Mathematical Biosciences 263 (2015) 133–142

Contents lists available at ScienceDirect

Mathematical Biosciences

journal homepage: www.elsevier.com/locate/mbs

Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation

Brenda N. Vo[a], Christopher C. Drovandi[a], Anthony N. Pettitt[a], Matthew J. Simpson[a,b,*]

[a] Mathematical Sciences, Queensland University of Technology (QUT), Brisbane, Queensland 4001, Australia
[b] Institute of Health and Biomedical Innovation, QUT, Brisbane, Australia

ARTICLE INFO

Article history:
Received 15 August 2014
Revised 9 January 2015
Accepted 25 February 2015
Available online 5 March 2015

Keywords:
Approximate Bayesian computation
Cell diffusivity
Cell proliferation
Random walk model
Collective cell spreading

ABSTRACT

Wound healing and tumour growth involve collective cell spreading, which is driven by individual motility and proliferation events within a population of cells. Mathematical models are often used to interpret experimental data and to estimate the parameters so that predictions can be made. Existing methods for parameter estimation typically assume that these parameters are constants and often ignore any uncertainty in the estimated values. We use approximate Bayesian computation (ABC) to estimate the cell diffusivity, D, and the cell proliferation rate, λ, from a discrete model of collective cell spreading, and we quantify the uncertainty associated with these estimates using Bayesian inference. We use a detailed experimental data set describing the collective cell spreading of 3T3 fibroblast cells. The ABC analysis is conducted for different combinations of initial cell densities and experimental times in two separate scenarios: (i) where collective cell spreading is driven by cell motility alone, and (ii) where collective cell spreading is driven by combined cell motility and cell proliferation. We find that D can be estimated precisely, with a small coefficient of variation (CV) of 2–6%. Our results indicate that D appears to depend on the experimental time, which is a feature that has been previously overlooked. Assuming that the values of D are the same in both experimental scenarios, we use the information about D from the first experimental scenario to obtain reasonably precise estimates of λ, with a CV between 4 and 12%. Our estimates of D and λ are consistent with previously reported values; however, our method is based on a straightforward measurement of the position of the leading edge whereas previous approaches have involved expensive cell counting techniques. Additional insights gained using a fully Bayesian approach justify the computational cost, especially since it allows us to accommodate information from different experiments in a principled way.

1. Introduction

Cell motility and cell proliferation play an important role in collective cell spreading, which is critical to many key biological processes, including wound healing [1–3] and tumour growth [4,5]. To improve our understanding of, and our ability to predict these processes, mathematical models have been formulated and calibrated using experimental data collected from various types of cell biology experiments. Modelling approaches that describe collective cell spreading can be grouped into two main categories: (i) continuum models, and (ii) discrete models [6]. Continuum models describe the behavior of populations of cells using reaction-diffusion equations, such as the generalized Fisher–Kolmogorov equation [6]:

$$\frac{\partial C}{\partial t} = D\nabla^2 C + \lambda C\left(1 - \frac{C}{K}\right), \tag{1}$$

where C is the cell density, D is the diffusion coefficient, λ is the intrinsic proliferation rate and K is the carrying capacity density. Continuum models provide insight into the relationships among the model parameters [2] and are relatively straightforward to solve numerically.

Discrete, individual-based models have been used for modelling biological systems in different contexts, including lattice-based and lattice-free frameworks [7–10]. Discrete models usually involve discretizing time and space, and they represent the behavior of individual cells by some kind of stochastic process. These models have several advantages: first, they are able to incorporate several important biological features such as heterogeneity [11] and fluctuations [12]; second, they produce image-based and movie-based information that is easy to compare with experimental data [13]. However, the likelihood function for discrete models is not available in an

* Corresponding author at: Discipline of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland 4001, Australia. Tel.: +617 3138 5241; fax: +617 3138 2310.
E-mail address: matthew.simpson@qut.edu.au (M.J. Simpson).

http://dx.doi.org/10.1016/j.mbs.2015.02.010
0025-5564/

analytical form, so a standard statistical inference for these models is challenging.

Previous approaches for estimating D and λ in Equation (1) include methods such as trial and error to find a best fit [14] and non-linear least squares estimation [2,15]. These approaches are limited in that they provide point estimates, and the uncertainty in the estimate is not always quantified. Sengers et al. [16] fitted solutions from the generalized Fisher–Kolmogorov equation to experimental cell density profiles to provide estimates of D and λ. Unfortunately, this approach is both experimentally and computationally expensive since it requires a cell labelling and cell counting technique to construct approximate cell density profiles.

Here, we propose an approximate Bayesian computation (ABC) approach to estimate parameters governing the spreading of a population of fibroblast cells (3T3 cells) in a circular barrier assay, using a discrete model. ABC does not require the specification of a likelihood function and the Bayesian approach allows us to incorporate information from multiple experiments. The ABC approach has been successfully applied in a wide range of biological problems [17–20]. It is possible that the Bayesian version of the parametric bootstrap method [21] may be an alternative to ABC for estimating posterior distributions but its properties are yet to be considered in the context of intractable likelihoods.

The aim of the ABC analysis presented here is to obtain approximate posterior distributions for D and λ, for three different levels of initial cell density in the barrier assays, and at three different experimental time periods. Through the ABC technique, the associated uncertainty in the parameter values is quantified and interpreted in terms of the coefficient of variation and probability intervals of the posterior distribution. For all cases, we choose the number of model simulations to be sufficiently large such that the Monte Carlo standard error (MCSE) of the target posterior distribution is at an acceptable level, less than 1% of its associated posterior mean [22].

The findings from this study emphasize the importance of considering the initial cell density and the experimental time when estimating D and λ. We are unaware of any previous study that compares estimates of D and λ for different experimental termination times. Therefore, this study is the first attempt to demonstrate that the estimated values of D may differ by a factor of two or more, depending on the experimental termination time and the initial number of cells. These results suggest that a more complicated model might be warranted, however this conclusion is not obvious without first exploring the suitability of the standard model that we consider. Furthermore, in an attempt to make our method as practical as possible, all parameter estimation is based solely on measuring the position of the leading edge since this technique is simple, inexpensive and can be applied retrospectively to previously published results [23,24].

This paper is organized as follows. Section 2.1 describes the data available for analysis and the leading edge algorithm. The discrete model is described in Section 2.2, and the ABC algorithm is given in Section 2.3. Results are presented in Section 3 and the discussion is given in Section 4.

2. Material and methods

2.1. Image analysis

The data for analysis consist of images showing the entire spreading populations at three particular time points [25]. Two different experimental scenarios are considered: (i) collective cell spreading driven by cell motility alone, where Mitomycin-C was applied to inhibit cell proliferation, and (ii) collective cell spreading driven by combined cell motility and cell proliferation [25]. For both scenarios, experiments were performed for three different initial numbers of cells inside the barrier (5000, 10,000 and 30,000 cells). For each initial cell density, experiments were terminated at three different times (24, 48 and 72 h). Furthermore, each experiment, for each initial cell density, and each termination time, was repeated three times. Thus, a $2 \times 3 \times 3$ balanced experimental design was conducted with three replicates, producing a total of 54 experimental images.

All experiments were performed in 24-well tissue culture plates, each with a diameter of 15.6 mm. To initiate each experiment, either 5000, 10,000 or 30,000 cells were approximately evenly distributed within a circular barrier, of diameter 6.0 mm, located at the centre of the well. After the barrier was lifted, images of the spreading population were recorded at 24, 48 or 72 h. All images were analysed using the Matlab Image Processing Toolbox [26]. The following steps were used to detect the position of the leading edge, and to compute the area enclosed by the leading edge:

Step 1. *Read and display image:* by function "imread" and "imshow" (Fig. 1a).

