + Y, \quad [X^*Y] \xrightarrow{k_9} X^* + Y^*, \\
Y^* + P_2 &\xrightarrow{k_{10}} [Y^*P_2], \quad [Y^*P_2] \xrightarrow{k_{11}} Y^* + P_2, \quad [Y^*P_2] \xrightarrow{k_{12}} Y + P_2,
\end{aligned}
\tag{30}
$$

where $k_1, k_2, \ldots, k_{12}$ are kinetic rate parameters, X^*, Y^* are the activated (phosphorylated) proteins, E is the enzyme involved in the activation of the X protein, and P_1, P_2 are phosphatase molecules that dephosphorylate X^*, Y^*. Finally note the two-step process where the activated X^* protein acts as an enzyme in the activation of Y.

In this case, we assume only activated proteins can be detected, therefore we consider the observation process

$$
\mathbf{y}_{\text{obs}}(t) \sim \mathcal{N}([X_t^*, Y_t^*]^{\mathsf{T}}, \sigma^2 \mathbf{I}),
\tag{31}
$$

where $\mathbf{I}$ the 2×2 identity matrix. Discrete observations are taken at regular four time unit intervals. $t_i = 4i$, for $i = 0, 1, \ldots 50$. Fig. 8(A) shows an example realisation of the two-step MAPK cascade model along with discrete observations of the activated proteins. Figs. 8(B)–(D) provide additional detail and highlight the complex dynamics of the unobserved chemical species. Given the very limited data in this realistic scenario, we do not have practical identifiability for all rate parameters. Based on the network structure, we only expect parameters k_3, k_6, k_9, and k_{12} to be identifiable as these rates correspond to reactions that change the copy numbers of the observed variables.

The target ABC inference problem we consider is to estimate the reverse rate parameter of the dephosphorylation reaction for the deactivation of Y^* and the marginal posterior distributions for the identifiable rate parameters

$$
\mathbb{E}\left[k_{11} \mid \mathbf{Y}_{\text{obs}}\right] \approx \int_{\mathbb{R}^8} k_{11} \, p(\boldsymbol{\theta} \mid \rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) \leq \epsilon) \, \mathrm{d}\boldsymbol{\theta},
$$

$$
\mathbb{P}(k_i < s_i \mid \mathbf{Y}_{\text{obs}}) \approx \int_{\mathbb{R}^8} \mathbb{1}_{(-\infty, s_i]}(k_i) \, p(\boldsymbol{\theta} \mid \rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) \leq \epsilon) \, \mathrm{d}\boldsymbol{\theta}, \quad i = 3, 6, 9, 12
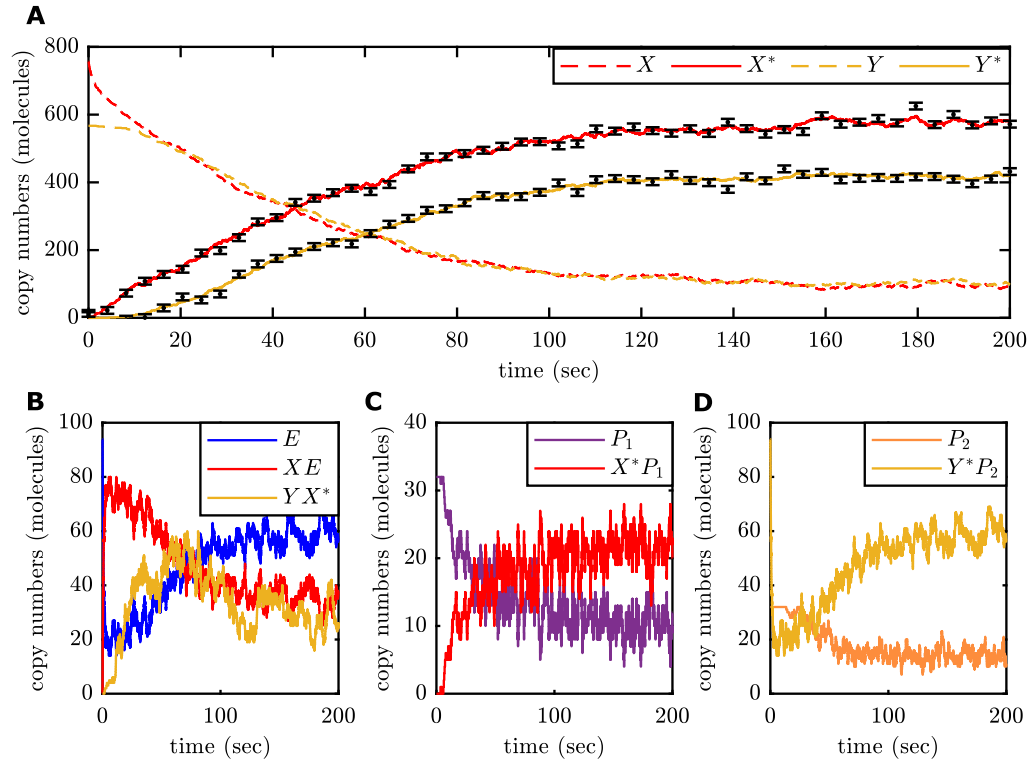\tag{32}
$$

Fig. 8. Example realisation of the two-step MAPK cascade reaction network model along with noisy observations of the two phosphorylated protein molecules, $\mathbf{y}_{\text{obs}}(t) \sim \mathcal{N}([X_t^*, Y_t^*]^{\mathsf{T}}, \sigma^2 \mathbf{I})$ (error bars indicate $\mathbf{y}_{\text{obs}}(t) \pm \sigma$). Here, the initial condition is $E_0 = 94$, $X_0 = 757$, $Y = 567$, $P_1 = P_2 = 32$, and $X^* = Y^* = EX = X^*P_1 = YX^* = Y^*P_2$, and the true rate parameters are $k_1 = 0.001$, $k_2 = k_1/120$, $k_3 = 0.18$, $k_4 = 0.001$, $k_5 = k_4/22$, $k_6 = 0.3$, $k_7 = 0.0001$, $k_8 = k_7/110$, $k_9 = 0.2$, $k_{10} = 0.001$, $k_{11} = k_{10}/22$, and $k_{12} = 0.3$, and observations are taken at $t_i = 4i$, for $i = 1, 2, \ldots 50$ with standard deviation $\sigma = 10$.

