Computational Statistics
https://doi.org/10.1007/s00180-025-01606-5

ORIGINAL PAPER

Approximate Bayesian inference in a model for self-generated gradient collective cell movement

Jon Devlin[1] · Agnieszka Borowska[2] · Dirk Husmeier[2] · John Mackenzie[1]

Received: 6 October 2023 / Accepted: 27 January 2025

Abstract
In this article we explore parameter inference in a novel hybrid discrete-continuum model describing the movement of a population of cells in response to a self-generated chemotactic gradient. The model employs a drift-diffusion stochastic process, rendering likelihood-based inference methods impractical. Consequently, we consider approximate Bayesian computation (ABC) methods, which have gained popularity for models with intractable or computationally expensive likelihoods. ABC involves simulating from the generative model, using parameters from generated observations that are "close enough" to the true data to approximate the posterior distribution. Given the plethora of existing ABC methods, selecting the most suitable one for a specific problem can be challenging. To address this, we employ a simple drift-diffusion stochastic differential equation (SDE) as a benchmark problem. This allows us to assess the accuracy of popular ABC algorithms under known configurations. We also evaluate the bias between ABC-posteriors and the exact posterior for the basic SDE model, where the posterior distribution is tractable. The top-performing ABC algorithms are subsequently applied to the proposed cell movement model to infer its key parameters. This study not only contributes to understanding cell movement but also sheds light on the comparative efficiency of different ABC algorithms in a well-defined context.

Keywords Approximate Bayesian computation · Model calibration · Drift-diffusion model · Stochastic differential equations · Chemotaxis

1 Introduction

Collective cell movement is an essential component of several important biological processes such as wound healing (Li et al. 2013), collective cell migration in embryonic development (Scarpa and Mayor 2016), the movement of leukocytes (white blood cells) to infections in immune response (De Oliveira et al. 2016) and cancer metastasis (Stuelten et al. 2018). Most of these processes depend on a type of

Extended author information available on the last page of the article

Published online: 08 March 2025

collective cell migration known as chemotaxis, the movement of cells along chemical gradients in response to a chemical stimulus. For example, it is well established that chemotaxis plays a key role in cancer metastasis (Roussos et al. 2011). Despite the obvious importance of chemotaxis, the sources of chemoattractants, and how these chemical gradients evolve in response to their depletion from cells, are often unknown (Tweedy et al. 2016).

Biophysical models have become an important and often essential tool in understanding complex biological processes, evidenced by the abundance of models in the literature (Tomlin and Axelrod 2007; Motta and Pappalardo 2013; Hori et al. 2021). These models can be used to help interpret experimental data and better understand the mechanisms underlying the observations. They can also be used to formulate hypotheses, make predictions under perturbations and allow certain aspects of the model to be added or removed to see its effect on the overall process, all of which can then be verified experimentally. We concentrate on quantitative models; those which describe and interpret results by linking mathematical models to quantitative data. There are many different types of quantitative models used within biology. For example, hybrid discrete-continuum models aim to combine different mathematical modelling approaches to try and account for often complicated biological behaviours (Osborne et al. 2010; Spill et al. 2015; Harrison and Yates 2016; Bardini et al. 2017). Whole-cell modelling aims to understand the inner working of cells by accounting for every gene and molecule within a cell (Purcell et al. 2013; Babtie and Stumpf 2017; Bhat and Balaji 2020). These models are often very high-dimensional and computationally expensive but very realistically capture the mechanisms underlying collective cell behaviour. In this paper, we consider using stochastic differential equations (SDEs) to model collective cell movement, an approach explored in a number of previous works (Hu et al. 2010; Shi et al. 2013; Tang et al. 2014; Giurghita and Husmeier 2018). SDEs can be used to describe the migration of individual cells, similar to individual-based models. SDE models can also be used to describe collective migration but work better for small population sizes. When the population size is taken much larger, SDE models can become computationally expensive and so partial differential equation (PDE) models are more suitable in that case.

Using biophysical models with physiologically relevant parameter values with the aim of replicating the results of an experiment is often called the forward problem. Equally important is the opposite: being able to estimate parameter values of a model from experimental data. This is known as the inverse problem or statistical inference, and it has a history of being used for biological problems (Wilkinson 2007; Secrier et al. 2009; Lillacci and Khammash 2010; Pullen and Morris 2014). However, statistical inference is seldom done in cell biology due to the complexity of the models and availability of the data. To the best of our knowledge, Ferguson et al. (Ferguson et al. 2016, 2017) were the first and only attempt at parameter inference for a PDE model describing self-generated gradient chemotaxis (Tweedy et al. 2016). These authors estimated the parameters of their PDE models of collective movement by numerical optimization with bootstrap (Ferguson et al. 2016) and Markov chain Monte Carlo (MCMC) (Ferguson et al. 2017). A related work is that of Devlin et al. (Devlin et al. 2019), who inferred drift and diffusion coefficients in a SDE model of a particle undergoing a directed random walk in the presence of static localization error. Their approach makes heavy use of specific analytical results and fits weighted least-squares to mean-square displacement (MSD) data.

There are three main contributions of this paper. First, we propose a novel, hybrid discrete-continuum model of a population of cells moving in response to a self-generated chemotactic gradient, as motivated by the experimental set-up in Tweedy et al. (Tweedy et al. 2016). To our knowledge, no one has used a drift-diffusion stochastic model to describe self-generated gradient chemotaxis. As the model is complex enough to render likelihood-based inference methods infeasible, our second contribution is to demonstrate how the class of approximate Bayesian computation (ABC) methods can be used to infer key parameters of interest. Our third contribution relates to the problem of algorithm selection. After testing the accuracy of the ABC methods on a related but much simpler drift-diffusion SDE (where the posterior is available in closed form), we compared the best methods from this study to inferring key parameters from our hybrid discrete-continuum model. Among the compared ABC methods, we considered an enhanced two-stage “residual" approach that, to the best of our knowledge, has not been used in an ABC setting. We note that ABC has been applied to SDE models before (Picchini 2014; Sun et al. 2015; Zhu et al. 2016; Picchini and Samson 2018; Picchini and Forman 2016; Kypraios et al. 2017; Maybank et al. 2017; Buckwar et al. 2020).

The structure of this paper is as follows. We present a new cell movement model in Sect. 2 and illustrate its ability to simulate self-generated cell chemotaxis. In the same section we also discuss the tractable toy problem based on the drift-diffusion dynamics as well as the mean-square displacement, which is a popular tool for analysing trajectories from SDEs and which we will use to form summary statistics for ABC. In Sect. 3 we revise popular ABC algorithms, where we also describe two enhanced algorithms. We discuss the ABC comparison results for the toy problem in Sect. 4. Results for parameter estimation using ABC for the cell movement model are reported in Sect. 5. Section 6 concludes with a discussion.

2 Model for self-generated gradient cell chemotaxis

Motivation for the development of a model of self-generated gradient chemotaxis comes from the experiment of Tweedy et al. (2016). *Dictyostelium discoideum* cells move within a two dimensional chamber of length L and height H. Initially, a saturating level of the chemoattractant, folic acid, is uniformly dissolved in an agarose gel. As the cells are introduced into a small well at the left hand side of the chamber it is observed that they gradually migrate away from the well by creating a self-generated gradient of the chemoattractant as depicted in Fig. 1. An analysis of cell migration data from (Tweedy et al. 2016) using Kolmogorov-Smirnov tests confirmed that the cell coordinates in the y-direction were not significantly different from samples from uniform distributions, indicating that there are no interesting features to be explained in the y-direction (Ferguson et al. 2017). To allow for efficient parameter inference, in the following subsections we therefore present a one-dimensional model for self-generated chemotaxis.

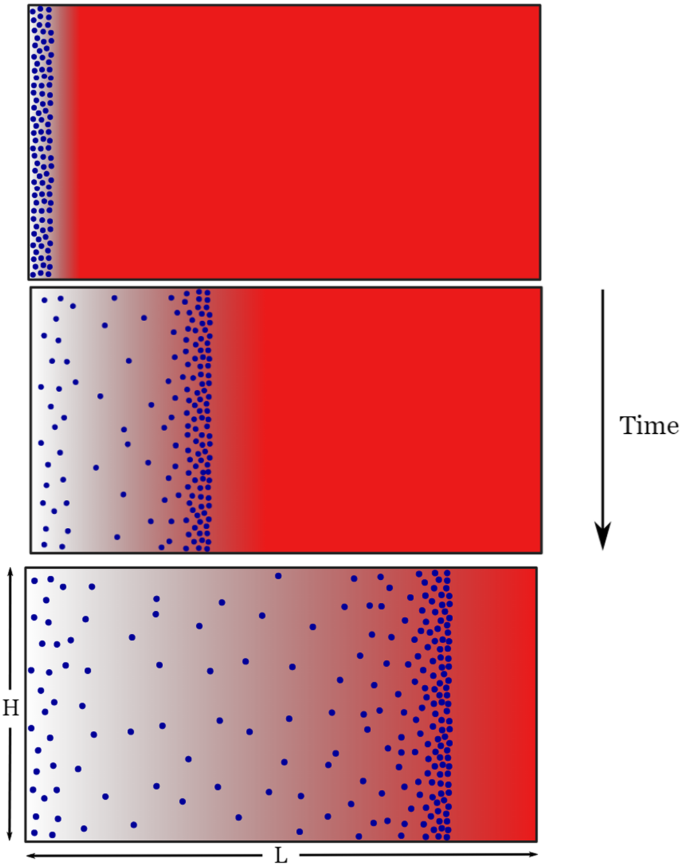


Fig. 1 Illustration of the experimental set-up: the blue circles represent cells and the red areas represent the chemical attractant

2.1 Model of the discrete cell movement

We model the movement of each individual cell by the one-dimensional drift-diffusion SDE

$$\mathrm{d}X_t = \nu\,\mathrm{d}l\,\frac{K_d}{(K_d+c)^2}\,\frac{\partial c}{\partial x}\,\mathrm{d}t + \sqrt{2D}\,\mathrm{d}W_t, \tag{1}$$

where X_t is the location of the cell at time t, ν is a parameter which converts the difference in receptor occupancy across a cell diameter to a cell velocity due to chemotaxis, dl is the diameter of the cell, K_d is the disassociation constant describing the interaction between the chemotactic ligand and its membrane bound receptor, $c \equiv c(x,t)$ is the concentration of the chemical at position x at time t, D is a measure of the random motion of the cells, assumed to be equal and constant for all cells, and W_t is a Wiener process. Here, the domain of X_t is [0, L], with the initial condition $X_0 = \upsilon$, where υ follows the uniform distribution $U(0, L/20)$. Note that this uniform distributions ensures that the cells begin in a small well of length $L/20\ \mu$m. To ensure that the cells remain in the chamber, we impose the boundary conditions $X_t = 0$, if at any stage $X_t < 0$, and $X_t = L$, if any cell is predicted to have $X_t > L$.

The chemotaxis velocity term in (1) is motivated by looking at receptor-ligand kinetics. First, imagine cells interacting with a chemical attractant. Over time, ligands begin to bind on and off the cell receptors. The rate at which ligands bind on to the receptors depends on the number of free receptors and the concentration of the chemical, while the rate at which they bind off the receptors depends on the number of bound receptors. From this, if we let ψ denote the number of bound receptors, then we have

$$\frac{\partial \psi}{\partial t} = k_1 c(R_{tot} - \psi) - k_{-1}\psi, \tag{2}$$

where k_1, k_{-1} are the rates at which the ligand binds on and off the receptors, respectively, and R_{tot} is the total receptor number. For simplicity, we assume that R_{tot} is constant. Denoting $R = \psi/R_{tot}$ as the fractional receptor occupancy, we can rewrite (2) so that

$$\frac{\partial R}{\partial t} = (k_1 c + k_{-1})\left(\frac{k_1 c}{k_1 c + k_{-1}} - R\right). \tag{3}$$

If the chemical concentration remains constant over a long enough time scale, then we can assume that the receptor occupancy reaches an equilibrium value where $\partial R/\partial t = 0$. Therefore, we get

$$R = \frac{c}{K_d + c}, \tag{4}$$

where $K_d = k_{-1}/k_1$ denotes the disassociation constant. From (4) we can see that the disassociation constant is the ligand concentration which results in half the total number of receptors being occupied. It is easy showed that equation (2) has solution

$$\psi = \frac{k_1 R_{tot} c}{k_1 c + k_{-1}} + \left(\psi_0 - \frac{k_1 R_{tot} c}{k_1 c + k_{-1}}\right)\exp(-(k_1 c + k_{-1})t), \tag{5}$$

where ψ_0 is the initial number of bound receptors. We can see therefore that the rate to reach equilibrium is determined by $k_1 c + k_{-1}$. We will assume that the initial background concentration $c \gg K_d$, and so

$$k_1 c + k_{-1} = (k_1 c - k_{-1}) + 2k_{-1} > k_1 c - k_{-1} \gg 0. \tag{6}$$

Therefore, the exponential term in (5) will decay rapidly, and so the timescale to reach equilibrium will be small compared to the other processes taking place (further justification about this assumption is given in Appendix A).

Denoting the difference in fractional receptor occupancy from the front to the back of the cell by ΔR, we can approximate this by

$$\Delta R \approx \mathrm{d}l \,\frac{K_d}{(K_d + c)^2}\, \frac{\partial c}{\partial x}. \tag{7}$$

If we assume that the chemotactic velocity is proportional to ΔR with velocity v, then we arrive at (1). The chemotactic term in (7) is similar to that used in Hillen and Painter (2009) and others (Segel 1977; Tyson et al. 1999). They looked at PDE chemotaxis models of advection–diffusion type, where the advection models the cell density movement.

It is instructive to consider the behaviour of this chemotatic term under different scenarios. For example, if we have a steady state relative concentration gradient, then

$$\frac{\Delta c}{c_0} \approx \frac{\mathrm{d}l}{c_0}\frac{\partial c}{\partial x} = \text{constant},$$

where Δc and c_0 denotes the difference and average concentration across the cell, respectively. In this situation we find that

$$\Delta R \propto \frac{cK_d}{(K_d + c)^2}. \tag{8}$$

We can see that the chemotactic term therefore decays to zero as the absolute concentration level tends to zero as expected. We also see that $\Delta R \to 0$ when $c \gg K_d$, as in this situation almost all of the cell's receptors are occupied and hence it is difficult for the cell to determine the gradient of the chemoattractant. It is easy to show that in fact ΔR is maximised when $c \approx K_d$. At this level of chemoattractant, roughly half of the cell's receptors are occupied at the front and the back of the cell.

2.2 Model of the continuous chemical concentration

We assume that the chemical concentration evolves according to a constant coefficient diffusion equation with moving point sinks to model the degradation of the chemical by membrane-bound enzymes on each cell. The governing equation is therefore

$$\frac{\partial c}{\partial t} = D_c \frac{\partial^2 c}{\partial x^2} - \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{j=1}^{N_S} \gamma(c(x^{(j)}, t)) \exp\left(\frac{-(x - x^{(j)})^2}{2\sigma^2}\right), \tag{9}$$

$$c(x, 0) = c_0, \quad t > 0, \tag{10}$$

where D_c is the diffusion coefficient of the chemical, $x^{(j)}$ is the location of the jth cell, σ^2 is variance of the Gaussian degradation term, c_0 is the initial concentration and $\gamma(c(x^{(j)}, t))$ denotes the rate of decay of the chemical at the jth cell. The strength of the cell degradation is modelled using a Michaelis-Menten formulation

$$\gamma(c(x^{(j)},t)) = \frac{V_{max}\, c(x^{(j)},t)}{K_m + c(x^{(j)},t)}, \tag{11}$$

where V_{max} is the maximum rate of degradation and K_m is the Michaelis-Menten constant.

2.3 Numerical discretisation

We assume that there are N_S cells which are simulated over the time interval $0 \le t \le T$. The total time of simulation T should be commensurate with the observational time over which experimental data is collected and as such is a possible experimental design parameter. The time interval is assumed to be partitioned uniformly by the N time points, $t_n = (n-1)T/(N-1) = (n-1)\Delta t, n = 1, \dots, N$. The position of the jth cell at the nth time point is given by $x_n^{(j)},\ 1 \le n \le N,\ 1 \le j \le N_S$.

The cells are moved by solving numerically the SDE (1) by the Euler-Maruyama method. This gives

$$x_{n+1}^{(j)} = x_n^{(j)} + v\,\mathrm{d}l \frac{K_d}{(K_d + c_n^{(j)})^2} \frac{\partial c_n}{\partial x} \Delta t + \sqrt{2D}\,\Delta W_n, \quad 1 \le n \le N, \quad 1 \le j \le N_S, \tag{12}$$

where $c_n^{(j)}$ is the chemical concentration evaluated at the location of the jth cell at the nth time point, $\partial c_n/\partial x$ is the chemical gradient evaluated at the location of the jth cell at the nth time point, and $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ follows a normal distribution of the form $\mathcal{N}(0, \Delta t)$.

Notice that equation (12) depends on the concentration and gradient of the concentration for each cell over all time. To estimate these quantities, we will use an implicit-explicit finite difference scheme to numerically solve (9). To do this, we split the spatial domain into $N_X + 1$ points, $x^i = (i-1)L/N_X = (i-1)h$, for $i = 1, \dots, N_X + 1$. Then, denoting the approximation of the concentration at the point x^i at time point t_k by c_k^i, we look to solve

$$\begin{aligned}\frac{c_{k+1}^i - c_k^i}{\Delta t} =& D_c \left(\frac{c_{k+1}^{i+1} - 2c_{k+1}^i + c_{k+1}^{i-1}}{h^2} \right) \\ &- \frac{1}{\sqrt{2\pi\sigma^2}} \sum_{j=1}^{N_S} \frac{V_{max}\, c(x_k^{(j)})}{K_m + c(x_k^{(j)})} \exp\left(\frac{-(x^i - x_k^{(j)})^2}{2\sigma^2} \right),\end{aligned} \tag{13}$$

for c_{k+1}^i, along with an approximation of the boundary conditions that $\partial c/\partial x = 0$ at $x = 0$ and $x = L$ which gives $c_{k+1}^0 = c_{k+1}^2$ and $c_{k+1}^{N_X-1} = c_{k+1}^{N_X+1}$. The updated concentration c_{k+1}^i, $i = 1, \dots, N_X + 1$ can be obtained by solving a tri-diagonal system of equations. Once we have calculated the concentration at the $N_X + 1$ spatial points, we use linear interpolation to estimate the concentration at the location of the cells. Similarly, we use a linear approximation of the gradient of the concentration so that $\partial c/\partial x \approx (c_{k+1}^{i+1} - c_{k+1}^{i-1})/2\,h$ when $x = x^i$, and again use linear interpolation to estimate its value at the location of the cells. The same size of time step is used to solve (13) as is used to moved the cells in (12).

Once we have solved numerically equation (12), we must ensure that the cells remain in the simulated chamber by imposing appropriate boundary conditions. This is done by assuming that if $x_{n+1}^{(j)} < 0$, then $x_{n+1}^{(j)} = 0$, and if $x_{n+1}^{(j)} > L$, then $x_{n+1}^{(j)} = L$.

2.4 Self-generated gradient simulations

The dataset of Tweedy et al. (2016) contains the coordinates of a group of *Dictyostelium discoideum* cells moving by self-generated gradients under a plate of agarose of length $L = 2500\,\mu$m. The time taken for the cells to traverse the majority of the plate length is $T = 5.5\,\text{h} = 19800\,\text{s}$. Initially, there is a uniform amount of folate of concentration $c_0 = 10\,\mu$M that covers the entire chamber.

Our mathematical model (1) and (9) is parameterised by a seven-dimensional vector $\boldsymbol{\theta} = (D, K_d, \mathrm{d}l, v, D_c, V_{max}, K_m)^T$. When available, the physical parameters in our model are set to literature values, see Table 1, with the following modifications: V_{max} is set to a slightly higher value to allow for the relatively low number of simulated cells ($N_S = 100$). The diffusion coefficient for folic acid, D_c, is based on the estimate in Kalimuthu and John (2009), but slightly reduced to allow for the fact that we do not have diffusion in solution, but in an agarose gel.