Step 2. *Detect the segmented edge:* This step enhances the contrast between the area occupied by the spreading population from the background area. First, the image was converted from violet to grayscale, "rgb2gray". A threshold value was calculated by "edge" and the Sobel operator ([∼, threshold]=edge (image, 'sobel')), then tuned and passed to "edge" again to obtain the binary gradient mask containing the segmented cell spreading area (f = edge (image, 'sobel', threshold×0.5)) (Fig. 1b).

Step 3. *Dilate the image:* The binary gradient image was dilated using linear structuring elements, functions "strel" and "imdilate", to trace the outline of the cell population (Fig. 1c).

Step 4. *Fill holes and remove unconnected objects:* The dilated gradient image showed the outline quite clearly, so the function "imfill" was used to fill any remaining holes inside the image (Fig. 1d).

Step 5. *Smooth and filter the object:* All objects not connected to the edge were removed by "imclearborder". The image was then eroded twice with a diamond structuring element to make the segmented object look smooth and natural. In these steps, only the cell spreading area is needed, that is the largest object, other small objects were removed by "bwareopen" (Fig. 1e).

Step 6. *Place an outline over the original image:* This was done using "bwperim" (Fig. 1f).

All images of the cell spreading process show that the population maintains an approximately circular shape for all times considered. Therefore, to quantify the degree of spreading we convert the area enclosed by the leading edge, A, into an equivalent circular diameter, $d = \sqrt{4A/\pi}$ (Supplementary material). We always ensured that we applied the exact same edge detection algorithm to both our experimental data and the images produced by the discrete random walk model.

2.2. Discrete model

The discrete model is an unbiased random walk that explicitly incorporates cell-to-cell crowding effects [13]. The experimental protocol was carefully designed to ensure that all experiments involved a single monolayer of cells [25]. Therefore, we model these experiments using a two-dimensional random walk on a square lattice, with lattice spacing Δ. For all results we set $\Delta = 25\,\mu$m, which corresponds to the average cell diameter [25]. The parameter $P_m \in [0, 1]$ is the probability that an agent will attempt to step a distance Δ within a time step of duration τ, and $P_p \in [0, 1]$ is the probability that an agent will attempt to proliferate within a time step of duration τ. Both the motility and proliferation mechanisms are unbiased, which means that a motile agent at (x, y), will attempt to step to either $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, with each target site chosen with equal probability. A proliferative agent at (x, y) will attempt to deposit a daughter agent at $(x \pm \Delta, y)$ or $(x, y \pm \Delta)$, with each target site chosen with equal

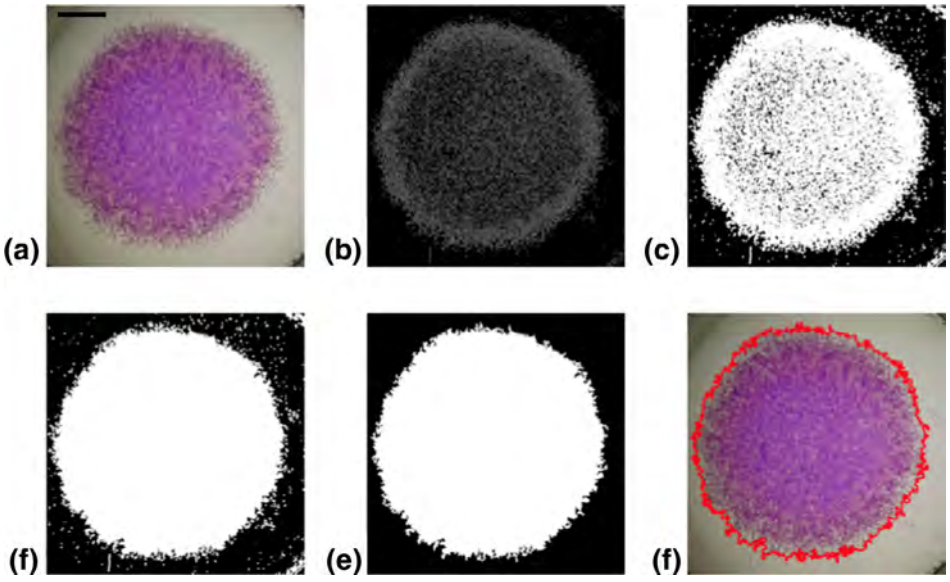

Fig. 1. Six steps in the leading edge detection algorithm for an experimental image corresponding to an experiment initialized with 30,000 cells and terminated after 72 h. Here, cell spreading is driven by combined motility and proliferation. Images show: (a) the original image, (b) the binary gradient mask, (c) the dilated gradient mask, (d) the binary image with fill holes, (e) the smoothed segmented image, (f) the original image with outline. The scale bar in (a) represents 2 mm.

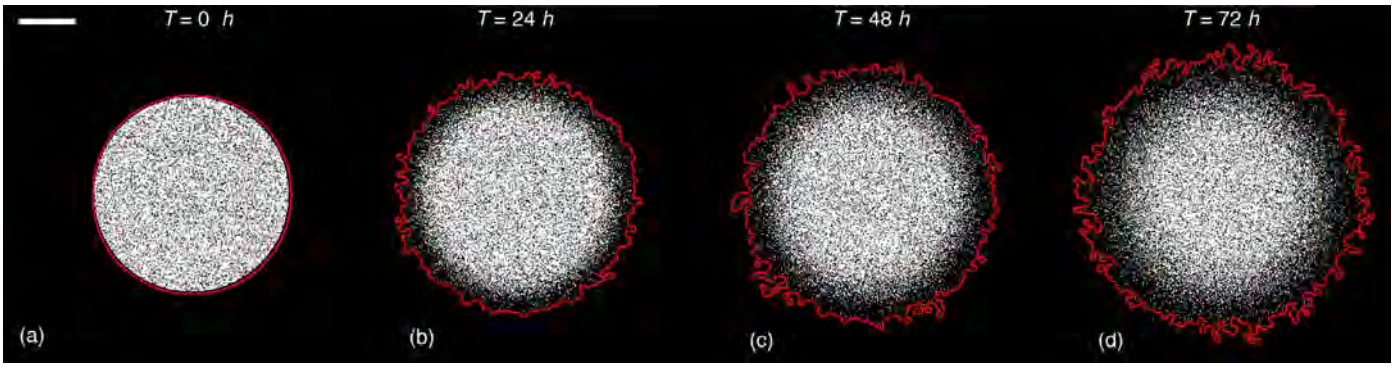

Fig. 2. Four typical realizations of the discrete random walk model with $C(0) = 30{,}000$, $P_m = 1$, $P_p = 0$, $\Delta = 25\ \mu$m and $\tau = 0.06$ h (or $D = 2604\ \mu\text{m}^2\,\text{h}^{-1}$). Images show the distribution of agents after: (a) $T = 0$ h, (b) $T = 24$ h, (c) $T = 48$ h, and (d) $T = 72$ h. The scale bar in (a) represents 2 mm.

probability. The random walk model is an exclusion process, with at most one agent per lattice site. Any attempted motility or proliferation event that would place an agent on an occupied site is aborted. The model does not incorporate any death mechanism since there was no experimental evidence of cell death [25].

The discrete model has four parameters: P_m, P_p, Δ and τ. These parameters are related to D and λ (Eq. (1)) using $D = P_m\Delta^2/(4\tau)$ and $\lambda = P_p/\tau$ [13]. Given that we have access to accurate estimates of Δ [25], we apply our ABC algorithm to obtain posterior distributions of τ and λ. We then convert the posterior distribution of τ into a posterior distribution of D.

