# Accelerating computational Bayesian inference for stochastic biochemical reaction network models using multilevel Monte Carlo sampling

David J. Warne[1], Ruth E. Baker[2], Matthew J. Simpson[1*]

**1** School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland 4001, Australia.

**2** Mathematical Institute, University of Oxford, Oxford, OX2 6GG, United Kingdom.

* Corresponding author: matthew.simpson@qut.edu.au

## Abstract

Investigating the behavior of stochastic models of biochemical reaction networks generally relies upon numerical stochastic simulation methods to generate many realizations of the model. For many practical applications, such numerical simulation can be computationally expensive. The statistical inference of reaction rate parameters based on observed data is, however, a significantly greater computational challenge; often relying upon likelihood-free methods such as approximate Bayesian computation, that require the generation of millions of individual stochastic realizations. In this study, we investigate a new approach to computational inference, based on multilevel Monte Carlo sampling: we approximate the posterior cumulative distribution function through a combination of model samples taken over a range of acceptance thresholds. We demonstrate this approach using a variety of discrete-state, continuous-time Markov models of biochemical reaction networks. Results show that a computational gain over standard rejection schemes of up to an order of magnitude is achievable without significant loss in estimator accuracy.

# Author Summary

We develop a new method to infer the reaction rate parameters for stochastic models of biochemical reaction networks. Standard computational approaches, based on numerical simulations, are often used to estimate parameters. These computational approaches, however, are extremely expensive, potentially requiring millions of simulations. To alleviate this issue, we apply a different method of sampling allowing us to find an optimal trade-off between performance and accuracy. Our approach is approximately one order of magnitude faster than standard methods, without significant loss in accuracy.

# Introduction

Stochastic models of biochemical reaction networks often provide a more accurate description of system dynamics than deterministic models [1]. In many cases, this is due to the inherent stochastic nature of many biochemical processes in which the system dynamics is significantly influenced by relatively low populations of certain chemical species [2]. 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 [3, 4].

A common approach to modeling biochemical systems is to consider a well-mixed collection of molecules that react according to some known chemical reactions. The well-mixed assumption simplifies the model by removing the spatial component [5, 6]. If the model is deterministic, evolution of the concentrations of each chemical species is governed by a system of ordinary differential equations (ODEs). Alternatively, a stochastic model will typically consider the evolution of copy numbers (i.e., the numbers of molecules) of each species, with each reaction occurring stochastically [6].

In the case of the stochastic model, the *probability density function* (PDF) for the state of such a system at time $t$ evolves according to a very large system of ODEs known as the *chemical master equation* (CME), which is in general intractable due to the very large, or countably infinite, number of possible system states [5, 7]. As a result, stochastic simulation techniques such as the exact *Gillespie direct method* (GDM) or approximations like the *tau-leaping* method are applied to study these models [8, 9]. However, accurate stochastic simulation is a computationally expensive task; the computation time for the GDM, for example, scales with the number of possible reactions yet the performance improvements gained using approximations can introduce approximation errors. Therefore, the development of efficient and accurate stochastic simulation algorithms is an area of active research [10–16].

In order to make quantitative predictions of real biochemical systems or to perform model validation, unknown reaction rate parameters must be determined through inference. The Bayesian approach to estimate an unknown parameter vector, $\boldsymbol{\theta}$, given some

observational data, $\mathcal{D}$, is based on Bayes' Theorem,

$$p\left(\boldsymbol{\theta} \,|\, \mathcal{D}\right) = \frac{p\left(\mathcal{D} \,|\, \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right)}{p\left(\mathcal{D}\right)}, \tag{1}$$

where $p\left(\boldsymbol{\theta}\right)$ is the *a priori* PDF of the unknown parameter, $p\left(\mathcal{D} \,|\, \boldsymbol{\theta}\right)$ is the likelihood of making observations $\mathcal{D}$ under the assumption of a particular value for $\boldsymbol{\theta}$, $p\left(\mathcal{D}\right)$ is often referred to as the evidence and $p\left(\boldsymbol{\theta} \,|\, \mathcal{D}\right)$ is the *a posterior* PDF of $\boldsymbol{\theta}$ given the observations [17]. Informally, Equation (1) represents the process of updating current understanding based on previous experience and observational data. The classical approach to inference is to maximize the right hand side of Equation (1) to determine the mode of the posterior. However, more generally, the Bayesian approach can be used to quantify the level of uncertainty associated with such parameter estimates.

Theoretically, given perfect observational data from a biochemical reaction network, it is possible to determine a closed form expression for the likelihood term in Equation (1) and the method of maximum likelihood may be directly applied [6]. In practice, however, such a process is sampled imperfectly and is subject to measurement errors; thus requiring solution of the CME to form the likelihood term. However, as we have noted, exact closed form solutions of the CME are rarely available for practical applications.

Approximate Bayesian computation (ABC) refers to a family of computational methods for performing inference for problems with intractable likelihoods [17,18]. As a result, ABC methods are routinely applied to practical inference problems [19–24]. The fundamental concept is to approximate the posterior PDF, $p\left(\boldsymbol{\theta} \,|\, \mathcal{D}\right)$, by $p\left(\boldsymbol{\theta} \,|\, \rho(\mathcal{D}_s, \mathcal{D}) < \epsilon\right)$, where $\mathcal{D}_s$ is simulated data, $\rho$ is a suitable distance function and $\epsilon$ is the acceptance threshold. If a simulation process exists for the prior distribution, $p\left(\boldsymbol{\theta}\right)$, and the underlying model of interest can be simulated to estimate $p\left(\mathcal{D}_s \,|\, \boldsymbol{\theta}\right)$, then the approximated posterior can be simulated using an ABC approach.

The computational overhead for ABC inference is significantly greater than that of stochastic simulation alone. This is because the computation time is inversely proportional to the probability of $\rho(\mathcal{D}_s, \mathcal{D}) < \epsilon$ over all possible parameter values $\boldsymbol{\theta}$; that is, many stochastic simulations are required for every sample computed from the posterior [17]. A considerable computational problem arises from this, because the acceptance rate decreases exponentially as the number of unknown parameters increases [25]. This

problem, referred to as the *curse of dimensionality*, is mitigated to some degree for certain classes of problems through the use of more advanced ABC techniques [26–29]. However, in general, the curse of dimensionality is an unresolved problem [17].

In this study, we propose, implement and analyze a method for accelerating ABC inference using a *multilevel Monte Carlo* (MLMC) approach [30]. MLMC is a framework for constructing computationally efficient and accurate estimators of system statistics for stochastic processes. MLMC was developed by Giles et al. [31, 32] as a stochastic variant of multigrid methods used for obtaining numerical solutions to differential equations. Since then many other applications have benefitted from MLMC including Markov process simulation [14, 33, 34], uncertainty quantification [35] and univariate probability distribution approximation [36, 37]. To the best of our knowledge, our work represents the first application of MLMC to full Bayesian inference with intractable likelihoods.

To summarize our approach, we construct an approximation to the posterior *cumulative distribution function* (CDF) using a MLMC estimator that is constructed from a sequence of ABC approximate posteriors. While we focus on stochastic biochemical reaction networks models, our inference method is applicable for any problem that could traditionally utilize ABC inference. We demonstrate that our method is guaranteed to out perform standard ABC methods under a few reasonable assumptions.

# Methods

In this section, we describe a commonly used stochastic approach to modeling biochemical reaction networks along with standard algorithms for both simulation and parameter inference. We also describe the fundamental concepts of MLMC. Finally, we present our multilevel approach to ABC inference and derive the asymptotic performance improvement of the method.

## Discrete-state, continuous-time Markov processes

Consider a biochemical reaction network involving $N$ chemical species with copy numbers $X_1, X_2, \ldots, X_N$ that react via a network of $M$ chemical reactions of the form

$$\sum_{i=1}^{N} \nu_{i,j}^{-} X_i \xrightarrow{k_j} \sum_{i=1}^{N} \nu_{i,j}^{+} X_i, \quad j = 1, 2, \ldots, M, \tag{2}$$

where $k_j$ is the kinetic rate constant of reaction $j$, and $\nu_{i,j}^{-}$ and $\nu_{i,j}^{+}$ are, respectively, the reactant and product stoichiometries for the species $X_i$ in reaction $j$. Under the assumption that the molecules are well-mixed, the probability that the $j$th reaction occurs in the time interval $(t, t+\Delta t]$ is given by $a_j(\mathbf{X}(t); k_j)\Delta t$ where $a_j$ is the propensity function of the reaction, and is given by

$$a_j(\mathbf{X}(t); k_j) = k_j \prod_{i=1}^{N} \nu_{i,j}^{-}! \binom{X_i(t)}{\nu_{i,j}^{-}}, \tag{3}$$

where $\mathbf{X}(t) = [X_1(t), X_2(t), \ldots, X_N(t)]^{\mathrm{T}}$ is the column vector of copy numbers representing the system state at time $t$.

We can model the reaction network defined by Equation (2) as a discrete-state, continuous-time (DSCT) Markov process in which each reaction channel is governed by a time-varying Poisson process with rate parameter $\lambda(t) = \int_0^t a_j(\mathbf{X}(s); k_j)\, \mathrm{d}s$. Such a DSCT Markov process can be represented according to the Kurtz representation [38] as

$$\mathbf{X}(t) = \mathbf{X}(0) + \sum_{j=1}^{M} Y_j \left( \int_0^t a_j\left(\mathbf{X}(s); k_j\right) \mathrm{d}s \right) \boldsymbol{\nu}_j, \tag{4}$$

where the $Y_j(\lambda)$ are unit time Poisson processes with rate parameter $\lambda$ and $\boldsymbol{\nu}_j$ is the state transition that results when reaction $j$ takes place, that is $\boldsymbol{\nu}_j = [\nu_{1,j}^{+} - \nu_{1,j}^{-}, \nu_{2,j}^{+} - \nu_{2,j}^{-}, \ldots, \nu_{N,j}^{+} - \nu_{N,j}^{-}]^{\mathrm{T}}$.

Let $p\left(\mathbf{x}, t \,|\, \mathbf{y}, s\right)$ denote the *transitional density function* of the DSCT Markov process given in Equation (4), that is, the probability $\mathbf{X}(t) = \mathbf{x}$ given $\mathbf{X}(s) = \mathbf{y}$ where $t > s$. Given an initial condition, $\mathbf{X}(0) = \mathbf{x}_0$, the evolution of $p\left(\mathbf{x}, t \,|\, \mathbf{x}_0, 0\right)$ is governed by the CME,

$$\frac{\mathrm{d}p\left(\mathbf{x}, t \,|\, \mathbf{x}_0, 0\right)}{\mathrm{d}t} = \sum_{j=1}^{M} a_j(\mathbf{x} - \boldsymbol{\nu}_j) p\left(\mathbf{x} - \boldsymbol{\nu}_j, t \,|\, \mathbf{x}_0, 0\right) - a_j(\mathbf{x}) p\left(\mathbf{x}, t \,|\, \mathbf{x}_0, 0\right). \tag{5}$$

It should be noted that Equation (5) is actually a system of ODEs that is potentially countably infinite since $\mathbf{x} \in \mathbb{N}^N$. With the exception of reaction networks involving only zeroth and first order reactions the CME has no closed form solution, and it is generally only computationally tractable when the number of possible states is small [7, 39].

## The chemical master equation and Bayesian inference

The solution of Equation (5) is of critical importance to the Bayesian approach of parameter inference for DSCT Markov processes [6]. Given a realization of Equation (4), $\mathbf{X}_d(t)$, observed at $N_t$ discrete points in time, $t_1 < t_2 < \cdots < t_{N_t}$, the inference problem is to determine the posterior PDF, $p\left(\boldsymbol{\theta} \,|\, \mathbf{X}_d(t_1), \mathbf{X}_d(t_2), \ldots, \mathbf{X}_d(t_{N_t})\right)$, for the kinetic rate parameters, $\boldsymbol{\theta} = [k_1, k_2, \ldots, k_M]$. If we denote $p\left(\boldsymbol{\theta}\right)$ as the prior PDF, that represents some prior knowledge about $\boldsymbol{\theta}$, then the posterior PDF is given through application of Bayes' Theorem (Equation (1)),

$$p\left(\boldsymbol{\theta} \,|\, \mathbf{X}_d(t_1), \mathbf{X}_d(t_2), \ldots, \mathbf{X}_d(t_{N_t})\right) \propto p\left(\mathbf{X}_d(t_1), \mathbf{X}_d(t_2), \ldots, \mathbf{X}_d(t_{N_t}) \,|\, \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right). \tag{6}$$

The likelihood, $p\left(\mathbf{X}_d(t_1), \mathbf{X}_d(t_2), \ldots, \mathbf{X}_d(t_{N_t}) \,|\, \boldsymbol{\theta}\right)$, can be expressed in terms of the transitional density function $p\left(\mathbf{x}, t \,|\, \mathbf{y}, s; \boldsymbol{\theta}\right)$; that is, the solution to the CME parameterized by the kinetic rate parameter vector $\boldsymbol{\theta}$. The likelihood term becomes,

$$p\left(\mathbf{X}_d(t_1), \mathbf{X}_d(t_2), \ldots, \mathbf{X}_d(t_{N_t}) \,|\, \boldsymbol{\theta}\right) = \prod_{i=1}^{N_t} p\left(\mathbf{X}_d(t_i), t_i \,|\, \mathbf{X}_d(t_{i-1}), t_{i-1}; \boldsymbol{\theta}\right), \tag{7}$$

where $t_0 = 0$ and $\mathbf{X}_d(0) = \mathbf{x}_0$.