As opposed to the other parameters, which refer to physical quantities that can in principle be directly measured, v and D characterise the collective cell movement and its interaction with the environment. This is a complex system that defies parameter estimation by direct measurement, and we therefore have to infer them based on the observed cell movement itself. The former does not have an equivalent literature value. This value controls how far along the domain the cells will travel. We have therefore chosen a value of $v = 31.57\mu\text{m/s}$, which allows the cells to move a similar distance as those from Tweedy et al. (2016).

Studies of tracks of cell movement in isotropic environments reveals a common feature that cells typically maintain their direction of motion over short time periods, but over longer periods the direction of movement becomes random. This type of motion is normally referred to as a persistent random walk. The short time period where cells maintain their direction is called the directional persistence time. An analysis of the mean squared displacement of a persistent random walk indicates

Table 1 Nominal model parameter values for the simulation of *Dictyostelium discoideum* cells moving in response to a self-generated gradient in the chemoattractant folic acid

Parameter	Dimensional	Reference
K_d	150 nM	Wurster and Butz (1980)
dl	$10\mu m$	Rivero et al. (1996)
D_c	$11.05\ \mu\text{m}^2/\text{s}$	Kalimuthu and John (2009)*
V_{max}	3×10^{-2} nM/s	Kakebeeke et al. (1980)*
K_m	5 μM	Kakebeeke et al. (1980)

The asterisk (*) indicates that the corresponding reference values from the literature were adjusted as discussed in the main text to match our model that an estimate for D can be obtained from the expression $D = t_p v^2/2$, where t_p is the directional persistence time and v is the speed of an individual cell (Dickinson and Tranquillo 1993). Li et al. (2008) carried out careful single-cell experiments on Dictyostelium cells and found $t_p = 8$ minutes and $v = 8\mu\text{m}/\text{minute}$. Therefore, we can estimate that $D \approx 3\mu\text{m}^2/s$. A similar value can be deduced from the gradient of a straight line fit to the long time mean-squared-displacement data in Bosgraaf and van Haastert (2009) for Dictyostelium cells migrating in the absence of a chemoattractant.

We calculate the location of the cells by (12) and the chemical concentration by (13). We take $N_X = 1000$, giving a spatial grid size of $h = 2.5\,\mu$m for the implicit-explicit finite difference scheme. Simulations are performed using $N_s = 100$ cells and the time interval is discretised using $N = 500$ time steps, and hence the time step $\Delta t = 19800/499 = 39.68\,$s. Initially, the cells are given the position $x_1^{(j)} = 125\upsilon$, where υ follows a standard uniform distribution $U(0, 1)$. Note that this condition ensures that the cells begin in the small well. To verify the correctness of the proposed numerical solution and empirically prove convergence, we analyse a progression of the location of the cells, chemical concentration profile and the cell location probability density function (PDF) at six equally spaced time points (see Fig. 2). We can see that the cells move from left to right as expected. We see a leading wave of cells, a key property of self-generated gradient chemotaxis. Tweedy et al. (2016) measure the chemical concentration profile at a single time point corresponding to the end of the experiment. They find that the chemical concentration is high in front of the cell wave and quickly drops off to near zero concentration at the location of the wave. We see very similar results with our simulated concentration profiles. Finally, we find a single mode in the cell location PDF corresponding with the cell wave, whereas the experiments done by Tweedy et al. (2016) find a bimodal distribution for the PDF. In their experiments, new cells continue to move into the chamber during the experiment, while in our simulated experiments, the number of cells in the chamber is constant from the start. We believe this is why we do not find a bimodal cell location PDF.

To test whether the time and space steps used in the Euler-Maruyama method and the implicit-explicit finite difference scheme give rise to accurate numerical approximations, we repeat the simulations which led to Fig. 2 with doubled values of N and N_X (which results in a halving of both the time step and the spatial grid size). The results shown in Fig. 3 are almost identical to those in Fig. 2, suggesting that the original values for N and N_X give rise to accurate numerical approximations.

2.5 Toy problem

Below we introduce a simple drift-diffusion SDE, which we refer to as the "toy problem". The purpose of this model is to facilitate selecting an appropriate ABC algorithm for inference in the proposed cell movement model. Comparing ABC methods directly on the cell movement model would be too computationally involved due to the complexity of that model.

We consider the following one-dimensional drift-diffusion SDE:

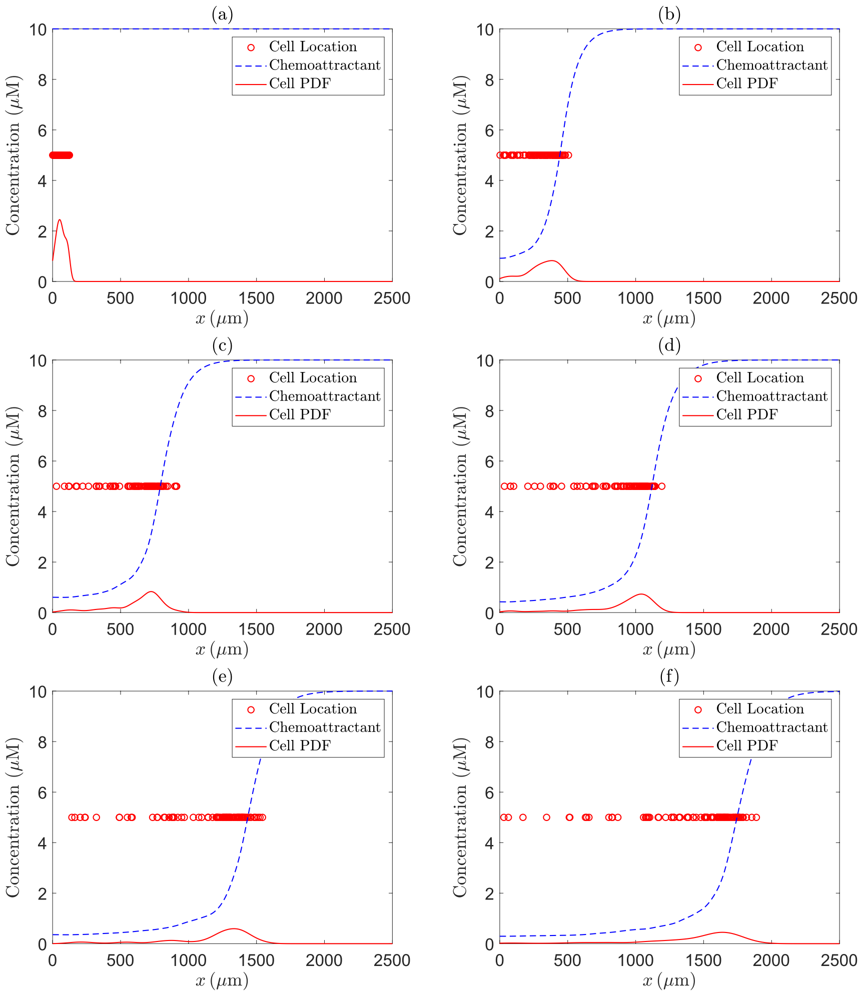


Fig. 2 The cell locations (red circles), the chemottractant concentration (dashed blue line) and the cell location PDF (solid red line) over time, where time progresses from (**a**) to (**f**)

$$dX_t = \alpha \, dt + \sqrt{2D} \, dW_t, \tag{14}$$

where $X_t \in \mathbb{R}$ denotes the true location of a particle[1] at time t, with the initial condition $X_0 = 0$, and dW_t is the increment of a Wiener process. The particles are assumed to move in an infinite domain, so there are no boundary conditions. The model parameters are α, the drift velocity, and D, the diffusion coefficient, which we collect in $\boldsymbol{\theta} = (\alpha, D)^T$. In the simulation study in Sect. 4, for both parameters we adopt the uniform prior distribution from 0 to 10, denoted $U(0, 10)$.

[1] For the sake of generality, we will refer to any entity which moves as a particle.

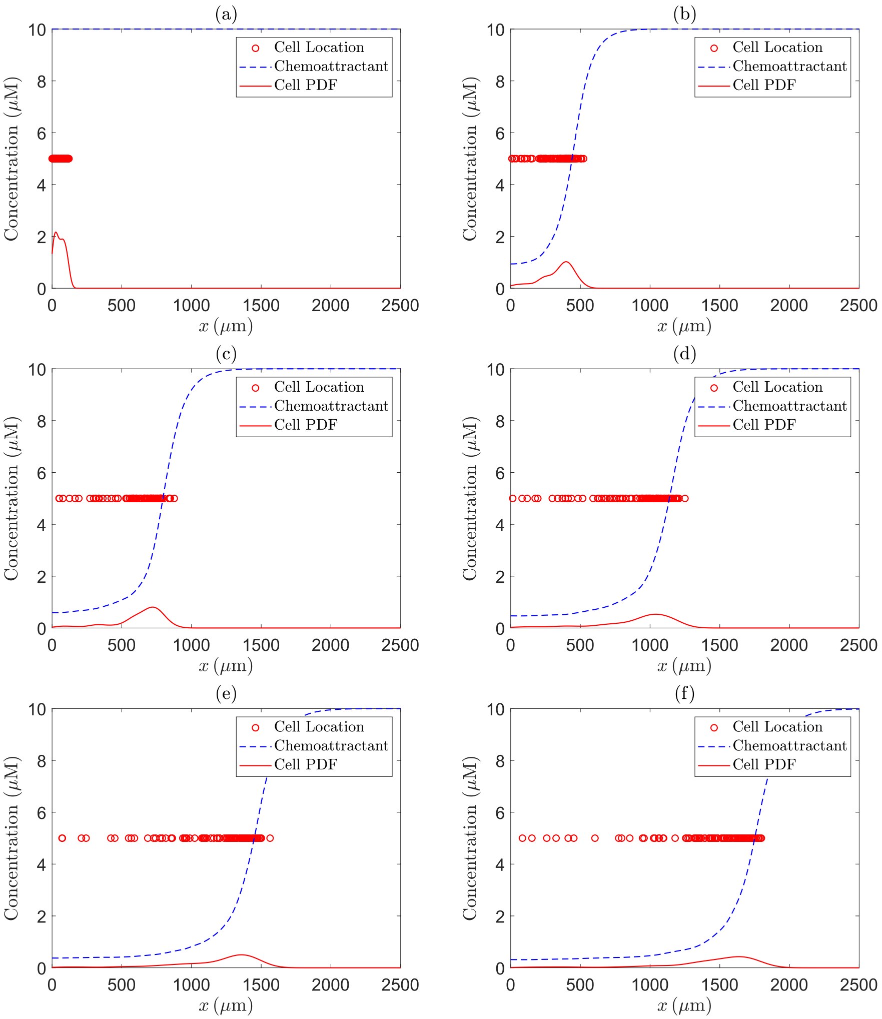


Fig. 3 The cell locations (red circles), the chemoattractant concentration (dashed blue line) and the cell location PDF (solid red line) over time, where time progresses from (**a**) to (**f**), for the same parameter values as in Fig. 2, except for the values of N and N_X, which we doubled to halve both the time step and the spatial grid size

2.5.1 Numerical solution

Due to its simplicity, model (14) admits an exact solution. However, to make our discussion of selecting an appropriate ABC algorithm (Sect. 4.3.1) general and applicable to more complex SDE models – that do require numerical methods for solving – we present a general solution to (14) based on the Euler-Maruyama method.

We assume $\boldsymbol{x}^{(j)}$, the jth trajectory generated from (14), $j = 1, \ldots, N_S$, is measured at N time points $t_n = (n-1)T/(N-1)$, $n = 1, \ldots, N$, covering the measurement time range [0, *T*], which we denote $\boldsymbol{x}^{(j)} = \{x_n^{(j)}\}_{n=1}^N$. We set $N = 100$ (the number of discretization time points) and $N_S = 100$ (the number of generated trajectories).

2.5.2 Exact posterior distribution

Under model (14), the likelihood for the *n*th time (measurement) point from the *j*th trajectory at time *t* is given by Codling et al. (2008)

$$p(x_n^{(j)}, t|\alpha, D) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(\frac{-(x_n^{(j)} - \alpha t)^2}{4Dt}\right), \tag{15}$$

which means that the likelihood for the whole trajectory $\boldsymbol{x}$ has the following form

$$\begin{aligned} L(\boldsymbol{x}^{(j)}|\alpha, D) =& \prod_{n=2}^{N} p(x_n^{(j)}, t|x_{n-1}^{(j)}, \alpha, D) \\ =&(4\pi D\mathrm{d}t)^{-\frac{N}{2}} \exp\left(\frac{-\sum_{n=2}^{N}(\Delta x_n^{(j)} - \alpha\Delta t)^2}{4D\Delta t}\right), \end{aligned} \tag{16}$$

where $\Delta t = T/(N-1)$ is the step size and $\Delta x_n^{(j)} = x_n^{(j)} - x_{n-1}^{(j)}$. We assume that each step from $x_n^{(j)}$ to $x_{n+1}^{(j)}$ is equivalent to taking a time step of size Δt starting from $x_n^{(j)} = 0$. The likelihood for the population data $\boldsymbol{y} = \{\boldsymbol{x}^{(1)}, \dots, \boldsymbol{x}^{(N_S)}\}$ of N_S independent trajectories is then a product of the likelihoods for individual trajectories

$$L(\boldsymbol{y}|\alpha, D) = \prod_{j=1}^{N_S} L(\boldsymbol{x}^{(j)}|\alpha, D). \tag{17}$$

Notice that the likelihood is tractable, which combined with uniform priors results in a closed form for the posterior.

2.6 Mean-square displacement

The MSD has been traditionally used to analyse trajectory data (Savin and Doyle 2005; Qian et al. 1991; Saxton and Jacobson 1997; Saxton 1997; Devlin et al. 2019). The MSD measures the spatial extent of a random process based on the deviation of the particle location with respect to a reference location (the 0 origin, in our case). The MSD is defined as

$$\rho(t) \equiv \mathbb{E}(|X_t|^2) = \int x^2 p(x, t|\alpha, D)dx, \tag{18}$$

where $p(x, t)$ is the pdf of the particle displacement at time *t* given in (15). For a one-dimensional system (14), the MSD can be derived analytically (Devlin et al. 2019) as

$$\rho(t) = \alpha^2 t^2 + 2Dt. \tag{19}$$

2.6.1 MSD estimation

In practice, we cannot use the theoretical, continuous-time formula (19) for the MSD and hence we need to estimate it based on discrete observations. The most popular method to do this is the time-average overlapping MSD (Michalet 2010), which, for the jth trajectory, $j = 1, \dots, N_S$, is computed as

$$\rho_n^{(j)} = \frac{1}{N+1-n} \sum_{i=1}^{N+1-n} (x_{i+n}^{(j)} - x_i^{(j)})^2, \quad n = 1, \dots, N. \tag{20}$$

Notice that $\rho_n^{(j)}$ is computed for each time lag $n\Delta t$ resulting in N values of the MSD per trajectory. To obtain more reliable estimates of the MSD we then average individual MSDs over trajectories to obtain the ensemble time-averaged MSD given by

$$\rho_n = \frac{1}{N_S} \sum_{j=1}^{N_S} \rho_n^{(j)}, \quad n = 1, \dots, N. \tag{21}$$

Figure 4 compares individual MSDs (20) with the ensemble MSD (21); the former are calculated at the $N-1$ non-zero time points for each of N_S cells, with $N = 500$ and $N_S = 100$ as in Sect. 2.4, and using the parameter values from Table 1. In Sect. 4.3.1 we will use (21) at $N = 100$ time points as the summary statistics for comparing the ABC algorithms on the toy problem. Note that using the ensemble time-averaged MSD (21) allows us to limit the number of summary statistics to a

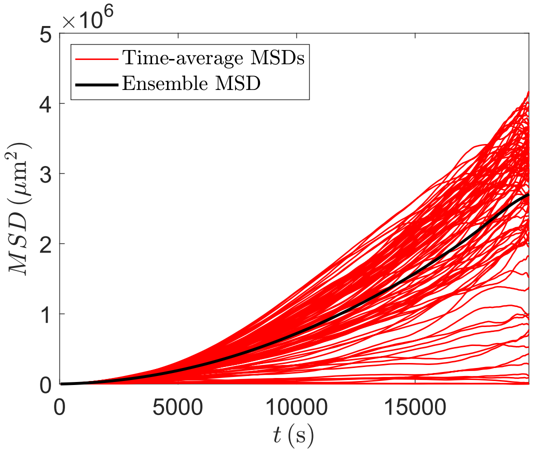


Fig. 4 A plot of the time-average overlapping MSDs for each individual cell (red lines) and the ensemble MSD (dashed black line) using the parameter values from Table 1

pre-selected value even with increasing number of data points as we take the mean of the whole distribution of the MSD values.

2.6.2 Properties of the MSD

Formula (19) reveals important properties of the MSD. First, it shows how the two parameters of the SDE model affect the MSD values for different values of t. Notice that the MSD is a quadratic function of t. For small values of t the linear term $2Dt$ dominates, so that the MSD value is mostly determined by the value of D, the diffusion coefficient. On the other hand, for large values of t, the quadratic term $\alpha^2 t^2$ prevails, and consequently the value of the drift coefficient α matters most. This will influence the informativeness of the MSD as the chosen summary statistic for inferring α and D for different values of T. Second, we note that the MSD increases quadratically with T, while its variance grows cubically with T (Devlin et al. 2019). The latter implies that the estimated MSD becomes less accurate a summary statistics as time increases.

3 Approximate Bayesian computation

In an inverse or statistical inference problem we are interested in the posterior distribution $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ of the unknown parameters $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^H$ of a model given the observed data $\boldsymbol{y}$, which by Bayes' theorem is given as

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{\theta})\,\pi(\boldsymbol{\theta})}{p(\boldsymbol{y})}, \tag{22}$$

where $p(\boldsymbol{y}|\boldsymbol{\theta})$ is the likelihood function, $\pi(\boldsymbol{\theta})$ is the prior distribution and $p(\boldsymbol{y})$ is the marginal likelihood. Application of Bayes' theorem requires computing the likelihood; however, this is not always feasible. For stochastic systems calculation of the likelihood depends on the solution of path integrals over all realisations of the latent state, which typically is analytically intractable. Likelihood-free methods are a common workaround for systems where the likelihood function is not available. The two common likelihood-free approaches are density estimation methods, which approximate the likelihood function numerically, e.g. the synthetic likelihood method (Wood 2010), and ABC, which compares observed and simulated data, or statistics of the data, through use of a distance measure. ABC has gained a considerable interest in recent years and has been used for parameter inference in a wide range of disciplines, from the biological sciences (Pritchard et al. 1999; Beaumont et al. 2002; Lintusaari et al. 2017; Lambert et al. 2018), through image analysis (Moores et al. 2015), epidemiology (Kypraios et al. 2017; McKinley et al. 2018), up to time series analysis (Toni et al. 2009; Drovandi et al. 2016; Martin et al. 2019; Tancredi 2019). We refer to Sisson et al. (2018) for a detailed treatment of ABC.

Below, we first discuss basic rejection ABC (Subsection 3.1). We then move to more advanced schemes, i.e. sequential Monte Carlo (SMC) ABC (Sisson et al. 2007; Prangle 2017) (Sect. 3.2) and semi-automatic ABC (Fearnhead and Prangle 2012) (Sect. 3.3). Finally, in Sect. 3.4 we discuss two novel ABC algorithms, based on standard techniques from computational statistics and machine learning, aimed at mitigating the problems faced by the previous semi-automatic ABC scheme. We present listings of the discussed algorithms in Appendix B.

3.1 Rejection ABC

In the basic modern rejection ABC algorithm (Tavaré et al. (1997), Pritchard et al. (1999); see Algorithm 2 in Appendix B) the prior distribution is sampled to obtain draws $\boldsymbol{\theta}^*$, which are used to simulate from a generative model $\boldsymbol{y}^* \sim f(\boldsymbol{y}|\boldsymbol{\theta}^*)$. The parameter value $\boldsymbol{\theta}^*$ is accepted if $||S(\boldsymbol{y}^*) - S(\boldsymbol{y})|| < \varepsilon$, where the norm $|| \cdot ||$, the summary statistics $S(\cdot)$, and the tolerance level ε must be specified. If the summary statistics used are sufficient, then the approximate posterior distribution approaches the exact posterior distributions as ε tends to zero. Beaumont et al. (2002) introduced smooth weighting to overcome the problem of the $S(\boldsymbol{y}^*)$ values being treated equally whenever $||S(\boldsymbol{y}^*) - S(\boldsymbol{y})|| < \varepsilon$, regardless of the exact value of $||S(\boldsymbol{y}^*) - S(\boldsymbol{y})||$.

