Calibration of a Voronoi cell-based model for tumour growth using approximate Bayesian computation

Xiaoyu Wang[1,3*], Adrianne L. Jenner[1,3], Robert Salomone[1,2] and Christopher Drovandi[1,3]

[1]School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD, Australia.
[2]School of Computer Science, Queensland University of Technology, Brisbane, QLD, Australia.
[3]Centre for Data Science, Queensland University of Technology, Brisbane, QLD, Australia.

*Corresponding author: x311.wang@hdr.qut.edu.au;

Abstract

Agent-based models (ABMs) are readily used to capture the stochasticity in tumour evolution; however, these models can pose a challenge in terms of their ability to be validated with experimental measurements. The Voronoi cell-based model (VCBM) is an off-lattice agent-based model that captures individual cell shapes using a Voronoi tessellation and mimics the evolution of cancer cell proliferation and movement. The ability of this type of model to capture tumour growth accurately remains unknown. In this paper, a likelihood-free method is employed to calibrate a VCBM to breast, ovarian and pancreatic cancer measurements *in vivo*. Our approach involves estimating the distribution of parameters that govern cancer cell proliferation and recovering the output from the model based on estimated parameters so that the outputs match the experimental data. Our results show that the VCBM can accurately model monophasic tumour growth but lacks the ability to capture biphasic or multiphasic tumour growth. Furthermore, we find that the time taken for a daughter cell to become a mature adult cell is highly informed by the data. Moving forward, this allows us to propose future refinements to the model to improve accuracy, whilst also making conclusions about the differences in cancer cell characteristics.

Keywords: Agent-based model, Cancer, Likelihood-free inference, Parameter estimation, Tumour growth

MSC Classification Code: 92-08 , 92-10 , 62-F15

1 Introduction

Cancer is a disease that arises through the progressive alteration of normal cells. Solid cancers form what is commonly referred to as tumours, which are populations of cancerous cells joined together with connective tissue (Weinberg and Weinberg, 2006). Over time, these tumours expand and take up residency in healthy tissue. In order to obtain a deep understanding of cancer's underlying mechanisms, experimentalists have focused on observing and analysing the dynamics of abnormal cells contained in tumours and how these play an important role in cancer development (Noble, 2002; Markowetz, 2017).

Mathematical modelling of cancer growth and development has been used to study the dynamical process of cancer cells for many years (Altrock et al., 2015; Beerenwinkel et al., 2015; Barbolosi et al., 2016; Tabassum et al., 2019). Deterministic approaches, such as systems of ordinary differential equations (ODEs) and systems of partial differential equations (PDEs), have been used successfully to model cancer growth (Villasana and Radunskaya, 2003; Yafia, 2011; Tao et al., 2014; Jenner et al., 2020b; Dehingia et al., 2021; Klowss et al., 2022; VandenHeuvel et al., 2022). While insightful, generally these models do not capture the heterogeneity that arises through stochastic processes of tumour growth, or consider the behaviour at an individual cancer cell level (Irurzun-Arana et al., 2020). In turn, such models often do not capture dynamical processes such as individual cell proliferation (Iyer et al., 2011; Sahoo et al., 2011) and cell movement (Groh and Louis, 2010).

More recently, agent-based models (ABMs) have become a popular approach for capturing the stochastic nature of tumour growth (Ozik et al., 2018; Metzcar et al., 2019; Macnamara, 2021; Klowss et al., 2022). ABMs model cells as agents, and the interactions of cells are governed by probabilistic rules that depend on physically-meaningful parameters (An, 2012). There are two popular paradigms of ABMs: on-lattice and off-lattice models (Railsback and Grimm, 2019). On-lattice models restrict movement of cells to neighbouring points on a pre-defined lattice. Whereas cell movement in off-lattice models are governed by interaction forces between neighbouring cells and the surrounding environment. Although, one drawback of ABMs is that model simulation can be more computationally expensive than solving a deterministic system of differential equations.

This work focuses on the Voronoi cell-based model (VCBM), an off-lattice ABM of tumour growth in 2-dimensions, which is proposed by Jenner et al. (2020a) with the concept originally introduced by Kansal et al. (2000b,a); Schaller and Meyer-Hermann (2005); Kempf et al. (2010); Fletcher et al.

(2013). Agents are defined as either healthy cells or tumour cells. The movement of cells through the domain is modelled off-lattice using a force-balance equation based on Hooke's law to account for cell adhesive and repulsive forces. Cell boundaries are then defined by a Voronoi tessellation.