We now introduce two models with convenient closed form solutions to the CME for theoretical purposes. Our aim in studying these two relatively simple models with closed form solutions is to illustrate the mathematical rigor of our analysis. Once we have established this, we will then apply our method to more practical examples where the CME is intractable.

### Example 1: Degradation

The simplest chemical reaction model one could conceive is the stochastic form of exponential decay; that is, the degradation model. This model has a single reaction,

$$X \xrightarrow{k} Y, \quad X(0) = x_0, \tag{8}$$

where $Y$ represents any chemical species that is not of interest, $k$ is the kinetic rate constant for the reaction, and $x_0$ is the initial condition. Figure 1(a) shows some typical realizations of this model generated using the GDM; note that $X$ can never increase in time.

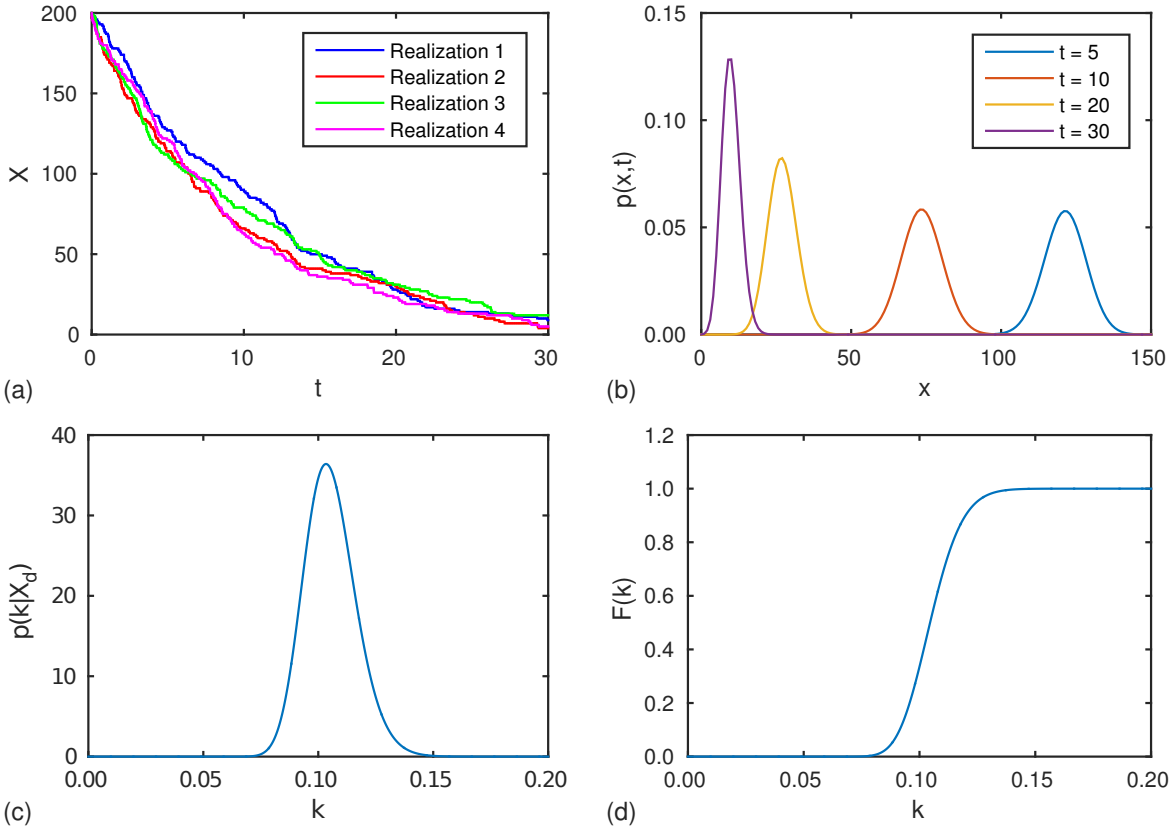


**Figure 1. Degradation model.** (a) Example realizations with $k = 0.1$ ($\text{sec}^{-1}$) and $x_0 = 200$. (b) Evolution of the CME solution, $p\,(x, t \,|\, x_0, 0; k)$. The posterior (c) PDF and (d) CDF, for $X(0) = 200$ and $X(30) = 9$.

Let $p\,(x, t \,|\, x_0, 0; k)$ be the transitional density function of the DSCT Markov process for Equation (8); that is, the solution to the CME parameterized by the kinetic rate parameter, $k$. The closed form solution to the CME can be obtained by noting that $p\,(x, t \,|\, x_0, 0; k) = 0$ for any $x > x_0$, hence the CME can be truncated into a finite system

of ODEs that may be solved through induction. The solution is given by [40]

$$p\left(x,t\,|\,x_0,0;k\right)=\binom{x_0}{x}(1-e^{-kt})^{x_0-x}e^{-kxt}. \tag{9}$$

The evolution of $p\left(x,t\,|\,x_0,0;k\right)$ is shown in Figure 1(b).

For the purposes of inference of $k$, if it is given that $X(t)=x$, we can view Equation (9) as the likelihood term in Equation (1) when the number of observation times is $N_t=1$. This enables the direct evaluation of the posterior PDF and *cumulative distribution function* (CDF) for the degradation model. Examples are given in Figure 1(c)-(d).

### Example 2: Degradation/Production

A natural extension to the degradation model (Equation (8)) is to incorporate a production reaction,

$$X\xrightarrow{k_1}Y,\quad Z\xrightarrow{k_2}X+Z,\quad X(0)=x_0,\quad Z(0)=1, \tag{10}$$

where $k_1$ and $k_2$ are, respectively, the degradation and production kinetic rate constants and $x_0$ is the initial condition. Again, $Y$ denotes any chemical species that is not of interest, and $Z$ represents a chemical species or process that produces $X$. We note that the copy number of $Z$ is constant, thus it is not required to be included in the CME. The example realizations that are shown in Figure 2(a) demonstrate the fact that $X$ is not strictly decreasing in time, thus the CME cannot be truncated into a finite system of ODEs without approximation.

In this case, we denote the transitional density function as $p\left(x,t\,|\,x_0,0;k_1,k_2\right)$. Despite the countably infinite nature of the CME in this case, it can also be solved analytically [41] to give

$$p\left(x,t\,|\,x_0,0;k_1,k_2\right)=e^{-|b(t)|}\sum_{i=0}^{\infty}\frac{b(t)^i}{i!}\binom{x_0}{|x-i|}\left(1-e^{-k_1t}\right)^{x_0-|x-i|}e^{-k_1|x-i|t}, \tag{11}$$

where $b(t)=k_2(e^{-k_1t}-1)/k_1$. The evolution of $p\left(x,t\,|\,x_0,0;k_1,k_2\right)$ approaches a steady state by approximately $t=30$ (sec) as shown in Figure 2(b). Just as with the degradation model, the exact posterior PDF can be derived using Equations (7) and (11) for the purposes of inference of both $k_1$ and $k_2$ given $X$ at discrete points in time. Figure 2(c)-(d) shows examples of the posterior PDF and CDF.

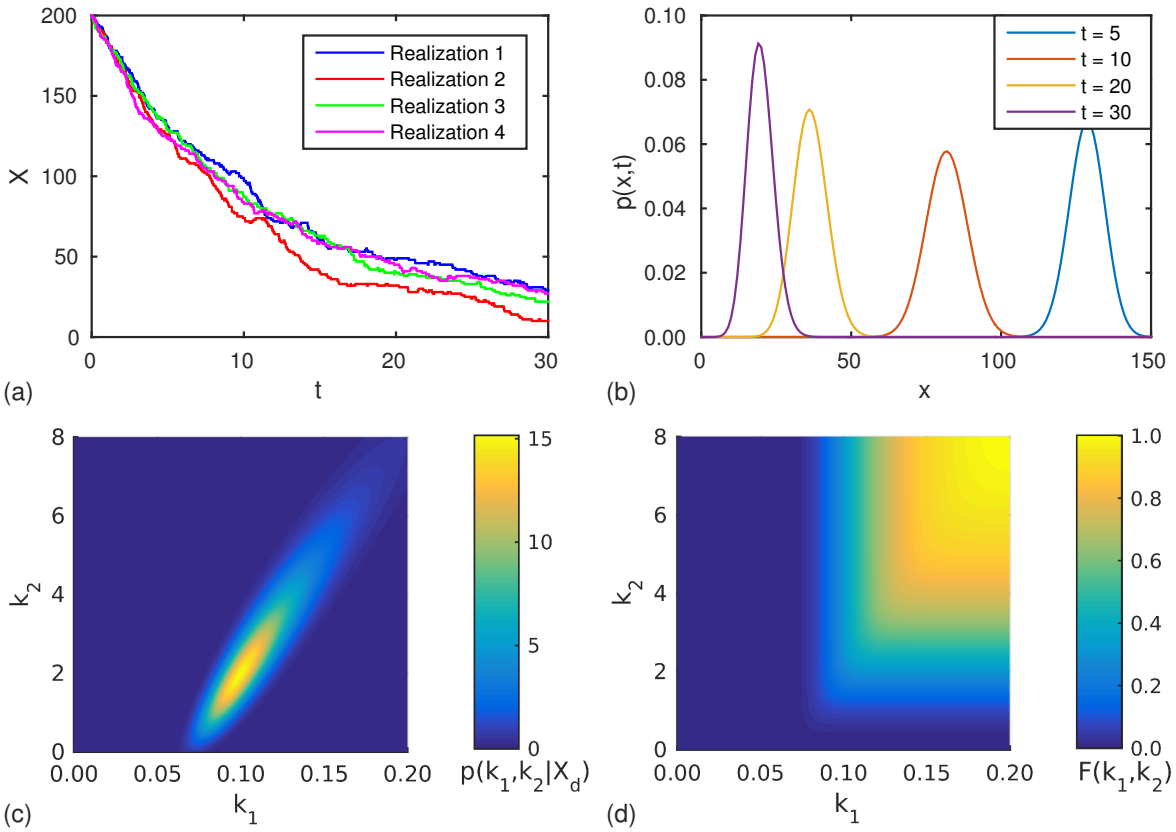


**Figure 2. Degradation/production model.** (a) Example realizations with $k_1 = 0.1$ ($\text{sec}^{-1}$), $k_2 = 1.0$ ($\text{sec}^{-1}$) and $x_0 = 200$. (b) Evolution of the CME solution, $p\left(x, t \,|\, x_0, 0; k_1, k_2\right)$. The posterior (c) PDF and (d) CDF, for $X(0) = 200$, $X(15) = 60$ and $X(30) = 29$.

## Stochastic simulation: Gillespie direct method

In general, only models dealing with zeroth and first order reactions have closed form solutions [41]. If we wish to study stochastic models of chemical reaction networks that have higher order reactions, then stochastic simulation must be utilized. The most well known *exact* stochastic simulation algorithm is the GDM [8].

The GDM arises naturally from Equation (4) by recalling that the time to the next event of a Poisson process with rate parameter $\lambda$ is exponentially distributed with parameter $\lambda$. Therefore at time $t$, if $a_0 = \sum_{j=1}^{M} a_j\left(\mathbf{X}(t)\right)$, then $\Delta t \sim \text{Exp}(a_0)$ where the next reaction occurs at $t + \Delta t$. The next reaction, $R$, to take place can be determined by sampling the probability mass function $p\left(R = j\right) = a_j / a_0$. The state vector can then be updated by adding $\boldsymbol{\nu}_j$. The result of repeating this process up to a given end time, $T$, is

the GDM, as shown in Algorithm 1.

**Algorithm 1** The Gillespie direct method

1. Initialize $\mathbf{X}$ and $t$.
2. Compute total propensity $a_0 \leftarrow \sum_{j=1}^{M} a_j(\mathbf{X})$.
3. Sample next reaction time $\Delta t \sim \text{Exp}(a_0)$.
4. If $t + \Delta t > T$, go to line 8, otherwise continue to line 5.
5. Select reaction $j$ with probability $a_j(\mathbf{X})/a_0$.
6. Update state vector, $\mathbf{X} \leftarrow \mathbf{X} + \boldsymbol{\nu}_j$.
7. Update time, $t \leftarrow t + \Delta t$.
8. Terminate and return $\mathbf{X}$.