Rejection ABC is subject to considerable computational inefficiencies, especially when dealing with high dimensional parameter spaces or continuous data (Lintusaari et al. 2017). It is also inefficient when the prior and posterior distribution are vastly different, requiring small tolerance values to obtain an accurate estimate of the posterior distribution, thereby increasing the computational cost.

3.2 Sequential Monte Carlo ABC

Sisson et al. (2007) propose an ABC method based on sequential Monte Carlo (SMC) with partial rejection control (Liu 2001). Their algorithm samples over a sequence of N_ε intermediary distributions, with decreasing tolerance values $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{N_\varepsilon}$ leading to closer approximations to $\pi(\boldsymbol{\theta}|S(\boldsymbol{y}))$. In each iteration samples are generated from the previously found intermediary distribution, except the initialization for which the prior distribution is used. The generated draws are then perturbed with Markov transition kernels and, to preserve convergence properties, weighted appealing to the importance sampling argument (Douc et al. 2007). The resulting approximation to the posterior is biased (Beaumont et al. 2009) and a number of unbiased alternatives has been proposed (Beaumont et al. 2009; Toni et al. 2009; Beaumont 2010) (a general version of one of them Beaumont 2010 is provided in Appendix B, Algorithm 3)

Sampling from distributions that are progressively becoming closer to the target distribution $\pi(\theta|S(y))$ reduces the number of parameter values drawn from low probability regions, thus enhancing computational efficiency compared to the rejection method. However, for low tolerance values, the probability of accepting parameter values can become small, even for good proposal distributions. This may result in the algorithm being ran for longer than needed with little improvement in the inference (Lintusaari et al. 2017).

Adaptive weighting

When the summary statistics are not standardised (Pritchard et al. 1999; Beaumont et al. 2002), i.e. they are out of scale with each other, the largest summary statistic can dominate over the others during the acceptance step. To overcome this problem Prangle (2017) proposes an adaptive approach to weighting summary statistics, focusing on the weighted Euclidean distance and adopting the median absolute deviation (MAD) for weighting. During the first ABC-SMC iteration, equal weights are used, while in all later iterations, the weights are based on the MAD of the accepted summary statistics from the previous iteration. The tolerance values are chosen automatically as a quantile of the previous accepted distance values (Drovandi and Pettitt 2011). Algorithm 4 in Appendix B summarises Prangle's (2017) algorithm.

3.3 Standard semi-automatic ABC

Fearnhead and Prangle (2012) propose a semi-automatic approach to selecting summary statistics, in which the posterior mean serves as the summary statistics. Fearnhead and Prangle (2012) demonstrate that this approach minimizes the quadratic loss measuring the estimation accuracy. The posterior mean is not known a priori and hence in practice it needs to be estimated using a regression model. Below we will refer to semi-automatic summary statistics as "second-stage" summary statistics, to distinguish them from "first-stage" summary statistics that are obtained directly as transformations of the data. Algorithm 1 presents a high-level description of semi-automatic ABC.

Algorithm 1 High-level description of semi-automatic ABC.

```
1: Determine a region of non-negligible posterior mass using a pilot run of ABC. (This step is optional, depending on the informativeness of the prior distribution.)
2: Use the training region from Step 1 to generate parameter values and the corresponding simulated data.
3: Calculate first-stage summary statistics for each dataset simulated in Step 2.
4: Regress the parameter values generated in Step 2 on the first-stage summary statistics calculated in Step 3 to estimate the second-stage summary statistics.
5: Run ABC (e.g. ABC-SMC) using the second-stage summary statistics from Step 4 as summary statistics.
```

Note that the regression model in Step 4 can be selected in different ways, which we discuss in Sects. 3.3.1 (multivariate linear regression), 3.3.2 (Gaussian process regression) and 3.4 (two modified approaches).

3.3.1 Linear regression approach

The original approach of Fearnhead and Prangle (2012) adopts multivariate linear regression (MVLR) of parameter values on first-stage summary statistics. Formally, suppose we have generated a set of M parameter values $\boldsymbol{\theta}^{(j)}$ and the corresponding observations $\boldsymbol{y}^{(j)}$, $j = 1, \dots, M$ (Step 2 in Algorithm 1). Let $g(\cdot)$ be a vector-valued function representing transformations of the data and $\mathbf{z}$ denote a vector of transformations under g, $\mathbf{z} := g(\boldsymbol{y}) \in \mathcal{X} \subseteq \mathbb{R}^G$. We refer to $\mathbf{z}$ as "first-stage summary statistics" (see Step 3 in Algorithm 1). As an illustration, consider an application by Fearnhead and Prangle (2012), who found that $g(\boldsymbol{y}) = [\boldsymbol{y}, \boldsymbol{y}^2, \boldsymbol{y}^3, \boldsymbol{y}^4]$, a vector of length $4n_y$ (where n_y is the length of the data vector $\boldsymbol{y}$) consisting of the data together with the second, third and fourth powers of individual data points, resulted in well-performing set of summary statistics. In our applications, $\mathbf{z}$ consists of ensemble MSD (21) calculated at either $N = 100$ points (in Sect. 4.3.1) or at the first and final ten MSD points (in Sect. 5).

In the semi-automatic ABC approach the posterior mean of the ith component of $\boldsymbol{\theta}$ is taken as the ith "second-stage summary statistic" (see Step 4 in Algorithm 1). Fearnhead and Prangle (2012) propose to estimate that posterior mean by regressing the parameter values θ_i on the linear of transformations of $\mathbf{z}$, i.e.

$$\theta_i = \mathbb{E}(\theta_i|\boldsymbol{y}) + \epsilon_i = \beta_0^{(i)} + \boldsymbol{\beta}^{(i)}\mathbf{z} + \epsilon_i, \tag{23}$$

where ϵ_i is a zero-mean noise term. We can use least-squares to fit (23); the fitted value $\hat{\beta}_0^{(i)} + \hat{\boldsymbol{\beta}}^{(i)}\mathbf{z}$ is an estimate for $\mathbb{E}(\theta_i|\boldsymbol{y})$.

3.3.2 Gaussian process enhanced ABC

Using MVLR in Step 4 of Algorithm 1 is likely to be suboptimal when the relationship between parameters and first-stage summary statistics is nonlinear. More accurate results are then likely to be obtained if we construct second-stage summary statistics using a flexible, nonlinear regression model e.g. using Gaussian processes (GPs) as in Borowska et al. (2021). We refer to this approach as GP-ABC.

GP regression is a nonparametric regression model where the values of the latent function f at the corresponding locations $\mathbf{z}$ explain the observed values $\mathbf{y}$. The unobserved function values are assumed to follow a GP, a stochastic process such that the joint distribution of any finite number of random variables from this process is Gaussian. A GP is fully specified by its mean and covariance function. Rasmussen and Williams (2006) provide an extensive treatment of GPs. Adopting the notation from Sect. 3.3.1, we can write the GP regression model as

$$\theta_i = f_i(\mathbf{z}) + \eta_i, \quad \eta_i \overset{iid}{\sim} \mathcal{N}(\cdot|0, \sigma_i^2), \tag{24}$$

where $\mathcal{N}(\cdot|\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2 (and *iid* stands for "independently and identically distributed"). The latent values $f_i(\mathbf{z})$ follow a GP $f_i(\mathbf{z}) \sim \mathcal{GP}(m_i(\mathbf{z}), k_i(\mathbf{z}, \mathbf{z}'))$, where $m_i(\mathbf{z}) = \mathbb{E}[f_i(\mathbf{z})]$ and $k_i(\mathbf{z}, \mathbf{z}') = \mathbb{E}[(f_i(\mathbf{z}) - m_i(\mathbf{z}))(f_i(\mathbf{z}') - m_i(\mathbf{z}'))]$ are the mean and the covariance function (kernel) of the process f_i, respectively. We follow a standard practice and assume that $m_i(\mathbf{z}) = 0$; this implies that the latent process $f_i(\mathbf{z})$ is fully specified by its kernel function.

For the collected covariates $\mathbf{Z} = [\mathbf{z}^{(1)}, \mathbf{z}^{(2)}, \dots, \mathbf{z}^{(M)}]^T$ the GP prior over $\mathbf{f}_i = [f_i(\mathbf{z}^{(1)}), f_i(\mathbf{z}^{(2)}), \dots, f_i(\mathbf{z}^{(M)})]^T$, the vector of stacked latent function values, is given by $p(\mathbf{f}_i|\mathbf{z}, \boldsymbol{\phi}_i) = \mathcal{N}(\mathbf{f}_i|\mathbf{0}, \mathbf{K}_i)$, where $\boldsymbol{\phi}_i$ denotes the GP kernel hyperparameters and $\mathbf{K}_i = k_i(\mathbf{Z}, \mathbf{Z})$. The likelihood can be expressed as $p(\boldsymbol{\Theta}_i|\mathbf{f}_i) = \mathcal{N}(\boldsymbol{\Theta}_i|\mathbf{f}_i, \sigma_i^2\mathbb{I})$, where $\boldsymbol{\Theta}_i = [\theta_i^{(1)}, \theta_i^{(2)}, \dots, \theta_i^{(M)}]^T$ and $\mathbb{I}$ is the identity matrix. Marginalising over the latent variables gives the formula for the marginal likelihood

$$p(\boldsymbol{\Theta}_i) = \mathcal{N}(\boldsymbol{\Theta}_i|\mathbf{0}, \mathbf{K}_i + \sigma_i^2\mathbb{I}).$$

Under the Gaussian observation model the conditional posterior distribution of the latent variables is also Gaussian and is given by

$$p(\mathbf{f}_i|\boldsymbol{\Theta}_i, \mathbf{Z}, \boldsymbol{\phi}_i) = \mathcal{N}(\mathbf{K}_i(\mathbf{K}_i + \sigma_i^2\mathbb{I})^{-1}\boldsymbol{\Theta}_i, \mathbf{K}_i - \mathbf{K}_i(\mathbf{K}_i + \sigma_i^2\mathbb{I})^{-1}\mathbf{K}_i).$$

Kernel specification

In our experiments in Sect. 4.3.1 we will adopt the squared exponential (SE) kernel, which is one of the most popular kernels in the literature, mostly because it results in a smooth prior on the latent function. We will also allow for automatic relevance determination (ARD) (Neal 2012; Rasmussen and Williams 2006), which provides a built-in method of variable selection. ARD kernels have a separate length scale per predictor; the ARD SE kernel has the following form

$$k(\mathbf{z}, \mathbf{z}') = \sigma_{se}^2 \exp\left(-\sum_{g=1}^{G} \frac{(z_g - z'_g)^2}{2l_g^2}\right),$$

where the vector of kernel hyperparameters to be estimated is given by $\boldsymbol{\phi} = (\sigma_{se}^2, l_1, \dots, l_G)^T$. The inverse of the length scale parameters l_g, $g = 1, \dots, G$, can be seen as the weight of the corresponding explanatory variable z_g, determining how relevant it is.

3.4 Modified semi-automatic GP-ABC

GPs are designed for interpolation and do not extrapolate well (Brynjarsdóttir and O'Hagan 2014). ABC-SMC algorithms (semi-automatic or not) perturb parameter samples, meaning that values could be sampled from outside the GP training region, requiring the GPs to extrapolate. This can result in inference problems, e.g. inferring false modes of the ABC-posterior distribution (which we report in Sect. 4.3.1). To overcome the extrapolation issues we investigate two variations of the basic GP setting and assess what influence they have on ABC inference.

3.4.1 GP-ABC with convex hulls

Our first approach is to restrict the sampling region for the GP-ABC algorithm to the convex hull of the GP training region. This ensures that the samples remain within the training region. Specifically, we first create a convex hull over the GP training set. Then, when perturbing the parameter samples during ABC-SMC, we reject and resample those draws that fall outside of the training region. We note, however, that restricting the sampling region of the GP-ABC algorithm to the GP training region could introduce a bias to the ABC results. The resulting algorithm is the same as Algorithm 1, except for potential rejection-resampling done in Step 5.

3.4.2 Residual approach

An alternative approach to alleviating the extrapolation problem is based on combining GP regression with MVLR in a two-stage regression. A similar approach is proposed by Conti et al. (2009), who integrate the regression parameters out.

We first fit MVLR (23) as in the linear regression semi-automatic ABC from Sect. 3.3.1. Then, we train a zero-mean GP regression model on the residuals from the first-stage regression as follows

$$\hat{\epsilon}_i = f_i(\mathbf{z}) + \xi_i, \quad \xi_i \overset{iid}{\sim} \mathcal{N}(\cdot|0, \omega_i^2), \tag{25}$$

where $\hat{\epsilon}_i = \theta_i - \hat{\theta}_i = \theta_i - (\hat{\beta}_0^{(i)} + \hat{\boldsymbol{\beta}}^{(i)}\mathbf{z})$ is the residual from (23), while f_i follows a GP (see Sect. 3.3.2).

4 ABC comparison results on the toy problem

In this section we compare the performance of the ABC algorithms discussed in Sect. 3 on the simple drift-diffusion SDE from Sect. 2.5. The aim of this analysis is to select the most appropriate ABC algorithm for inference in the cell movement model proposed in Sects. 2.1 and 2.2. As already emphasised, an exhaustive comparative evaluation of the ABC methods directly on the complex cell movement model is practically infeasible due to the high computational costs.

4.1 Simulation setting

We generate synthetic data by solving numerically (14) by the Euler-Maruyama method as discussed in Sect. 2.5.1 with the model parameters set to $D = 2$ and $\alpha = 1$. Hence, the observed data $\boldsymbol{y} = \{\boldsymbol{x}^{(j)}\}_{j=1}^{N_S}$ consist of $N_S = 100$ trajectories generated from (14).

Because Devlin et al. (2019) demonstrated that the inference accuracy for a MVLR model depends crucially on the value of T, we want to investigate whether

Table 2 Rejection ABC: tolerance values for different values of T

T	ε values			
0.05	2	1	0.5	0.25
5	200	100	50	25
500	200000	100000	50000	25000

Table 3 ABC-SMC: tolerance schedules for different values of T

T	ε_1	ε_2	ε_3	ε_4	ε_5	ε_6
0.05	8	4	2	1	0.5	0.25
5	800	400	200	100	50	25
500	800000	400000	200000	100000	50000	25000

the performance of the ABC schemes reveals a similar dependence for our SDE model 14. Therefore, we consider three values of T: $T = 0.05$, 5 and 500.

4.2 Implementation of ABC algorithms

Each ABC algorithm is run with $N_a = 1000$ parameter acceptances. The semi-automatic approaches from Sects. 3.3 and 3.4 utilise pilot runs with calibration sets of $M = 1000$ draws. That training set comes from the last intermediary distribution of a pilot run of the ABC-SMC with adaptive weights of Prangle (2017) with three iterations.

For the rejection ABC and ABC-SMC algorithms we use the l^2 norm, while for the semi-automatic algorithms we adopt the l^1 norm as the distance function. The reason for using different norms is somewhat arbitrary and motivated by different lengths of summary statistics vectors. In the former case we compare MSD vectors of length 100, while in the latter case there are only two second-stage summary statistics corresponding to posterior means for the two parameters. Below we report configurations for specific algorithms.

4.2.1 Rejection algorithm

We set higher values of the tolerance parameter ε for larger T values, which is motivated by the properties of the MSD, see Sect. 2.6.2. Moreover, we investigate how the ABC-posterior distributions change for a schedule of decreasing tolerance values reported in Table 2.

4.2.2 ABC-SMC

We consider the ABC-SMC algorithm of Beaumont (2010) with the deterministic tolerance schedules reported in Table 3. As for the rejection algorithms, we condition the tolerance values on the value of T. We initialise with a fairly large value as is typically done for ABC-SMC algorithms.

4.2.3 ABC-SMC with adaptive distance function

We run the algorithm with eight iterations. In the first iteration we set the tolerance value large enough to accept all parameter samples. In subsequent iterations we use the τth quantile of the previous accepted distance values as tolerance values. For the examples tested by Prangle (2017) the quantile value $\tau = 0.5$ performed best. In our experimentations, however, such a small value led to too low acceptance rates causing the algorithm to be very slow (most likely due to using the MSD as our summary statistics). Therefore, we take the smallest value of τ that allows for reasonable run time, for each of the three different values of T: for $T = 0.05$, 5 and 500, we set $\tau = 0.9$, 0.5 and 0.6, respectively.

4.2.4 Semi-automatic ABC

In the main run of the semi-automatic approach we adopt the ABC-SMC with adaptive weights algorithm of Prangle (2017). Hence, as before, we weight each summary statistics by an estimate of its MAD. We also take the same quantile values τ (in the first iteration we set the tolerance at the τth quantile of the draws from the last iteration of the pilot run – the one used to fit the regression model).

4.2.5 Residual approach

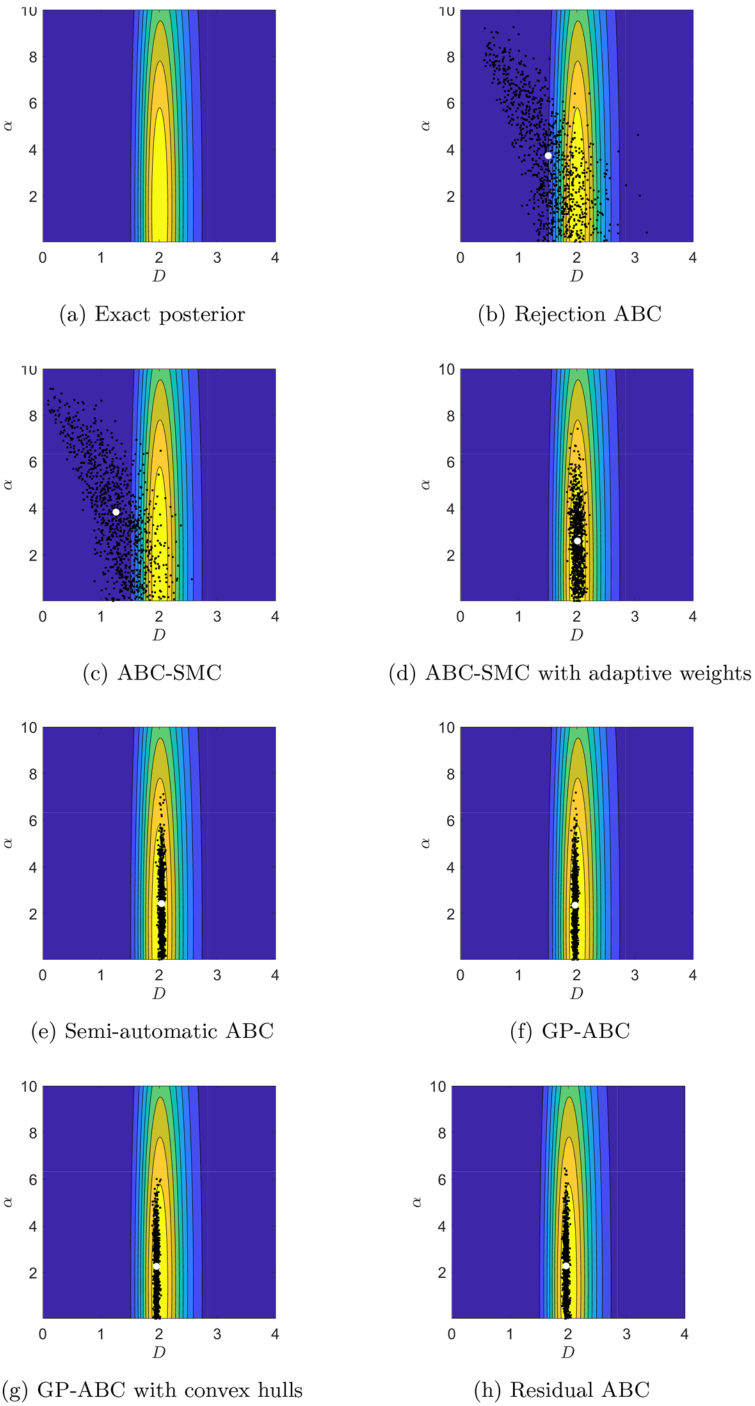
We initialise the residual approach in the same way we initialise semi-automatic ABC. Having fitted MVLR, we retrieve the residuals of the parameter values, to which we then fit a GP regression model. The second-stage summary statistics are then calculated as a sum of the fitted values from the MVLR and the GP regression model. Note that for the semi-automatic and the residual approach we found good acceptance rates could be obtained by setting $\tau = 0.5$.