where $\boldsymbol{\theta} = [k_2, k_3, k_5, k_6, k_8, k_9, k_{11}, k_{12}]$ is the vector of unknown rate parameters, $\mathbf{Y}_{\text{obs}} = [\mathbf{y}_{\text{obs}}(t_0), \ldots, \mathbf{y}_{\text{obs}}(t_{50})]$ are noisy observations of the activated protein copy numbers at discrete times (Equation (31)), $\mathbf{Y}_s \sim s(\cdot \mid \boldsymbol{\theta})$ is simulated data of the two-step MAPK model using the Gillespie direct method and simulating the observation process (Equation (31)), ϵ is the discrepancy threshold, and the discrepancy metric is $\rho(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s) = \|\mathbf{Y}_{\text{obs}} - \mathbf{Y}_s\|_2$ where $\| \cdot \|_2$ is the Euclidean norm. Independent uniform priors are used with $k_2 \sim \mathcal{U}(0, k_1)$, $k_3 \sim \mathcal{U}(0, 1)$, $k_5 \sim \mathcal{U}(0, k_4)$, $k_6 \sim \mathcal{U}(0, 1)$, $k_8 \sim \mathcal{U}(0, k_7)$, $k_9 \sim \mathcal{U}(0, 1)$, $k_{11} \sim \mathcal{U}(0, k_{10})$, $k_{12} \sim \mathcal{U}(0, 1)$. We treat the rate parameters of the complex binding in all Michaelis–Menten reactions as known with $k_1 = k_4 = 0.001$, and $k_7 = k_{10}$.

We apply MF-MLMC-ABC to this problem with $L = 7$, $\tau = 0.5$, $\epsilon_1 = 1600$, $\epsilon_L = 300$, $\epsilon_\ell = \epsilon_{\ell-1}/m$ for all $\ell = 2, \ldots, L$ and $m \approx 1.32$, $\tilde{\epsilon}_\ell = \epsilon_\ell$ for $\ell = 1, \ldots, L$, $\rho^\tau(\mathbf{Y}_{\text{obs}}, \mathbf{Y}_s^\tau) = \|\mathbf{Y}_{\text{obs}} - \mathbf{Y}_s^\tau\|_2$ with $\mathbf{Y}_s^\tau \sim s^\tau(\cdot \mid \boldsymbol{\theta})$ the approximate stochastic simulation process using the tau-leaping method (Algorithm 2), and $\tau_\ell = \tau$ for $\ell = 1, 2, \ldots, L$. Just as with the repressilator, we observe significant improvements (Fig. 9) against MLMC-ABC, MF-ABC, and ABC rejection sampling. We also show the estimated marginal posterior densities for the identifiable parameters, k_3, k_6 k_9, and k_{12}, in Fig. 10.

In this case, MF-MLMC-ABC exceeds two orders of magnitude improvement over both MF-ABC and ABC rejection sampling. It is also clear that the MLMC-ABC approach is still very effective, with almost two orders of magnitude improvement over ABC rejection sampling. However, just as with the repressilator model, the MF-MLMC-ABC approach achieves the improved convergence rate of MLMC and reduced computation time for each term in the telescoping summation, notwithstanding a cost reduction in the MLMC tuning overhead.

4. Discussion

In this work, we have introduced a new approach to ABC-based parameter inference for partially observed stochastic processes. Our approach combines the benefits of MLMC variance reduction techniques [51,61,62] with the benefits of a multifidelity method for reducing stochastic simulation costs [46,47,82]. We have developed the formulation by applying the multifidelity weighting scheme directly to the MLMC telescoping summation for ABC inference. Our practical implementation of the new algorithm demonstrates how various components of the algorithm are tuned and optimised, leading to the acceleration of ABC inference by two orders of magnitude for realistic inference problems in systems biology.