In this study, we restrict our experimentation and discussion to the GDM for stochastic simulation to ensure the only source of approximation error is due to our inference method. As a result, we do not consider approximations such as the tau-leaping method [9].

## Parameter inference: ABC rejection

ABC methods provide a means of sampling an approximation to the posterior when the stochastic process of interest has an intractable likelihood but can be simulated [6, 17]. In the context of biochemical reaction networks, this means that repeated sampling of the model using the GDM replaces the explicit solution of the CME.

Given a data set, $\mathcal{D}$, and a prior distribution for the parameter of interest, $\boldsymbol{\theta}$, we approximate Equation (1) with

$$p\left(\boldsymbol{\theta} \,|\, \rho(\mathcal{D}_s, \mathcal{D}) < \epsilon\right) \propto p\left(\rho(\mathcal{D}_s, \mathcal{D}) < \epsilon \,|\, \boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right). \tag{12}$$

In Equation (12), $\mathcal{D}_s$ is simulated data from the model of interest, $\rho$ is a suitable distance metric and $\epsilon$ is a sufficiently small acceptance threshold, such that $p\left(\boldsymbol{\theta} \,|\, \rho(\mathcal{D}_s, \mathcal{D}) < \epsilon\right) \approx$

$p(\boldsymbol{\theta} \,|\, \mathcal{D})$. The most direct approach to sampling the posterior in Equation (12), given in Algorithm 2, is the *ABC rejection sampling* method [6, 17, 18].

**Algorithm 2** ABC rejection sampling: Generates $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta} \,|\, \rho(\mathcal{D}_s, \mathcal{D}) < \epsilon)$

1. Sample the prior $\boldsymbol{\theta}^* \sim p(\boldsymbol{\theta})$.
2. Simulate stochastic processes $\mathcal{D}_s^* \sim p(\mathcal{D}_s \,|\, \boldsymbol{\theta}^*)$.
3. If $\rho(\mathcal{D}_s^*, \mathcal{D}) < \epsilon$, accept $\boldsymbol{\theta}^*$ as a posterior sample and terminate, otherwise go to line 1.

For this study, the data set, $\mathcal{D}$, for the model of interest is assumed to be the observation of a single realization, $\mathbf{X}_d(t)$, observed at $N_t$ uniformly spaced time points $t_i = i\Delta t$ with $i = 1, 2, \ldots, N_t$. That is, $\mathcal{D}(\mathbf{X}_d) = [\mathbf{X}_d(\Delta t), \mathbf{X}_d(2\Delta t), \ldots, \mathbf{X}_d(N_t\Delta t)]$. Similarly, simulated data is given by $\mathcal{D}(\mathbf{X}_s) = [\mathbf{X}_s(\Delta t), \mathbf{X}_s(2\Delta t), \ldots, \mathbf{X}_s(N_t\Delta t)]$, where $\mathbf{X}_s(t)$ is a sample path generated by the GDM using the candidate parameter vector, $\boldsymbol{\theta}^*$, that has been sampled from the prior distribution $p(\boldsymbol{\theta})$. A natural distance metric, $\rho$, for such data is

$$\rho(\mathcal{D}(\mathbf{X}_s), \mathcal{D}(\mathbf{X}_d)) = \left( \frac{1}{N_t} \sum_{i=0}^{N_t} \frac{|\mathbf{X}_s(i\Delta t) - \mathbf{X}_d(i\Delta t)|_2^2}{|\mathbf{X}_d(i\Delta t)|_2^2} \right)^{\frac{1}{2}}, \tag{13}$$

where $|\cdot|_2$ is the $\ell_2$-norm.

## Multilevel Monte Carlo sampling

To explain the basics of MLMC, consider the task of computing $\mathbb{E}[f(X)]$ for a random variable $X$ with unknown distribution and a suitably defined functional $f$. If we have another random variable $Y$ that approximates $X$ then we have,

$$\mathbb{E}[f(X)] = \mathbb{E}[f(Y)] + \mathbb{E}[f(X) - f(Y)]. \tag{14}$$

That is, an unbiased estimator for $\mathbb{E}[f(Y)]$ will be a biased estimator for $\mathbb{E}[f(X)]$. If it is possible to simulate $X$, then we can correct for this bias by using an estimator for the bias $\mathbb{E}[f(X) - f(Y)]$ [31]. If $Y$ can be constructed in such a way that the estimator

for $\mathbb{E}[f(Y)]$ and the bias $\mathbb{E}[f(X) - f(Y)]$ can be computed more efficiently than the estimator for $\mathbb{E}[f(X)]$ then we have a net computational gain.

MLMC extends this idea to the case when there exists a sequence random variables $\{Y_\ell\}_{0 \leq \ell \leq L}$, such that as $\ell$ increases the simulation time for $\mathbb{E}[f(Y_\ell)]$ increases and the bias $\mathbb{E}[f(X) - f(Y_\ell)]$ decreases at a suitable rate [31,37]. The resulting recursive application of Equation (14) yields the telescoping sum

$$\mathbb{E}[f(X)] = \mathbb{E}[f(Y_0)] + \left[\sum_{\ell=1}^{L} \mathbb{E}[f(Y_\ell) - f(Y_{\ell-1})]\right] + \mathbb{E}[f(X) - f(Y_L)]. \tag{15}$$

Under certain conditions, constructing an estimator based on Equation (15) is significantly faster than estimating $\mathbb{E}[f(X)]$ directly [31,33,34].

## A multilevel approach to inference

In this section, we develop a new approach to ABC inference using MLMC and provide key theoretical results about the performance gain using our approach. The method, which we refer to as *multilevel approximate Bayesian computation* (MLABC), combines ABC rejection sampling using a sequence of acceptance thresholds to approximate the CDF of the posterior to within a prescribed level of accuracy defined in terms of the root mean squared error (RMSE). For brevity, we only present the key features of the analysis here; for detailed analysis of the method, based on the work of Giles et al. [37], see Appendices S1–S3.

### Derivation

We present, for the sake of simplicity, our MLABC method in terms of the degradation/production model (Equation (10)). However, we note that applying the ideas to different models with different numbers of unknown parameters is straightforward extension of the degradation/production method. Given observed trajectory data, $\mathcal{D}(\mathbf{X}_d)$, the task is to approximate, at a point $(s_1, s_2) \in \mathbb{R}^2$, the posterior CDF given by

$$F(s_1, s_2) = \int_0^{s_1} \int_0^{s_2} p(k_1, k_2 \,|\, \mathcal{D}(\mathbf{X}_d))\, \mathrm{d}k_2 \mathrm{d}k_1. \tag{16}$$

We can reformulate Equation (16) as the expectation,

$$F(s_1, s_2) = \mathbb{E}\left[\mathbb{1}_{(0,s_1]\times(0,s_2]}(k_1, k_2)\right], \tag{17}$$

where $\mathbb{1}_{(0,s_1]\times(0,s_2]}$ is the indicator functional,

$$\mathbb{1}_{(0,s_1]\times(0,s_2]} = \begin{cases} 1 & (k_1, k_2) \in (0, s_1] \times (0, s_2] \\ 0 & \text{otherwise.} \end{cases}$$

Assuming distance metric $\rho$, as defined in Equation (13), along with a suitably chosen acceptance threshold, $\epsilon$, the standard Monte Carlo estimator is

$$F(s_1, s_2) \approx \hat{F}(s_1, s_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(0,s_1]\times(0,s_2]}\left((k_1, k_2)_{\epsilon}^{(i)}\right), \tag{18}$$

where the number of samples, $n$, is sufficiently large and $(k_1, k_2)_{\epsilon}^{(i)}$ is the $i$th accepted sample from $p\left(k_1, k_2 \,|\, \rho(\mathcal{D}(\mathbf{X}_s), \mathcal{D}(\mathbf{X}_d)) < \epsilon\right)$ using ABC rejection.

Now, consider a geometric sequence of acceptance thresholds $\epsilon_\ell = \epsilon_0 K^{-\ell}$ for integer $K \geq 2$ and $\ell = 0, 1, 2, \ldots, L$ and denote $(k_1, k_2)_{\epsilon_\ell}$ as the random vector distributed according the approximation $p\left(k_1, k_2 \,|\, \rho(\mathcal{D}(\mathbf{X}_s), \mathcal{D}(\mathbf{X}_d)) < \epsilon_\ell\right)$. Using this sequence of posterior distribution approximations, a multilevel estimator for the posterior CDF at the point $(s_1, s_2)$ can be determined as

$$F(s_1, s_2) \approx \hat{F}(s_1, s_2) = \frac{1}{n_0}\sum_{i=1}^{n_0} \mathbb{1}_{(0,s_1]\times(0,s_2]}\left((k_1, k_2)_{\epsilon_0}^{(i)}\right) + \sum_{\ell=1}^{L} \mu_\ell^{\text{bias}}, \tag{19}$$

with

$$\mu_\ell^{\text{bias}} = \frac{1}{n_\ell}\sum_{i=1}^{n_\ell}\left[\mathbb{1}_{(0,s_1]\times(0,s_2]}\left((k_1, k_2)_{\epsilon_\ell}^{(i)}\right) - \mathbb{1}_{(0,s_1]\times(0,s_2]}\left((k_1, k_2)_{\epsilon_{\ell-1}}^{(i)}\right)\right],$$

where the sample sizes, $n_\ell$, are sufficiently large to ensure the estimator variance is below some predetermined value. In terms of bias, the multilevel estimator in Equation (19) is equivalent to the standard Monte Carlo estimator with $\epsilon = \epsilon_L$ in Equation (18). By evaluating Equation (19) at a set of points in $\mathbb{R}^2$ combined with a suitable interpolation scheme (see Appendix S3), an approximation of the entire posterior CDF can be constructed.

The estimator given in Equation (19) is the essence of the MLABC method. Although we present the method in terms of the degradation/production model with two unknown

parameters, an equivalent general estimator can be formed for any stochastic model with $M$ unknown parameters by evaluating the indicator functional over the region $(-\infty, s_1] \times (-\infty, s_2] \times \cdots \times (-\infty, s_M]$ . We now aim to analyze and demonstrate that, under certain reasonable assumptions, this approach can always provide a computational gain over standard ABC rejection sampling for a sufficiently small target bias level.

## Assumptions

There are three main assumptions required for the analysis of Equation (19) [37]. The first two assumptions are related to the convergence rates of the posterior approximation as $\ell \to \infty$. The third is a condition on the computation time for generating a sample pair $\left[(k_1, k_2)_{\epsilon_\ell}, (k_1, k_2)_{\epsilon_{\ell-1}}\right]$ which is required when computing the bias correction term $\mu_\ell^{\text{bias}}$. Such a sample pair is a sample from the joint distribution of $(k_1, k_2)_{\epsilon_\ell}$ and $(k_1, k_2)_{\epsilon_{\ell-1}}$.

More specifically, given a geometric sequence of acceptance thresholds $\epsilon_\ell = \epsilon_0 K^{-\ell}$ for integer $K \geq 2$, we assume there exist constants $\alpha > 0$, $\beta > 0$ and $\gamma > 0$ such that:

1 as $\ell \to \infty$, $\mathbb{1}_{(0,s_1]\times(0,s_2]}\left((k_1, k_2)_{\epsilon_\ell}\right)$ converges weakly to $\mathbb{1}_{(0,s_1]\times(0,s_2]}\left(k_1, k_2\right)$ with order $\alpha$. That is, $\left|\mathbb{E}\left[\mathbb{1}_{(0,s_1]\times(0,s_2]}\left((k_1, k_2)_{\epsilon_\ell}\right)\right] - \mathbb{E}\left[\mathbb{1}_{(0,s_1]\times(0,s_2]}\left(k_1, k_2\right)\right]\right| = O(\epsilon_\ell^\alpha)$;

2 as $\ell \to \infty$, $(k_1, k_2)_{\epsilon_\ell}$ converges strongly to $(k_1, k_2)$ with order $\beta$. That is, $\mathbb{E}\left[\left|(k_1, k_2)_{\epsilon_\ell} - (k_1, k_2)\right|_1\right] = O(\epsilon_\ell^\beta)$. Here, $|\cdot|_1$ is the $\ell_1$-norm;

3 the computation time for sampling the distribution of $\left[(k_1, k_2)_{\epsilon_\ell}, (k_1, k_2)_{\epsilon_{\ell-1}}\right]$ is $O(\epsilon_\ell^{-\gamma})$.

Assumptions 1 and 2 are reasonable due to the convergence properties of ABC rejection itself [18]. However, obtaining exact values or estimates for $\alpha$ and $\beta$ is generally difficult. In practice, exact values of $\alpha$ and $\beta$ is not required, though we assume these are known in order to analyze the asymptotic computational gain.