4.3 Results

4.3.1 Exact posterior parameter distribution

We start our analysis by discussing the results for the gold-standard, the exact posterior distribution, which we plot on a uniform mesh of 1000×1000 points over the prior domain. At each of the mesh points we calculate the population likelihood (17) for the N_S trajectories, the shape of which matches that of the population posterior due to uniform priors for D and α.

Figures 5a, 6a and 7a present contour plots of the likelihood for the values $T = 0.05$, 5 and 500, respectively. As expected, the shape of the likelihood heavily depends on T. Similar to Devlin et al. (2019) for a small value of T, α is difficult to infer accurately, with the likelihood being very dispersed in α. As we increase T, the likelihood becomes more isometric and concentrated around the true parameter values. However, increasing T indefinitely will not improve inference as too high a T



(a) Exact posterior

(b) Rejection ABC

(c) ABC-SMC

(d) ABC-SMC with adaptive weights

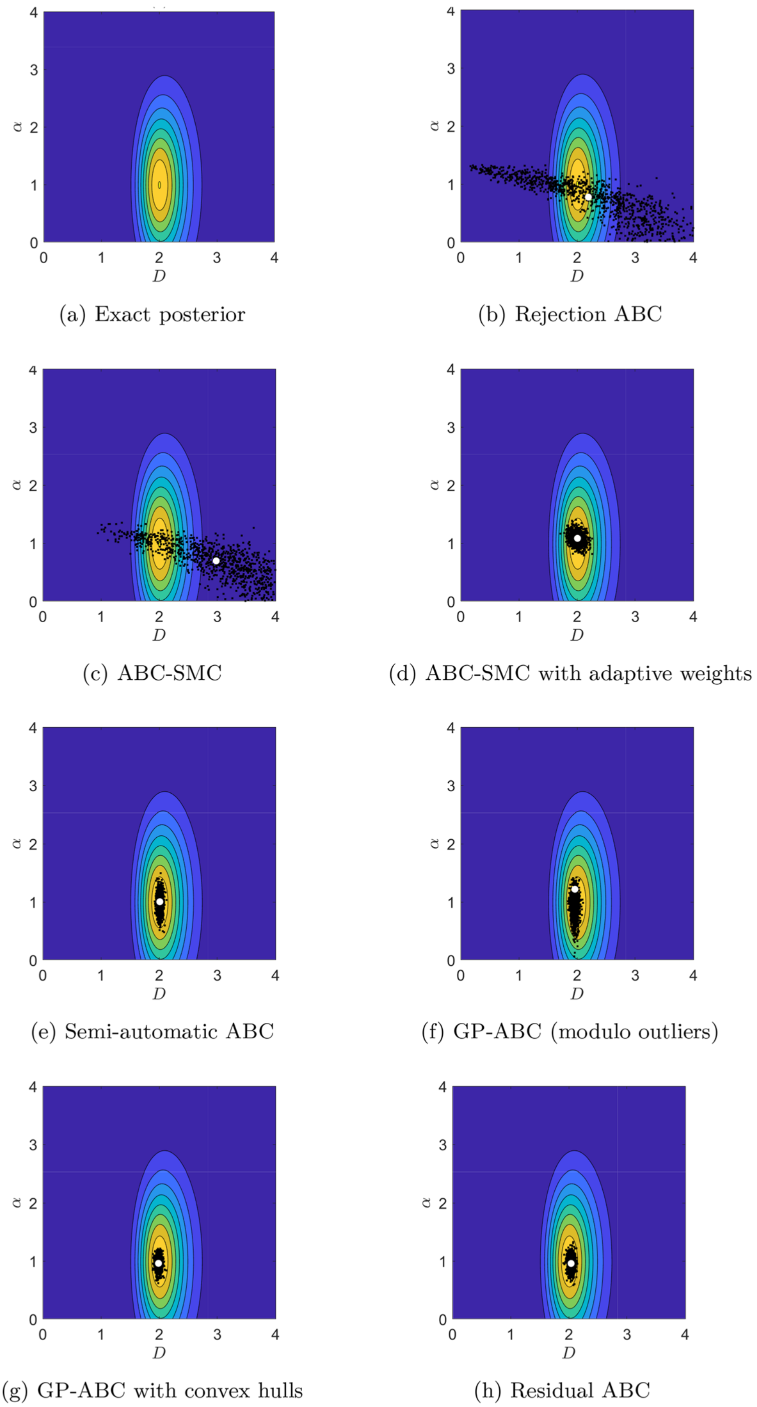
(e) Semi-automatic ABC

(f) GP-ABC

(g) GP-ABC with convex hulls

(h) Residual ABC

Fig. 5 Contour plots of the exact posterior distribution, superimposed with draws from the ABC algorithms at T=0.05 s. White dot: ABC posterior mean



(a) Exact posterior

(b) Rejection ABC

(c) ABC-SMC

(d) ABC-SMC with adaptive weights

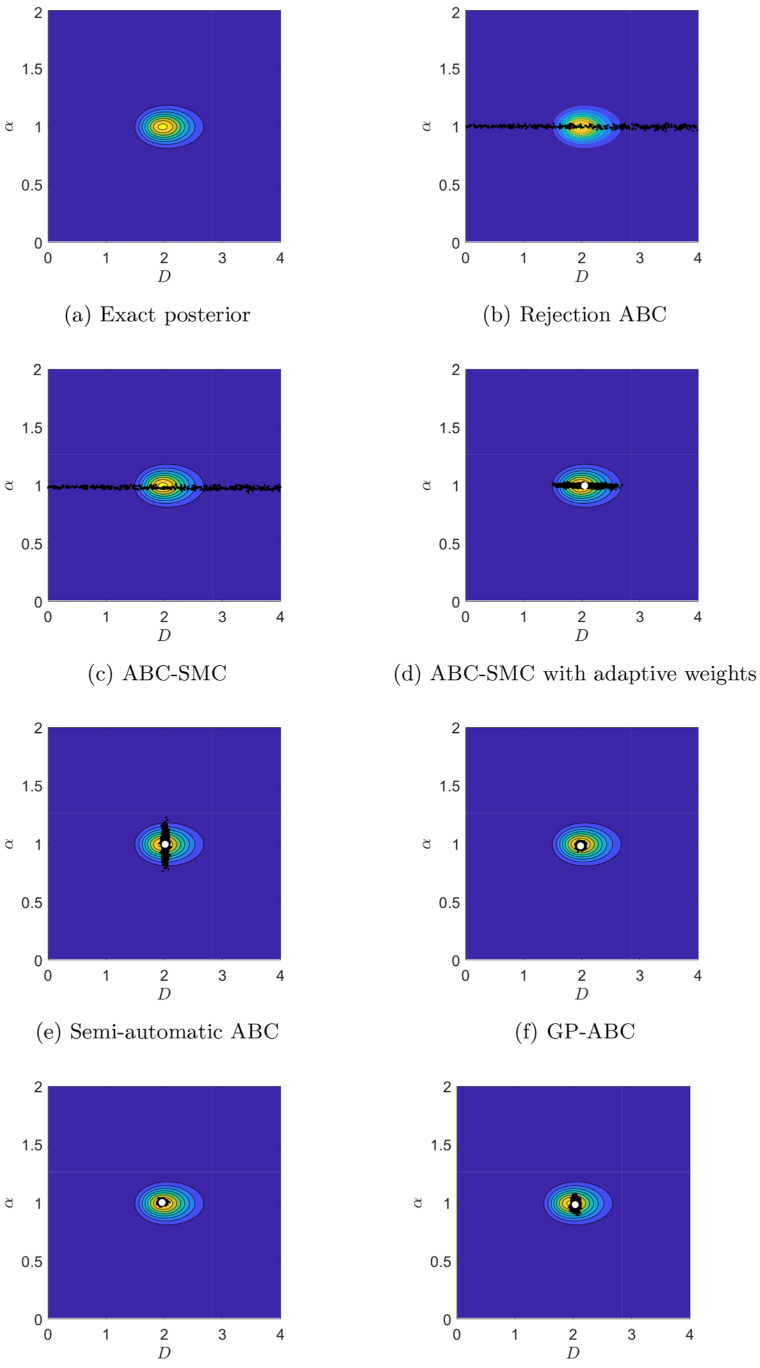
(e) Semi-automatic ABC

(f) GP-ABC (modulo outliers)

(g) GP-ABC with convex hulls

(h) Residual ABC

Fig. 6 Contour plots of the exact posterior distribution, superimposed with draws from the ABC algorithms at T=5 s. White dot: ABC posterior mean



(a) Exact posterior

(b) Rejection ABC

(c) ABC-SMC

(d) ABC-SMC with adaptive weights

(e) Semi-automatic ABC

(f) GP-ABC

(g) GP-ABC with convex hulls

(h) Residual ABC

Fig. 7 Contour plots of the exact posterior distribution, superimposed with draws from the ABC algorithms at *T*=500 s. White dot: ABC posterior mean

value will lead to identification problems for D as the MSD will be dominated by the drift term, as we demonstrate in Appendix C.1. The chosen large value for T of 500 is still moderate in this regard so D is practically identifiable in this case.

We focus on comparing the results from the ABC algorithms with the exact posterior visually, by superimposing the draws from each algorithm with the contour plots of the exact posterior in Figs. 5, 6 and 7 for $T = 0.05$, 5 and 500, respectively. For all the values of T the sampling domain is $[0, 10]^2$, however, for the middle and large value of T the posterior distributions become very concentrated, so we "zoom in" and present the results on $[0, 4]^2$. Table 4 quantitatively summarises the results in terms of percentage relative errors of means and variances from ABC with respect to the corresponding values for the exact posterior. To provide ballpark figures of the uncertainty associated with the presented error estimates we report their approximate standard deviation based on Gaussian error propagation. We assume no correlation between the moments of the ABC posteriors and those of the exact posterior (this assumption can only lead to slightly overestimated uncertainty of the presented errors).

4.3.1.1 Rejection ABC Figures 5b, 6b and 7b present the draws from the rejection ABC algorithm with the lowest tolerance values from Table 2 for $T = 0.05$, 5 and 500, respectively. Figures 14, 15 and 16 in Appendix C illustrate the results for all the considered tolerance values from Table 2. Not surprisingly, rejection ABC leads to poor approximations of the exact posterior distributions, as confirmed by the extremely large errors reported in Table 4. This might be due to using the MSD as the summary statistics, which might not be informative enough.

4.3.1.2 ABC-SMC Samples from the final intermediary distributions of the ABC-SMC algorithm for $T = 0.05$, 5 and 500 are shown in Figs. 5c, 6c and 7c, respectively. We refer the reader to Figs. 17, 18 and 19 in Appendix C for an overview of the convergence over the sequence of intermediary distributions with the tolerance schemes from Table 3. The final ABC posterior distributions are very similar to those from the rejection algorithm – as we would expect, given the same ε values used in both cases. There is still a noticeable discrepancy between the ABC and true posterior distributions, resulting in substantial errors (see Table 4).

4.3.1.3 ABC-SMC with adaptive distance function Figures 5d, 6d and 7d show the draws from the ABC-SMC algorithm with adaptive weighting (Prangle 2017) for $T = 0.05$, 5 and 500, respectively. We find that scaling each MSD value by an estimate of its MAD generally improves inference compared to the standard SMC approach, resulting in much lower errors (see Table 4). For the small value of T, the correlation between α and D is better captured, while for the large value of T we are able to practically identify α and D, as demonstrated by negligible errors for the posterior means of both parameters. For the middle value of T, both α and D are inferred relatively accurately. However, for all the values of T the algorithm struggles to capture the true parameter uncertainty, with the ABC-posterior variances being much larger than the true ones (see Table 4).

4.3.1.4 Semi-automatic ABC with MVLR Figures 5e, 6e, 7e illustrate the samples from semi-automatic ABC for $T = 0.05$, 5 and 500, respectively. We see that using the second-stage summary statistics (estimates of the posterior means) improves the inference over all the previous approaches based on the first-stage summary statistics (MSD), especially for D. Still, it remains problematic to accurately estimate parameter uncertainty, with the ABC-posterior variances generally being either too disperse (for α) or too tight (for D). For the latter, however, the standard errors are larger than the absolute values of the corresponding estimates, signalling that the apparent underdispersion of the ABC posteriors is not significant. Compared with ABC-SMC with adaptive distance function, the variance of the semi-automatic ABC posteriors is higher for α but lower for D (see Table 4). This may be expected as the MSD is linear in D while quadratic in α, and so we would expect using linear regression to extract more reliable values for D than for α.

4.3.1.5 GP-ABC The samples from GP-ABC for $T = 0.05$, 5 and 500 are given in Figs. 5f, 6f and 7f, respectively. For the middle value of T there was a small group of outlier draws concentrated around $D = 2$, $\alpha = 7$, which is not illustrated in Fig. 6f for the scale consistence with other figures. This explains the inflated error for the ABC-posterior variance of α for $T = 5$. For the remaining values of T the difference with respect to the MVLR semi-automatic algorithm is moderate, with the GP-ABC algorithm tending to underestimate the posterior variance of D a bit more often (see Table 4). We note, however, that the standard errors of the variance estimates for D are substantial, meaning that the recorded underdispersion is not significant. For the middle and large values of T, the location of the GP-ABC posterior for α is better than for the MVLR semi-automatic ABC approach.

4.3.1.6 GPs with convex hulls Figures 5g, 6g and 7g present the samples obtained with GP-ABC with convex hulls for $T = 0.05$ s, 5 and 500, respectively. We find that restricting the sampled parameter values using convex hulls removes the outliers present in the GP-ABC posterior distributions observed for $T = 5$. This confirms our conjecture that the problem with practical identification was caused by unsatisfactory extrapolation from the GP training region. However, the ABC-posteriors are still underdispersed in D and, for the large T, also in α compared to the exact posterior (see Table 4). However, again, the underestimated variance for D is subject to considerable standard errors – for the large T even greater then the absolute value of the corresponding estimate – suggesting that there is substantial simulation noise affecting the recorded results.

4.3.1.7 The residual approach The samples from the residual ABC algorithm for $T = 0.05$ s, 5 and 500 are given in Figs. 5h, 6h and 7h, respectively. The ABC-posterior distributions for the small and middle value of T seem rather similar to those from GP-ABC with convex hulls, while for the large value of T the ABC-posterior is more dispersed in α. The uncertainty in D remains underestimated for the middle and large T, but is captured relatively well for the small value of T (see Table 4). This suggests that the potential bias induced by introducing the convex hull truncation might be negligible in our case. As for semi-automatic ABC with MVLR the standard errors for the variances for D are larger than the absolute values of the corresponding estimates, which signals that the underestimation of the variance is likely due to the simulation noise.

4.3.2 Predictive distributions

Next, we inspect the predictive posterior distributions in output space. The out-of-sample posterior distribution of x for a future timepoint t^*, $x[t^*]$, is given by

$$p(x[t^*]|\boldsymbol{y}) = \int p(x[t^*]|\alpha, D)p(\alpha, D|\boldsymbol{y}) \tag{26}$$

where $\boldsymbol{y}$ is the set of data used for training. For a conjugate prior the posterior distribution

$$p(\alpha, D|\boldsymbol{y}) = \frac{L(\boldsymbol{y}|\alpha, D)p(\alpha, D)}{\int L(\boldsymbol{y}|\alpha, D)p(\alpha, D)d\alpha dD} \tag{27}$$

and the integral in equation (26) can be worked out analytically. However, for a uniform prior, it is easier to compute the integral in equation (26) numerically with the Monte Carlo method:

$$p(x[t^*]|\boldsymbol{y}) = \frac{p(x[t^*]|\alpha_i, D_j)L(\boldsymbol{y}|\alpha_i, D_j)}{\sum_{i,j} p(x[t^*]|\alpha_i, D_j)L(\boldsymbol{y}|\alpha_i, D_j)} \tag{28}$$

where the summation $\sum_{i,j}$ extends over a dense mesh in parameter space and the likelihood is obtained from equation (17). When the likelihood is unknown (or assumed unknown, for method evaluation), the predictive distribution is approximated by

$$p(x[t^*]|\boldsymbol{y}) \approx \tilde{p}(x[t^*]|\boldsymbol{y}) = \frac{1}{M}\sum_{i=1}^{M} p(x[t^*]|\alpha_i, D_i) \tag{29}$$

where $\{\alpha_i, D_i\}_{i=1}^{M}$ is our approximate sample from the unknown posterior distribution $p(\alpha, D|\boldsymbol{y})$, obtained with the various ABC samplers discussed earlier. Inserting the expression from equation (15), we get:

$$p(x[t^*]|\boldsymbol{y}) \approx \frac{1}{M}\sum_{i=1}^{M} \frac{1}{\sqrt{4\pi D_i t^*}} \exp\left(\frac{-(x[t^*] - \alpha_i t^*)^2}{4D_i t^*}\right). \tag{30}$$

We have repeated the training simulations three times, for different time intervals $[0, T]$ with $T \in \{0.05, 5, 500\}$, and made out-of-sample predictions 10% above that time interval, that is, we set $t^* = 1.1T$. The results are shown in Fig. 8.