These promising results open new avenues for accelerating various ABC-based schemes. While we focus on the acceleration of ABC rejection sampling, this is chosen as a first step toward other schemes. For example, Jasra et al. [61] demonstrate an MLMC approach to SMC for ABC, and Prescott and Baker [47] develop an SMC-based implementation of multifidelity ABC.
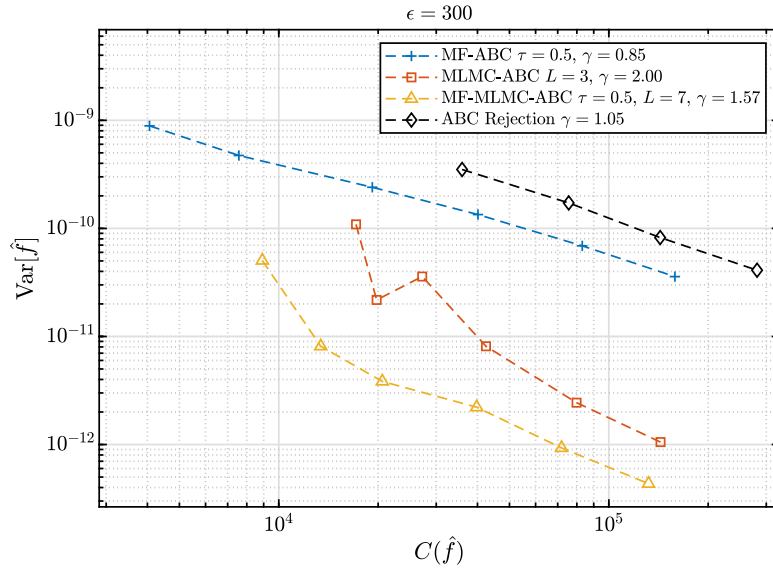
Fig. 9. Comparison of convergence rates for MF-MLMC-ABC (yellow triangles) with MF-ABC (blue crosses), MLMC-ABC (red squares), and ABC rejection sampling (black diamonds) using the two-step MAPK model with threshold $\epsilon = 300$. Rates are estimated by fitting $\mathrm{Var}\left[\hat{f}\right] \propto C(\hat{f})^{-\gamma}$ to benchmark data using least squares.
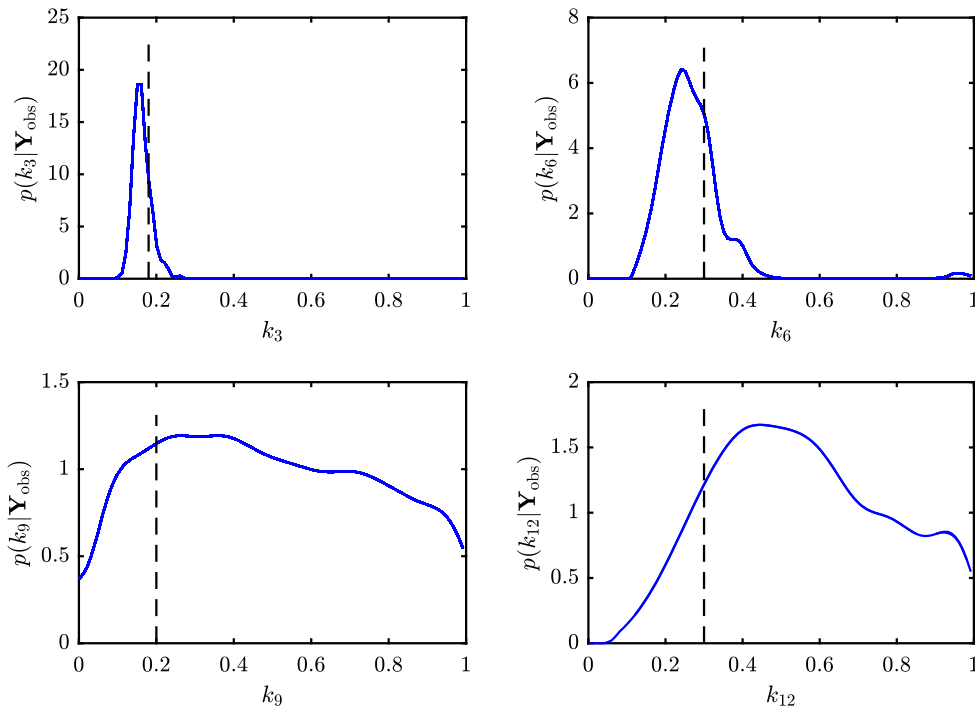

Fig. 10. Marginal posterior probability density functions (blue lines) estimated by the for identifiable parameters from the two-stage MAPK cascade model. True parameter values are indicated (black dashed lines). Densities were estimated using the MF-MLMC-ABC method with $L = 7$, $\tau = 0.5$ and $\epsilon_1 = 1600$ and $\epsilon_7 = 300$.

Both of these methods improve upon SMC for ABC and future work combining these methods could produce a multifidelity MLMC version of SMC. Further, ABC is not the only likelihood-free method that relies heavily on many stochastic simulations from the model, pseudo-marginal methods [23–25] and BSL [28,95] also use simulations to either estimate the likelihood function or construct a Gaussian approximation to the likelihood. There may be opportunities to combine MLMC and multifidelity methods for both these approaches to likelihood-free inference. For example, Jasra et al. [96] consider a MLMC approach to particle MCMC and it may be possible to obtain additional benefit from multifidelity methods.

We have also focussed our attention on ABC inference for partially observed discrete-state Markov processes since ABC methods are widely applied for such models in systems biology [21,93,97,98], epidemiology [99–103], ecology [104,105], and physics [106–108]. Consequently, the high-fidelity and low-fidelity simulations within the MF-ABC framework are, respectively, assumed to be the natural choices of Gillespie's direct method and the tau-leaping method parameterised by τ. However, our methodology supports many other variations, including differences in discrepancy metrics, summary statistics, and discrepancy thresholds, that could provide improvements in the ability for the low-fidelity simulation to predict the outcome of the high-fidelity simulation and thus reduce the optimal continuation probabilities. Furthermore, the approach is generally applicable to any ABC inference problem in which an appropriate approximate simulation scheme can be identified. For example, there is a wide range of potential choices for approximations that will lead to efficient multifidelity sampling [83], such as model reduction [65,109], mean-field or linear approximations [45,66,110,111], or surrogate models that act as emulators [112–114]. Finally, recent advances in deep learning provide potential for the automatic construction of surrogates [29–31].

An additional novel modification we develop here, with further generalisations developed in Prescott et al. [63], is an adaptive scheme for iteratively updating the continuation probabilities (Supplementary Material). This approach is more robust than previous implementations [46], however, it relies upon a fixed choice of the time-step, τ, in the approximation. While this scheme will identify poor choices in τ with $\eta_1, \eta_2 \to 1$, future work should investigate adaptive schemes to optimise τ along with η_1, η_2; this would lead to a near automatic tuning step for MF-MLMC-ABC. Furthermore, a similar adaptive optimisation approach could be used to adapt the optimal MLMC sample size sequence to reduce the MLMC tuning overhead. However, a more pressing issue with MLMC-based approaches to ABC is the assumption of a fixed sequence of discrepancy thresholds [47,61,62], while the state-of-the-art in SMC-ABC is to adaptively select this sequence [38,115]. Such adaptive schemes may be possible for MLMC-based ABC especially if applied in an SMC setting. Finally, other variance reduction techniques, such as array randomised quasi-Monte Carlo methods for Markov chains, could be applied to accelerate estimation of the multifidelity expectations even further [116–118].

Our implementations have been developed efficiently using a high-level programming environment using a single CPU core. While this is sufficient to demonstrate the computational benefits of our new algorithm, there are many additional optimisation techniques that could also be applied here. In particular, one substantial advantage of utilising MF-MLMC-ABC based upon rejection sampling, is that there are few synchronisation steps. Therefore, most of the stochastic simulations (especially the approximate simulations) can exploit many parallel computing architectures, such as general-purpose graphics processing units (GPGPUs) [119–121], single instruction multiple data (SIMD) CPU processors [122,123], and recent advances in AI hardware [124]. This leads to methods in which the statistical efficiency also directly scales to state-of-the-art massively parallel computing.