The constant $\gamma$ depends on the dimensionality of the parameter space, since the computation time is inversely proportional to the average acceptance rate of the ABC rejection method. The general proof of Assumption 3 is given in Appendix S1. For the degradation/production model we have $\gamma = 2$.

### Theoretical performance gain

In this section, we present asymptotic bounds on both the RMSE of the standard ABC inference and our MLABC method. We then express the asymptotic computation time given a target RMSE, $h$. This allows us to construct an expression for the asymptotic computational gain in terms of the convergence parameters $\alpha$ and $\beta$. We do this in the context of a two parameter inference problem, such as the degradation/production model (Equation (10)), leaving more general analysis to Appendix S2.

Assume that the posterior PDF, $p\left(k_1, k_2 \,|\, \mathbf{X}_d\right)$, has compact support in the region $R = [s_1^{min}, s_1^{max}] \times [s_2^{min}, s_2^{max}]$. Such an assumption is reasonable for a sufficiently large number of observation times $N_t$; for the degradation/production model, this assumption is valid for $N_t \geq 2$. Now apply a regular discretization to $R$ using $D$ divisions in each dimension, resulting in $(D+1)^2$ grid points $\{(s_1^i, s_2^j)\}_{0 \leq i,j \leq D}$, where $(s_1^i, s_2^j) = ((i/D)(s_1^{max} - s_1^{min}) + s_1^{min}, (j/D)(s_2^{max} - s_2^{min}) + s_2^{min})$.

By applying estimators Equation (18) or Equation (19) we obtain a discrete approximation, $\mu_{i,j} = \hat{F}(s_1^i, s_2^j)$, to the true posterior CDF. The RMSE of this approximation is

$$\text{RMSE}(\mu) = \sqrt{\mathbb{E}\left[\max_{0 \leq i,j \leq D} \left|F(s_1^i, s_2^j) - \mu_{i,j}\right|^2\right]}. \tag{20}$$

For simplicity, we will ignore the interpolation of $\mu_{i,j}$ for a continuous approximation to $F$ over the entire region $R$. It is important to note, however, our analysis (Appendices S2 and S3) accounts for this, and is crucial to our results.

The standard Monte Carlo estimator is given by $\mu = \{\hat{F}(s_1^i, s_2^j)\}_{0 \leq i,j \leq D}$ where $\hat{F}(s_1^i, s_2^j)$ is computed using Equation (18) with $\epsilon = \epsilon_L$. If the convergence parameter $\alpha$ is known, the RMSE for the standard Monte Carlo estimator, $\mu$, is bounded, for some constant $c_1$, by

$$\text{RMSE}(\mu) \leq c_1 \sqrt{K^{-2\alpha L} + \frac{v_L}{n} \log_e (D+1)^2}, \tag{21}$$

where $v_L$ is the variance of $(k_1, k_2)_\epsilon$ and $n$ is the number of Monte Carlo samples.

Under Assumption 3, and by taking $L = (1/\alpha) \log_K (1/h)$, the cost of using standard Monte Carlo to construct an approximation, $\mu$, to the posterior CDF with $\text{RMSE}(\mu) =$

$O(h)$ is bounded, for some constant $c_2$, by

$$\text{cost}(\mu) \leq c_2 h^{-(2+\gamma/\alpha)} \log_e h^{-1}. \tag{22}$$

This bound is identified by bounding the right hand side of Equation (20) by $h$ and solving for a lower bound on the number of accepted samples, $n$.

Bounding the RMSE for the multilevel estimator, $\mu_{ML} = \{\hat{F}(s_1^i, s_2^j)\}_{0 \leq i,j \leq D}$ with $\hat{F}(s_1^i, s_2^j)$ computed using Equation (19), is a relatively straightforward application of the method for obtaining Equation (20) along with invoking Assumption 2. The equivalent result for the multilevel estimator, for some constant $c_3$, is

$$\text{RMSE}(\mu_{ML}) \leq c_3 \sqrt{K^{-2\alpha L} + \left( \frac{v_0}{n_0} + \sum_{\ell=1}^{L} \frac{M^{-\beta\ell}}{n_\ell} \right) \log_e (D+1)^2}, \tag{23}$$

where $v_0$ is the variance of $(k_1, k_2)_{\epsilon_0}$ and $n_\ell$ are the Monte Carlo sample numbers for each level.

The upper bound on the computation time for $\mu_{ML}$ depends on the choice of the number of samples for each level, $n_\ell$. We can use a Lagrange multiplier method to choose the $n_\ell$ such that the asymptotic cost is minimized under the constraint of $\text{RMSE}(\mu_{ML}) = O(h)$. The choice of $n_\ell$ is

$$n_\ell \geq \frac{\log_e (D+1)^2}{h^2} \left( 1 + \frac{K^{(\gamma-\beta)/2} - K^{(L+1)(\gamma-\beta)/2}}{1 - K^{(\gamma-\beta)/2}} \right) K^{-\ell(\gamma+\beta)}. \tag{24}$$

Using Assumption 3 and the optimal $n_\ell$ given by Equation (24), we arrive at an expression for the asymptotic bound on the computation time for evaluating $\mu_{ML}$. In the case of $\beta < \gamma$, then there exists some constant $c_4$ such that

$$\text{cost}(\mu_{ML}) \leq c_4 \left( 1 - h^{-(\gamma-\beta)/2\alpha} + h^{-(\gamma-\beta)/\alpha} \right) h^{-2} \log_e h^{-1}. \tag{25}$$

The results from Equation (22) and Equation (25) directly yield a reduction in the order of the computation time upper bound from using the multilevel method. We denote the reduction as the asymptotic computational gain, given by

$$\text{AG}(\mu, \mu_{ML}) = \frac{h^{-(2+\gamma/\alpha)} \log_e h^{-1}}{h^{-(2+(\gamma-\beta)/\alpha)} \log_e h^{-1}} = h^{-\beta/\alpha}. \tag{26}$$

Note that Equation (26) indicates that, assuming the upper bounds are achieved asymptotically, it is always possible to choose a target RMSE, $h$, such that the multilevel

approach achieves some desired level of computational gain. The rate of growth of this gain as a function of $h$ depends on the convergence characteristics of the sequence of ABC posterior approximations, however it is always an improvement since $\alpha > 0$ and $\beta > 0$.

## Practical application of MLABC

In practice, the convergence parameters $\alpha$ and $\beta$ are unknown. In this section, we present a practical implementation of an algorithm for MLABC that does not depend on explicit knowledge of $\alpha$ and $\beta$. We find that our algorithm can obtain up to an order of magnitude performance improvement while still maintaining a desired level of accuracy.

### Removing dependence on convergence parameters

First, note that when ABC methods such as ABC rejection are used in practice, there are certain assumptions made about the weak convergence rate, $\alpha$. If we expect $n$ accepted samples using acceptance threshold $\epsilon$ to provide an acceptable estimator of the real posterior CDF, then we are implicitly assuming $\alpha \geq \left(\log_e \sqrt{(\log_e(D+1)^2)/n}\right) / (\log_e \epsilon)$. Therefore, we can determine, for a given scale factor, $K$, and base level threshold, $\epsilon_0$, the number of levels, $L$, required to match the final bias of the standard ABC rejection method. As a result, we can treat $\alpha$ as a known parameter in the sense of equivalence to the standard ABC rejection estimator.

Second, consider the role of $\beta$ in determining the $n_\ell$ in Equation (24). While the details are given in Appendix S2, we informally state that the summation in Equation (24) can be derived by showing that Assumption 2 implies, for some constant $c_5$,

$$\mathrm{Var}\left[\mathbb{1}_{(0,s_1]\times(0,s_2]}((k_1,k_2)_{\epsilon_\ell}) - \mathbb{1}_{(0,s_1]\times(0,s_2]}((k_1,k_2)_{\epsilon_{\ell-1}})\right] \leq c_5 K^{-\beta\ell}. \tag{27}$$

That is, $K^{-\beta\ell}$ is used to bound the variances of the bias correction terms in Equation (19). It is important to note that the rigorous version of Equation (27) requires a smoothing process to be applied to the indicator functional to ensure the Lipschitz continuity conditions discussed in the supporting information (Appendix S2 and S3).

If we knew exactly, for each level $\ell$, the computation time to generate a posterior

**Algorithm 3** Sampling method for variance reduction of bias correction term $\mu_\ell^{\text{bias}}$

1. Let $n_d$ denote the number of samples accepted at levels $\ell$ and $\ell-1$.
2. Initialize $n_d \leftarrow 0$.
3. Set sample counter $j \leftarrow 1$.
   - 3.1 Sample the prior $(k_1, k_2)^* \sim p\,(k_1, k_2)$.
   - 3.2 Generate $\mathbf{X}_s(t)$ using the GDM with kinetic rates $(k_1, k_2)^*$.
   - 3.3 If $\rho(\mathcal{D}(\mathbf{X}_s), \mathcal{D}(\mathbf{X}_d)) < \epsilon_\ell$, accept $(k_1, k_2)^*$ as both a level $\ell$ and a level $\ell-1$ sample, increment $n_d \leftarrow n_d + 1$ and go to line 3.5.
   - 3.4 If $\rho(\mathcal{D}(\mathbf{X}_s), \mathcal{D}(\mathbf{X}_d)) < \epsilon_{\ell-1}$, accept $(k_1, k_2)^*$ as a level $\ell-1$ sample only.
   - 3.5 If $j < n_\ell$, increment $j \leftarrow j+1$ and go to line 3.1, otherwise go to line 4.
4. Reset sample counter $j \leftarrow 1$.
   - 4.1 Sample the prior $(k_1, k_2)^* \sim p\,(k_1, k_2)$.
   - 4.2 Generate $\mathbf{X}_s(t)$ using the GDM with kinetic rates $(k_1, k_2)^*$.
   - 4.3 If $\rho(\mathcal{D}(\mathbf{X}_s), \mathcal{D}(\mathbf{X}_d)) < \epsilon_\ell$, accept $(k_1, k_2)^*$ as a level $\ell$.
   - 4.4 If $j + n_d < n_\ell$, increment $j \leftarrow j+1$ and go to line 4.1, otherwise go to line 5.
5. Terminate and return all accepted sample pairs.

realization, $c_\ell$, and the variance, $v_\ell$, of each bias correction term, $\mu_\ell^{bias}$, then we could calculate the optimal $n_\ell$ using the same Lagrange multiplier approach as for Equation (24). The result would be,

$$n_\ell = \frac{\sqrt{v_\ell}}{h^2\sqrt{c_\ell}} \sum_{m=0}^{L} \sqrt{v_m c_m}. \tag{28}$$

In practice we can can only estimate $c_\ell$ and $v_\ell$. Using the approach used by Anderson et al. [33] and Lester et al. [34] we generate a relatively small number of trial samples at each level to obtain estimates for $c_\ell$ and $v_\ell$ that work well in practice. Using this approach we do not have the same theoretical bounds on the RMSE. However, accurate approximations for the variances will result in estimators with a RMSE that is close to

the target in practice.

**Algorithm 4** MLABC (Multilevel approximate Bayesian computation): Estimates the posterior CDF $F(s_1, s_2)$

1. Let $n$, $\epsilon$ and $D$ be parameters from equivalent ABC estimator.
2. Calculate number of levels, $L$, using $L \leftarrow \log_K(\sqrt{D+1})/\log_\epsilon \sqrt{(\log_e(D+1)^2/n)}$.
3. Calculate $n_0, n_1, \ldots, n_L$ using Equation (28) with $c_\ell$ and $v_\ell$ estimated using 100 samples on each level.
4. Generate $n_0$ samples $\{(k_1, k_2)_{\epsilon_0}^{(j)}\}_{j=1}^{j=n_0}$ from $p\left(k_1, k_2 \,|\, \rho(\mathcal{D}(\mathbf{X}_s), \mathcal{D}(\mathbf{X}_d)) < \epsilon_0\right)$ using ABC rejection.
5. Generate $n_\ell$ correlated sample pairs $\{((k_1, k_2)_{\epsilon_\ell}^{(j)}, (k_1, k_2)_{\epsilon_{\ell-1}}^{(j)})\}_{j=1}^{j=n_\ell}$ for $\ell = 1, 2, \ldots, L$ using Algorithm 3.
6. Repeat steps 6.1-6.4 for every point $(s_1, s_2) \in G$ where, $G$ is a grid of $(D+1)^2$ points over the support of the posterior.
   - 6.1 Initialize base level estimator, $E_0 \leftarrow 0$, and bias level estimators $B_\ell \leftarrow 0$.
   - 6.2 Compute base level estimator, $E_0 \leftarrow \frac{1}{n_0} \sum_{j=1}^{n_0} \mathbb{1}_{[0,s_1]\times[0,s_2]}\left((k_1, k_2)_{\epsilon_0}^{(j)}\right)$.
   - 6.3 Compute bias correction estimators for every $\ell = 1, 2, \ldots, L$, $B_\ell \leftarrow \frac{1}{n_\ell} \sum_{j=1}^{n_\ell} \mathbb{1}_{[0,s_1]\times[0,s_2]}\left((k_1, k_2)_{\epsilon_\ell}^{(j)}\right) - \mathbb{1}_{[0,s_1]\times[0,s_2]}\left((k_1, k_2)_{\epsilon_{\ell-1}}^{(j)}\right)$.
   - 6.4 Compute CDF point estimate $\hat{F}(s_1, s_2) \leftarrow E_0 + \sum_{\ell=1}^{L} B_\ell$.