A comparison of Figs. 5, 6, 7 and 8 shows very good agreement. Rejection ABC and ABC-SMC, for which the ABC posterior samples are not well aligned

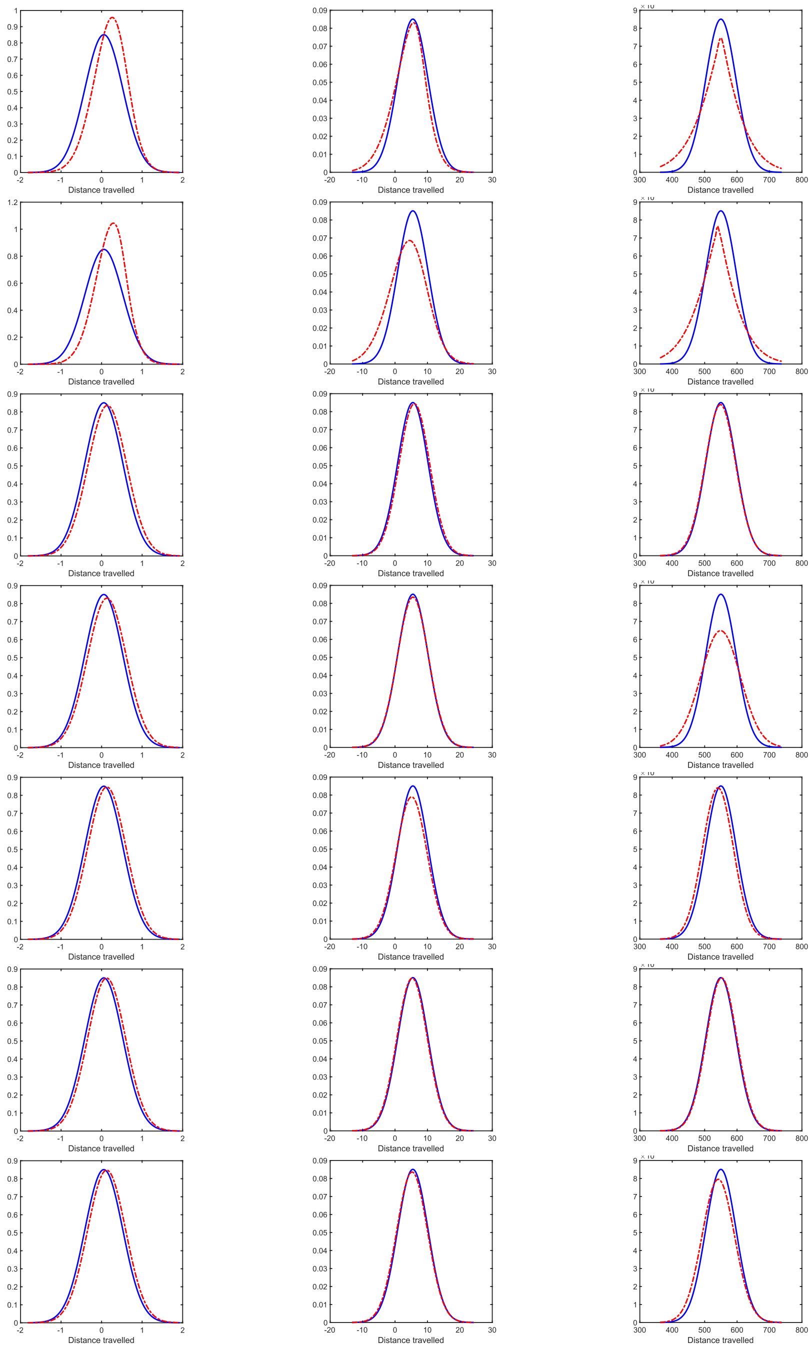


Fig. 8 True (blue solid lines) and predictive distributions (red dashed lines) at $T = 0.055$s (left column), $T = 5.5$s (centre column) and $T = 550$s (right column).The different rows correspond to the different ABC methods, from top to bottom: Rejection ABC, ABC-SMC, ABC-SMC with adaptive weights, Semi-automatic ABC, GP-ABC, GP-ABC with convex hulls and Residual ABC

Table 4 Relative percentage errors between the ABC posterior means and variances and those for the exact posterior for D and α, together with the corresponding standard errors (SE)

Method	Mean D		Variance D		Mean α		Variance α	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
T=0.05 s								
Rejection	−27.92%	2.64%	23454.60%	308.50%	−19.93%	5.56%	7097.26%	170.53%
SMC	−39.85%	2.69%	20242.61%	286.69%	**−17.79%**	5.33%	6691.40%	165.65%
SMC adj	**−4.05%**	0.34%	328.27%	41.60%	−44.41%	4.14%	2673.38%	105.86%
Semi-auto	−4.53%	0.20%	−14.35%	18.60%	−30.88%	4.22%	3458.04%	119.90%
GP	−5.00%	0.19%	−25.05%	17.40%	−41.61%	3.94%	**2524.57%**	102.98%
GP conv	−5.00%	0.19%	−35.76%	16.11%	−43.33%	4.06%	2605.74%	104.56%
Residual	−5.00%	0.20%	**−3.64%**	19.73%	−41.40%	4.10%	2754.55%	107.39%
T=5 s								
Rejection	4.48%	4.18%	85163.16%	586.94%	−36.57%	3.42%	1900.36%	89.90%
SMC	41.52%	3.15%	65163.16%	513.51%	−43.08%	3.27%	1536.66%	81.32%
SMC adj	−4.54%	0.43%	636.84%	54.56%	−12.17%	1.04%	81.85%	27.11%
Semi-auto	−3.12%	0.20%	**−15.79%**	18.45%	−29.25%	1.75%	445.55%	46.95%
GP	−3.12%	0.19%	−26.32%	17.25%	**−4.85%**	6.43%	10629.22%	208.21%
GP conv	**−2.65%**	0.19%	−26.32%	17.25%	−12.99%	0.96%	**45.48%**	24.24%
Residual	**−2.65%**	0.20%	**−15.79%**	18.45%	−15.43%	0.61%	−83.63%	8.13%
T=500 s								
Rejection	125.83%	8.92%	835029.31%	1836.92%	−0.65%	0.20%	261.45%	38.22%
SMC	114.75%	8.60%	738046.55%	1726.97%	−2.66%	0.17%	140.96%	31.20%
SMC adj	−1.29%	1.29%	7443.10%	174.58%	**0.35%**	0.14%	**20.48%**	22.06%
Semi-auto	**1.12%**	0.21%	**7.76%**	20.87%	8.38%	0.92%	10743.37%	209.31%
GP	**1.12%**	0.19%	−24.57%	17.46%	**0.35%**	0.10%	−90.36%	6.24%
GP conv	**1.12%**	0.20%	−13.79%	18.66%	**0.35%**	0.10%	−89.16%	6.62%
Residual	**1.12%**	0.20%	−13.79%	18.66%	**0.35%**	0.14%	**20.48%**	22.06%

Compared methods are: rejection ABC (Rejection), ABC-SMC (SMC), ABC-SMC with adaptive weights (SMC adj), semi-automatic ABC with MVLR (Semi-auto), GP-ABC (GP), GP-ABC with convex hulls (GP conv), residual ABC (Residual). Negative numbers mean that the ABC value is smaller than the corresponding exact posterior value. Lowest estimate values (for each quantity, for each value of T) in bold

with the parameter domain of high likelihood (Panels b-c in Figs. 5, 6 and 7) also show the largest deviation between the estimated and true predictive distributions (top two rows in Fig. 8). The other ABC algorithms, whose posterior parameter samples are much better aligned with the regions of high likelihood, also achieve a much better agreement between the estimated and true predictive distributions. Of those the poorest agreement has been found for the semi-automatic ABC algorithm at the largest value of T (Fig. 8, 4th row, right column). This tallies with Fig. 7e, which shows several sampled α values falling into the tails of the high likelihood region.

4.3.3 Coverage of HDI intervals

We have also estimated the coverage of the highest posterior density intervals (HDI) for the best performing methods. To this end we have repeated the ABC simulations $K = 100$ times. For each simulation we have estimated the 95% highest density interval using the package bayestestR (Makowski et al. 2019). We have then established how often the true parameters were included in the central 95% interval. The coverage is given by the percentage of simulations that included the true parameters in this central 95% posterior interval. For a reliable sampler, the coverage should be close to 95%. We also recorded the width of the central 95% posterior interval; this gives an indication of how compact the 95% posterior support is. The results can be found in Table 5. We note that in a few cases where the posterior distribution is bimodal, this measure is misleading as it does not allow for the low posterior probability region between the modes; see e.g. Figure 2.2 in Gelman et al. (2014) for an illustration. This explains some of the outliers in Table 5. However, the coverage is, in general, close to the target value of 95%, suggesting that our selected ABC samplers achieve a reliable approximation of the true posterior distribution.

4.3.4 Practical identification and accuracy measures

In Sect. 4.3.1 we have analysed the performance of ABC algorithms compared with the gold-standard exact posterior distribution. However, as we have seen in Sect. 4.3.1, even for the exact posterior whether both parameters are practically identifiable depends on the value of T. For ABC these issues with practical identification can be further amplified by choosing the MSD as summary statistics. Moreover, we have observed that substantial simulation noise can make capturing the true uncertainty of both parameters challenging, even for more advanced ABC schemes (ABC-SMC with adaptive weights and semi-automatic ABC algorithms). This raises different questions to that of comparing ABC with the exact posterior. How good a job do ABC algorithms do in practically identifying the parameters? And,

Table 5 Percentage coverage of 95% highest posterior density regions C and the mean of the widths of the 95% highest density intervals (HDI) for D and α

Method	$T = 0.05$			$T = 5$			$T = 500$		
	$\langle \lvert HDI_D \rvert \rangle$	$\langle \lvert HDI_\alpha \rvert \rangle$	C	$\langle \lvert HDI_D \rvert \rangle$	$\langle \lvert HDI_\alpha \rvert \rangle$	C	$\langle \lvert HDI_D \rvert \rangle$	$\langle \lvert HDI_\alpha \rvert \rangle$	C
Semi-auto	0.88	6.22	99	0.11	0.58	84	0.11	0.30	90
GP	0.11	4.17	90	0.16	4.94	98	0.16	0.21	96
GP conv	0.10	4.21	93	0.14	0.37	88	0.14	0.05	97
Residual	0.18	4.71	94	0.15	0.40	90	0.14	0.13	100

Compared methods are: Semi-automatic ABC with MVLR (Semi-auto), GP-ABC (GP), GP-ABC with convex hulls (GP conv), and residual ABC (Residual). For $T = 5$ the increased width of the credible interval for GP is caused by the bimodality of the corresponding ABC posterior distribution; see main text, Sect. 4.3.3

how to measure practical identification accuracy in the first place? To address these questions, in Appendix D we introduce two accuracy measures, which we then use to compare ABC algorithms in terms of their practical identification performance.

5 Self-generated gradient model simulation study

This section presents our main results on inferring key parameters of the proposed cell movement model (see Sect. 2) using ABC. As we discuss previously, we focus on two parameters of the model, D and v, as they are physiologically most relevant.

5.1 ABC implementation

To illustrate the performance of the ABC methods we will focus initially on the residual ABC algorithm (see Sect. 3.4.2), as it was found to be one of the best approaches among those tested in Sect. 4. We set the nominal values for D and v to 3 $\mu\text{m}^2/\text{s}$ and 31.57 $\mu\text{m}/\text{s}$, respectively, and adopt a triangular distribution[2] as the prior on D and v, with maximum and minimum values equal to double and half the nominal values, respectively. For the vector of summary statistics used during the initial run of ABC-SMC with adaptive distance function, we use the ensemble MSD calculated at the $N-1$ non-zero time points. The observed vector of summary statistics will be the ensemble MSD calculated using the nominal values from Table 1.

5.2 ABC results

Figure 9 presents samples from the estimated joint posterior distribution for D and v after 8 iterations of the residual ABC algorithm, where the dashed lines indicate the true values of D and v. Although we do not have access to the exact joint posterior distribution, we find that both D and v are inferred relatively well from the final estimated joint posterior distribution.

The results from our toy problem simulation study in Sect. 4.3.1 showed that taking a smaller value of T generally improved the inference of the diffusion coefficient, while taking a larger value of T generally improves the inference of the drift velocity. Here, we test re-running the residual ABC algorithm where we calculate the ensemble MSD as before, but the summary statistics will be taken to be the first and final ten MSD points. If the MSD for the self-generated gradient problem is similar in form to that for the toy problem, then the behaviour of the MSD for short time intervals will be dominated by the random motility parameter D, but over longer time intervals the behaviour of the MSD is determined by the chemotaxis velocity parameter v. We would expect using the first ten MSD points to improve the inference of D and the final ten MSD points to improve the inference of v. Figure 10

[2] The triangular distribution is described in Evans et al. (2000) and is characterised by three values (a; b; c), where b is the nominal value of the parameter, and a and c are the minimum and maximum values, respectively

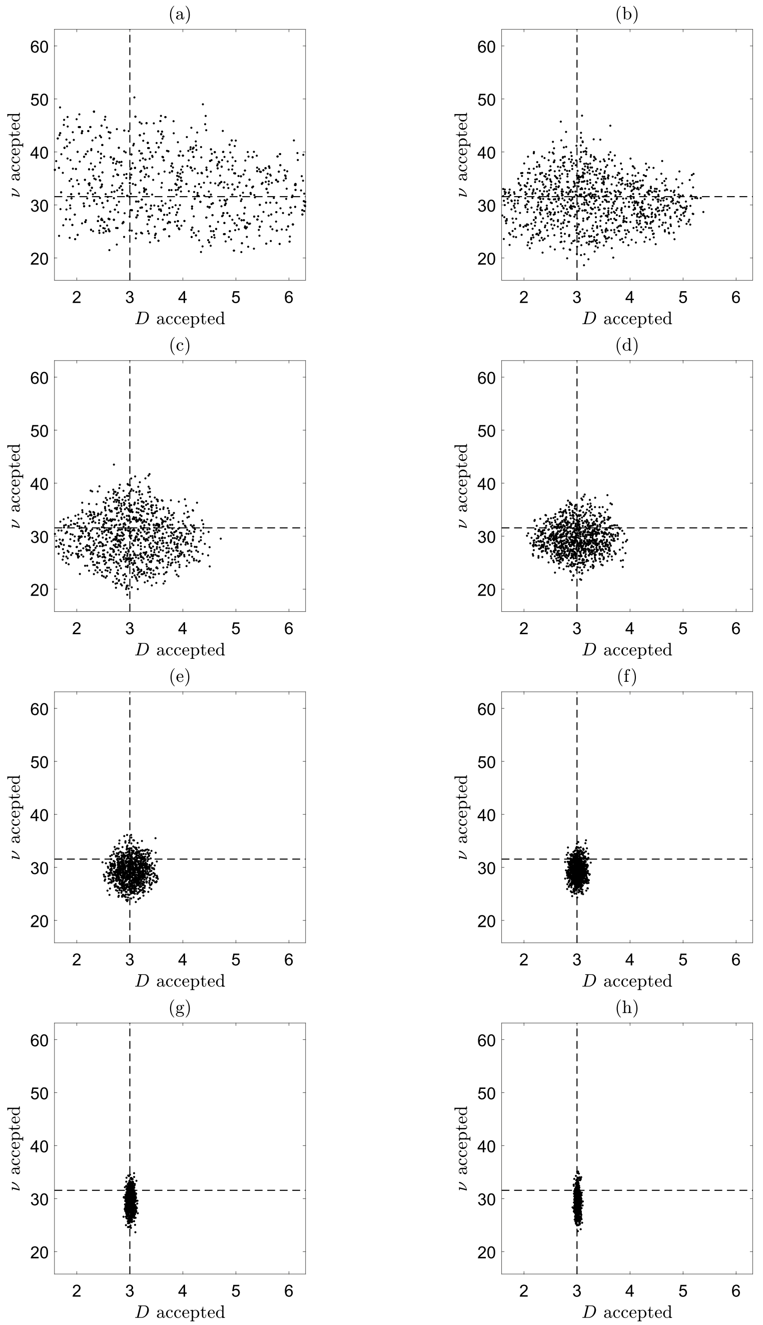


Fig. 9 Samples from the joint posterior distributions for D and ν using the residual ABC algorithm. The value of ε decreases as we progress through (**a**) to (**h**). The dashed lines correspond with the true values of D and ν. These experiments were for $N_S = 100$, $N = 500$ and $T = 19800\,\mathrm{s}$, with the parameter values given in Table 1

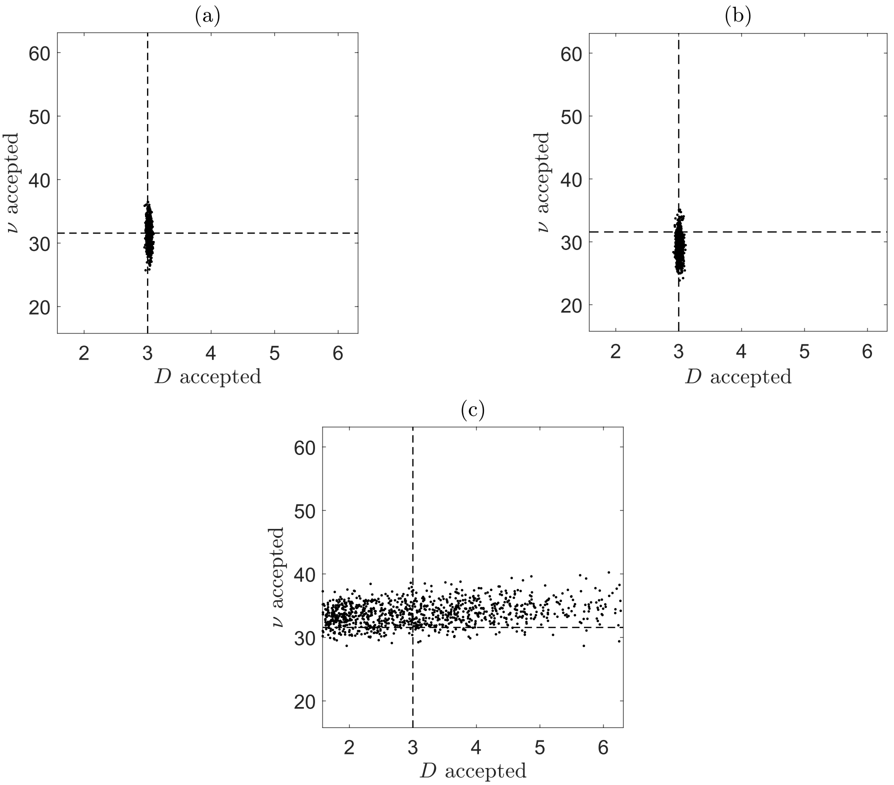


Fig. 10 Samples from the joint posterior distributions for D and ν using the residual ABC algorithm, where the summary statistics is taken to be the first ten MSD points (**a**), all the MSD points (**b**), and the final ten MSD points (**c**). The dashed lines correspond with the true values of D and ν. These experiments were for $N_S = 100$, $N = 500$ and $T = 19800$ s, with the parameter values given in Table 1

shows a plot of the residual ABC results when we use the first ten MSD points, all the MSD points (note that this is the same plot as Fig. 9h), and the last ten MSD points as the summary statistics. When we use the final ten MSD points, D becomes unidentifiable as we would expect. For ν, however, the marginal posterior distributions appear similar using the three different summary statistics and we do not find an improvement in the inference of ν when we use the final ten MSD points. This is likely due to the chemotactic term of our chemotaxis model being much more complex than the simple drift term considered in the toy model problem.

In Table 6 and Fig. 11 the results using the residual ABC algorithm are compared to those obtained using the three other best performing methods identified in Sect. 4.3.1. In Table 6 we observe that the relative error in the posterior mean estimate for D is generally lower than that for the chemotaxis parameter ν. This is likely to to due to the more complex nature of this term in the self-generated gradient model. In general, all four methods perform well in terms of accuracy and there is no evidence of any systematic bias in the estimates of the model

Table 6 Percentage errors in ABC posterior means and relative MSE for self-generated gradient study

Method	*D*		ν		
	Estimate	SE	Estimate	SE	MSE
Semi-auto	0.37%	1.01×10^{-3} %	−6.79%	5.51×10^{-3} %	8.22×10^{-3}
GP	−0.67%	6.59×10^{-4} %	5.72%	3.82×10^{-3} %	4.28×10^{-3}
GP conv	−0.10%	7.11×10^{-4} %	0.21%	3.84×10^{-3} %	1.53×10^{-3}
Residual	0.36%	9.64×10^{-4} %	−6.17%	5.21×10^{-3} %	7.66×10^{-3}

Compared methods are: Semi-automatic ABC with MVLR (Semi-auto), GP-ABC (GP), GP-ABC with convex hulls (GP conv), and residual ABC (Residual)

parameters. While our focus has been on the relative accuracy of the considered ABC algorithms, our observations are that the four methods considered in this section produce their estimated posterior distributions over a similar time period.

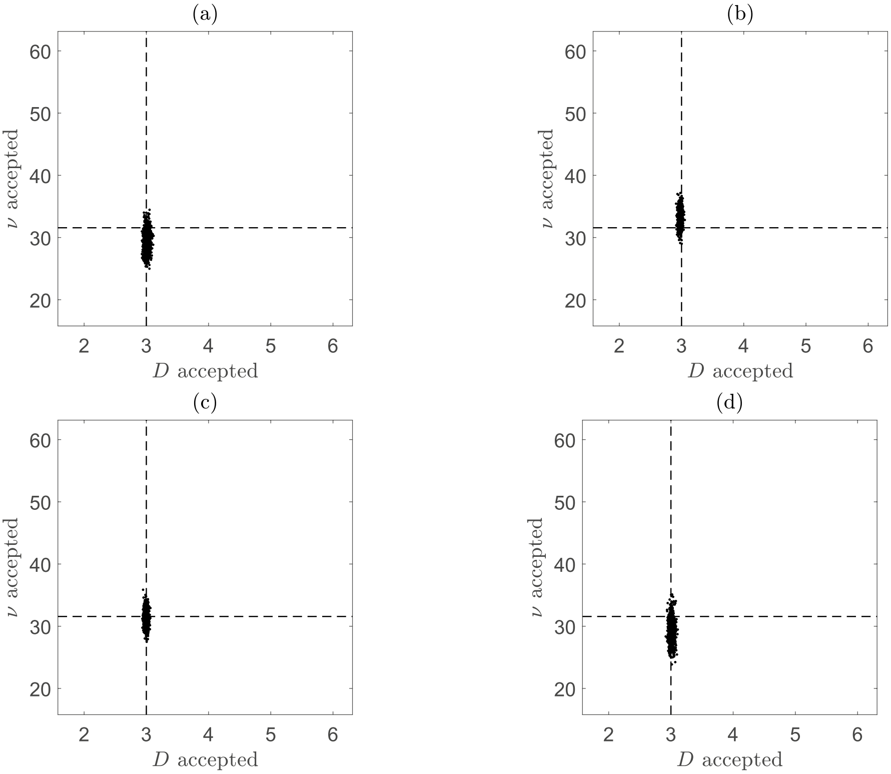


Fig. 11 Samples from the joint posteriors using **a** Semi-automatic **b** GP **c** GP-Convex Hulls and **d** Residual ABC methods

6 Discussion

In this paper we have first proposed a novel discrete-continuum biophysical model describing self-generated collective cell chemotaxis induced by the diffusion of a local chemoattractant. The cells move collectively according to the drift-diffusion SDE (1) and the evolution of the concentration of the chemoattractant is modelled using the PDE (9). Second, we have studied the inverse problem, inferring the two model parameters most critical for collective cell movement, D and ν. Since the complexity of our biophysical model renders likelihood-based statistical inference infeasible, we resort to the class of approximate Bayesian computation algorithms. To select the best algorithm for our problem, we have compared different ABC methods on a simpler analytically tractable drift-diffusion SDE model, referred to as the "toy problem". Comparing ABC methods directly on the cell movement model is practically infeasible due to its computational complexity.