Finally, our numerical results demonstrate that high performance and high precision Bayesian inference can be performed for challenging partially observed stochastic processes such as those that arise in systems biology. We demonstrate this high performance for a large network with 12 parameters and 11 chemical species without any dimensionality reduction using summary statistics that is almost always required for ABC inference to be viable. With computational improvements of up to two orders of magnitude, our method is a significant advance in the use of approximations to accelerate ABC inference without incurring accuracy penalties. Furthermore, the success of the multifidelity and MLMC approach to inference will enable more realistic and complex models to be used to analyse modern, high resolution data.

Software Availability

Matlab source code with example implementations and demonstrations for all numerical examples presented in this work is available on GitHub at https://github.com/davidwarne/MLMCandMultifidelityForABC).

CRediT authorship contribution statement

David J. Warne: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft. **Thomas P. Prescott:** Conceptualization, Formal analysis, Funding acquisition, Methodology, Software, Validation, Writing – review & editing. **Ruth E. Baker:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Matthew J. Simpson:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jcp.2022.111543.

References

[1] F. van der Meulen, M. Schauer, Bayesian estimation of incompletely observed diffusions, Int. J. Probab. Stochast. Process. 90 (5) (2017) 641–662, https://doi.org/10.1080/17442508.2017.1381097.

[2] P. Dellaportas, N. Friel, G.O. Roberts, Bayesian model selection for partially observed diffusion models, Biometrika 93 (4) (2006) 809–825, https://doi.org/10.1093/biomet/93.4.809.

[3] A. Golightly, D.J. Wilkinson, Bayesian sequential inference for nonlinear multivariate diffusions, Stat. Comput. 16 (4) (2006) 323–338, https://doi.org/10.1007/s11222-006-9392-x.

[4] M.B. Elowitz, A.J. Levine, E.D. Siggia, P.S. Swain, Stochastic gene expression in a single cell, Science 297 (2002) 1183–1186, https://doi.org/10.1126/science.1070919.

[5] A. Raj, A. van Oudenaarden, Nature, nurture, or chance: stochastic gene expression and its consequences, Cell 135 (2) (2008) 216–226, https://doi.org/10.1016/j.cell.2008.09.050.

[6] D.J. Wilkinson, Stochastic modelling for quantitative description of heterogeneous biological systems, Nat. Rev. Genet. 10 (2009) 122–133, https://doi.org/10.1038/nrg2509.

[7] N. Fedoroff, W. Fontana, Small numbers of big molecules, Science 297 (2002) 1129–1131, https://doi.org/10.1126/science.1075988.

[8] W.J. Blake, M. Kaern, C.R. Cantor, J.J. Collins, Noise in eukaryotic gene expression, Nature 422 (2003) 633–637, https://doi.org/10.1038/nature01546.

[9] S. Braichenko, J. Holehouse, R. Grima, Distinguishing between models of mammalian gene expression: telegraph-like models versus mechanistic models, J. R. Soc. Interface 18 (2021) 20210510, https://doi.org/10.1098/rsif.2021.0510.

[10] L. Gammaitoni, P. Hanggi, P. Jung, F. Marchesoni, Stochastic resonance, Rev. Mod. Phys. 70 (1) (1998) 223–287, https://doi.org/10.1103/RevModPhys.70.223.

[11] C.B. Muratov, E. Vanden-Eijnden, E. Weinan, Self-induced stochastic resonance in excitable systems, Physica D 210 (3–4) (2005) 227–240, https://doi.org/10.1016/j.physd.2005.07.014.

[12] T. Wellens, V. Shatokhin, A. Buchleitner, Stochastic resonance, Rep. Prog. Phys. 67 (1) (2004) 45–105, https://doi.org/10.1088/0034-4885/67/1/R02.

[13] J. Paulsson, O.G. Berg, M. Ehrenberg, Stochastic focusing: fluctuation-enhanced sensitivity of intracellular regulation, Proc. Natl. Acad. Sci. 97 (13) (2000) 7148–7153, https://doi.org/10.1073/pnas.110057697.

[14] F. Schlögl, Chemical reaction models for non-equilibrium phase transitions, Z. Phys. 253 (2) (1972) 147–161, https://doi.org/10.1007/BF01379769.

[15] J. Schnakenberg, Simple chemical reaction systems with limit cycle behaviour, J. Theor. Biol. 81 (3) (1979) 389–400, https://doi.org/10.1016/0022-5193(79)90042-0.

[16] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, D.B. Rubin, Bayesian Data Analysis, 3rd edition, Chapman & Hall/CRC, 2014.

[17] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (6) (1953) 1087–1092, https://doi.org/10.1063/1.1699114.

[18] W.K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1) (1970) 97–109.

[19] P. Del Moral, A. Doucet, A. Jasra, Sequential Monte Carlo samplers, J. R. Stat. Soc., Ser. B, Stat. Methodol. 68 (3) (2006) 411–436, https://doi.org/10.1111/j.1467-9868.2006.00553.x.

[20] S.A. Sisson, Y. Fan, M. Beaumont, Handbook of Approximate Bayesian Computation, Chapman and Hall/CRC Press, 2018.

[21] M. Sunnåker, A.G. Busetto, E. Numminen, J. Corander, M. Foll, C. Dessimoz, Approximate Bayesian computation, PLoS Comput. Biol. 9 (1) (2013) e1002803, https://doi.org/10.1371/journal.pcbi.1002803.

[22] D.J. Warne, R.E. Baker, M.J. Simpson, Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art, J. R. Soc. Interface 16 (151) (2019) 20180943, https://doi.org/10.1098/rsif.2018.0943.

[23] C. Andrieu, G.O. Roberts, The pseudo-marginal approach for efficient Monte Carlo computations, Ann. Stat. 37 (2) (2009) 697–725.

[24] C. Andrieu, A. Doucet, R. Holenstein, Particle Markov chain Monte Carlo methods, J. R. Stat. Soc., Ser. B, Stat. Methodol. 72 (3) (2010) 269–342.