Cancer cell proliferation is largely the driving factor of tumour growth and is impacted by spatial limitations and nutrient sources (Iyer et al., 2011; Sahoo et al., 2011; Jenner et al., 2020a). As such, the key mechanism of tumour growth in the VCBM is cancer cell proliferation (Cheng et al., 2012; Benzekry et al., 2014).

Calibrating the cellular proliferation parameters to tumour growth measurements would provide a better understanding of the differences between tumour types. However, obtaining parameter estimates for an ABM like this is not trivial. This work proposes a Bayesian framework to calibrate cancer cell proliferation in the VCBM and capture the tumour growth in ovarian, pancreatic and breast tumour cell lines implanted *in vivo*.

ABMs are simulation-based models, for which inference can be challenging as the likelihood function for such models is often intractable. Hence, approximate Bayesian computation (ABC) (Sisson et al., 2018) is proposed to bypass the likelihood function and estimate the model parameters. ABC requires that the model can be simulated, but does not require evaluation of its corresponding likelihood function. ABC has been previously used to calibrate biological ABMs. For example, Lambert et al. (2018) apply ABC to a lattice model and Ross et al. (2017) use ABC to improve the experimental design of a wound-healing assay by calibrating a cell-cell adhesion model (Khain et al., 2007; Ross et al., 2015). Rocha et al. (2021) use ABC to calibrate key parameters in their model to identify cell motility using biological experiments in their ABM developed in PhysiCell (Ghaffarizadeh et al., 2018). However, they use simple ABC algorithms which require a very large number of model simulations, and are thus computationally expensive when the model simulation is not computationally trivial. We calibrate the VCBM to tumour growth time series data using sequential Monte Carlo (SMC)-ABC, specifically the algorithm of Drovandi and Pettitt (2011), since it is much more efficient than the standard ABC rejection algorithm, and it can easily take advantage of parallel computing resources.

Understanding the limitations and flexibility of the VCBM is important to be able to determine its reliability as a predictor of tumour growth. The main contribution of this paper is to develop an efficient Bayesian approach for fitting off-lattice stochastic models of tumour growth to time series data of tumour volumes. As a case study, we consider the VCBM of Jenner et al. (2020a), which has not been previously calibrated to data in the literature. By using an efficient Bayesian algorithm, we are also able to thoroughly investigate if the VCBM is flexible enough to recover real tumour volume time series data across a range of cancers. Specifically, we calibrate the model to *in vivo* measurements for breast cancer, ovarian cancer and pancreatic cancer in mice obtained from Wade (2019); Wade et al. (2020) and Kim et al. (2011). Based

on the calibration results, we find that that the VCBM is only able to describe monophasic tumour growth behaviour. The results also show that, for several of the datasets we analyse, as the tumour evolves with time cells require a shorter time to proliferate, which causes the tumour to switch from one growth phase to another. As revealed by our calibration approach, we find that the VCBM can not describe the biphasic tumour growth exhibited by the mice, and suggest that further development of the model is required.

We now describe the structure of the paper. In Section 2, we describe the data we analyse, explain the VCBM that we calibrate to the data and provide details of the ABC calibration method that we adopt for fitting the VCBM to data. Section 3 presents the results of the calibration process. In Section 4, we discuss the findings further, the limitations of our study, and directions for future research.

2 Methods

2.1 Experimental measurements: *in vivo* tumour volume

Measurements from three independent *in vivo* cancer datasets (Fig 1) are used to calibrate parameters in the VCBM: breast cancer cells (Kim et al., 2011), ovarian cancer cells (Kim et al., 2011), and pancreatic cancer cells (Wade, 2019; Wade et al., 2020). In these experiments, tumour volume (mm^3) is approximated using caliper measurements in two-dimensions:

$$\text{volume} = \text{width} \times \frac{\text{length}^2}{2}, \tag{1}$$

where width (mm) is the longest tumour measurement and length (mm) is the tumour measurement along a perpendicular axis (Kim et al., 2011). We also create a synthetic dataset similar in certain aspects to that of each cancer type, which we use to calibrate the VCBM and test its ability to recover known parameter values. The synthetic and real-world experimental datasets are briefly described below. Full details of the real data can be found in Kim et al. (2011) and Wade (2019); Wade et al. (2020).

2.1.1 Breast cancer data

Kim et al. (2011) measure the *in vivo* breast cancer tumour volume over 66 days, with measurements being recorded every second day. Each mouse is injected with Her2/neu-expressing human breast cancer cells MDA-MB435, and tumour volume measurements commenced when the seeded tumours reach $100 - 120\ mm^3$. Fig 1a depicts tumour growth data for four mice with breast cancer. We remove measurements from mouse two from day 58 as the tumour volume began to exhibit a decline, which we deem biologically infeasible for our chosen model assumptions. For more details, see Kim et al. (2011).