To implement the random walk algorithm we let $C(t)$ be the number of agents in the model at time t and T be the termination time. Simulations are performed on a square lattice of size 624×624, so that the width of the lattice corresponds to the diameter of the experimental cell culture well, 15.6 mm (15,600 μm/25 μm = 624). To initialise the simulations, $C(0)$ agents are uniformly distributed inside a circle which has diameter of 240 lattice sites, corresponding to the 6 mm diameter of the circular barrier (6,000 μm/25 μm = 240). For the first experimental scenario, where cell spreading is driven by cell motility alone, $C(t)$ remains constant with time. We use an approximate random sequential update (RSU) algorithm [27] which involves dividing time into small uniform intervals, each of duration τ. To step from time t to time $t + \tau$, $C(t)$ agents are sampled, at random, one at a time, with replacement, and given the opportunity to move with probability P_m. Given the termination time, T, the model requires T/τ time steps. For the second experimental scenario, where cell spreading is driven by cell motility and proliferation, $C(t)$ is non-decreasing and to simulate proliferation we sample $P_p \times C(t)$ agents, at random, one at a time, with replacement, and give each chosen agent the opportunity to proliferate with probability of one. If $P_p \times C(t)$ is not an integer, an additional step is required (Supplementary material). We also checked that our approximate RSU algorithm produced results that were indistinguishable from the more sophisticated, but computationally demanding, exact Gillespie algorithm [28] (results not shown). Results in Fig. 2 show typical snapshots of the simulated collective cell spreading together with the position of the leading edge.

2.3. Approximate Bayesian computation

In the Bayesian framework, model parameters are considered as random variables and the uncertainty about them is updated using observed data. Before the data are collected, information about the model parameters is encoded within prior distributions, which are often based on expert knowledge, previous studies, or both. After observing the data, the information from the prior is updated by the likelihood, to produce the posterior distribution of the parameters. All inferences about the model parameters such as point estimates,

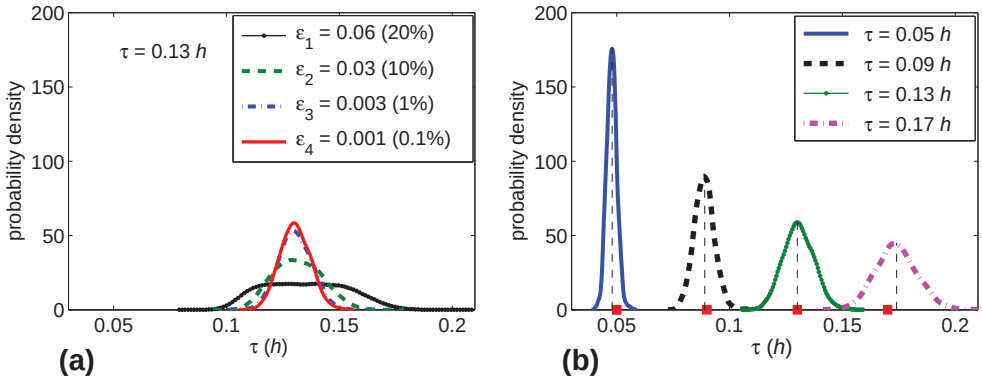

Fig. 3. (a) ABC posterior distributions for τ for different tolerance values, $\epsilon_j, j = 1, \ldots, 4$, corresponding to the 20%, 10%, 1% and 0.1% quantiles, respectively, using the synthetic data set with $\tau = 0.13$ h and a uniform prior for τ, $\tau \sim U(0.01, 0.3)$ h. (b) Comparison of ABC posterior distribution estimates of τ in four separate synthetic data sets with $C(0) = 5000$, $T = 24$ h and $P_p = 0$. Each posterior distribution consists of $N = 1000$ parameter samples and the posterior means are plotted as black vertical dashed lines. In the four posterior distributions, the tolerances are $\epsilon = 0.004, 0.002, 0.001$ and 0.0013, respectively. The true parameter values are plotted as red squares. The horizontal axis corresponds to τ, whereas the vertical axis is an estimate of the dimensionless probability density. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

probability intervals and tests of hypotheses are made from the posterior distributions. Thus, in Bayesian inference, evaluation of the likelihood function is one of the critical steps to computing or estimating the posterior distribution. However, for numerous complex stochastic models arising from the ecological, medical and biological sciences, the likelihood functions are computationally intractable. ABC is a method of inference for such models [17,29].

The focus of this work is to apply an ABC technique to estimate the posterior distribution for τ, which can be used to estimate D, for the first experimental scenario. We then approximate the joint posterior distribution for τ and λ for the second experimental scenario. We use the most straightforward ABC approach, ABC rejection sampling [17]. We could use a more efficient technique [19,29–31]; however, given that this is the first time that ABC has been used to infer the parameter values for a model describing collective cell spreading using leading edge information, it is reasonable to implement the most straightforward ABC approach. ABC rejection sampling also allows simulations to be performed in parallel [32], thus reducing the required computational time.

Let S_{obs} and S_{sim} represent the summary statistics of the observed and the simulated data, $\theta = (\tau, \lambda)$ represent the vector of unknown parameters, $\pi(\theta)$ be the prior distribution and ρ denote the distance function that compares S_{obs} and S_{sim}, $\rho = \rho(S_{\text{obs}}, S_{\text{sim}})$. The ABC sampling algorithm is given by Algorithm 1.

When a sufficiently large number, N, of samples (θ_i, ρ_i) is obtained, the set $(\theta_i, \rho_i)_{i=1}^N$ is sorted via the computed discrepancy ρ such that $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_N$. A set of tolerance values $(\epsilon_1 > \epsilon_2 > \epsilon_3 > \epsilon_4)$ are then computed based on the 20%, 10%, 1% and 0.1% quantiles of the calculated discrepancies, respectively. For each value of $\epsilon_j, j = 1, \ldots, 4$, the ABC posterior sample consists of the set $(\theta_i | \rho_i \leq \epsilon_j)_{i=1}^N$. The selected sub-set forms the posterior distribution for θ. A small value of ϵ_j will force the simulated data to be closer to the observed data, but will also lead to fewer samples being retained, and as such increase the MCSE of estimation from the true ABC posterior [33]. To overcome this we always choose $N = 10^6$ to ensure that there are at least 1000 samples in the target ABC posterior.

Algorithm 1: ABC sampling.

1 Draw $\theta_i \sim \pi(\theta)$;
2 Simulate data as per observed data structure from the model with θ_i;
3 Compute S_{sim_i};
4 Compute $\rho_i = \rho(S_{\text{obs}}, S_{\text{sim}_i})$ and store samples (θ_i, ρ_i);
5 Repeat steps 1, 2, 3, 4 until N samples are simulated.

The ABC analysis was conducted to infer the key parameter values that govern cell spreading experiments for each experimental time and initial cell density combination. For each experiment, at each initial cell density, at a particular termination time, the experimental data include three diameters, d_1, d_2 and d_3, which corresponds to three identically prepared experimental replicates. We considered the mean statistics, $S = (1/3)\sum_{i=1}^3 d_i$, and the distance function, $\rho(S_{\text{obs}}, S_{\text{sim}}) = |S_{\text{obs}} - S_{\text{sim}}|$. We also repeated all analysis using order statistics and found that this gave similar results (not shown). We now test the performance of our ABC algorithm using synthetically generated data for which we know the true parameter values.