[25] D.J. Warne, R.E. Baker, M.J. Simpson, A practical guide to pseudo-marginal methods for computational inference in systems biology, J. Theor. Biol. 496 (2020) 110255, https://doi.org/10.1016/j.jtbi.2020.110255.

[26] Z. An, D.J. Nott, C. Drovandi, Robust Bayesian synthetic likelihood via a semi-parametric approach, Stat. Comput. 30 (3) (2019) 543–557, https://doi.org/10.1007/s11222-019-09904-x.

[27] V.M.H. Ong, D.J. Nott, M.-N. Tran, S.A. Sisson, C.C. Drovandi, Variational Bayes with synthetic likelihood, Stat. Comput. 28 (4) (2017) 971–988, https://doi.org/10.1007/s11222-017-9773-3.

[28] L.F. Price, C.C. Drovandi, A. Lee, D.J. Nott, Bayesian synthetic likelihood, J. Comput. Graph. Stat. 27 (1) (2017) 1–11, https://doi.org/10.1080/10618600.2017.1302882.

[29] J.-M. Lueckmann, P.J. Goncalves, G. Bassetto, K. Öcal, M. Nonnenmacher, J.H. Macke, Flexible statistical inference for mechanistic models of neural dynamics, Adv. Neural Inf. Process. Syst. 30 (2017), https://proceedings.neurips.cc/paper/2017/file/addfa9b7e234254d26e9c7f2af1005cb-Paper.pdf.

[30] G. Papamakarios, D. Sterratt, I. Murray, Sequential neural likelihood: fast likelihood-free inference with autoregressive flows, Proc. Mach. Learn. Res. 89 (2019) 837–848, https://proceedings.mlr.press/v89/papamakarios19a.html.

[31] K. Cranmer, J. Brehmer, G. Louppe, The frontier of simulation-based inference, Proc. Natl. Acad. Sci. USA 117 (48) (2020) 30055–30062, https://doi.org/10.1073/pnas.1912789117.

[32] J.K. Pritchard, M.T. Seielstad, A. Perez-Lezaun, M.W. Feldman, Population growth of human y chromosomes: a study of y chromosome microsatellites, Mol. Biol. Evol. 16 (12) (1999) 1791–1798, https://doi.org/10.1093/oxfordjournals.molbev.a026091.

[33] S. Tavaré, D.J. Balding, R.C. Griffiths, P. Donnelly, Inferring coalescence times from DNA sequence data, Genetics 145 (2) (1997) 505–518.

[34] M.A. Beaumont, W. Zhang, D.J. Balding, Approximate Bayesian computation in population genetics, Genetics 162 (4) (2002) 2025–2035.

[35] P. Marjoram, J. Molitor, V. Plagnol, S. Tavaré, Markov chain Monte Carlo without likelihoods, Proc. Natl. Acad. Sci. USA 100 (26) (2003) 15324–15328, https://doi.org/10.1073/pnas.0306899100.

[36] S.A. Sisson, Y. Fan, M.M. Tanaka, Sequential Monte Carlo without likelihoods, Proc. Natl. Acad. Sci. USA 104 (6) (2007) 1760–1765, https://doi.org/10.1073/pnas.0607208104.

[37] P. Del Moral, A. Doucet, A. Jasra, An adaptive sequential Monte Carlo method for approximate Bayesian computation, Stat. Comput. 22 (5) (2012) 1009–1020, https://doi.org/10.1007/s11222-011-9271-y.

[38] C.C. Drovandi, A.N. Pettitt, Estimation of parameters for macroparasite population evolution using approximate Bayesian computation, Biometrics 67 (1) (2011) 225–233, https://doi.org/10.1111/j.1541-0420.2010.01410.x.

[39] J.J. Bon, A. Lee, C. Drovandi, Accelerating sequential Monte Carlo with surrogate likelihoods, Stat. Comput. 31 (2021) 62, https://doi.org/10.1007/s11222-021-10036-4, arXiv:2009.03699.

[40] M. Banterle, C. Grazian, A. Lee, C.P. Robert, Accelerating Metropolis-Hastings algorithms by delayed acceptance, Foundations Data Sci. 1 (2) (2019) 103–128, https://doi.org/10.3934/fods.2019005.

[41] R.G. Everitt, P.A. Rowińska, Delayed acceptance ABC-SMC, J. Comput. Graph. Stat. (2020) 1–12, https://doi.org/10.1080/10618600.2020.1775617.

[42] A. Golightly, D.A. Henderson, C. Sherlock, Delayed acceptance particle MCMC for exact inference in stochastic kinetic models, Stat. Comput. 25 (5) (2014) 1039–1055, https://doi.org/10.1007/s11222-014-9469-x.

[43] D. Prangle, Lazy ABC, Stat. Comput. 26 (1) (2014) 171–185, https://doi.org/10.1007/s11222-014-9544-3.

[44] M.D. Parno, Y.M. Marzouk, Transport map accelerated Markov chain Monte Carlo, SIAM/ASA J. Uncert. Quantific. 6 (2) (2018) 645–682, https://doi.org/10.1137/17m1134640.

[45] D.J. Warne, R.E. Baker, M.J. Simpson, Rapid Bayesian inference for expensive stochastic models, J. Comput. Graph. Stat. (2021) (in press), arXiv:1909.06540.

[46] T.P. Prescott, R.E. Baker, Multifidelity approximate Bayesian computation, SIAM/ASA J. Uncert. Quantific. 8 (1) (2020) 114–138, https://doi.org/10.1137/18m1229742.

[47] T.P. Prescott, R.E. Baker, Multifidelity approximate Bayesian computation with sequential Monte Carlo parameter sampling, SIAM/ASA J. Uncert. Quantific. 9 (2) (2021) 788–817, https://doi.org/10.1137/20m1316160.

[48] B. Peherstorfer, K. Willcox, M. Gunzburger, Optimal model management for multifidelity Monte Carlo estimation, SIAM J. Sci. Comput. 38 (5) (2016) A3163–A3194, https://doi.org/10.1137/15M1046472.