### Improving performance with variance reduction

In Equation (28), note that $n_\ell \propto \sqrt{v_\ell}$. Furthermore, note that for the bias correction terms in Equation (19), our only concern is the expected difference between indicator function values at each level rather than the expected values themselves. As a result, if we can introduce some correlation in the generation of accepted samples at each level, then the variance of the estimators and hence the number of samples required will decrease. In this work, we implement a simple method to ensure such correlation. Given a simulated

trajectory, $\mathbf{X}_s(t)$, then $\rho(\mathcal{D}(\mathbf{X}_s), \mathcal{D}(\mathbf{X}_d)) < \epsilon_\ell$ implies $\rho(\mathcal{D}(\mathbf{X}_s), \mathcal{D}(\mathbf{X}_d)) < \epsilon_{\ell-1}$, since $\epsilon_\ell < \epsilon_{\ell-1}$. As a result, we can sample level $\ell - 1$ first and keep track of any samples that can also be accepted at level $\ell$; such samples can be validly taken as samples from both levels. A pair constructed this way will make no contribution to the bias correction term, hence reducing the variance. We arrive at the method presented in Algorithm 3 for reducing the variance when sampling for the bias correction term $\mu_\ell^{\text{bias}}$.

### The multilevel ABC algorithm

By combining Equation (28) with Algorithm 3, we construct a practical implementation of the MLABC estimator (Equation (19)). This MLABC method, presented in Algorithm 4 for the degradation/production model, is implemented as a prototype using the C programming language. The prototype (Code S3) is specific to biochemical reaction network models, however, only minor changes are required to change the target application.

## Results

In this section, we examine the accuracy and performance of MLABC using the previously presented degradation and degradation/production models because we can directly evaluate the estimator RMSE using the exact solution to the CME. We then compare MLABC with standard ABC rejection using two more complex biochemical reaction networks.

### Estimating convergence parameters

The asymptotic computational gain using MLABC is presented in Equation (26). To validate this theoretical result, we need to estimate $\alpha$ and $\beta$. We use the ABC rejection method (Algorithm 2) and linear regression to estimate $\alpha$ and $\beta$ for the degradation model and the degradation/production model using data observed at $N_t$ discrete points in time for $N_t = 2, 3, \ldots, 10$. For the degradation model and degradation/production model, respectively, data is generated using $k = 0.1$ and $(k_1, k_2) = (0.1, 1)$. A univariate uniform prior distribution with support $\{k : k \in [0, 1]\}$ is utilized for the degradation

model and a bivariate uniform prior with support $\{(k_1, k_2) : (k_1, k_2) \in [0, 1] \times [0, 10]\}$ is utilized for the degradation/production model. Estimates are produced using $1,000$ accepted samples for threshold levels $\epsilon_\ell = \epsilon_0 K^{-\ell}$, $\ell = 1, 2, \ldots, L$ with $K = 2$, $\epsilon_0 = 1$ and $L = 5$. These estimates are shown in Figure 3(a)-(b).

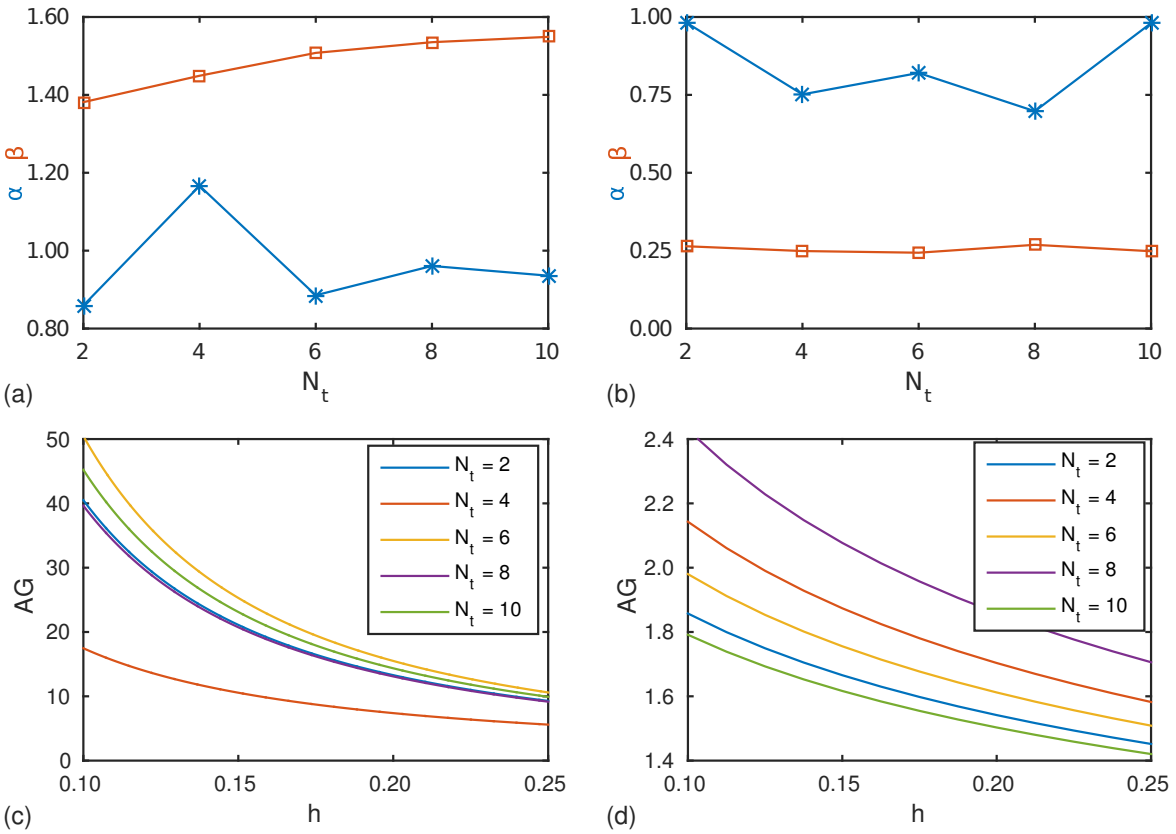


**Figure 3. Estimated convergence rates.** Least squares approximation of $\alpha$ and $\beta$ for the (a) degradation model with $k = 0.1$ and (b) degradation/production model with $(k_1, k_2) = (0.1, 1)$ using observation times $N_t \in [2, \ldots, 10]$. Asymptotic computational gain functions $\text{AG} = O(h^{-\beta/\alpha})$ for the (c) degradation and (d) degradation/production models.

Using these estimates of $\alpha$ and $\beta$ it is possible to predict the asymptotic growth of computational gain by using MLABC based on sample numbers chosen according to Equation (24). These computational gains are given in Figure 3(c)-(d) for $N_t = 2, 3, \ldots, 10$. By using Equation (24), we expect minimal increase in the RMSE compared with standard ABC rejection.

## Performance using estimated convergence parameters

We now compare numerical simulation results against the theoretical performance and error analysis. Throughout we refer to the measured computational gain of our MLABC approach that is given by

$$\mathrm{MG}(\mu, \mu_{ML}) = \frac{C(\mu)}{C(\mu_{ML})}, \tag{29}$$

where $\mu$ and $\mu_{ML}$ are the standard ABC rejection and MLABC estimators, respectively, and $C(\cdot)$ is the averaged measured computation time to evaluate the estimator.

For the degradation model, 20 independent MLABC and ABC rejection CDF estimators are computed for target RMSE $h \in [0.1, 0.25]$. The CDF is approximated over the interval $[0, 0.5]$ on a grid of $1,000$ nodes using data with $N_t = 8$. The sample numbers, $n_\ell$, are computed according to Equation (24) using the estimated convergence parameter values of $\alpha \approx 0.96$ and $\beta \approx 1.54$. As shown in Figure 4(a), the increase in computational gain (Equation (29)) is the same order of magnitude as the theoretical asymptotic prediction (Equation (26)), albeit at a smaller absolute scale. In Figure 4(b), we see that the target of RMSE $\leq h$ is achieved with the exception of $h = 0.1$, however, the fact that there is also an increase in RMSE at $h = 0.1$ for the standard ABC estimator indicates that this is a feature of the problem that can be probably be improved with the application of smoothing to the indicator functional (see Appendix S3).

For the degradation/production model, 20 independent MLABC and ABC rejection CDF estimators are computed for target RMSE $h \in [0.1, 0.25]$. The CDF is approximated over the region $[0, 1] \times [0, 10]$ on a grid of $100 \times 100$ nodes using data with $N_t = 4$. The sample numbers, $n_\ell$, are computed according to Equation (24) using the estimated convergence parameter values of $\alpha \approx 0.75$ and $\beta \approx 0.25$. Figure 4(c)-(d) provides, for the production/degradation model, the computational gain using MLABC over ABC rejection and the RMSE versus the target RMSE. Compared to the degradation model, the computational gain is significantly less; however, it is consistent with the theoretical results. Similarly, Figure 4(d) shows practically no increase in the RMSE of the CDF estimator.

An important remark must be made about these results. They do not represent

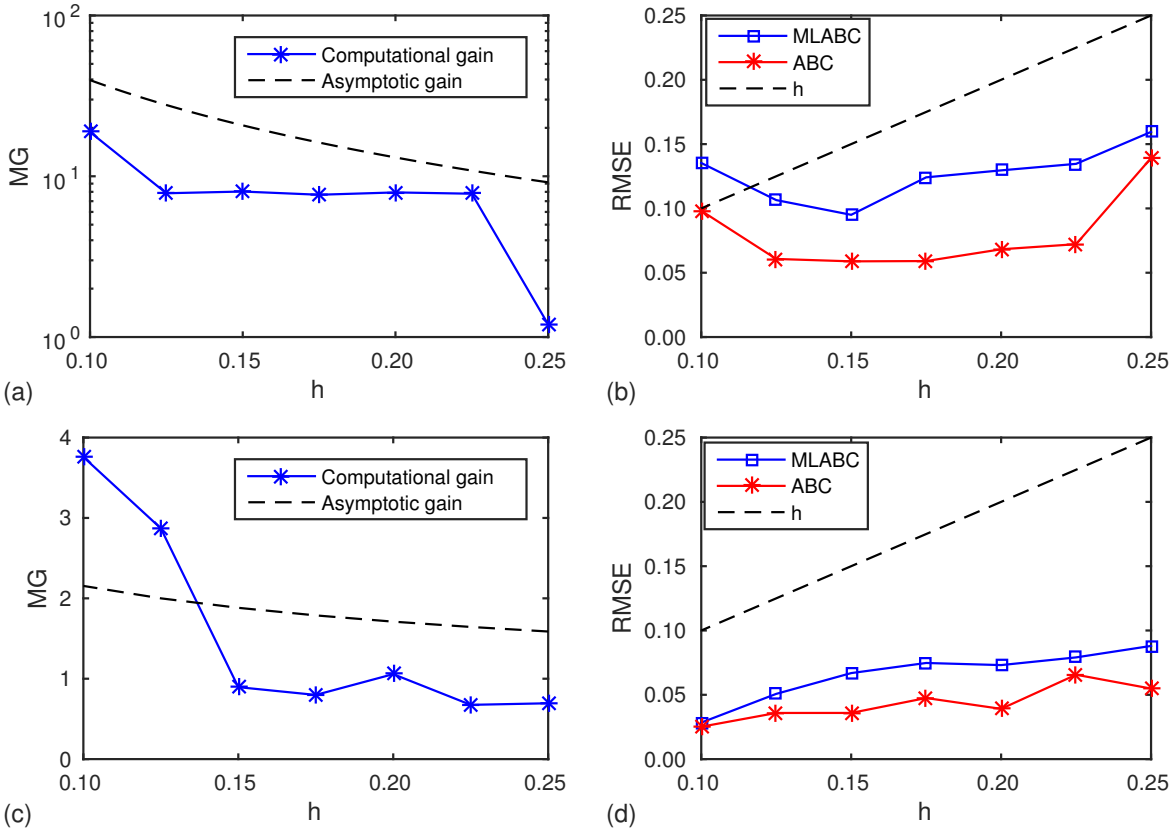


**Figure 4. Measured performance gain and error.** (a) Computational gain and (b) RMSE for the degradation model results using data with $N_t = 8$ observation times and convergence rate estimates $\alpha \approx 0.96$ and $\beta \approx 1.54$. (c) Computational gain and (d) RMSE for the degradation/production model results using data with $N_t = 4$ observation times and convergence rate estimates $\alpha \approx 0.75$ and $\beta \approx 0.25$.

the best computational gain that can be achieved in practice, but rather they provide experimental validation for our theory. If the values of $\alpha$ and $\beta$ are known, then it is possible to predict the asymptotic computational gain available whilst maintaining control on the RMSE of the estimator. Furthermore, we note that, as predicted by the theory, the computational gain grows proportionally to a power of $h^{-1}$.