2.1.2 Ovarian cancer data

Kim et al. (2011) measure the *in vivo* ovarian cancer tumour volume over 25 days, with measurements being recorded daily in mice injected with SK-OV3 cells. The tumour volume measurements commence when the seeded tumours reach $100 - 120\ mm^3$. Fig 1b displays tumour growth data for three mice with ovarian cancer. For more details, see Kim et al. (2011).

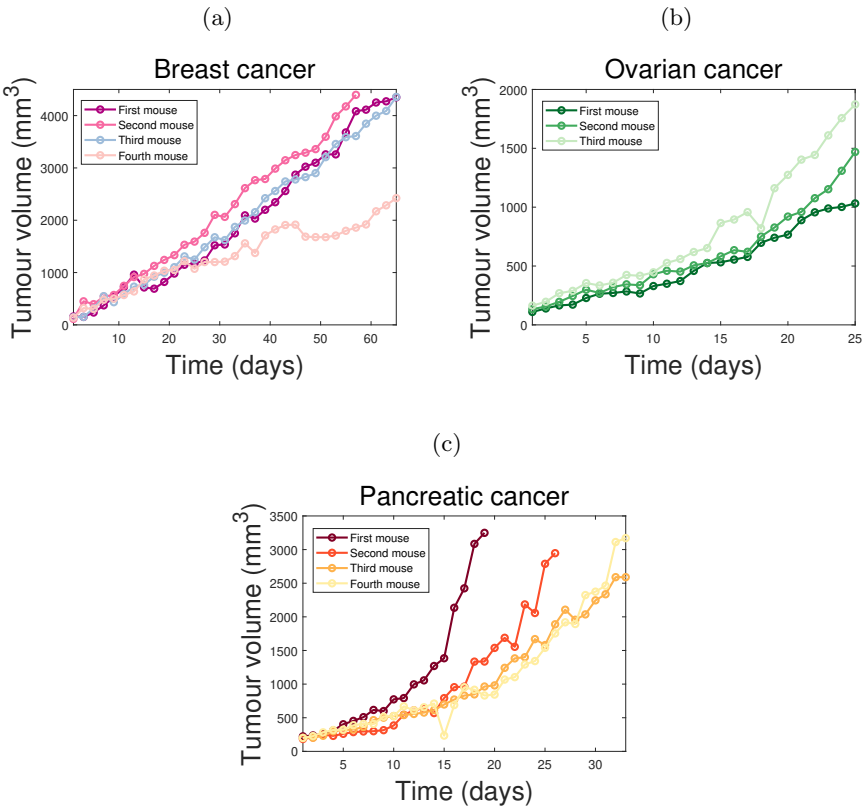

Fig. 1: *In vivo* tumour volume measurements. Tumour growth is measured *in vivo* for (a) breast cancer cells ($n = 4$) (Kim et al., 2011), (b) ovarian cancer cells ($n = 3$) (Kim et al., 2011) and (c) pancreatic cancer cells ($n = 4$) (Wade, 2019; Wade et al., 2020). Each mouse is identified by a unique colour and the specific data measurements are represented by solid points. The ordering of the naming convention for the mice was chosen randomly.

2.1.3 Pancreatic cancer data

Tumour growth measurements in mice seeded with pancreatic cancer cells Mia-PaCa-2 are collected by Wade (2019); Wade et al. (2020). Fig 1c displays the four lines representing the tumour volume for four different mice over 33 days which were recorded every day. The cessation of tumour volume measurements occurred once the volume exceeded a given threshold. Hence, the final observation time varies between mice.

2.1.4 Synthetic cancer data

Three synthetic data sets are generated from our model using given values of the model parameters (see Section 2.2 for more details on the model and the parameters we wish to estimate). We generate two synthetic cancer data sets measuring tumour growth every day for 32 days and one measuring tumour growth every two days for 64 days (to match the characteristics of the breast cancer dataset). To ensure that the synthetic datasets resemble the general features of the real data, we assume that the synthetic tumour growth do not exceed the maximum volume of real tumour growth data, i.e. 5000 mm^3. This was achieved by controlling the model parameter relating to the probability of cell invasion, p_{psc}. For more details, see Section 3.1.