[49] S. Heinrich, Monte Carlo complexity of global solution of integral equations, J. Complex. 14 (2) (1998) 151–175, https://doi.org/10.1006/jcom.1998.0471.

[50] M.B. Giles, Multilevel Monte Carlo methods, Acta Numer. 24 (2015) 259–328, https://doi.org/10.1017/S09624929.

[51] M.B. Giles, Multilevel Monte Carlo path simulation, Oper. Res. 56 (3) (2008) 607–617, https://doi.org/10.1287/opre.1070.0496.

[52] D.F. Anderson, D.J. Higham, Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics, Multiscale Model. Simul. 10 (1) (2012) 146–179, https://doi.org/10.1137/110840546.

[53] C. Lester, R.E. Baker, M.B. Giles, C.A. Yates, Extending the multi-level method for the simulation of stochastic biological systems, Bull. Math. Biol. 78 (8) (2016) 1640–1677, https://doi.org/10.1007/s11538-016-0178-9.

[54] C. Lester, C.A. Yates, M.B. Giles, R.E. Baker, An adaptive multi-level simulation algorithm for stochastic biological systems, J. Chem. Phys. 142 (2) (2015) 024113, https://doi.org/10.1063/1.4904980.

[55] T.J. Dodwell, C. Ketelsen, R. Scheichl, A.L. Teckentrup, A hierarchical multilevel Markov chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow, SIAM/ASA J. Uncert. Quantific. 3 (1) (2015) 1075–1108, https://doi.org/10.1137/130915005.

[56] T.J. Dodwell, C. Ketelsen, R. Scheichl, A.L. Teckentrup, Multilevel Markov chain Monte Carlo, SIAM Rev. 61 (3) (2019) 509–545, https://doi.org/10.1137/19m126966x.

[57] Y. Efendiev, B. Jin, P. Michael, X. Tan, Multilevel Markov chain Monte Carlo method for high-contrast single-phase flow problems, Commun. Comput. Phys. 17 (1) (2015) 259–286, https://doi.org/10.4208/cicp.021013.260614a.

[58] A. Beskos, A. Jasra, K. Law, R. Tempone, Y. Zhou, Multilevel sequential Monte Carlo samplers 127 (2017) 1417–14440, https://doi.org/10.1016/j.spa.2016.08.004.

[59] J. Latz, I. Papaioannou, E. Ullmann, Multilevel sequential^2 Monte Carlo for Bayesian inverse problems, J. Comput. Phys. 368 (2018) 154–178, https://doi.org/10.1016/j.jcp.2018.04.014.

[60] N. Guha, X. Tan, Multilevel approximate Bayesian approaches for flows in highly heterogeneous porous media and their applications, J. Comput. Appl. Math. 317 (2017) 700–717, https://doi.org/10.1016/j.cam.2016.10.008.

[61] A. Jasra, S. Jo, D. Nott, C. Shoemaker, R. Tempone, Multilevel Monte Carlo in approximate Bayesian computation, Stoch. Anal. Appl. 37 (3) (2019) 346–360, https://doi.org/10.1080/07362994.2019.1566006.

[62] D.J. Warne, R.E. Baker, M.J. Simpson, Multilevel rejection sampling for approximate Bayesian computation, Comput. Stat. Data Anal. 124 (2018) 71–86, https://doi.org/10.1016/j.csda.2018.02.009.

[63] T.P. Prescott, D.J. Warne, R.E. Baker, Efficient multifidelity likelihood-free Bayesian inference with adaptive computational resource allocation, http://arxiv.org/abs/2112.11971, 2021.

[64] V. Dhananjaneyulu, V.N.S. P, G. Kumar, G.A. Viswanathan, Noise propagation in two-step series MAPK cascade, PLoS ONE 7 (5) (2012) e35958, https://doi.org/10.1371/journal.pone.0035958.

[65] K.S. Brown, C.C. Hill, G.A. Calero, C.R. Myers, K.H. Lee, J.P. Sethna, R.A. Cerione, The statistical mechanics of complex signaling networks: nerve growth factor signaling, Phys. Biol. 1 (3) (2004) 184–195, https://doi.org/10.1088/1478-3967/1/3/006.

[66] D. Schnoerr, G. Sanguinetti, R. Grima, Approximation and inference methods for stochastic biochemical kinetics—a tutorial review, J. Phys. A, Math. Theor. 50 (9) (2017) 093001.

[67] D.J. Higham, Modeling and simulating chemical reactions, SIAM Rev. 50 (2) (2008) 347–368, https://doi.org/10.1137/060666457.

[68] D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (25) (1977) 2340–2361, https://doi.org/10.1021/j100540a008.

[69] T.G. Kurtz, The relationship between stochastic and deterministic models for chemical reactions, J. Chem. Phys. 57 (7) (1972) 2976–2978, https://doi.org/10.1063/1.1678692.

[70] R. Erban, S.J. Chapman, P.K. Maini, A practical guide to stochastic simulation of reaction-diffusion processes, http://arxiv.org/abs/0704.1908, 2007.

[71] D.T. Gillespie, A rigorous derivation of the chemical master equation, Physica A 188 (1992) 404–425, https://doi.org/10.1016/0378-4371(92)90283-V.

[72] M.A. Gibson, J. Bruck, Efficient exact stochastic simulation of chemical systems with many species and many channels, J. Phys. Chem. 104 (9) (2000) 1876–1889, https://doi.org/10.1021/jp993732q.

[73] D.F. Anderson, A modified next reaction method for simulating chemical systems with time dependent propensities and delays, J. Chem. Phys. 127 (2007) 214107, https://doi.org/10.1063/1.2799998.

[74] D.T. Gillespie, Approximate accelerated simulation of chemically reacting systems, J. Chem. Phys. 115 (4) (2001) 1716–1733, https://doi.org/10.1063/1.1378322.

[75] D.T. Gillespie, The chemical Langevin equation, J. Chem. Phys. 113 (1) (2000) 297–306, https://doi.org/10.1063/1.481811.

[76] T. Tian, K. Burrage, Binomial leap methods for simulating stochastic chemical kinetics, J. Chem. Phys. 121 (21) (2004) 10356–10364, https://doi.org/10.1063/1.1810475.