## Performance using empirical sample numbers

We now look to the more practical approach to MLABC, as presented in Algorithm 4. Here we specifically focus on the degradation/production model. In the first experiment, 20 independent MLABC (using Algorithm 4) and ABC rejection CDF estimators of the

degradation/production model are computed for target RMSE $h \in [0.1, 0.25]$. The CDF is approximated over the support region $\{(k_1, k_2) : (k_1, k_2) \in [0, 1] \times [0, 10]\}$ on a grid of $100 \times 100$ nodes using data with $N_t = 4$.

Results in Figure 5 are analogous to the results in Figure 4(c)-(d). By using Algorithm 4, we have achieved greater computational gain, whilst maintaining reasonable control over the RMSE (i.e., still within target $h$). The main advantage is that values for $\alpha$ and $\beta$ have not been required. While the performance results in Figure 5(a) are an improvement over those in Figure 4(c), the new results still show the same order of magnitude increase in performance. However, we now demonstrate that Algorithm 4 can out perform the asymptotic results by an order of magnitude.

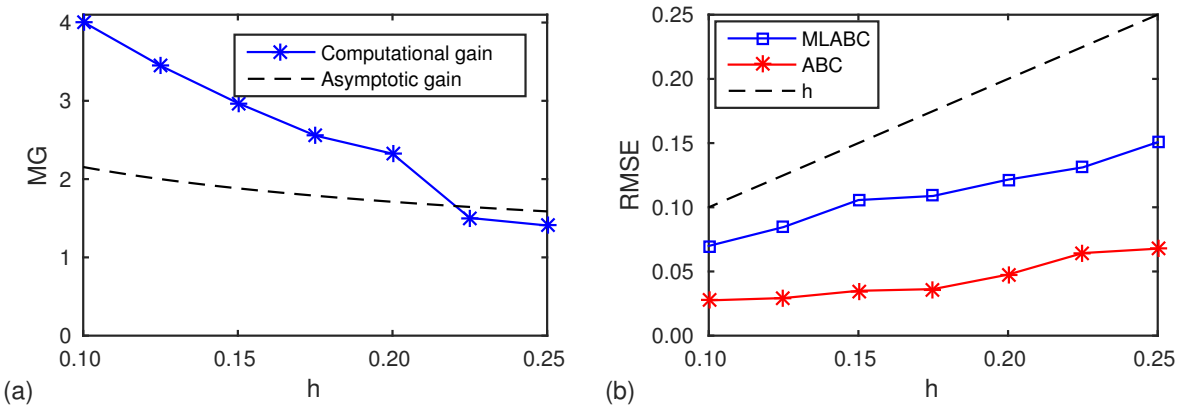


**Figure 5. Measured performance gain and error for the degradation/production model.** Using data with $N_t = 4$ observation times and 100 samples at each level to select sample numbers $n_\ell$. (a) Measured computational gain and (b) RMSE as a functions of $h$.

In the second experiment, the target RMSE is kept constant at $h = 0.2$. MLABC and ABC rejection estimators for the posterior CDF are computed using data with $N_t = 2, 4, 6, \ldots, 20$. For each value of $N_t$, 20 simulations are executed, with computation times being the average. Figure 6(a) demonstrates a significant improvement in computation time for MLABC over ABC rejection in this case. Note that the computational gain shown in Figure 6(b) is an order of magnitude greater than the asymptotic analysis shown in Figure 3(d) predicts for $N_t = 10$.

We now compare the quality of the posteriors computed by MLABC and ABC rejection. For this we consider the posterior mean and the marginal distribution 90% credible

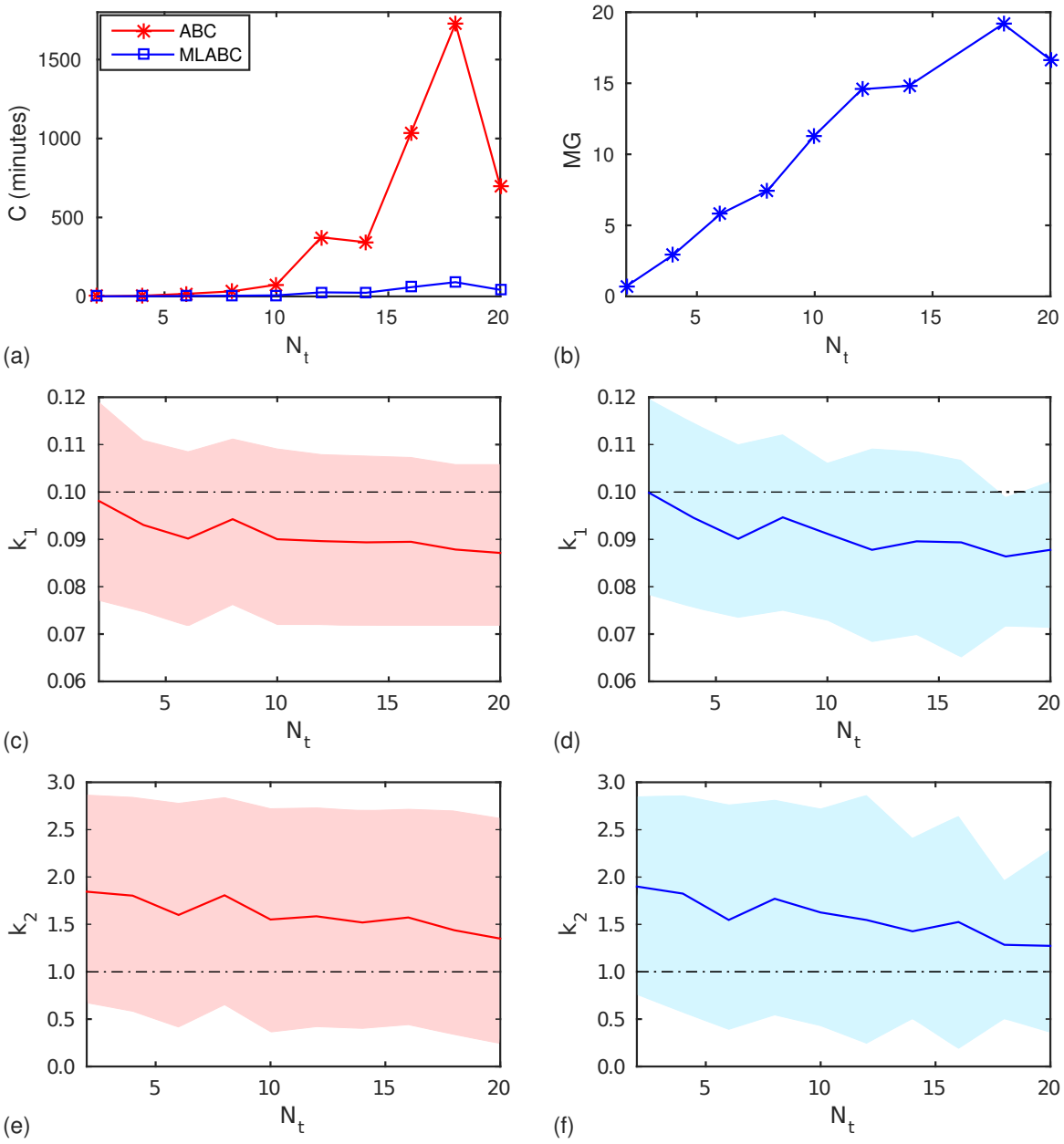


**Figure 6. Comparison of ABC and MLABC: degradation/production.**

Performance of ABC and MLABC for the degradation/production model as $N_t$ increases. (a) Computation time. (b) Computational gain. (c) and (e) ABC parameter estimates. (d) and (f) MLABC parameter estimates. The true parameter values, $(k_1, k_2) = (0.1, 1.0)$, are indicated with dashed lines, posterior means and 90% credible intervals are indicated with solid lines and shaded areas, respectively.

intervals. The mean, representing the central tendency of the posterior, represents a likely parameter candidate, $(k_1^m, k_2^m)$, given by

$$k_i^m = \iint_{\mathbb{R}^2} k_i p\left(k_1, k_2 \,|\, \rho(\mathcal{D}_s, \mathcal{D}) < \epsilon_L\right) \mathrm{d}k_1 \mathrm{d}k_2. \tag{30}$$

Given the joint CDF, $F(s_1, s_2)$, the marginal CDFs, $F_1(s)$ and $F_2(s)$, can be determined using

$$F_1(s) = \lim_{s_2 \to \infty} F(s_1, s_2), \quad F_2(s) = \lim_{s_1 \to \infty} F(s_1, s_2). \tag{31}$$

For some significance level, $a$, the $(1-a)100\%$ credible interval for parameter $k_i$, denoted by $[l(k_i), u(k_i)]$, is

$$l(k_i) = \sup\left\{s \in \mathbb{R} : F_i(s) < a/2\right\}, \quad u(k_i) = \inf\left\{s \in \mathbb{R} : F_i(s) > 1 - a/2\right\}. \tag{32}$$

The credible interval provides a measure of uncertainty around the parameter estimates.

Figures 6(c) and 6(e) present the estimates produced by standard ABC for different values of $N_t$. These are to be compared with the estimates produced by MLABC as shown in Figures 6(d) and 6(f). These results show that, from a practical perspective, the MLABC method is just as appropriate as ABC rejection. That is, the true parameter values lie within the credible region of the posteriors and these credible intervals are almost the same for MLABC compared with ABC rejection. Both methods yield very similar mean and credible interval values.

## Higher dimensional and higher order models

We now investigate the validity of our MLABC approach to inference for higher dimensional and higher order models. In the first instance, we investigate the inference problem in four-dimensional parameter spaces for first order reactions. We then investigate a more biologically inspired enzyme kinetics model that includes a second order reaction.

### Mono-molecular chains

A general mono-molecular chain biochemical reaction network has the form

$$Z \xrightarrow{k_1} X_1 + Z,\ X_1 \xrightarrow{k_2} X_2,\ \cdots,\ X_{i-1} \xrightarrow{k_i} X_i,\ \cdots,\ X_{M-2} \xrightarrow{k_{M-1}} X_{M-1},\ X_{M-1} \xrightarrow{k_M} Y. \tag{33}$$

As shown by Jahnke et al. [41], the CME for Equation (33) has a closed form solution, however, it is non-trivial to evaluate. For the purposes of this study, we suppose the CME of the mono-molecular chain to be intractable.

When $M = 4$ we have the four reaction mono-molecular chain

$$Z \xrightarrow{k_1} X_1 + Z, \quad X_1 \xrightarrow{k_2} X_2, \quad X_2 \xrightarrow{k_3} X_3, \quad X_3 \xrightarrow{k_4} Y, \tag{34}$$

with initial conditions

$$X_1(0) = x_{1,0}, \quad X_2(0) = x_{2,0}, \quad X_3(0) = x_{3,0}, \quad Z(0) = 1. \tag{35}$$

We compute 20 independent MLABC and ABC rejection CDF estimators for the four reaction mono-molecular chains (Equation (34)) using data with $N_t = [2, 4, \ldots, 20]$ observation points. The CDFs are approximated over the region $\{(k_1, k_2, k_3, k_4) : (k_1, k_2, k_3, k_4) \in [0, 3] \times [0, 0.3] \times [0, 0.03] \times [0, 0.03]\}$ using the target RMSE $h = 0.2$. Computation times are averaged over the 20 samples.

For the four reaction mono-molecular chain model, we note in Figure 7(b) that a peak computational gain of approximately five times is achieved before a reduction back to four times. Figures 7(c)-(j) also provides the resulting parameter estimates using ABC and MLABC. In this case, the MLABC and standard ABC estimates are in very close agreement across all four parameters. The MLABC estimator displays more variability in $u(k_i)$ for $N_t \geq 10$, however these oscillations follow the same trend as the ABC estimates. We note that for a three reaction mono-molecular chain the results (not shown) are very similar, however, a peak of 10 times computational gain is observed.