3. Results

3.1. Validation with synthetic data

We simulate four independent data sets with $\tau = 0.05, 0.09, 0.13$ and 0.17 h. All synthetically generated experiments are initialized with $C(0) = 5000$ and terminated after $T = 24$ h, where cell spreading is driven by cell motility alone. Each simulated experiment is repeated three times. We apply the ABC algorithm to each synthetic data set separately and we choose a uniform prior for τ, $\tau \sim U(0.01, 0.3)$ h. Since we know, in advance, that previous estimates of D for these experiments are in the range $1300 < D < 3000$ μm^2 h^{-1} [25], we always chose our prior parameter distributions so that our ABC algorithm sampled well-beyond this range. Given values of τ drawn from the prior distribution, we performed $N = 10^6$ discrete model simulations with $P_m = 1$, $P_p = 0$ and $\Delta = 25$ μm.

For the synthetic data set with $\tau = 0.13$ h, we compare the ABC posterior distribution of τ based on four different choices of tolerance, $\epsilon_j, j = 1, \ldots, 4$ in Fig. 3a. We observe that posterior distributions for ϵ_3 and ϵ_4 are almost indistinguishable so we do not reduce the tolerance any further. For other synthetically generated data sets we find that taking the 0.1% quantile of the discrepancies resulted in a sufficiently small tolerance value.

A comparison of the posterior distributions for all four synthetic data sets are shown in Fig. 3b, indicating that all the ABC posterior means are remarkably close to the true values and the distributions have a narrow spread. The uncertainty of the parameter estimates is quantified in terms of the coefficient of variation, CV= $\sigma / E[\tau]$, where σ is the ABC estimate of the posterior standard deviation. For all four data sets, the CV is small, approximately 4–6%. In summary, our examination of the synthetic data suggests that our ABC algorithm

Table 1
ABC posterior summary for τ and D for all experiments in the first experimental scenario. Results shown include the posterior mean, CV and the 90% CI. For all these estimates of $E[\tau]$, MCSE $< 0.002 \times E[\tau]$.

$C(0)$	T (h)	$E[\tau]$ (h)	CV (τ) (%)	90% CI (τ)	$E[D]$ $\mu m^2\,h^{-1}$	CV (D) (%)	90% CI (D) $\mu m^2\,h^{-1}$
5000	24	0.125	5.2	(0.115, 0.136)	1300	5.1	(1100, 1400)
	48	0.093	4.8	(0.085, 0.099)	1700	4.7	(1500, 1800)
	72	0.062	5.7	(0.056, 0.068)	2500	5.7	(2400, 2600)
10,000	24	0.110	3.6	(0.105, 0.117)	1400	3.5	(1300, 1500)
	48	0.077	3.0	(0.073, 0.081)	2200	3.1	(1900, 2100)
	72	0.059	2.9	(0.055, 0.061)	2700	2.8	(2600, 2800)
30,000	24	0.092	2.3	(0.089, 0.095)	1700	2.1	(1600, 1800)
	48	0.064	1.9	(0.063, 0.068)	2400	1.9	(2300, 2500)
	72	0.057	1.8	(0.055, 0.058)	2760	1.8	(2700, 2800)

with the information from the leading edge as a summary statistic allows us to recover the parameters precisely. We now apply the same approach to the experimental data.

3.2. First experimental scenario: Collective cell spreading driven by cell motility alone

We present results for τ for all experimental conditions, a comparison of τ for different experimental times, and a comparison of τ for different initial cell numbers. The ABC estimate of the posterior expected value of τ, $E[\tau]$, the 90% credible interval (CI) and the CVs, from all experimental conditions, are given in Table 1. To assess the accuracy of our resulting estimates from the true ABC posteriors, we computed the MCSE for $E[\tau]$ in all experimental conditions, MCSE $= \sigma/\sqrt{\text{ESS}}$ [34], where ESS $= 1000$ is the effective sample size. For all cases, we found that the MCSE for estimates of $E[\tau]$ is small, less than 0.2% of the estimate of $E[\tau]$. We report values of $E[\tau]$ with up to three significant figures [22]. To obtain an ABC posterior distribution for D, we transformed all the values of the ABC posterior samples for τ using $D = P_m\Delta^2/(4\tau)$.

From Table 1, for all conditions, we see that the CV for τ is small, approximately 2–6%, implying that τ is estimated precisely. Comparing the CV results indicates that the posterior inference for τ is more precise for experiments with a larger numbers of cells. In general, we found that $E[\tau]$ was smaller for those experiments initiated with a larger number of cells. This implies that those cells in the experiments that are initiated with a larger number of cells appear to have a higher D than those cells in the experiments that are initiated with a smaller number of cells.

Previous estimates of D have been obtained for a range cell types under various experimental conditions. For example, Maini et al. [2] estimated D for human peritoneal mesothelial cells in a scratch assay, and found that $D \approx 1400\,\mu m^2\,h^{-1}$. Sengers et al. [16] studied osteoblast MG63 and human bone marrow stromal cells, and estimated $D \approx 1800$ and $D \approx 3060\,\mu m^2\,h^{-1}$, respectively. Simpson et al. [25] used a combination of a discrete random walk model and the solution of the generalized Fisher–Kolmogorov equation, on the same data set that we have studied here, and found that $D \approx 1500, 1700$ and $2900\,\mu m^2\,h^{-1}$ for experiments initialized with 5000, 10,000 and 30,000 cells, respectively. Comparing estimates of $E[D]$ in Table 1 with these previous estimates indicates that our technique produces similar estimates. However, the real advantage of our approach is that we obtain a posterior distribution of D, which can be used to quantify the uncertainty in the estimates.

Fig. 4 compares the ABC posterior densities of τ for different T and $C(0)$. We observe that all of the posterior distributions of τ are approximately symmetric. Results in Fig. 4(a)–(c) correspond to the experiments initiated with 5000, 10,000 and 30,000 cells, respectively. The value of $E[\tau]$ is largest for the experiments terminated after 24 h, and progressively decreases for the experiments terminated after 48 h and 72 h, suggesting that D depends on the termination time, T. The putative relationship between D and T could have several explanations. For example, it seems reasonable to assume that once the cells are placed into the experimental apparatus, some amount of time could be required for the cells to adjust to their new environment and hence the cell motility could be reduced during this transition phase. We note that most previous studies neglect to consider any relationship between D and T [3,16,35], and therefore we suggest that future studies ought to consider how estimates of D might depend on T.

Results in Fig. 4(d)–(f) compare posterior distributions of τ for the experiments terminated after 24, 48 and 72 h. In Fig. 4(d), at time $T = 24$ h, $E[\tau]$ is largest for the experiment with $C(0) = 5000$, whereas $E[\tau]$ is smallest for the experiment with $C(0) = 30{,}000$. Results in Fig. 4(e) and (f) show a similar trend after $T = 48$ h and $T = 72$ h, respectively. However, the differences between the estimates of the posterior distributions of τ are less pronounced at $T = 72$ h. In summary, these findings suggest that D depends on $C(0)$. That is, more initial crowding leads to larger D. While these results are consistent with those identified by Simpson et al. [25], they are of interest since they are contrary to the results reported by Cai et al. [3] and Tremel et al. [35] who found that D at low cell density is larger than D at high cell density. Although we cannot give any definite explanation for the observed relationship between D and $C(0)$, it seems possible that if the cells produce some kind of chemical signal which enhances migration this could provide a potential explanation of these results. Further experimental and modelling work would be required to examine this hypothesis.

3.3. Second experimental scenario: Collective cell spreading driven by combined cell motility and cell proliferation

We consider two approaches to examine the second set of experiments: (i) we make limited assumptions about the values of τ and λ, essentially assuming that τ in the second experimental scenario could be completely unrelated to τ from the first experimental scenario, and (ii) we assume that the values of τ from the first experimental scenario are equal to those of τ in the second experimental scenario. In the latter approach, we therefore use the posterior distribution of τ from the first scenario as the prior for τ in the second scenario.