[77] Y. Cao, H. Li, L. Petzold, Efficient formulation of the stochastic simulation algorithm for chemically reacting systems, J. Chem. Phys. 121 (2004) 4059–4067, https://doi.org/10.1063/1.1778376.

[78] D.F. Anderson, A. Ganguly, T.G. Kurtz, Error analysis of tau-leap simulation methods, Ann. Appl. Probab. 21 (6) (2011) 2226–2262, https://doi.org/10.1214/10-AAP756.

[79] D. Wilson, R.E. Baker, Multi-level methods and approximating distribution functions, AIP Adv. 6 (2016) 075020, https://doi.org/10.1063/1.4960118.

[80] M.B. Giles, T. Nagapetyan, K. Ritter, Multilevel Monte Carlo approximation of distribution functions and densities, SIAM/ASA J. Uncert. Quantific. 3 (1) (2015) 267–295, https://doi.org/10.1137/140960086.

[81] P. Fearnhead, D. Prangle, Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation, J. R. Stat. Soc., Ser. B, Stat. Methodol. 74 (3) (2012) 419–474, https://doi.org/10.1111/j.1467-9868.2011.01010.x.

[82] C.-H. Rhee, P.W. Glynn, Unbiased estimation with square root convergence for SDE models, Oper. Res. 63 (5) (2015) 1026–1043, https://doi.org/10.1287/opre.2015.1404.

[83] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, SIAM Rev. 60 (3) (2018) 550–591, https://doi.org/10.1137/16m1082469.

[84] D.J. Warne, Computational inference in mathematical biology: methodological developments and applications, PhD Thesis, Queensland University of Technology, 2020, https://doi.org/10.5204/thesis.eprints.202835.

[85] L. Michaelis, M.L. Menten, Die Kinetik der Invertinwirkung, Biochem. Z. 49 (1913) 333–369.

[86] C.V. Rao, A.P. Arkin, Stochastic chemical kinetics and the quasi-steady-state assumption: application to the Gillespie algorithm, J. Chem. Phys. 118 (11) (2003) 4999–5010, https://doi.org/10.1063/1.1545446.

[87] B. Finkenstädt, E.A. Heron, M. Komorowski, K. Edwards, S. Tang, C.V. Harper, J.R.E. Davis, M.R.H. White, A.J. Millar, D.A. Rand, Reconstruction of transcriptional dynamics from gene reporter data using differential equations, Bioinformatics 24 (24) (2008) 2901–2907, https://doi.org/10.1093/bioinformatics/btn562.

[88] M.A.J. Iafolla, M. Mazumder, V. Sardana, T. Velauthapillai, K. Pannu, D.R. McMillen, Dark proteins: effect of inclusion body formation on quantification of protein expression, Proteins, Struct. Funct. Bioinform. 72 (4) (2008) 1233–1242, https://doi.org/10.1002/prot.22024.

[89] B. Bajar, E. Wang, S. Zhang, M. Lin, J. Chu, A guide to fluorescent protein FRET pairs, Sensors 16 (9) (2016) 1488, https://doi.org/10.3390/s16091488.

[90] M.J. Simpson, R.E. Baker, S.T. Vittadello, O.J. Maclaren, Practical parameter identifiability for spatio-temporal models of cell invasion, J. R. Soc. Interface 17 (164) (2020) 20200055, https://doi.org/10.1098/rsif.2020.0055.

[91] H.-O. Georgii, Stochastics, De Gruyter, 2013.

[92] M.B. Elowitz, S. Leibler, A synthetic oscillatory network of transcriptional regulators, Nature 403 (6767) (2000) 335–338, https://doi.org/10.1038/35002125.

[93] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, M.P.H. Stumpf, Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems, J. R. Soc. Interface 6 (2009) 187–202, https://doi.org/10.1098/rsif.2008.0172.

[94] K. Oda, Y. Matsuoka, A. Funahashi, H. Kitano, A comprehensive pathway map of epidermal growth factor receptor signaling, Phys. Biol. 1 (1) (2005), https://doi.org/10.1038/msb4100014.

[95] J.W. Priddle, S.A. Sisson, D.T. Frazier, I. Turner, C. Drovandi, Efficient Bayesian synthetic likelihood with whitening transformations, J. Comput. Graph. Stat. (2021) 1–27, https://doi.org/10.1080/10618600.2021.1979012.

[96] A. Jasra, K. Kamatani, K. Law, Y. Zhou, Bayesian static parameter estimation for partially observed diffusions via multilevel Monte Carlo, SIAM J. Sci. Comput. 40 (2) (2018) A887–A902, https://doi.org/10.1137/17M1112595.

[97] J. Liepe, P. Kirk, S. Filippi, T. Toni, C.P. Barnes, M.P.H. Stumpf, A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation, Nat. Protoc. 9 (2) (2014) 439–456, https://doi.org/10.1038/nprot.2014.025.

[98] Q. Wu, K. Smith-Miles, T. Tian, Approximate Bayesian computation schemes for parameter inference of discrete stochastic models using simulated likelihood density, BMC Bioinform. 15 (S12) (2014), https://doi.org/10.1186/1471-2105-15-s12-s3.

[99] M. Chinazzi, J.T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A.P. y Piontti, K. Mu, L. Rossi, K. Sun, C. Viboud, X. Xiong, H. Yu, M.E. Halloran, I.M. Longini, A. Vespignani, The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak, Science 368 (6489) (2020) 395–400, https://doi.org/10.1126/science.aba9757.

[100] T.J. McKinley, I. Vernon, I. Andrianakis, N. McCreesh, J.E. Oakley, R.N. Nsubuga, M. Goldstein, R.G. White, Approximate Bayesian computation and simulation-based inference for complex stochastic epidemic models, Stat. Sci. 33 (1) (2018), https://doi.org/10.1214/17-sts618.

[101] A. Minter, R. Retkute, Approximate Bayesian computation for infectious disease modelling, Epidemics 29 (2019) 100368, https://doi.org/10.1016/j.epidem.2019.100368.