The drift term in the SDE of the collective cell movement model is derived by considering receptor-ligand kinetics. By looking at the rate at which ligands bind on and off the cell receptors, assuming a constant total receptor concentration, we arrive at the chemotactic term given by (7). This term allows cells to chemotax proportionally to the chemical gradient when the concentration is small relative to the disassociation constant K_d, and it induces random cell movement when the concentration is large relative to the disassociation constant. The evolution of the chemoattractant is described using the diffusion equation with constant diffusion coefficient, along with a Gaussian-like term that models the degradation of the chemical by the cells. The strength of the cell degradation is assumed to have a Michaelis-Menten form.

To numerically simulate the movement of the cells and the evolution of the chemoattractant, we have retrieved experimental quantities for the self-generated gradient data from Tweedy et al. (2016) and values for the model parameters from the literature. To simulate the movement of the cells, we solve our drift-diffusion SDE numerically by the Euler-Maruyama method. The evolution of the chemoattractant is simulated on a uniform background grid using an implicit-explicit finite difference scheme to numerically solve (9). The updated concentration is then found by solving a tri-diagonal system of equations. Linear interpolation is used to estimate the chemical concentration at the location of the cells, as well as using a linear approximation for the chemical gradient at the location of the cells. Our simulations have shown a leading wave of cells, a key property of self-generated gradient, and concentration profiles which matched that found in Tweedy et al. (2016). This demonstrates that our drift-diffusion model has the flexibility to give rise to self-generated gradient chemotaxis.

We have compared different ABC approaches to inferring the parameters of the synthetic benchmark problem, the diffusion coefficient D and the drift velocity α. The adopted model specification renders the exact posterior tractable, allowing us to compare results from ABC algorithms against the known gold-standard. We have investigated the performance of five popular ABC algorithms: rejection

ABC, ABC-SMC, ABC-SMC with adaptive distance function of Prangle (2017), semi-automatic ABC of Fearnhead and Prangle (2012), and GP-enhanced semi-automatic ABC of Borowska et al. (2021). To further improve upon the obtained results, we have proposed two modified algorithms: an extension of the GP-ABC algorithm utilising convex hulls, and a two-stage semi-automatic approach, in which a GP regression model is fitted to the residuals from a first stage linear regression model. In all our experiments we have used the MSD calculated at $N = 100$ discretisation points as the first-stage summary statistics; for semi-automatic ABC algorithms we then used estimates of the posterior mean for D and α as the second-stage summary statistics.

The main finding from our synthetic benchmark study is that regardless of the algorithm used, the ABC-posterior distributions depend crucially on the value of the measurement time interval T, as does the exact posterior distribution. For small values of T, the exact posterior distribution is relatively tight in D and considerably dispersed in α, meaning that α is hard to identify. In contrast, for large values of T, the exact posterior distribution is more dispersed in D compared with α, which may lead to identification problems for D. However, there are intermediate values of T for which the exact posterior distribution is more isotropic, which facilitates identification and precise inference simultaneously for D and α. These findings mirror the results reported by Devlin et al. (2019), who show that an intermediate value of T balances the accuracy of the inference of both parameter values.

Compared to standard approaches based on first-stage summary statistics, the semi-automatic algorithms considered, including the two new ones proposed in this paper, provide more accurate inference of the location of the posterior distribution. Moreover, these methods generally do a (relatively) good job of estimating the uncertainty of D, in contrast to the basic methods, such as rejection ABC or ABC-SMC, which tend to produce considerably overdispersed ABC-posterior distributions. We have therefore used the proposed residual ABC approach for inferring D and v of the cell movement model. As expected, both parameters are inferred relatively accurately.

One prominent challenge, universal for most ABC methods, relates to the use of summary statistics. Our results demonstrate that the semi-automatic approaches, in which second-stage summary statistics are obtained as predictions from a pre-fitted regression model, provide an advantage over standard algorithms based on first-stage summary statistics. However, the semi-automatic methods still rely on manually selected and constructed transformations of the data (first-stage summary statistics). ABC methods without summary statistics seem a promising strand of research (Sousa et al. 2009). More recently, applying the Wasserstein distance to directly convert high dimensional data into a one-dimensional distance value has been proposed (Bernton et al. 2019). However, even such seemingly fully automatic methods still require user inputs, e.g. as to what mapping from the data space to the distance space to use. In other words, application-specific intuition behind selecting summary statistics is traded for intuition behind selecting an appropriate space filling curve.

We have used the MSD as first-stage summary statistics either directly in rejection ABC or ABC-SMC, or indirectly in semi-automatic algorithms, where it served as the explanatory variable in regression models. We note that the 100 time points at which we calculate MSD are uniformly spread over the time domain [0, T]. This might be sub-optimal, especially given the relationship between the MSD and the two parameters of interest. Generally, the quadratic nature of the MSD means for small values of T, the quadratic term in formula (21) essentially disappears so we cannot estimate α well; on the other hand, for large values of T, the relative weight of the linear term becomes negligible so that D becomes hard to estimate accurately. Selecting MSDs uniformly intensifies this issue. Such insights can inspire better choices and designs of summary statistics, but that requires domain knowledge that may not always be available.

Appendix A: Kinetics of receptor-ligand binding

In this section we consider receptor-ligand binding dynamics to justify the assumption that these occur quickly compared to the other physical processes involved in cell motility. In Fig. 12 we have plotted the evolution of fractional receptor occupancy $R(t) = \psi(t)/R_{tot}$, where $\psi(t)$ is the number of occupied receptors and R_{tot} is the total (occupied and and unoccupied) number of receptors. We have assumed initially that all receptors are unoccupied, the ligand-receptor disassociation parameter $K_d = 150$nM (Wurster and Butz 1980) and the ligand-receptor on rate $k_1 = 0.1s^{-1}$. For a range of concentrations c, we can see that the fractional receptor occupancy rapidly increases to its equilibrium value $R = c/(K_d + c)$, and the time to reach equilibrium decreases with increasing concentration c. In all situations, equilibrium is

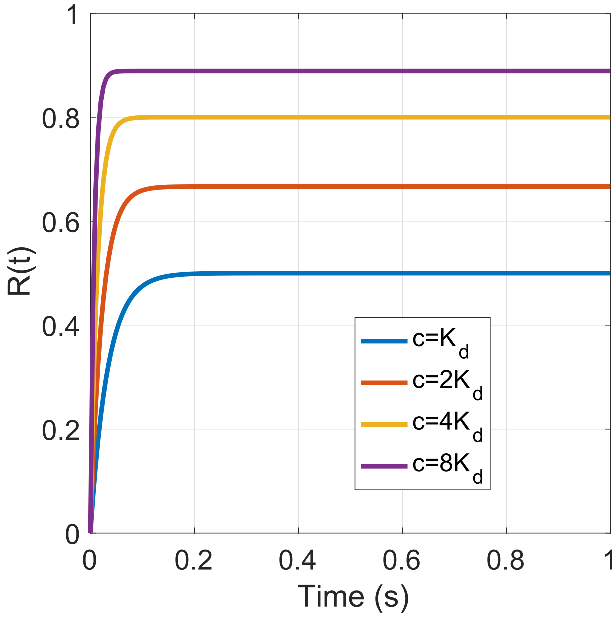


Fig. 12 Kinetics of ligand-receptor binding

reached rapidly and hence this justifies the use of equilibrium conditions for the fractional receptor occupancy term in the modelling of chemotaxis in the SDE (1).

Appendix B: ABC algorithms

Algorithm 2 ABC rejection algorithm (Pritchard et al. 1999)

```
Input: Observed summary statistic(s) of the data $S(\boldsymbol{y})$, prior distribution for the parameters $\pi(\boldsymbol{\theta})$, a generative model $f(\boldsymbol{y}|\boldsymbol{\theta})$, the number of parameter acceptances $N_a$ and the tolerance $\varepsilon$.
Output: Samples $\boldsymbol{\theta}_i,\ i = 1, \ldots, N_a$ from the approximate posterior distributions $\pi_\varepsilon(\boldsymbol{\theta}|S(\boldsymbol{y}))$.
Set the number of accepted parameter values $i = 0$.
while $i < N_a$ do
    Sample $\boldsymbol{\theta}^*$ from the prior $\pi(\boldsymbol{\theta})$.
    Simulate data using the generative model $\boldsymbol{y}^* \sim f(\boldsymbol{y}|\boldsymbol{\theta}^*)$.
    Calculate the summary statistic(s) of the simulated data $S(\boldsymbol{y}^*)$.
    if $||S(\boldsymbol{y}^*) - S(\boldsymbol{y})|| < \varepsilon$ then
        accept and store the parameters values $\boldsymbol{\theta}_i = \boldsymbol{\theta}^*$ and set $i = i + 1$.
    end if
end while
```

Algorithm 3 ABC-SMC algorithm (Beaumont 2010)

```
**Input**: Observed summary statistic(s) of the data $S(\boldsymbol{y})$, prior distribution for the parameters $\pi(\boldsymbol{\theta})$, a generative model $f(\boldsymbol{y}|\boldsymbol{\theta})$, the number of parameter acceptances $N_a$ and a schedule of tolerances $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{N_\varepsilon}$.
**Output**: Samples $\boldsymbol{\theta}_i$, $i = 1, \dots, N_a$ from the approximate posterior distributions $\pi_\varepsilon(\boldsymbol{\theta}|S(\boldsymbol{y}))$.
for $t = 1, \dots, N_\varepsilon$ do
    Set the number of accepted parameter values $i = 0$.
    while $i < N_a$ do
        Sample $\boldsymbol{\theta}^*$ from
        $q_t(\boldsymbol{\theta}) = \begin{cases} \pi(\boldsymbol{\theta}), & \text{if } t = 1, \\ \sum_{i=1}^{N_a} w_i^{t-1} K_t(\boldsymbol{\theta}|\boldsymbol{\theta}_i^{t-1}) / \sum_{i=1}^{N_a} w_i^{t-1}, & \text{otherwise.} \end{cases}$
        Simulate data using the generative model $\boldsymbol{y}^* \sim f(\boldsymbol{y}|\boldsymbol{\theta}^*)$.
        Calculate the summary statistic(s) of the simulated data $S(\boldsymbol{y}^*)$.
        if $||(S(\boldsymbol{y}^*) - S(\boldsymbol{y}))|| < \varepsilon_t$ then
            accept the parameters values $\boldsymbol{\theta}_i^t = \boldsymbol{\theta}^*$,
            set $i = i + 1$.
        end if
    end while
    Calculate the importance weights $w_i^t = \pi(\boldsymbol{\theta}_i^t)/q_t(\boldsymbol{\theta}_i^t)$.
    Set $t = t + 1$.
end for
```

Algorithm 4 ABC-SMC with adaptive weighting (Prangle 2017)

```
    Input: Observed summary statistic(s) of the data $S(\boldsymbol{y})$, prior distribution for the parameters $\pi(\boldsymbol{\theta})$, a generative model $f(\boldsymbol{y}|\boldsymbol{\theta})$, the number of parameter acceptances $N_a$, a stopping tolerance $\varepsilon_s$, the quantile value $\tau$.
    Output: Samples $\boldsymbol{\theta}_i$, $i = 1, \ldots, N_a$ from the approximate posterior distributions $\pi_\varepsilon(\boldsymbol{\theta}|S(\boldsymbol{y}))$.
 1: Set $t = 1$, $\varepsilon_1 = \infty$ and $\omega_j^1 = 1$.
 2: while $\varepsilon_t > \varepsilon_s$ do
 3:     Set the number of accepted parameter values $i = 0$.
 4:     while $i < N_a$ do
 5:         Sample $\boldsymbol{\theta}^*$ from
            $q_t(\boldsymbol{\theta}) = \begin{cases} \pi(\boldsymbol{\theta}), & \text{if } t = 1 \text{ or } 2, \\ \sum_{i=1}^{N_a} w_i^{t-1} K_t(\boldsymbol{\theta}|\boldsymbol{\theta}_i^{t-1}) / \sum_{i=1}^{N_a} w_i^{t-1}, & \text{otherwise.} \end{cases}$
 6:         Simulate data using the generative model $\boldsymbol{y}^* \sim f(\boldsymbol{y}|\boldsymbol{\theta}^*)$.
 7:         Calculate the summary statistic(s) of the simulated data $S(\boldsymbol{y}^*)$.
 8:         if $d^t = ||(S(\boldsymbol{y}^*) - S(\boldsymbol{y}))/\omega_j^t|| < \varepsilon_t$ then
 9:             accept and store the parameters values $\boldsymbol{\theta}_i^t = \boldsymbol{\theta}^*$ and set $i = i+1$.
10:         end if
11:     end while
12:     Calculate the MAD, $M_1^t, M_2^t, \ldots$, for each summary statistic in Step 8.
13:     Calculate the distance function weights $\omega_j^{t+1} = M_j^t$.
14:     Calculate the importance weights $w_i^t = \pi(\boldsymbol{\theta}_i^t)/q_t(\boldsymbol{\theta}_i^t)$.
15:     Let $\varepsilon_{t+1}$ be the $\tau$th quantile of the accepted $d^t$ values.
16:     Set $t = t + 1$.
17: end while
```

Appendix C: Additional results

C.1 Exact posterior for a very large *T* value

Figure 13 illustrates the shape of the exact posterior distribution for a value of T much higher than in our simulations in Sect. 4.3.1 and equal to $50{,}000\,$s. Notice that the marginal for α becomes concentrated around the true value of $\alpha = 1$, however the marginal for D becomes essentially diffuse and is not centred around the true value $D = 2$. One of the reasons for this practical unidentifiability of D is related to keeping the number of discretisation points N fixed (which is dictated by keeping the computational costs equal between different simulation configurations). However, as investigated by Michalet (2010) (see Fig. 4 there), for diffusion processes increasing the number of fitting points helps to infer diffusion

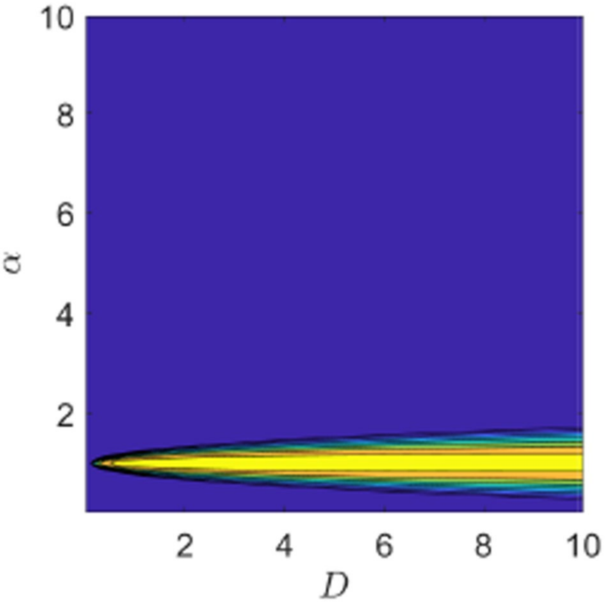


Fig. 13 Contour plots of the exact posterior distribution for $T = 50,000$ s

coefficients to a certain point, beyond which it becomes detrimental to use a higher number of fitting points.

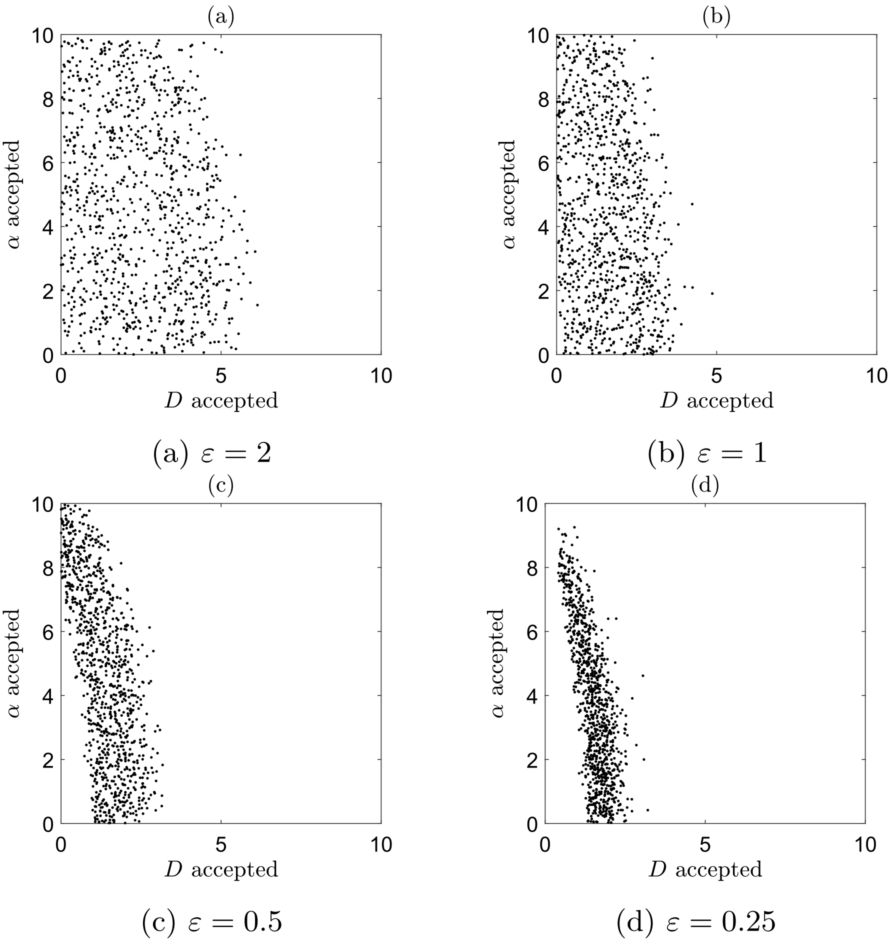
C.2 The impact of the tolerance value on rejection ABC and ABC-SMC

C.2.1 Rejection ABC

Figures 14, 15 and 16 show samples from the joint posterior distributions for $T = 0.05$ s, 5 s and 500 s, respectively. As we decrease ε, the joint posterior distributions get slowly closer to the exact posterior distributions for all values of T. We see a larger improvement for a small and middle value of T, while for a large value of T, the joint posterior distribution in D covers the entire prior distribution.

C.2.2 ABC SMC

See Figs. 17, 18 and 19.