3.3.1. Uniform priors for τ and λ

We first specify a uniform prior for both τ ($\tau \sim U(0.01, 0.3)\,h$) and λ ($\lambda \sim U(0, 0.3)\,h^{-1}$). The uniform prior for λ covers a very wide range of cell doubling times, $t_d = \ln(2)/\lambda$, from 2.3 h to an infinite doubling time. This is conservative since a typical estimate of the doubling time is approximately 10–30 h [24]. The ABC posterior distributions for τ and λ for the experiments with $C(0) = 5000$ and $T = 24$ h are presented in Fig. 5. Our results show that we infer τ reasonably well; however, we obtain far less information about λ. Thus, we only report

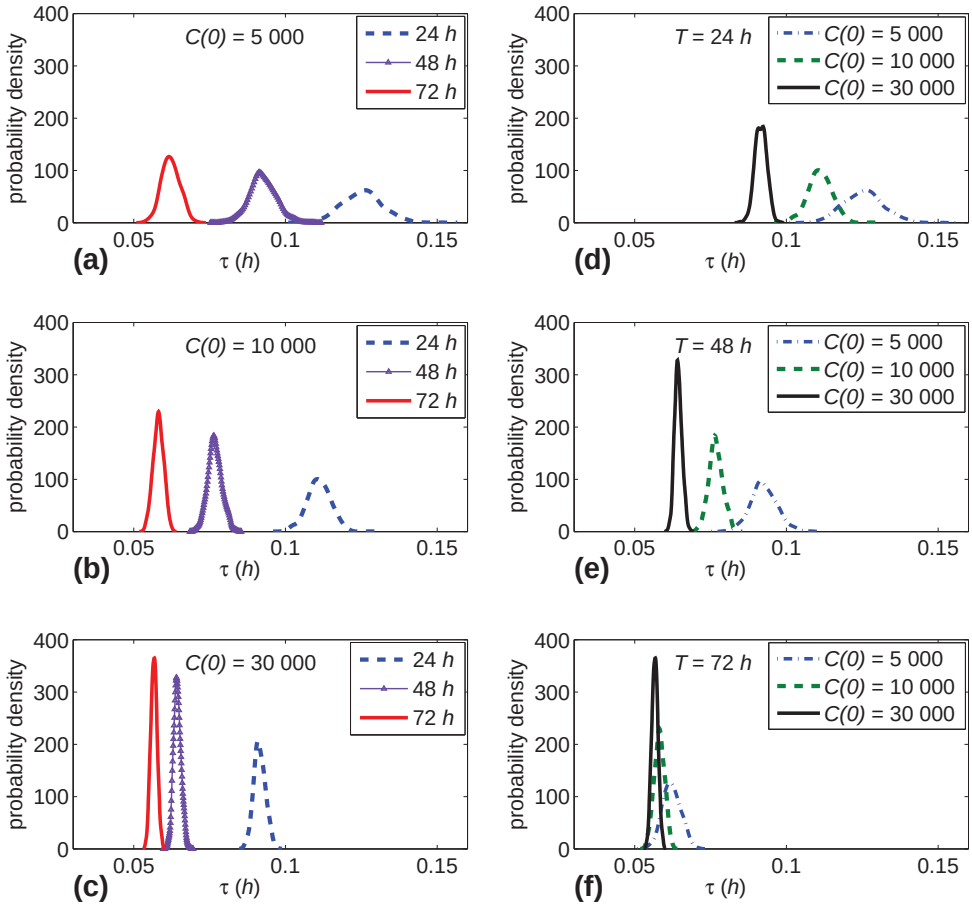

Fig. 4. Comparisons of ABC posterior distribution estimates of τ for all combinations of $C(0)$ and T, in the first experimental scenario, where collective cell spreading is driven by cell motility alone. Results in (a)–(c) correspond to the estimated posterior distributions of τ at three time points for the experiments initialized with 5000, 10,000 and 30,000 cells, respectively. Results in (d)–(f) correspond to the estimated posterior distributions of τ for all three initial cell numbers after a termination time of 24, 48 and 72 h, respectively. All results correspond to a uniform prior for τ, $\tau \sim U(0.01, 0.3)$ h.

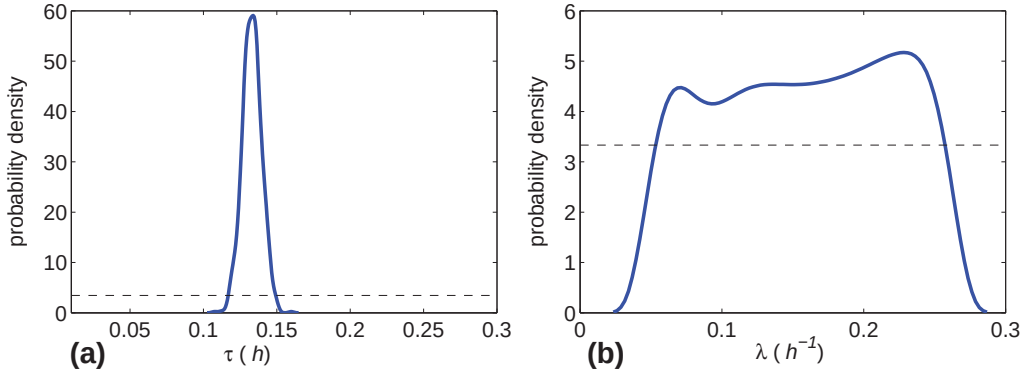

Fig. 5. The ABC posterior distribution estimates of τ and λ for experiments with $C(0) = 5000$, $T = 24$ h, in the second experimental scenario where cell spreading is driven by combined cell motility and cell proliferation. The horizontal dashed lines in (a) and (b) correspond to the uniform priors for τ and λ as specified in Section 3.3.1, respectively.

the posterior summaries for τ (converted into D) in Table 2, and comparisons of τ with respect to different experimental times and different initial numbers of cells are given in Fig. 6. We find that the MCSE for estimates of $E[\tau]$ is sufficiently small, less than 0.2% of the estimated value of $E[\tau]$. Therefore, all values for $E[\tau]$ (and $E[D]$) are reported using up to three significant figures.

Comparing estimates of $E[D]$ in Table 2 with the previous estimates in Table 1 indicates that the values of $E[D]$ are slightly higher for experiments when proliferation is suppressed. Therefore, it seems possible that proliferation could influence cell motility, and this effect is consistent for all combinations of cell densities and experimental times. We note that, for the second experimental scenario, our estimates of the CV for τ are also small, between 2 and 5 %, indicating that we obtain precise estimates of τ.

The ABC posterior distributions of τ are compared with respect to different experimental termination times in Fig. 6(a)–(c). We observe a similar trend to that observed in Fig. 4 for the first experimental scenario. In summary, in both experimental scenarios, we observe a consistent time-dependence in our estimate of τ (or D), suggesting that the longer the experiment, the higher the value of D.

Results in Fig. 6(d)–(f) show the posterior distributions for τ, with each sub-figure corresponding to a particular termination time. The

Table 2
ABC posterior summary for τ and D for all experiments in the second experimental scenario, using uniform priors (Section 3.3.1). Results are given for the posterior mean, CV and the 90% CI. For all these estimates of $E[\tau]$, MCSE $< 0.002 \times E[\tau]$.