[102] J.N. Walker, A.J. Black, J.V. Ross, Bayesian model discrimination for partially-observed epidemic models, Math. Biosci. 317 (2019) 108266, https://doi.org/10.1016/j.mbs.2019.108266.

[103] D.J. Warne, A. Ebert, C. Drovandi, W. Hu, A. Mira, K. Mengersen, Hindsight is 2020 vision: a characterisation of the global response to the COVID-19 pandemic, BMC Public Health 20 (1) (2020), https://doi.org/10.1186/s12889-020-09972-z.

[104] M.A. Beaumont, Approximate Bayesian computation in evolution and ecology, Annu. Rev. Ecol. Evol. Syst. 41 (1) (2010) 379–406, https://doi.org/10.1146/annurev-ecolsys-102209-144621.

[105] J. Sirén, L. Lens, L. Cousseau, O. Ovaskainen, Assessing the dynamics of natural populations by fitting individual-based models with approximate Bayesian computation, Methods Ecol. Evol. 9 (5) (2018) 1286–1295, https://doi.org/10.1111/2041-210x.12964.

[106] J. Akeret, A. Refregier, A. Amara, S. Seehars, C. Hasner, Approximate Bayesian computation for forward modeling in cosmology, J. Cosmol. Astropart. Phys. 2015 (08) (2015) 043, https://doi.org/10.1088/1475-7516/2015/08/043.

[107] D. Barajas-Solano, A. Tartakovsky, Approximate Bayesian model inversion for PDEs with heterogeneous and state-dependent coefficients, J. Comput. Phys. 395 (2019) 247–262, https://doi.org/10.1016/j.jcp.2019.06.010.

[108] J.D. Christopher, N.T. Wimer, C. Lapointe, T.R.S. Hayden, I. Grooms, G.B. Rieker, P.E. Hamlington, Parameter estimation for complex thermal-fluid flows using approximate Bayesian computation, Phys. Rev. Fluids 3 (10) (2018) 104602, https://doi.org/10.1103/physrevfluids.3.104602.

[109] M.K. Transtrum, P. Qiu, Model reduction by manifold boundaries, Phys. Rev. Lett. 113 (9) (2014) 098701, https://doi.org/10.1103/physrevlett.113.098701.

[110] A.P. Browning, D.J. Warne, K. Burrage, R.E. Baker, M.J. Simpson, Identifiability analysis for stochastic differential equation models in systems biology, J. R. Soc. Interface 17 (173) (2020) 20200652, https://doi.org/10.1098/rsif.2020.0652.

[111] Z. Cao, R. Grima, Linear mapping approximation of gene regulatory networks with stochastic dynamics, Nat. Commun. 9 (1) (2018), https://doi.org/10.1038/s41467-018-05822-0.

[112] E.O. Buzbas, N.A. Rosenberg, AABC: approximate approximate Bayesian computation for inference in population-genetic models, Theor. Popul. Biol. 99 (31) (2015) 42, https://doi.org/10.1016/j.tpb.2014.09.002.

[113] R.K. Tripathy, I. Bilionis, Deep UQ: learning deep neural network surrogate models for high dimensional uncertainty quantification, J. Comput. Phys. 375 (2018) 565–588, https://doi.org/10.1016/j.jcp.2018.08.036.

[114] A. Borowska, D. Giurghita, D. Husmeier, Gaussian process enhanced semi-automatic approximate Bayesian computation: parameter inference in a stochastic differential equation system for chemotaxis, J. Comput. Phys. 429 (2021) 109999, https://doi.org/10.1016/j.jcp.2020.109999.

[115] M.A. Beaumont, J.-M. Cornuet, J.-M. Marin, C.P. Robert, Adaptive approximate Bayesian computation, Biometrika 96 (4) (2009) 983–990.

[116] P. L'Ecuyer, C. Lécot, A. L'Archevêque-Gaudet, On array-RQMC for Markov chains: mapping alternatives and convergence rates, in: Monte Carlo and Quasi-Monte Carlo Methods 2008, Springer, Berlin Heidelberg, 2009, pp. 485–500.

[117] F. Puchhammer, A.B. Abdellah, P. L'Ecuyer, Variance reduction with array-RQMC for tau-leaping simulation of stochastic biological and chemical reaction networks, Bull. Math. Biol. 83 (8) (2021), https://doi.org/10.1007/s11538-021-00920-5.

[118] C.H.L. Beentjes, R.E. Baker, Quasi-Monte Carlo methods applied to tau-leaping in stochastic biological systems, Bull. Math. Biol. 81 (8) (2018) 2931–2959, https://doi.org/10.1007/s11538-018-0442-2.

[119] A. Lee, C. Yau, M.B. Giles, A. Doucet, C.C. Holmes, On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods, J. Comput. Graph. Stat. 19 (4) (2010) 769–789, https://doi.org/10.1198/jcgs.2010.10039.

[120] G. Klingbeil, R. Erban, M. Giles, P.K. Maini, STOCHSIMGPU: parallel stochastic simulation for the systems biology toolbox 2 for MATLAB, Bioinformatics 27 (8) (2011) 1170–1171, https://doi.org/10.1093/bioinformatics/btr068.

[121] S. Hurn, K. Lindsay, D.J. Warne, A heterogeneous computing approach to maximum likelihood parameter estimation for the Heston model of stochastic volatility, ANZIAM J. 57 (2016) 364, https://doi.org/10.21914/anziamj.v57i0.10425.

[122] D.J. Warne, S.A. Sisson, C. Drovandi, Vector operations for accelerating expensive Bayesian computations – a tutorial guide, Bayesian Anal. 17 (2) (2021) 593–622, https://doi.org/10.1214/21-ba1265.

[123] A.S. Mahani, M.T. Sharabiani, SIMD parallel MCMC sampling with applications for big-data Bayesian analytics, Comput. Stat. Data Anal. 88 (2015) 75–99, https://doi.org/10.1016/j.csda.2015.02.010.

[124] S. Kulkarni, A. Tsyplikhin, M.M. Krell, C.A. Moritz, Accelerating simulation-based inference with emerging AI hardware, in: 2020 International Conference on Rebooting Computing (ICRC), IEEE, 2020.