One final remark on the results for mono-molecular chains: note that the degradation/production model is actually a mono-molecular chain with $M = 2$. In light of this, it is interesting to note that the peak computational gain achieved for the degradation/production model in Figure 6(b) is around 20 times ($M = 2$), for the three reaction mono-molecular chain it is around 10 times ($M = 3$)(results not shown) and for the four reaction mono-molecular chain it is around five times ($M = 4$) (Figure 7(b)). This could indicate that the ratio of the convergence rates, $\beta/\alpha$, is inversely proportional to the number of reactions, $M$. There could be other factors at work here, but given the

curse of dimensionality inversely affects the order of the acceptance rate, $\gamma$, it is logical to conclude that there is also an additional influence on the convergence rates $\alpha$ and $\beta$. More experimental and theoretical work is needed to analyze the relationship between these convergence rates and the dimensionality of the parameter space.

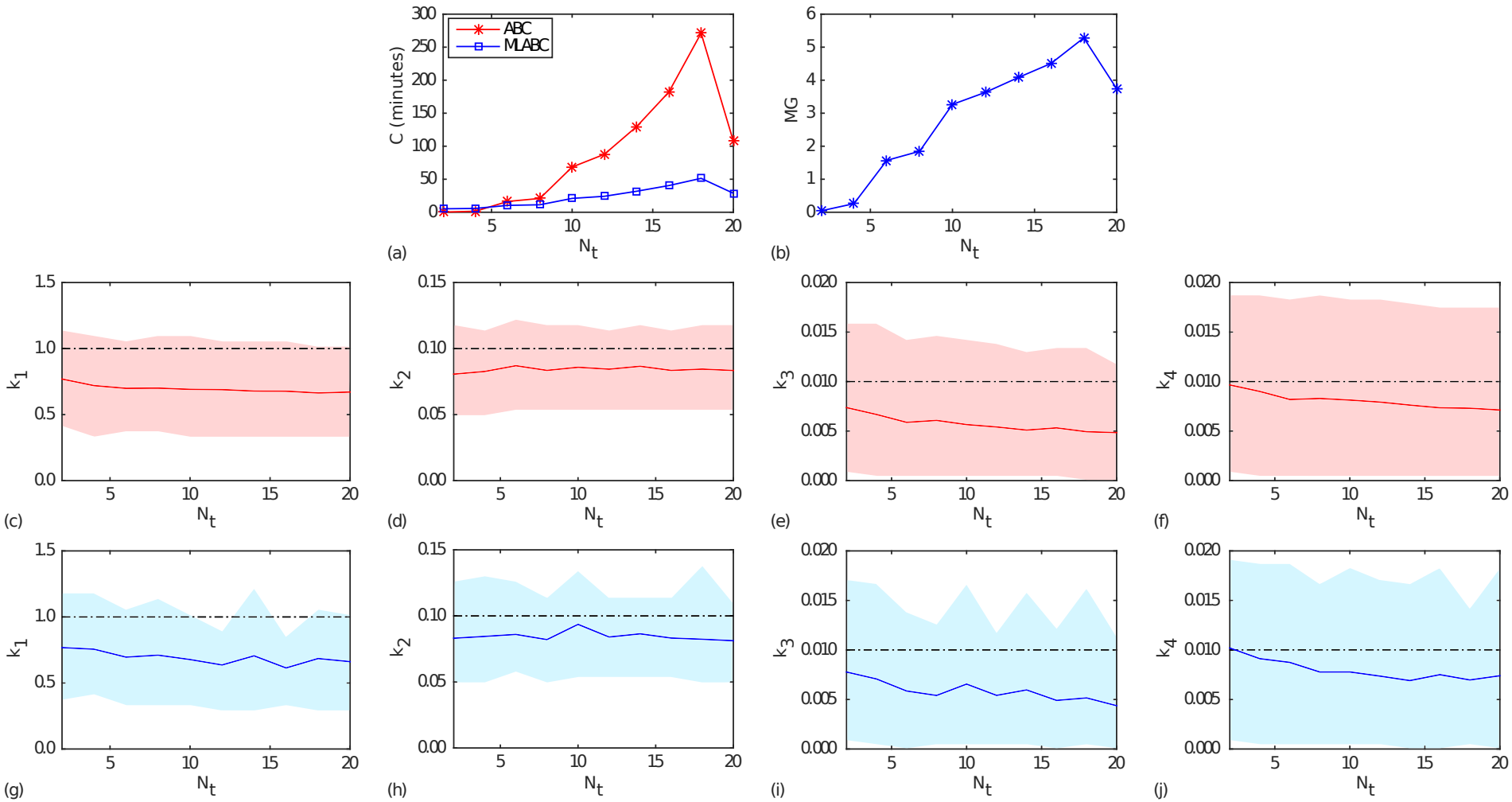


**Figure 7. Comparison of ABC and MLABC: Four reaction mono-molecular chain.** Performance of ABC and MLABC for the four reaction mono-molecular chain as the number of observation times, $N_t$, increases. (a) Computation time. (b) Computational gain. (c)-(f) ABC parameter estimates. (g)-(j) MLABC parameter estimates. The true parameter values, $(k_1, k_2, k_3, k_4) = (1, 0.1, 0.01, 0.01)$, are indicated with dashed lines, posterior means and 90% credible intervals are indicated with solid lines and shaded areas, respectively.

### Higher order models

We now test the applicability of MLABC to more general biochemical reaction networks with higher order reactions. Such networks rarely yield a tractable solution to the CME [7, 39]. As a result, such higher order models are practical target applications for MLABC. We consider a Michaelis-Menten enzyme kinetic model [42], which describes the dynamics of an enzyme-catalyzed reaction of a substrate $S$ into a product $P$ with the enzyme $E$ acting as a catalyst. A three reaction Michaelis-Menten model is given by

$$S + E \xrightarrow{k_1} ES, \quad ES \xrightarrow{k_2} S + E, \quad ES \xrightarrow{k_3} P + E, \tag{36}$$

with initial conditions

$$S(0) = s_0, \quad E(0) = e_0, \quad ES(0) = 0, \quad P(0) = 0. \tag{37}$$

An example realization demonstrating the additional complexity of the dynamics of this model is provided in Figure 8. The realization displays the conversion of $S$ into $P$. Note that as $S(t) \to (k_2+k_3)/k_1$, the propensity function of the third reaction $a_3(ES(t); k_3) \to k_3 e_0/2$, so the shape of the $ES$ curve depends crucially on this ratio of parameters. As a result, more observations time points are required to ensure the assumption of compact support for the posterior is reasonable.

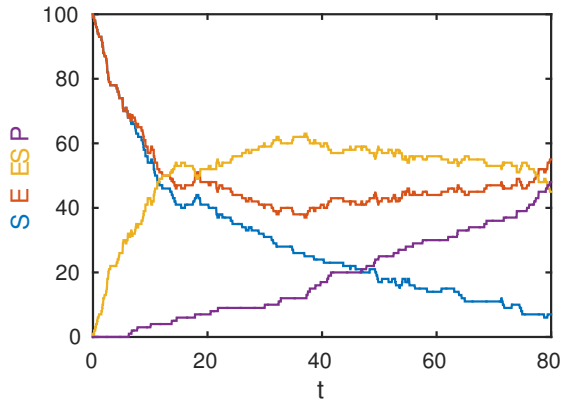


**Figure 8. Michaelis-Menten model realization.** Example realization with $k_1 = 0.001$ (sec$^{-1}$), $k_2 = 0.005$ (sec$^{-1}$), $k_3 = 0.01$ (sec$^{-1}$) and $s_0 = e_0 = 100$.

The Michaelis-Menten model given in Equation (36) is the first we investigate with an intractable CME. As a result we investigate how the performance of MLABC is affected by both the choice of target RMSE, $h$, and the choice of the number of observation points, $N_t$.

First, we consider the computational gain achieved for the inference on the Michaelis-Menten model as the target RMSE, $h$, decreases. This time, the CDF estimators are constructed using ABC and MLABC with a uniform prior with support $\{(k_1, k_2, k_3) : (k_1, k_2, k_3) \in [0, 0.003] \times [0, 0.015] \times [0, 0.03]\}$ and fixed number of observation point, $N_t = 12$. Estimators are computed for target RMSEs of $h = 0.125, 0.15, \ldots, 0.2$. Computation times are averaged over 20 independent simulations. Results are shown in Figure 9, confirming the growth in computational gain as $h \to 0$.

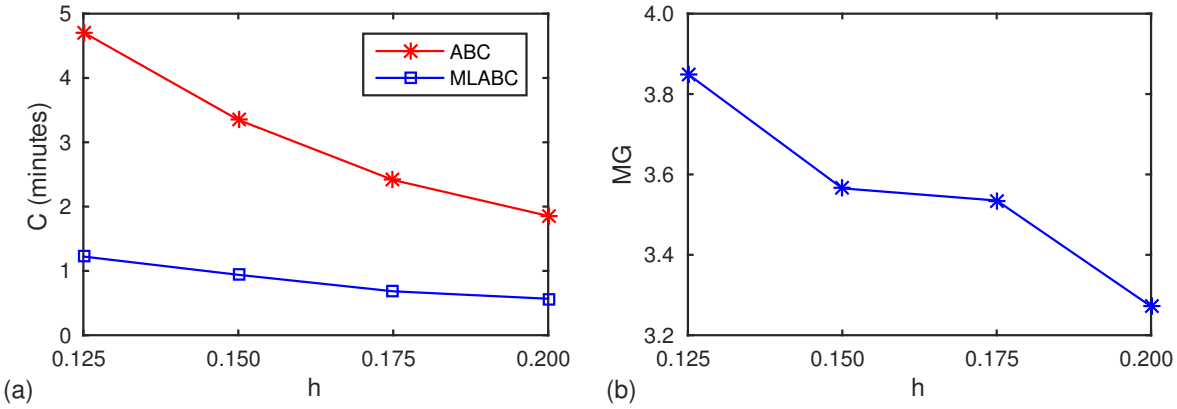


**Figure 9. Comparison of ABC and MLABC: Michaelis-Menten model.** The performance of MLABC as the target error $h$ decreases. (a) Computation time. (b) Computational gain.

For our last experiment, we take data over a much larger interval $N_t = [2, 12, \ldots 192]$. We choose this interval to highlight the fact that as $N_t$ is increased the computational gain reaches a maximum then plateaus. We conjecture that this reflects a point at which little information is gained through the additional observations. Just as with our other models we compute 20 independent CDF estimators using ABC and MLABC for each value of $N_t$ with target RMSE $h = 0.2$. In this case, we take a uniform prior with support $\{(k_1, k_2, k_3) : (k_1, k_2, k_3) \in [0, 0.005] \times [0, 0.025] \times [0, 0.05]\}$. The result is a peak computational gain of six times followed by a plateau between four and six times, as shown in Figure 10(b).

The associated parameter estimates are given in Figure 10(c)-(h). The MLABC estimator follows very closely the ABC estimator in terms of the mean for all parameters however the uncertainties in the MLABC estimate for $k_2$ are nearly double that of the ABC estimator for some values of $N_t$. Interestingly, this seems to occur for values of $N_t$