$C(0)$	T (h)	$E[\tau]$ (h)	CV (τ) (%)	90% CI (τ) (h)	$E[D]$ $\mu m^2\,h^{-1}$	CV (D) (%)	90 % CI (D) $\mu m^2\,h^{-1}$
5000	24	0.133	5.0	(0.122, 0.144)	1170	5.1	(1090, 1280)
	48	0.105	3.9	(0.099, 0.112)	1490	4.0	(1390, 1580)
	72	0.075	4.5	(0.069, 0.081)	2090	5.0	(1920, 2270)
10,000	24	0.120	3.2	(0.115, 0.126)	1300	3.0	(1250, 1360)
	48	0.083	2.9	(0.079, 0.087)	1880	2.8	(1810, 1970)
	72	0.062	3.0	(0.059, 0.065)	2520	2.9	(2410, 2650)
30,000	24	0.095	2.3	(0.092, 0.099)	1650	2.3	(1580, 1710)
	48	0.068	2.2	(0.065, 0.071)	2300	2.2	(2220, 2390)
	72	0.059	3.3	(0.056, 0.063)	2610	3.3	(2470, 2760)

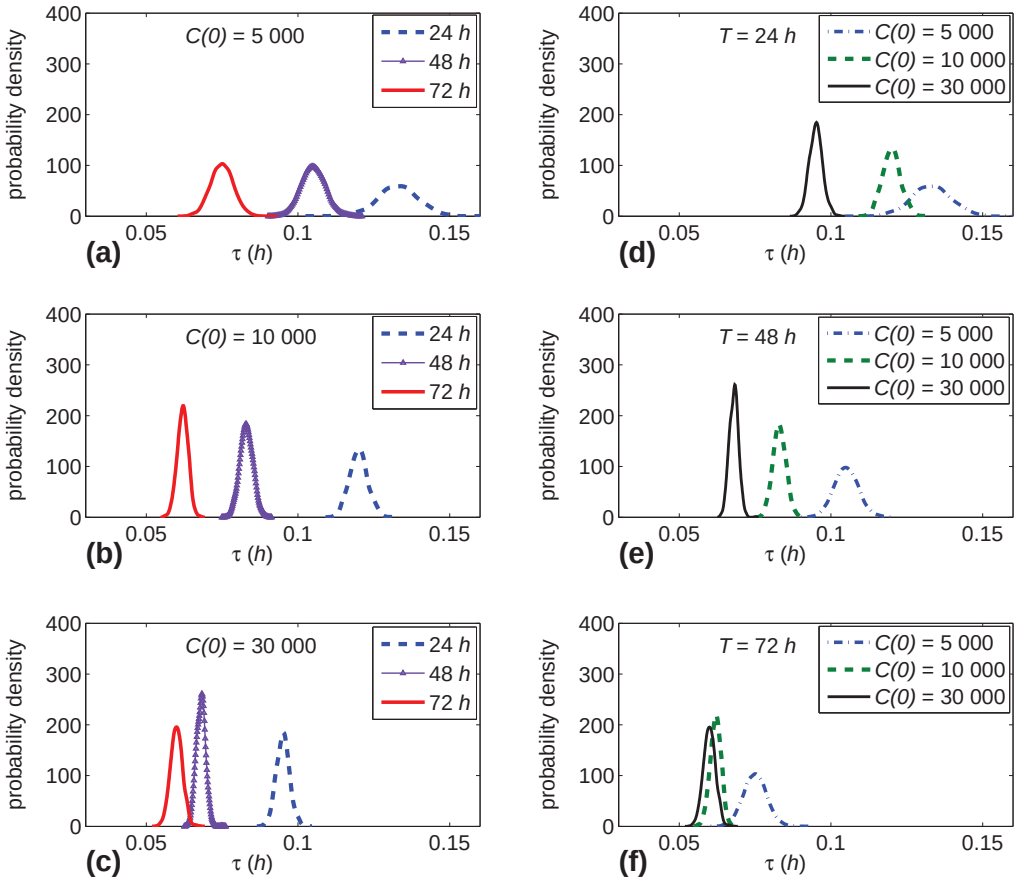

Fig. 6. Comparisons of ABC posterior distribution estimates of τ for all combinations of $C(0)$ and T, in the second experimental scenario, where collective cell spreading is driven by combined cell motility and cell proliferation. Results in (a)–(c) correspond to the estimated posterior distributions of τ at three time points for the experiments initialized with 5000, 10,000 and 30,000 cells, respectively. Results in (d)–(f) correspond to the estimated posterior distributions of τ for all three initial cell numbers after a termination time of 24, 48 and 72 h, respectively. All results correspond to the uniform priors specified in Section 3.3.1.

posterior distributions of τ for experiments initialized with 5000 and 30,000 cells have little overlap, indicating that the cell motility is very different. This apparent density-dependent mechanism is consistent with the results previously reported for the first experimental scenario; that is, higher initial cell densities lead to larger D.

The limitation regarding the determination of λ (Fig. 5b) could be explained by the correlation between τ and λ. We found that there are many pairs of (τ, λ) which lead to the same diameter of the cell spreading population at any particular time point. To overcome this, one could incorporate additional information from the experimental data such as counting the number of cells or computing the distance between cells in some sub-regions of the cell populations. Instead of dealing with these kinds of more detailed experimental data, which may not always be available, we take a different approach and make use of our Bayesian framework by incorporating information from the first experimental scenario into our analysis of the second experimental scenario.

3.3.2. *Informative Gamma prior for τ and uniform prior for λ*

We use an accurate ABC posterior of τ from the first experimental scenario as the prior for τ for the corresponding experiments in the second experimental scenario. The Bayesian sequential learning approach allows us to incorporate information from previous experiments. This amounts to assuming that D in the first experimental scenario is the same as D in the second experimental scenario, and this is similar to the approach used by