(a) $\varepsilon = 2$

(b) $\varepsilon = 1$

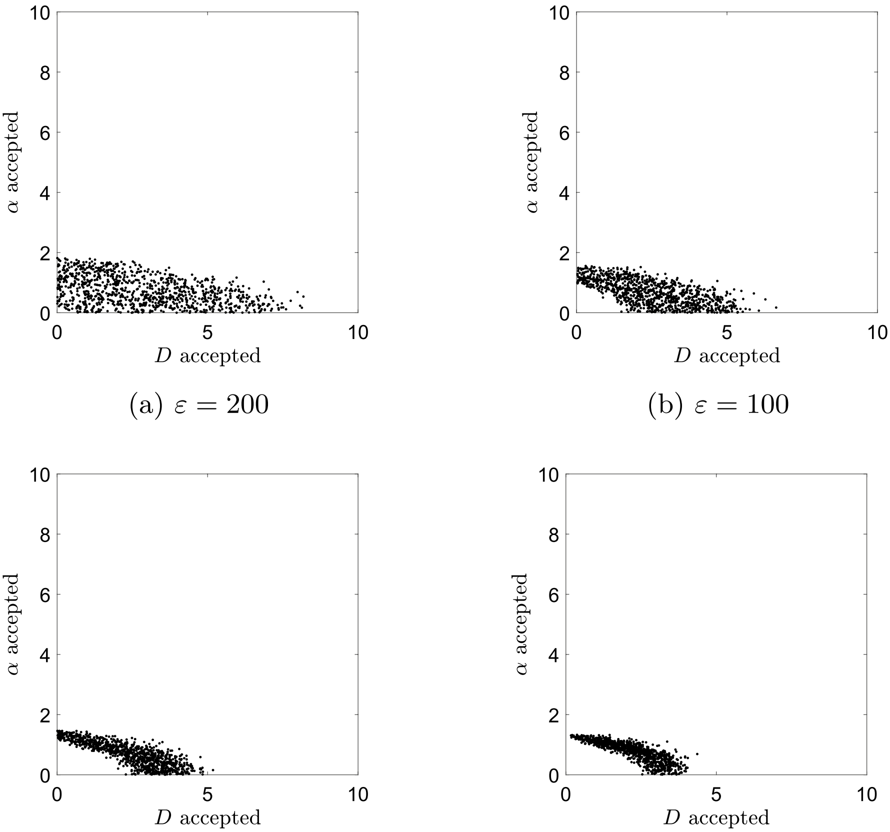
(c) $\varepsilon = 0.5$

(d) $\varepsilon = 0.25$

Fig. 14 Rejection ABC: samples from the ABC-posterior distributions for $T = 0.05$ s and different ε values

Appendix D: Accuracy measures

As discussed in Sect. 4.3.4, for models with practical identification problems, such as SDE (14), one may be interested in assessing how well an ABC algorithm practically identifies each parameter, or both parameters simultaneously. To this end, we need to define accuracy measures that take the ground-true values of parameters into account. In this Section, we first introduce two such measures, based on mean squared errors (MSEs), and then we use them to compare ABC algorithms. We focus on the five ABC algorithms that we have demonstrated to perform relatively well (see Sect. 4.3.1), i.e. ABC-SMC with adaptive weights (Prangle 2017) and four semi-automatic schemes (based on MVLR, GP regression, GP regression with convex hulls, and the residual approach).



(a) $\varepsilon = 200$

(b) $\varepsilon = 100$

(c) $\varepsilon = 50$

(d) $\varepsilon = 25$

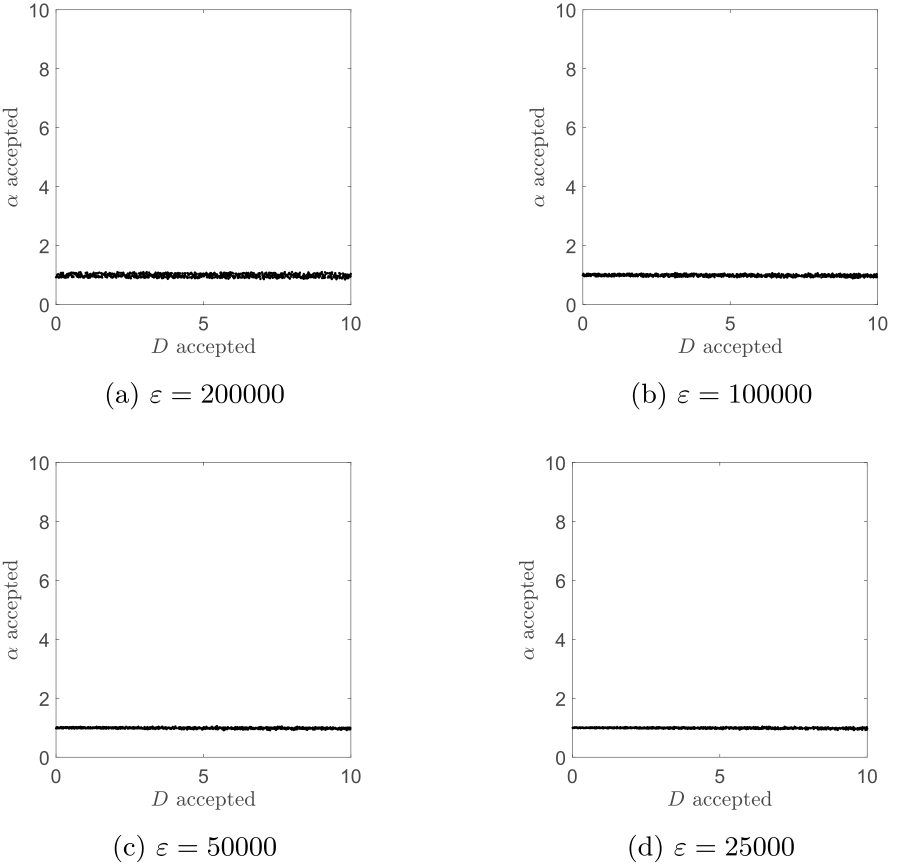
Fig. 15 Rejection ABC: samples from the ABC-posterior distributions for $T = 5$ s and different ε values

D.1 Definitions

Harmonic accuracy

We first introduce the harmonic accuracy that is predominately determined by the change in the strongly (practically) identifiable parameter (which parameter it is depends on the value of T). To define it, we calculate the posterior mean for D and α for the ith tolerance value, denoted by $\overline{D}_i$ and $\overline{\alpha}_i$, respectively, as well as the variance of the accepted values of D and α, denoted by $(\sigma_D^2)_i$ and $(\sigma_\alpha^2)_i$, respectively. The MSE for D and α for the ith tolerance value is then given by

$$(MSE_D)_i = (D - \overline{D}_i)^2 + (\sigma_D^2)_i, \quad i = 1, \dots, N_\varepsilon,$$
$$(MSE_\alpha)_i = (\alpha - \overline{\alpha}_i)^2 + (\sigma_\alpha^2)_i, \quad i = 1, \dots, N_\varepsilon,$$



(a) $\varepsilon = 200000$

(b) $\varepsilon = 100000$

(c) $\varepsilon = 50000$

(d) $\varepsilon = 25000$

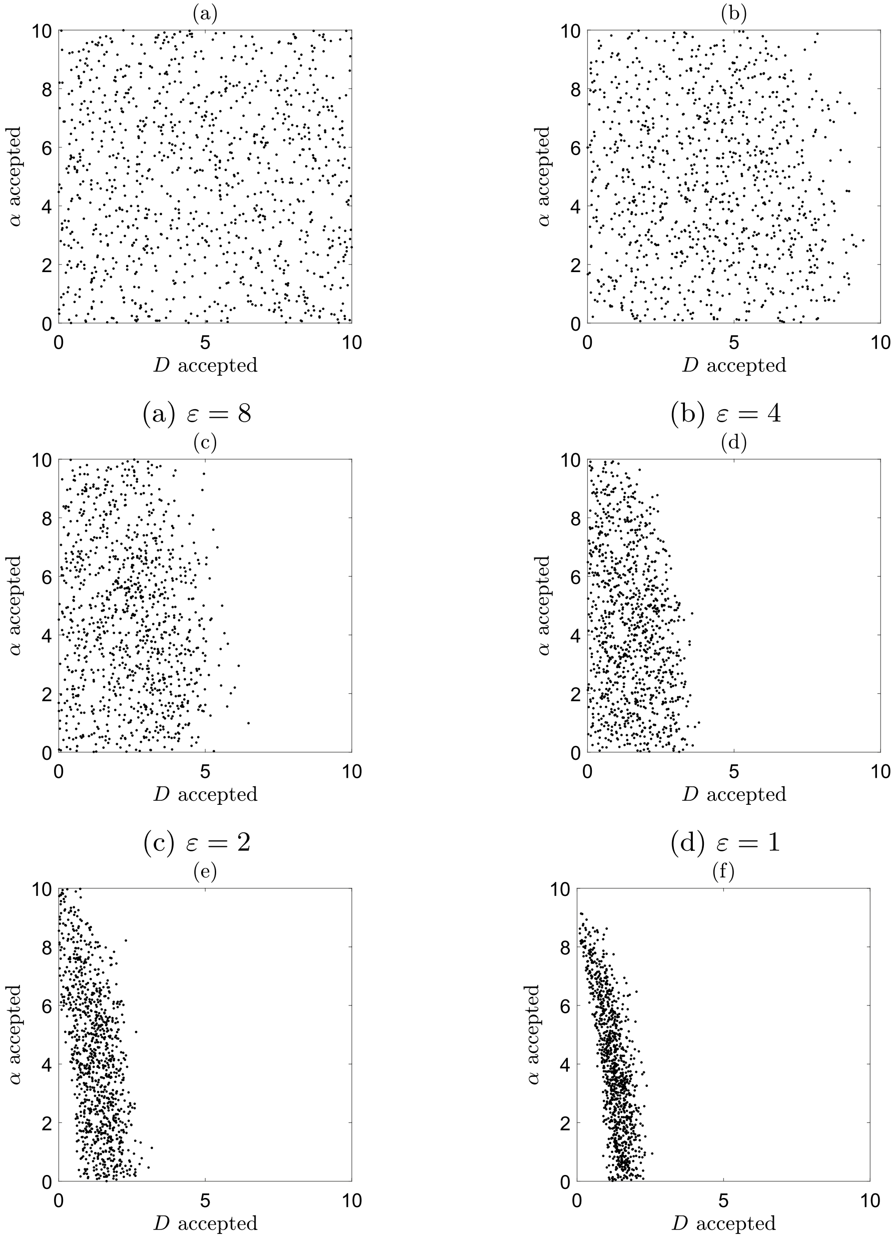
Fig. 16 Rejection ABC: samples from the ABC-posterior distributions for $T = 500$ s and different ε values

where D and α are the true parameter values, and N_ε is the number of tolerance values. The harmonic accuracy for the ith tolerance value is then given by

$$\mathcal{A}_i^{\mathcal{H}} = \frac{1}{(MSE_D)_i} + \frac{1}{(MSE_\alpha)_i}, \quad i = 1, \ldots, N_\varepsilon. \tag{31}$$

Circular accuracy

We will define the second accuracy measure so that a higher level of accuracy corresponds with estimating both parameters well, rather than only the strongly (practically) identifiable parameter. To do this, we need to think about what it means for both parameters to be estimated well. In two dimensions, we could think of this as the shape of the level sets of joint posterior distribution being circular/elliptical around the true parameter values an appropriately small radius.



(a) $\varepsilon = 8$

(b) $\varepsilon = 4$

(c) $\varepsilon = 2$

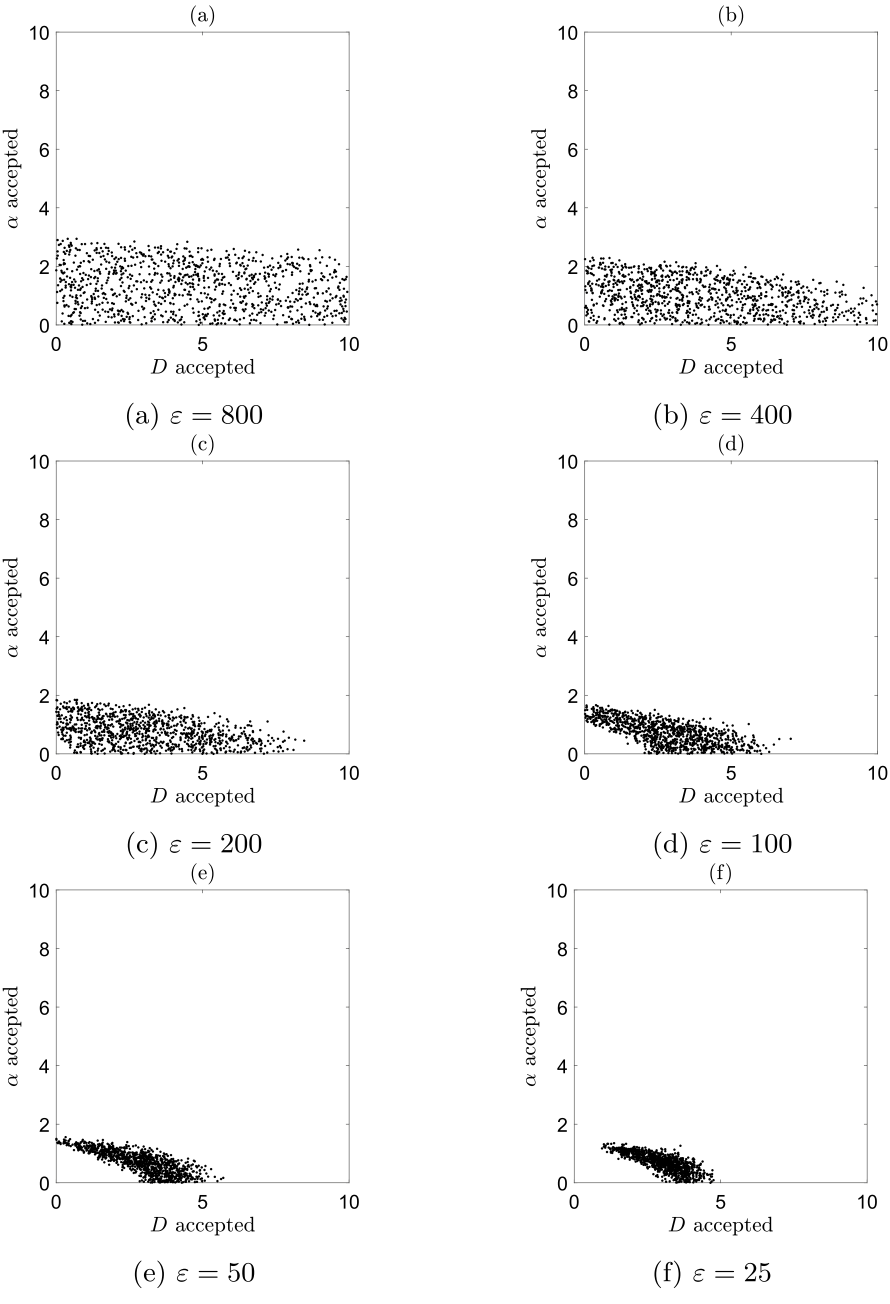
(d) $\varepsilon = 1$

(e) $\varepsilon = 0.5$

(f) $\varepsilon = 0.25$

Fig. 17 ABC-SMC: samples from subsequent intermediary joint distributions for $T = 0.05$ s

To implement this idea, we need a measure of circularity of a general shape. One simple measure would be the compactness measure [?], defined as $C = 4\pi A/p^2$,



(a) $\varepsilon = 800$

(b) $\varepsilon = 400$

(c) $\varepsilon = 200$

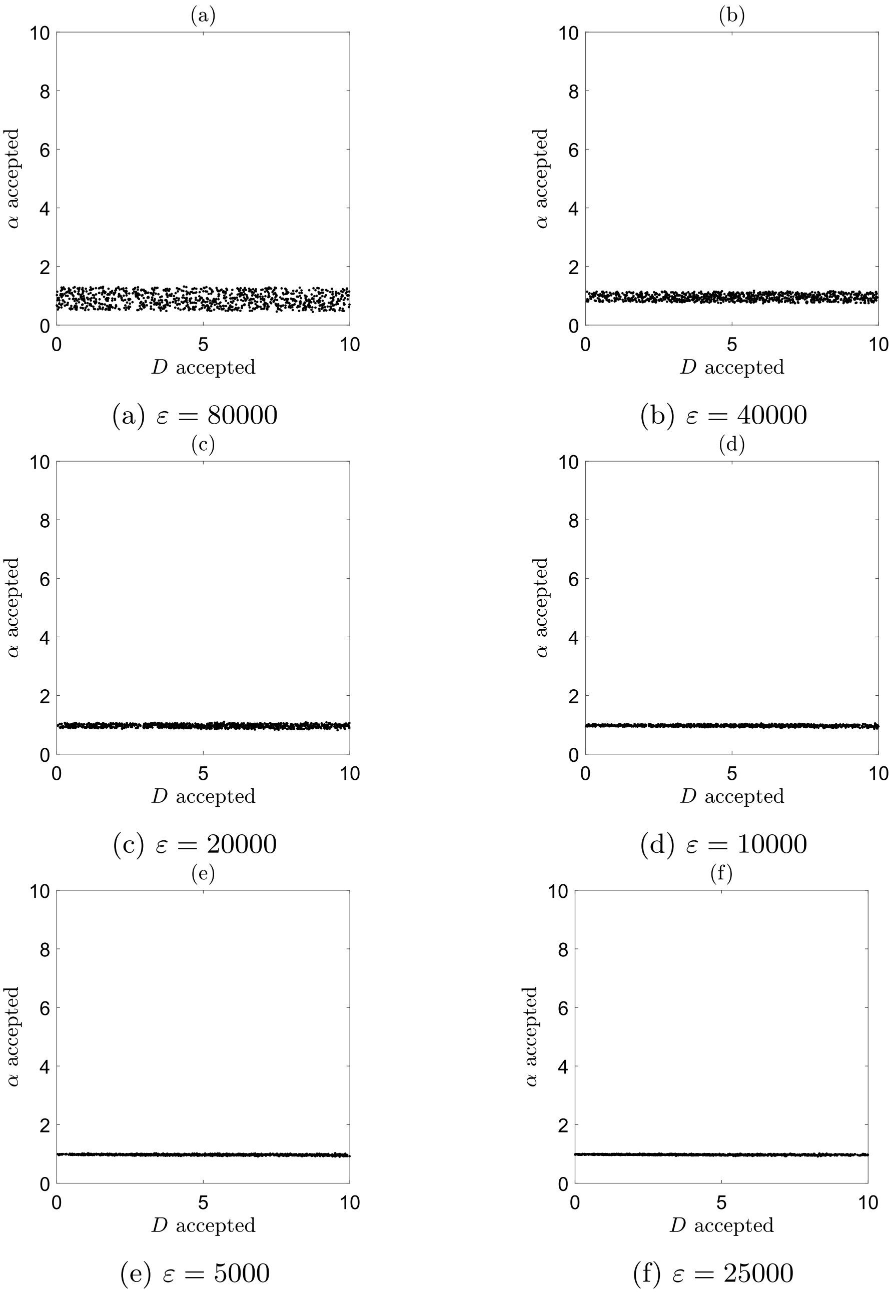
(d) $\varepsilon = 100$

(e) $\varepsilon = 50$

(f) $\varepsilon = 25$

Fig. 18 ABC-SMC: samples from subsequent intermediary joint distributions for $T = 5$ s

where A is the area enclosed and p is the perimeter of the shape. If the shape is a circle then $C = 1$ and $C < 1$ otherwise.



(a) $\varepsilon = 80000$

(b) $\varepsilon = 40000$

(c) $\varepsilon = 20000$

(d) $\varepsilon = 10000$

(e) $\varepsilon = 5000$

(f) $\varepsilon = 25000$

Fig. 19 ABC-SMC: samples from subsequent intermediary joint distributions for $T = 500$ s

For posteriors roughly rectangular in level set plots, we set $A = MSE_D \times MSE_\alpha$ and $p = 2(MSE_D + MSE_\alpha)$. This gives an approximate value of

$C \approx \pi(MSE_D \times MSE_\alpha)/(MSE_D + MSE_\alpha)^2$. Since C is maximised when the area is close to a circle, dividing by C puts more emphasis on estimating both parameters well, rather than just one of them. A possible error measure could then be specified as $\sqrt{A}/C \approx (MSE_D + MSE_\alpha)^2/\sqrt{MSE_D \times MSE_\alpha}$. We then define our second accuracy measure, which we call the circular accuracy, as the inverse of the error, giving

$$\mathcal{A}_i^C = \frac{\sqrt{(MSE_D)_i \times (MSE_\alpha)_i}}{((MSE_D)_i + (MSE_\alpha)_i)^2}, \quad i = 1, \dots, N_\varepsilon. \tag{32}$$

Note that this measure of accuracy could be extended to higher dimensional parameter spaces by using higher-dimensional analogues for the area and perimeter.