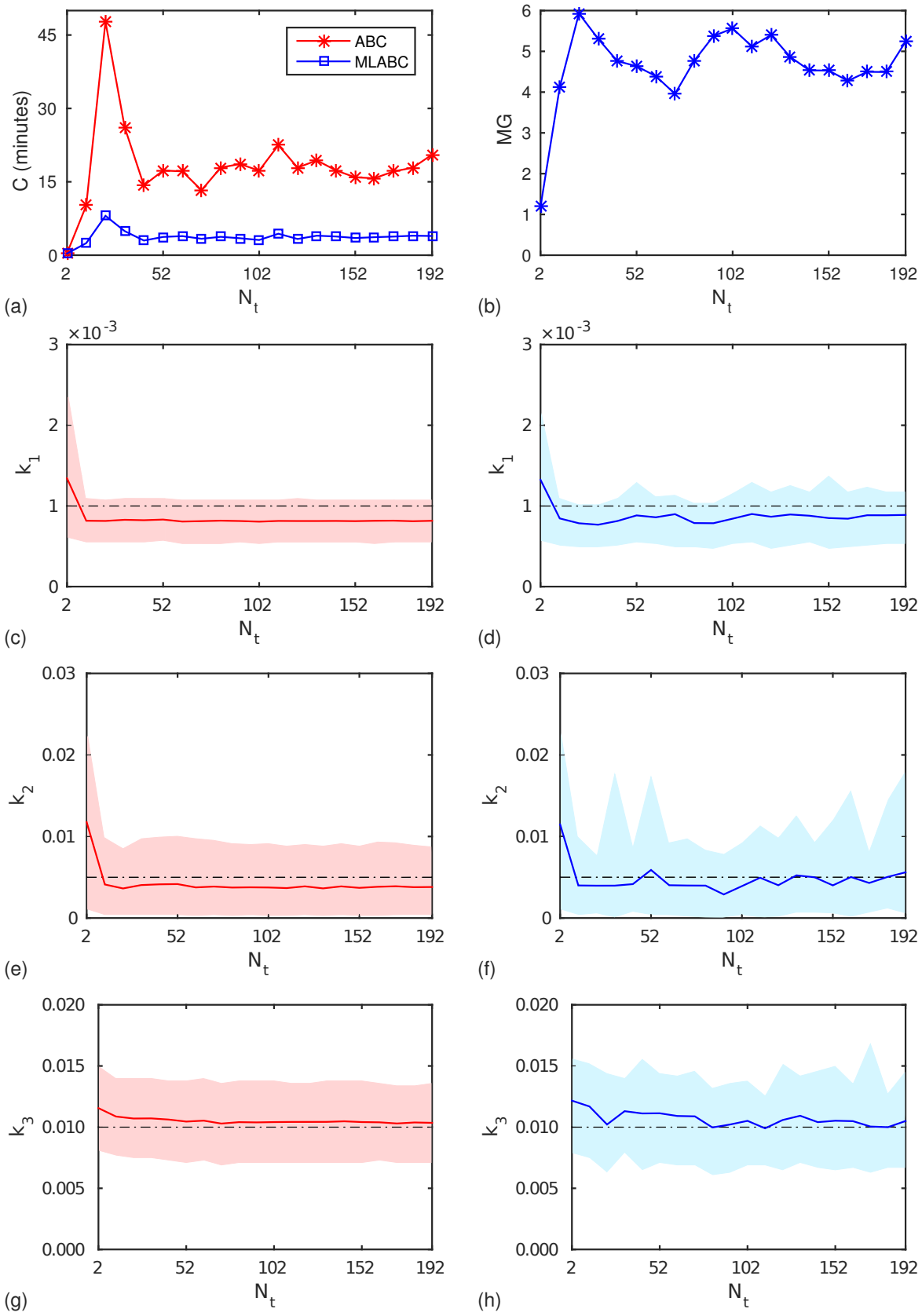


**Figure 10. Comparison of ABC and MLABC: Michaelis-Menten model.** Performance of ABC and MLABC for the Michaelis-Menten model as the number of observation times, $N_t$, increases. (a) Computation time. (b) Computational gain. (c), (e) and (g) ABC parameter estimates. (d), (f) and (h) MLABC parameter estimates. The true parameter values, $(k_1, k_2, k_3) = (0.005, 0.025, 0.05)$, are indicated with dashed lines, posterior means and 90% credible intervals are indicated with solid lines and

in which the computational gain is lower. Further investigation is required to explain the reasons for this.

# Conclusion

In this study, we present a new approach to computational Bayesian inference using MLMC sampling. We perform a general analysis based on the approximation of posterior CDFs using MLMC techniques developed by Giles et al. [37], to show that under our convergence assumptions, asymptotically, a net computational gain is always achievable for some sufficiently small value of RMSE, $h$, and simulation results confirm this prediction.

We also develop a practical implementation of the MLABC method that does not require the convergence rate parameters to be known *a priori*. Numerical estimates of the posterior CDF over a range of models suggest that a computational gain of four to 20 times can often be achieved over standard ABC rejection, with larger data set dimensionality improving this gain, up to some maximum. Under the right conditions, such as one-dimensional inference problems, a computational gain of up to 60 times is achievable.

Though the target application of this work is parameter inference for stochastic biochemical reaction network models, the MLABC method is as general as ABC rejection. From a practical perspective, MLABC can be used in place of ABC rejection if the standard assumptions on weak and strong convergence hold. Minor modifications to the provided prototype code are required to achieve this.

While our current approach is a step towards dealing with the curse of dimensionality, there is still much work to be done. For the purposes of this initial investigation, we use ABC rejection for our benchmark inference method and as the basis of the MLABC method. The most natural extension of this work is the application of the MLMC framework to other ABC methods. We acknowledge that more advanced approaches such *likelihood-free Markov chain Monte Carlo* [26] and *likelihood-free sequential Monte Carlo* [27] will generally deal with higher dimensional models and data more efficiently

than ABC rejection. Our MLABC framework, however, is sufficiently general that it is not intimately tied to ABC rejection and there will be future opportunities to apply this approach to these more advanced methods.

## Acknowledgments

This project utilized the high performance computing (HPC) facility at the Queensland University of Technology (QUT). We thank Mike Giles for useful advice.

## References

1. Wilkinson DJ. Stochastic modelling for quantitative description of heterogeneous biological systems. Nature Reviews Genetics. 2009;10:122–133. doi:10.1038/nrg2509.

2. Fedoroff N, Fontana W. Small numbers of big molecules. Science. 2002;297:1129–1131. doi:10.1126/science.1075988.

3. Blake WJ, Kaern M, Cantor CR, Collins JJ. Noise in eukaryotic gene expression. Nature. 2003;422:633–637. doi:10.1038/nature01546.

4. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. Science. 2002;297:1183–1186. doi:10.1126/science.1070919.

5. Gillespie DT. A rigorous derivation of the chemical master equation. Physica A. 1992;188:404–425. doi:10.1016/0378-4371(92)90283-V.

6. Wilkinson DJ. Parameter inference for stochastic kinetic models of bacterial gene regulation: a Bayesian approach to systems biology. In: Bernardo JM, Bayarri MJ, Berger JO, editors. Bayesian Statistics 9: Proceedings of the Ninth Valencia International Meeting. Oxford University Press; 2011. p. 679–706.

7. Burrage K, Hegland M, MacNamara S, Sidje RB. A Krylov-based finite state projection algorithm for solving the chemical master equation arising in the discrete

modelling of biological systems. In: Langville AN, Stewart WJ, editors. Proceedings of the Markov 150th Anniversary Conference. Charleston, South Carolina: Boston Books; 2006. p. 21–38.

8. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry. 1977;81(25):2340–2361. doi:10.1021/j100540a008.

9. Gillespie DT. Approximate accelerated simulation of chemically reacting systems. The Journal of Chemical Physics. 2001;115(4):1716–1733. doi:10.1063/1.1378322.

10. Anderson DF. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. The Journal of Chemical Physics. 2007;127:214107. doi:10.1063/1.2799998.

11. Anderson DF, Ganguly A, Kurtz TG. Error analysis of tau-leap simulation methods. Annals of Applied Probability. 2011;21(6):2226–2262. doi:10.1214/10-AAP756.

12. Cao Y, Li H, Petzold L. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. The Journal of Chemical Physics. 2004;121:4059–4067. doi:10.1063/1.1778376.

13. Gibson MA, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. The Journal of Physical Chemistry. 2000;104(9):1876–1889. doi:10.1021/jp993732q.

14. Lester C, Yates CA, Giles MB, Baker RE. An adaptive multi-level simulation algorithm for stochastic biological systems. The Journal of Chemical Physics. 2015;142(2):024113. doi:10.1063/1.4904980.

15. MacNamara S, Bersani AM, Burrage K, Sidje RB. Stochastic chemical kinetics and the total quasi-steady-state assumption: Application to the stochastic simulation algorithm and chemical master equation. The Journal of Chemical Physics. 2008;129(9):095105. doi:10.1063/1.2971036.

16. McCollum JM, Peterson GD, Cox CD, Simpson ML, Samatova NF. The sorting direct method for stochastic simulation of biochemical systems with varying reac-

tion execution behavior. Computational Biology and Chemistry. 2006;30(1):39–49. doi:10.1016/j.compbiolchem.2005.10.007.

17. Sunnaker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. Approximate Bayesian computation. PLOS Computational Biology. 2013;9(1):e1002803. doi:10.1371/journal.pcbi.1002803.

18. Beaumont MA, Zhang W, Balding DJ. Approximate Bayesian computation in population genetics. Genetics. 2002;162(4):2025–2035.

19. Johnston ST, Simpson MJ, McElwain DLS, Binder BJ, Ross JV. Interpreting scratch assays using pair density dynamics and approximate Bayesian computation. Open Biology. 2014;4(9):140097. doi:10.1098/rsob.140097.

20. Ratmann O, Donker G, Meijer A, Fraser C, Kelle K. Phylodynamic inference and model assessment with approximate Bayesian computation: Influenza as a case study. PLOS Computational Biology. 2012;8(12):e1002835. doi:10.1371/journal.pcbi.1002835.

21. Stumpf MPH. Approximate Bayesian inference for complex ecosystems. F1000Prime Reports. 2014;6:60. doi:10.12703/P6-60

22. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MPH. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of the Royal Society Interface. 2009;6:187–202. doi:10.1098/rsif.2008.0172

23. Vo BN, Drovandi CC, Pettit AN, Simpson MJ. Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation. Mathematical Biosciences. 2015;263:133–142. doi:10.1016/j.mbs.2015.02.010.

24. Vo BN, Drovandi CC, Pettit AN, Pettet GJ. Melanoma cell colony expansion parameters revealed by approximate Bayesian computation. PLOS Computational Biology. 2015;11(2):e1004635. doi:10.1371/journal.pcbi.1004635.

25. Fearnhead P, Prangle D. Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation.

Journal of the Royal Statistical Society Series B (Statistical Methodology). 2012;74(3):419–474. doi:10.1111/j.1467-9868.2011.01010.x.

26. Marjoram P, Molitor J, Plagnol V, Tavare S. Markov chain Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences of the United States of America. 2003;100(26):15324–15328. doi:10.1073/pnas.0306899100.

27. Sisson SA, Fan Y, Tanaka MM. Sequential Monte Carlo without likelihoods. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(6):1760–1765. doi:10.1073/pnas.0607208104.

28. Pooley CM, Bishop SC, Marion G. Using model-based proposals for fast parameter inference on discrete state space, continuous-time Markov processes. Journal of the Royal Society Interface. 2015;12:20150225. doi:10.1098/rsif.2015.0225.

29. Prangle D. Lazy ABC. Statistics and Computing. 2014;26(1):171–185. doi:0.1007/s11222-014-9544-3.

30. Giles MB. Multilevel Monte Carlo methods. Acta Numerica. 2015;24:259–328. doi:10.1017/S09624929.

31. Giles MB. Multilevel Monte Carlo path simulation. Operations Research. 2008;56(3):607–617. doi:10.1287/opre.1070.0496.

32. Giles MB, Higham DJ, Mao X. Analysing multi-level Monte Carlo for options with non-globally Lipschitz payoff. Finance and Stochastics. 2009;13(3):403–413. doi:10.1007/s00780-009-0092-1.

33. Anderson DF, Higham DJ. Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics. Multiscale Modeling & Simulation. 2012;10(1):146–179. doi:10.1137/110840546.

34. Lester C, Baker RE, Giles MB, Yates CA. Extending the multi-level method for the simulation of stochastic biological systems. Bulletin of Mathematical Biology. 2016; (in press) arXiv:1412.4069v3.

35. Efendiev Y, Jin B, Michael P, Tan X. Multilevel Markov chain Monte Carlo method for high-contrast single-phase flow problems. Communications in Computational Physics. 2015;17(1):259–286. doi:10.4208/cicp.021013.260614a.

36. Bierig C, Cherov A. Approximation of probability density functions by the multilevel Monte Carlo maximum entropy method. Journal of Computational Physics. 2016;314:661–681. doi:10.1016/j.jcp.2016.03.027.

37. Giles MB, Nagapetyan T, Ritter K. Multilevel Monte Carlo approximation of cumulative distribution function and probability densities. SIAM/ASA Journal on Uncertainty Quantification. 2015;3:267–295. doi:10.1137/140960086.

38. Kurtz TG. The relationship between stochastic and deterministic models for chemical reactions. The Journal of Chemical Physics. 1972;57(7):2976–2978. doi:10.1063/1.1678692.

39. Liao S, Vejchodsky T, Erban R. Tensor methods for parameter estimation and bifurcation analysis of stochastic reaction networks. Journal of the Royal Society Interface. 2015;12(108):201550233. doi:10.1098/rsif.2015.0233.

40. Erban R, Chapman SJ, Maini PK. A practical guide to stochastic simulation of reaction-diffusion processes. ArXiv e-prints. 2007;arXiv:0704.1908v2 .

41. Jahnke T, Huisinga W. Solving the chemical master equation for monomolecular reaction systems analytically. Journal of Mathematical Biology. 2007;54(1):1–26. doi:10.1007/s00285-006-0034-x.

42. Michaelis L, Menten ML. Die Kinetik der Invertinwirkung. Biochem Z. 1913;49:333–369.

# Supporting Information

## S1. Derivation of average acceptance rate.

## S2. Multilevel ABC asymptotic performance analysis.

## S3. Smoothing and extension.

## S3. MLABC prototype code.