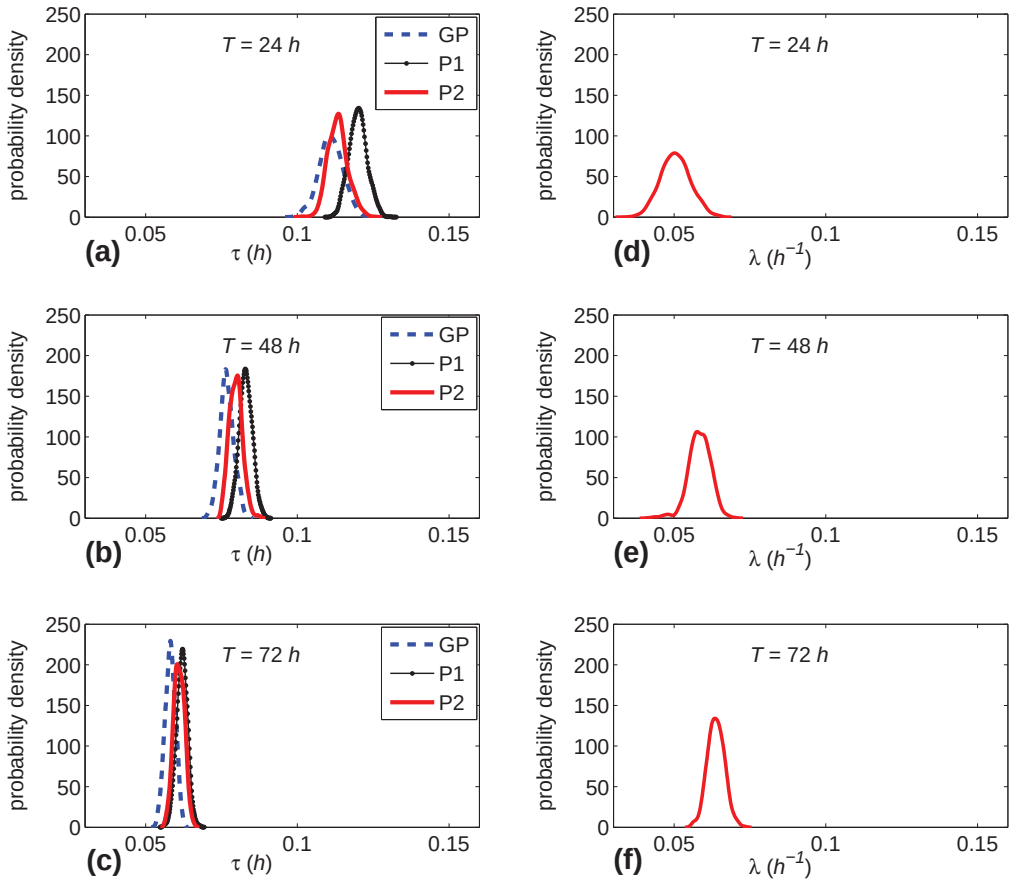

Fig. 7. ABC posterior densities for τ and λ for the experiments with $C(0) = 10{,}000$, in the second experimental scenario. Results in (a)–(c) correspond to the ABC posterior distributions for τ, whereas the results in (d)–(f) summarise the ABC posterior distributions for λ. The blue (dashed) curves in (a)–(c) corresponds to the fitted Gamma prior distributions, GP, for τ; the black (dotted) curves, P1, correspond to the approach outlined in Section 3.3.1 (using a uniform prior for both τ and λ), while the red (solid) curves, P2, correspond to the approach outlined in Section 3.3.2 (using a Gamma prior for τ and a uniform prior for λ). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
ABC posterior summary for τ and λ for all experiments in the second scenario, where collective cell spreading is driven by combined cell motility and cell proliferation. In this case we use the technique outlined in Section 3.3.2 and use fitted Gamma distributions as the prior distributions for τ. Results include the posterior mean (and the 90% CI in the parentheses), CV and the correlation coefficient, r. For all these estimates of $E[\tau]$, MCSE $< 0.002 \times E[\tau]$.

$C(0)$	T (h)	$E[\tau]$ (h)	CV (τ) (%)	$E[\lambda]$ (h^{-1})	CV (λ) (%)	r
5 000	24	0.129 (0.120, 0.139)	4.9	0.046 (0.037, 0.055)	11.5	0.60
	48	0.096 (0.089, 0.102)	4.2	0.057 (0.050, 0.063)	7.1	0.74
	72	0.067 (0.062, 0.073)	4.8	0.059 (0.054, 0.065)	5.2	0.86
10,000	24	0.113 (0.108, 0.119)	3.5	0.050 (0.042, 0.058)	9.7	0.50
	48	0.079 (0.076, 0.083)	3.0	0.059 (0.054, 0.064)	6.4	0.56
	72	0.061 (0.058, 0.064)	2.9	0.064 (0.058, 0.068)	4.4	0.47
30,000	24	0.092 (0.089, 0.095)	2.0	0.069 (0.060, 0.078)	7.9	0.40
	48	0.064 (0.062, 0.066)	1.6	0.070 (0.065, 0.076)	4.7	0.45
	72	0.058 (0.056, 0.059)	1.3	0.071 (0.067, 0.076)	3.8	0.43

Finkenstädt et al. [36]. We fitted a Gamma distribution to each ABC posterior density reported in Section 3.3, for all combinations of $C(0)$ and T (Supplementary material), and use this Gamma distribution as a prior for τ in the corresponding experiment for the second scenario.

The ABC posterior distributions for τ and λ for the experiments with $C(0) = 10{,}000$ are presented in Fig. 7. Results in Fig. 7(a)–(c) compare posterior distributions of τ obtained using the approach in Section 3.3.1 and those results obtained by specifying a uniform prior for λ ($\lambda \sim U(0, 0.3)\,\mathrm{h}^{-1}$) and an informative Gamma prior for τ. Our results show that we infer τ reasonably well for both sets of priors. Values of $E[\tau]$ and the 90% CI for τ are very similar regardless of these choices of priors.

The ABC posterior estimates of λ at 24, 48 and 72 h are given in Fig. 7(d)–(f), respectively. We obtain reasonably precise estimates of λ, with the CV between 4 and 12%, when we specify an informative Gamma prior for τ. Our estimates of λ and τ are correlated, with the correlation coefficient, $r \approx 0.5$. In summary, these results suggest that if we are given some information about τ, via the informative Gamma prior, we can obtain precise information about λ. However,

without some prior information about τ, it is difficult to obtain precise information about λ using leading edge data.

Our estimates of $E[\lambda]$, using the Gamma prior for τ, are consistent with previously reported estimates by Simpson et al. [25]. However, these previously reported estimates are based on a cell labelling and cell counting technique, which can be experimentally and computationally time consuming. In comparison, our approach is far simpler since we only rely on measuring the position of the leading edge. A summary of the ABC posterior for τ and λ, using all combinations of $C(0)$ and T is presented in Table 3 (a summary of posteriors for D are given in the Supplementary material). Plots of the posterior densities for τ and λ for the experiments with $C(0) = 5000$ and $C(0) = 30{,}000$ are also reported in the Supplementary material.

4. Discussion and conclusion

In this work we present an ABC approach to estimate D and λ, which describe the cell motility and the cell proliferate rates, respectively, in a discrete model of collective cell spreading. We estimate D and λ by applying the model to a detailed set of cell spreading experiments in which we consider two distinct sets of experiments: (i) where collective cell spreading is driven by cell motility alone ($D > 0, \lambda = 0$), and (ii) where cell spreading is driven by combined cell motility and cell proliferation ($D > 0, \lambda > 0$). Even with relatively crude leading edge data, the ABC approach can provide precise inferences for both D and λ. One particular finding from our analysis is that D appears to depend on the experimental time, which is a feature that has been previously overlooked. Furthermore, we also found that our estimates of D depend on the initial number of cells present in the assay, and this effect is consistent in both experimental scenarios. These results imply that a more sophisticated model might be warranted. However, it is worthwhile to note that these conclusions would not have been obvious had we not attempted to fit our simpler model to this data set.

Our results suggest we can obtain precise estimates of D and λ only if we have access to a detailed experimental data set (Fig. 7). In particular, our data set involves repeating each experiment twice. In the first experimental scenario, the experiment is performed by treating the cells with a drug to block proliferation, allowing us to estimate D. In the second scenario, each experiment is repeated without the drug so that cell proliferation occurs. Using our estimate of D from the first experiment, we then use the second set of experimental results to estimate λ. If, however, we did not have access to this kind of detailed experimental data set, we would have to estimate D and λ from the second set of experiments alone, and our results for this approach show that we obtain precise estimates for D, but far less information about λ (Fig. 5). Therefore, if the aim of performing these kinds of experiments is to obtain precise estimates of both D and λ, we suggest that the more detailed experimental approach is necessary.

We anticipate that our approach could also be applied to more detailed mathematical models such as lattice free models [9], models incorporating cell-to-cell adhesion [37], models incorporating nutrient dynamics, and models of three dimensional cell spreading [38]. However, for a more detailed mathematical model, we also anticipate that we would need more informative summary statistics and an improved ABC algorithm [19,29] to reduce the computational time. Finally, we also expect that ABC techniques could be used to investigate questions regarding optimal experimental designs, such as when to terminate the experiments and how many replicates to use [39].

Acknowledgments

We thank Katrina Treloar for access to the experimental data, and we appreciate support from the Australian Research Council (FT130100148 and DP110100159). We acknowledge the assistance of the high performance computing team at QUT. We thank the six anonymous referees for their helpful and supportive comments.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.mbs.2015.02.010.