D.2 Results

A comparison of the harmonic and circular accuracy against the simulation count for ABC-SMC with adaptive weights and the four semi-automatic ABC algorithms (semi-automatic ABC with MVLR, GP-ABC, GP-ABC with convex hulls, and the residual approach) is given in Fig. 20.

Small *T*

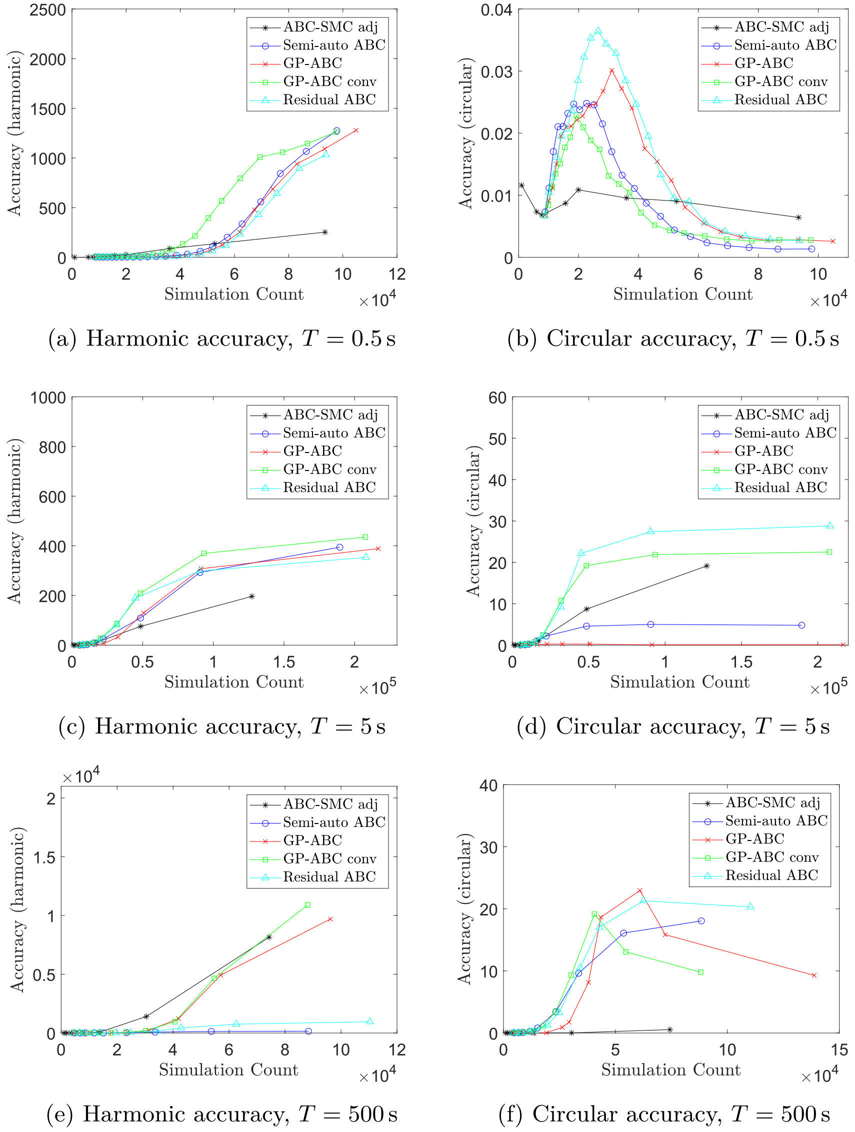
For $T = 0.05$ s, the harmonic accuracy for all the semi-automatic approaches is comparable, with GP-ABC with convex hulls achieving the highest value the quickest. The semi-automatic approaches considerably outperform ABC-SMC with adaptive weight, which corresponds to the narrower ABC-posterior distribution in the practically identifiable parameter D for the former compared to the latter (see Fig. 5).

As for the circular accuracy, for all the algorithms it initially increase as both parameters are inferred more accurately, but then falls as the joint posterior distributions remain wide in α. ABC-SMC with adaptive weights performs best in this case, as the corresponding ABC-posterior distribution is most circular. Semi-automatic ABC with MVLR results in the lowest circular accuracy, which is due the ABC-posterior of α being the most dispersed among those for the four semi-automatic algorithms (see Table 4).

Medium *T*

For $T = 5$ s, the harmonic accuracy exhibits a similar patter as for $T = 0.05$ s, with GP-ABC with convex hulls performing the best and ABC-SMC with adaptive weight performing worst.

This time, however, the circular accuracy monotonically increases in the simulation count, which corresponds to the ABC-posterior distributions becoming more and more isotropic. The residual approach results in the highest circular accuracy, with GP-ABC with convex hulls coming second. The two standard semi-automatic algorithms perform relatively poor in this case.



(a) Harmonic accuracy, $T = 0.5\,\mathrm{s}$

(b) Circular accuracy, $T = 0.5\,\mathrm{s}$

(c) Harmonic accuracy, $T = 5\,\mathrm{s}$

(d) Circular accuracy, $T = 5\,\mathrm{s}$

(e) Harmonic accuracy, $T = 500\,\mathrm{s}$

(f) Circular accuracy, $T = 500\,\mathrm{s}$

Fig. 20 Accuracies against the simulation count for ABC-SMC with adaptive weights (black line with asterisks), semi-automatic ABC with MVLR (blue line with circles), GP-ABC (red line with crosses), GP-ABC with convex hulls (green line with squares) and residual ABC (cyan line with triangles) for the three different values of T

Large *T*

For $T = 500\,\mathrm{s}$, GP-ABC with convex hulls again delivers the highest harmonic accuracy, with the standard GP-ABC being second-best. Interestingly, ABC-SMC with adaptive weights works relatively well in this case. This is due to the ABC-posterior for this method being relatively tight in α, especially compared with semi-automatic ABC with MVLR (also see Table 4).

As far as the circular accuracy is concerned, the semi-automatic approaches clearly outperform ABC-SMC with adaptive weights. Among the former, the residual approach results in the highest circular accuracy overall, which reflects the most isotropic ABC-posterior distribution for this algorithm.

Conclusions

We find that the two semi-automatic schemes proposed in this paper (Sect. 3.4) result in the highest accuracy measures for the three values of T. When it comes to accurately inferring the practically identifiable parameter, GP-ABC with convex hulls performs best, as quantified by the highest values of the harmonic accuracy for this algorithm. As for inferring both parameters simultaneously, the residual approach is generally superior, with the largest circular accuracy values (except for long simulations for $T = 0.05$ s, where it becomes overtaken by the ABC-SMC algorithms with adaptive weights).

Acknowledgements This work was funded by CRUK, Grant No. C22713/A21462, and the UK Engineering and Physical Sciences Research Council (EPSRC) Grant No. EP/N014642/1. AB and DH acknowledge funding from EPSRC, Grant No. EP/R018634/1 and EP/T017899/1.

References

Babtie AC, Stumpf MP (2017) How to deal with parameters for whole-cell modelling. J R Soc Interface 14(133):20170237

Bardini R, Politano G, Benso A, Di Carlo S (2017) Multi-level and hybrid modelling approaches for systems biology. Comput Struct Biotechnol J 15:396–402

Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. Annu Rev Ecol Evol Syst 41:379–406

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

Beaumont MA, Cornuet J, Marin J, Robert CP (2009) Adaptive approximate Bayesian computation. Biometrika 96(4):983–990

Bernton E, Jacob PE, Gerber M, Robert CP (2019) Approximate Bayesian computation with the Wasserstein distance. J Royal Stat Soc: Series B (Stat Methodol) 81(2):235–269

Bhat NG, Balaji S (2020) Whole-cell modeling and simulation: a brief survey. N Gener Comput 38(1):259–281

Borowska A, Giurghita D, Husmeier D (2021) Gaussian process enhanced semi-automatic approximate Bayesian computation: parameter inference in a stochastic differential equation system for chemotaxis. J Comput Phys 429:109999

Bosgraaf L, van Haastert PJM (2009) The ordered extension of pseudopodia by Amoeboid cells in the absence of external cues. PLoS ONE 4(4):5253

Brynjarsdóttir J, O'Hagan A (2014) Learning about physical parameters: the importance of model discrepancy. Inverse Prob 30(11):114007

Buckwar E, Tamborrino M, Tubikanec I (2020) Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs. Stat Comput 30(3):627–648

Codling EA, Plank MJ, Benhamou S (2008) Random walk models in biology. J R Soc Interface 5(25):813–834

Conti S, Gosling JP, Oakley JE, O'Hagan A (2009) Gaussian process emulation of dynamic computer codes. Biometrika 96(3):663–676

De Oliveira S, Rosowski EE, Huttenlocher A (2016) Neutrophil migration in infection and wound repair: going forward in reverse. Nat Rev Immunol 16(6):378–391

Devlin J, Husmeier D, Mackenzie JA (2019) Optimal estimation of drift and diffusion coefficients in the presence of static localization error. Phys Rev E 100:022134

Dickinson RB, Tranquillo RT (1993) Potimal estimation of cell movement indices from the statistical analysis of cell tracking data. AIChE J 39(12):1995–2010

Douc R, Guillin A, Marin J-M, Robert CP (2007) Convergence of adaptive mixtures of importance sampling schemes. Ann Stat 35(1):420–448

Drovandi CC, Pettitt AN (2011) Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. Biometrics 67(1):225–233

Drovandi CC, Pettitt AN, McCutchan RA (2016) Exact and approximate Bayesian inference for low integer-valued time series models with intractable likelihoods. Bayesian Anal 11(2):325–352

Evans M, Hastings N, Peacock B (2000) Ch40 in statistical distributions. Wiley, New York, pp 187–188

Fearnhead P, Prangle D (2012) Constructing summary statistics for approximate Bayesian computation: semi-automatic approximate Bayesian computation. J Royal Stat Soc: Series B (Stat Methodolo) 74(3):419–474

Ferguson EA, Matthiopoulos J, Insall RH, Husmeier D (2016) Inference of the drivers of collective movement in two cell types: dictyostelium and melanoma. J R Soc Interface 13(123):20160695

Ferguson EA, Matthiopoulos J, Insall RH, Husmeier D (2017) Statistical inference of the mechanisms driving collective cell movement. J Roy Stat Soc: Ser C (Appl Stat) 66(4):869–890

Gelman A, Carlin JB, Stern HS, Dunson DB, Vhtari A, Rubin DB (2014) Bayesian Data Analysis, 3rd edn. CRC Press, Boca Raton, Florida, USA

Giurghita D, Husmeier D (2018) Statistical modelling of cell movement. Stat Neerl 72(3):265–280

Harrison JU, Yates CA (2016) A hybrid algorithm for coupling partial differential equation and compartment-based dynamics. J R Soc Interface 13(122):20160335

Hillen T, Painter KJ (2009) A user's guide to PDE models for chemotaxis. J Math Biol 58(1):183–217

Hori SS, Tong L, Swaminathan S, Liebersbach M, Wang J, Gambhir SS, Felsher DW (2021) A mathematical model of tumor regression and recurrence after therapeutic oncogene inactivation. Sci Rep 11(1):1–14

Hu B, Fuller D, Loomis WF, Levine H, Rappel W-J (2010) Phenomenological approach to eukaryotic chemotactic efficiency. Phys Rev E 81(3):031906

Kakebeeke PI, de Wit RJ, Konijn TM (1980) Folic acid deaminase activity during development in Dictyostelium discoideum. J Bacteriol 143(1):307–312

Kalimuthu P, John SA (2009) Selective electrochemical sensor for folic acid at physiological pH using ultrathin electropolymerized film of functionalized thiadiazole modified glassy carbon electrode. Biosens Bioelectron 24(12):3575–3580

Kypraios T, Neal P, Prangle D (2017) A tutorial introduction to Bayesian inference for stochastic epidemic models using Approximate Bayesian Computation. Math Biosci 287:42–53

Lambert B, MacLean AL, Fletcher AG, Combes AN, Little MH, Byrne HM (2018) Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. J Math Biol 76(7):1673–1697

Li L, Nørrelykke SF, Cox EC (2008) Persistent cell motion in the absence of external signals: a search strategy for Eukaryotic cells. PLoS ONE 3(5):2093

Li L, He Y, Zhao M, Jiang J (2013) Collective cell migration: implications for wound healing and cancer invasion. Burns Trauma 1(1):2321–3868

Lillacci G, Khammash M (2010) Parameter estimation and model selection in computational biology. PLoS Comput Biol 6(3):1000696

Lintusaari J, Gutmann MU, Dutta R, Kaski S, Corander J (2017) Fundamentals and Recent developments in Approximate Bayesian computation. Syst Biol 66(1):66–82

Liu JS (2001) Monte Carlo Strategies in Scientific Computing. Springer, New York City, New York, USA

Makowski D, Ben-Shachar MS, Lüdecke D (2019) bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. J Open Sour Softw 4(40):1541

Martin GM, McCabe BP, Frazier DT, Maneesoonthorn W, Robert CP (2019) Auxiliary likelihood-Based approximate Bayesian computation in state space models. J Comput Graph Stat 28(3):508–522

Maybank P, Bojak I, Everitt RG (2017) Fast approximate Bayesian inference for stable differential equation models. arXiv preprint arXiv:1706.00689

McKinley TJ, Vernon I, Andrianakis I, McCreesh N, Oakley JE, Nsubuga RN, Goldstein M, White RG (2018) Approximate Bayesian computation and simulation-based inference for complex stochastic epidemic models. Stat Sci 33(1):4–18

Michalet X (2010) Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. Phys Rev E 82(4):041914

Moores MT, Drovandi CC, Mengersen K, Robert CP (2015) Pre-processing for approximate Bayesian computation in image analysis. Stat Comput 25(1):23–33

Motta S, Pappalardo F (2013) Mathematical modeling of biological systems. Brief Bioinform 14(4):411–422

Neal RM (2012) Bayesian Learning for Neural Networks, 118th edn. Lecture Notes in Statistics. Springer, New York City, New York, USA

Osborne JM, Walter A, Kershaw S, Mirams G, Fletcher A, Pathmanathan P, Gavaghan D, Jensen O, Maini P, Byrne H (2010) A hybrid approach to multi-scale modelling of cancer. Philos Trans Royal Soc A Math Phys Eng Sci 368(1930):5013–5028

Picchini U (2014) Inference for SDE models via approximate Bayesian computation. J Comput Graph Stat 23(4):1080–1100

Picchini U, Forman JL (2016) Accelerating inference for diffusions observed with measurement error and large sample sizes using approximate Bayesian computation. J Stat Comput Simul 86(1):195–213

Picchini U, Samson A (2018) Coupling stochastic EM and approximate Bayesian computation for parameter inference in state-space models. Comput Stat 33(1):179–212

Prangle D (2017) Adapting the ABC distance function. Bayesian Anal 12(1):289–309

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. Mol Biol Evol 16(12):1791–1798

Pullen N, Morris RJ (2014) Bayesian model comparison and parameter inference in systems biology using nested sampling. PLoS ONE 9(2):88419

Purcell O, Jain B, Karr JR, Covert MW, Lu TK (2013) Towards a whole-cell modeling approach for synthetic biology. Chaos Interdiscip J Nonlinear Sci 23(2):025112

Qian H, Sheetz MP, Elson EL (1991) Single particle tracking. Analysis of diffusion and flow in two-dimensional systems. Biophys J 60(4):910–921

Rasmussen CE, Williams CKI (2006) Gaussian Processes for Machine Learning. MIT Press, Cambridge, Massachusetts, USA

Rivero F, Köppel B, Peracino B, Bozzaro S, Siegert F, Weijer CJ, Schleicher M, Albrecht R, Noegel AA (1996) The role of the cortical cytoskeleton: F-actin crosslinking proteins protect against osmotic stress, ensure cell size, cell shape and motility, and contribute to phagocytosis and development. J Cell Sci 109(11):2679–2691

Roussos ET, Condeelis JS, Patsialou A (2011) Chemotaxis in cancer. Nat Rev Cancer 11(8):573–587

Savin T, Doyle PS (2005) Static and dynamic errors in particle tracking microrheology. Biophys J 88(1):623–638

Saxton MJ (1997) Single-particle tracking: the distribution of diffusion coefficients. Biophys J 72(4):1744–1753

Saxton MJ, Jacobson K (1997) Single-particle tracking: applications to membrane dynamics. Annu Rev Biophys Biomol Struct 26(1):373–399

Scarpa E, Mayor R (2016) Collective cell migration in development. J Cell Biol 212(2):143–155

Secrier M, Toni T, Stumpf MP (2009) The abc of reverse engineering biological signalling systems. Mol BioSyst 5(12):1925–1935

Segel LA (1977) A theoretical study of receptor mechanisms in bacterial chemotaxis. SIAM J Appl Math 32(3):653–665

Shi C, Huang C-H, Devreotes PN, Iglesias PA (2013) Interaction of motility, directional sensing, and polarity modules recreates the behaviors of chemotaxing cells. PLoS Comput Biol 9(7):1003122

Sisson SA, Fan Y, Tanaka MM (2007) Sequential Monte Carlo without likelihoods. Proc Natl Acad Sci 104(6):1760–1765

Sisson SA, Fan Y, Beaumont M (2018) Handbook of Approximate Bayesian Computation. Chapman & Hall/CRC Press, London, UK

Sousa VC, Fritz M, Beaumont MA, Chikhi L (2009) Approximate Bayesian computation without summary statistics: the case of admixture. Genetics 181(4):1507–1519

Spill F, Guerrero P, Alarcon T, Maini PK, Byrne H (2015) Hybrid approaches for multiple-species stochastic reaction-diffusion models. J Comput Phys 299:429–445

Stuelten CH, Parent CA, Montell DJ (2018) Cell motility in cancer invasion and metastasis: insights from simple model organisms. Nat Rev Cancer 18(5):296–312

Sun L, Lee C, Hoeting JA (2015) Parameter inference and model selection in deterministic and stochastic dynamical models via approximate Bayesian computation: modeling a wildlife epidemic. Environmetrics 26(7):451–462

Tancredi A (2019) Approximate Bayesian inference for discretely observed continuous-time multi-state models. Biometrics 75(3):966–977

Tang M, Wang M, Shi C, Iglesias PA, Devreotes PN, Huang C-H (2014) Evolutionarily conserved coupling of adaptive and excitable networks mediates eukaryotic chemotaxis. Nat Commun 5(1):1–13

Tavaré S, Balding DJ, Griffiths RC, Donnelly P (1997) Inferring coalescence times from DNA sequence data. Genetics 145(2):505–518

Tomlin CJ, Axelrod JD (2007) Biology by numbers: mathematical modelling in developmental biology. Nat Rev Genet 8(5):331–340

Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP (2009) Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J R Soc Interface 6(31):187–202

Tweedy L, Knecht DA, Mackay GM, Insall RH (2016) Self-generated chemoattractant gradients: attractant depletion extends the range and robustness of chemotaxis. PLoS Biol 14(3):1002404

Tyson R, Lubkin SR, Murray JD (1999) A minimal mechanism for bacterial pattern formation. Proc R Soc Lond B 266(1416):299–304

Wilkinson DJ (2007) Bayesian methods in bioinformatics and computational systems biology. Brief Bioinform 8(2):109–116

Wood SN (2010) Statistical inference for noisy nonlinear ecological dynamic systems. Nature 466(7310):1102–1104

Wurster B, Butz U (1980) Reversible binding of the chemoattractant folic acid to cells of Dictyostelium discoideum. Eur J Biochem 109(2):613–618

Zhu W, Marin JM, Leisen F (2016) A bootstrap likelihood approach to Bayesian computation. Aust New Zealand J Stat 58(2):227–244

Authors and Affiliations

Jon Devlin[1] · Agnieszka Borowska[2] · Dirk Husmeier[2] · John Mackenzie[1]

✉ John Mackenzie
j.a.mackenzie@strath.ac.uk

Jon Devlin
jon.devlin@strath.ac.uk

Agnieszka Borowska
agnieszka.borowska@glasgow.ac.uk

Dirk Husmeier
dirk.husmeier@glasgow.ac.uk

[1] Department of Mathematics and Statistics, University of Strathclyde, Glasgow, UK

[2] School of Mathematics and Statistics, University of Glasgow, Glasgow, UK