References

[1] P.D. Dale, P.K. Maini, J.A. Sherratt, Mathematical modelling of corneal epithelial wound healing, Math. Biosci. 124 (2) (1994) 127–147.
[2] P.K. Maini, D.L.S. McElwain, D.I. Leavesley, Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells, Tissue Eng. 10 (3) (2004) 475–482.
[3] A.Q. Cai, K.A. Landman, B.D. Hughes, Multi-scale modeling of a wound-healing cell migration assay, J. Theor. Biol. 245 (3) (2007) 576–594.
[4] K.R. Swanson, C. Bridge, J.D. Murray, E.C. Alvord Jr, Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion, J. Neurol. Sci. 216 (1) (2003) 1–10.
[5] K.R. Swanson, Quantifying glioma cell growth and invasion in vitro, Math. Comput. Model. 47 (2008) 638–648.
[6] J. Murray, Mathematical Biology I: An Introduction, third ed., Springer, Heidelberg, Germany, 2002.
[7] E.A. Codling, M.J. Plank, S. Benhamou, Random walk models in biology, J. Roy. Soc. Interf. 5 (2008) 813–834.
[8] C.A. Yates, Discrete and continuous models for tissue growth and shrinkage, J. Theor. Biol. 350 (2014) 37–48.
[9] M.J. Plank, M.J. Simpson, Models of collective cell behaviour with crowding effects: Comparing lattice-based and lattice-free approaches, J. Roy. Soc. Interf. 9 (2012) 2983–2996.
[10] A.M. Middleton, C. Fleck, R. Grima, A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion, J. Theor. Biol. 359 (2014) 220–232.
[11] M.J. Simpson, P. Haridas, D.L.S. McElwain, Do pioneer cells exist? PLOS ONE 9(1) (2014) e85488.
[12] J.P. Murray, A. Walter, A.G. Fletcher, C.M. Edwards, M.J. Tindall, P.K. Maini, Comparing a discrete and continuum model of the intestinal crypt, Phys. Biol. 8 (2) (2011) 026011.
[13] M.J. Simpson, K.A. Landman, B.D. Hughes, Cell invasion with proliferation mechanisms motivated by time-lapse data, Phys. A: Stat. Mech. Appl. 389 (18) (2010) 3779–3790.
[14] K. Takamizawa, S. Niu, T. Matsuda, Mathematical simulation of unidirectional tissue formation: In vitro transanastomotic endothelialization model, J. Biomater. Sci., Polym. Ed. 8 (4) (1997) 323–334.
[15] U. Savla, L.E. Olson, C.M. Waters, Mathematical modeling of airway epithelial wound closure, J. Appl. Physiol. 96 (2) (2004) 566–574.
[16] B.G. Sengers, C.P. Please, R.O.C. Oreffo, Experimental characterization and computational modelling of two-dimensional cell spreading for skeletal regeneration, J. Roy. Soc. Interf. 4 (2007) 1107–1117.
[17] M.A. Beaumont, W. Zhang, D.J. Balding, Approximate Bayesian computation in population genetics, Genetics 162 (4) (2002) 2025–2035.
[18] J.K. Pritchard, M.T. Seielstad, A. Perez-Lezaun, M.W. Feldman, Population growth of human Y chromosomes: A study of Y chromosome microsatellites., Mol. Biol. Evol. 16 (12) (1999) 1791–1798.
[19] S.A. Sisson, Y. Fan, M.M. Tanaka, Sequential Monte Carlo without likelihoods, Proc. Natl. Acad. Sci. 104 (6) (2007) 1760–1765.
[20] C.C. Drovandi, A.N. Pettitt, Using approximate Bayesian computation to estimate transmission rates of nosocomial pathogens, Stat. Commun. Infect. Dis. 3 (1) (2011) 2.
[21] B. Efron, Bayesian inference and the parametric bootstrap, Ann. Appl. Stat. 6 (4) (2012) 1971–1997.
[22] J.M. Flegal, M. Haran, G.L. Jones, Markov chain Monte Carlo: Can we trust the third significant figure? Stat. Sci. 23 (2) (2008) 250–260.
[23] K.K. Treloar, M.J. Simpson, Sensitivity of edge detection methods for quantifying cell migration assays, PLOS ONE 8 (6) (2013) e67389.
[24] S.T. Johnston, M.J. Simpson, D.L.S. McElwain, How much information can be obtained from tracking the position of the leading edge in a scratch assay? J. Roy. Soc. Interf. 11 (2014) 20140325.
[25] M.J. Simpson, K.K. Treloar, B.J. Binder, P. Haridas, K.J. Manton, D.I. Leavesley, D.L.S. McElwain, R.E. Baker, Quantifying the roles of cell motility and cell proliferation in a circular barrier assay, J. Roy. Soc. Interf. 10 (2013) 20130007.
[26] MathWorks, Image processing toolbox, 2012. Retrieved from January 5, 2015.
[27] D. Chowdhury, A. Schadshcneider, K. Nishinari, Physics of transport and traffic phenomena in biology: from molecular motors and cells to organisms, Phys. Life Rev. 2 (2005) 318–352.
[28] D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (25) (1977) 2340–2361.
[29] C.C. Drovandi, A.N. Pettitt, Likelihood-free Bayesian estimation of multivariate quantile distributions, Comput. Stat. Data Anal. 55 (9) (2011) 2541–2556.
[30] P. Marjoram, J. Molitor, V. Plagnol, S. Tavaré, Markov chain Monte Carlo without likelihoods, Proc. Natl. Acad. Sci. 100 (26) (2003) 15324–15328.

[31] P. Bortot, S.G. Coles, S.A. Sisson, Inference for stereological extremes, J. Amer. Stat. Assoc. 102 (477) (2007) 84–92.
[32] A. Sottoriva, S. Tavaré, Integrating approximate Bayesian computation with complex agent-based models for cancer research, in: Y. Lechevallier, G. Saporta (Eds.), Proceedings of COMPSTAT'2010, Physica-Verlag HD, 2010, pp. 57–66.
[33] P. Fearnhead, D. Prangle, Constructing summary statistics for approximate Bayesian computation: Semi-automatic ABC (with discussion), J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.) 74 (3) (2012) 419–474.
[34] C.J. Geyer, Practical Markov chain Monte Carlo (with discussion), Stat. Sci. 7 (1992) 473–511.
[35] A. Tremel, A.Q. Cai, N. Tirtaatmadja, B.D. Hughes, G.W. Stevens, K.A. Landman, A.J. O'Connor, Cell migration and proliferation during monolayer formation and wound healing, Chem. Eng. Sci. 64 (2) (2009) 247–253.
[36] B. Finkenstädt, D.J. Woodcock, M. Komorowski, C.V. Harper, J.R.E. Davis, M.R.H. White, D.A. Rand, Quantitying intrinsic and extrinsic noise in gene transcription using the linear noise approximation: An application to single cell data, Ann. Appl. Stat. 7 (4) (2013) 1960–1982.
[37] K.K. Treloar, M.J. Simpson, P. Haridas, K.J. Manton, D.I. Leavesley, D.L.S. McElwain, R.E. Baker, Multiple types of data are required to identify the mechanisms influencing the spatial expansion of melanoma cell colonies, BMC Syst. Biol. 7 (2013) 137.
[38] M.H. Zaman, R.D. Kamm, P. Matsudaira, D.A. Lauffenburger, Computational model for cell migration in three-dimensional matrices, Biophysical Journal 89 (2) (2005) 1389–1397.
[39] C.C. Drovandi, A.N. Pettitt, Bayesian experimental design for models with intractable likelihoods, Biometrics 69 (2013) 937–948.