2.2 Voronoi Cell-based Model (VCBM)

ABMs are popular computational models that model cells as agents with associated probabilistic sets of rules for the purpose of capturing stochastic cellular dynamics. In this study, we choose to model the growth of a tumour using an off-lattice ABM because it permits us to model cellular movement more flexibly. Although on-lattice ABMs (Lundh, 2007; Lowengrub et al., 2009; Cristini and Lowengrub, 2010; Voss-Böhme, 2012; Wang et al., 2015; Poleszczuk et al., 2016; Norton et al., 2019) are less expensive to simulate, off-lattice ABMs can be more effective at simulating the heterogeneous nature of the tumour microenvironment.

Here, we describe the VCBM, an off-lattice model developed by Jenner et al. (2020a, 2022) that simulates the 2-dimensional cellular dynamics of tumour formation. In this study, we simplify their model to capture control *in vivo* tumour growth only, omitting the treatment-related agents of virus-infected cancer cells, dead cancer cells, and empty space. As a result, we only investigate three primary dynamics governing tumour growth: cancer cell proliferation, cell movement, and cell invasion into healthy tissue. We assume that no cells die during the simulation and that healthy cells are pushed outward solely by the force of multiplying cancer cells.

2.2.1 Voronoi tessellation

The 2-dimensional spatial position of the ith cell agent at time t is defined by a center point $\boldsymbol{r}_i(t) \in \mathbb{R}^2$. We consider a lattice as the set of cell agent points

$\boldsymbol{r}(t)$, where the size of $\boldsymbol{r}(t)$ depends on the number of cells in the domain at time t. We consider the area of a cell to be the region of space enclosed by a Voronoi tessellation formed by the cell centre point. A Voronoi tessellation is used to define the edges of a cell (see Fig 2), i.e. the boundary of cell i is the line equidistant between the cell's centre point and one of its connected cell's centre points. Voronoi cells on the boundary have an infinite area, and these cells are not explicitly modelled as contributing to the VCBM dynamics. Compared to on-lattice models, the VCBM does not require fixed cell positions. We define the ith cell's neighbourhood, $Neig(i)$, as the set of cells connected to that cell by a Delaunay trangulation (Van Liedekerke et al., 2015).

2.2.2 Cancer cell movement

In the VCBM, cancer cell proliferation is the primary driver of cell movement. The network of forces interacting on cell i is modelled by Hooke's Law (Fig 3a), which computes the effective displacement of the cell and we use this to update the cell's position. The force between cell i and its connected neighbour cell j is modelled by a damped spring (Fig 3b), and the spring connecting cell i and cell j has a rest length $\mathrm{s}_{i,j}(t)$, where we assume the rest length can be a function of time t. The displacement of the ith cell is given by

$$m_i \frac{\mathrm{d}^2 \boldsymbol{r}_i}{\mathrm{d}t^2} = \sum_{j \in \mathrm{Neig}(i)} \mathbf{F}^{\mathrm{I}}_{i,j} + \mathbf{F}^{\mathrm{V}}_i, \tag{2}$$

where m_i is the mass of cell i, $\boldsymbol{r}_i$ is the position of ith cell's centre point, $\mathbf{F}^{\mathrm{I}}_{i,j}$ is the interaction force between a neighbourhood cell j connected to cell i and $\mathbf{F}^{\mathrm{V}}_i$ is the viscous force acting on cell i.

Following Jenner et al. (2020a); Kansal et al. (2000a,b); Schaller and Meyer-Hermann (2005); Kempf et al. (2010); Fletcher et al. (2013), the total interaction force $\mathbf{F}^{\mathrm{I}}_i(t)$ acting on cells i at time t is given by the sum of the individual forces from the neighbourhood connected cells:

$$\mathbf{F}^{\mathrm{I}}_i(t) = \sum_{j \in \mathrm{Neig}(i)} \mathbf{F}^{\mathrm{I}}_{i,j} = \mu \sum_{\forall j} \frac{\boldsymbol{r}_{i,j}(t)}{\|\boldsymbol{r}_{i,j}(t)\|} (s_{i,j}(t) - \|\boldsymbol{r}_{i,j}(t)\|), \tag{3}$$

where μ is the spring constant, $\boldsymbol{r}_{i,j}(t)$ is the vector from the ith point to jth point at time t, $s_{i,j}(t)$ is the spring rest length between cell i and j at time t and $\|\boldsymbol{r}_{i,j}(t)\|$ is the Euclidean norm of $\boldsymbol{r}_{i,j}(t)$. The viscous force acting on cell i is

$$\mathbf{F}^{\mathrm{I}}_i = -\mathbf{F}^{\mathrm{V}}_i = \mu \boldsymbol{v}_i, \tag{4}$$

(Galle et al., 2005; Macklin et al., 2012; Jenner et al., 2020a) where $\boldsymbol{v}_i$ is the velocity of the ith point. By taking a small time interval Δt, the effective

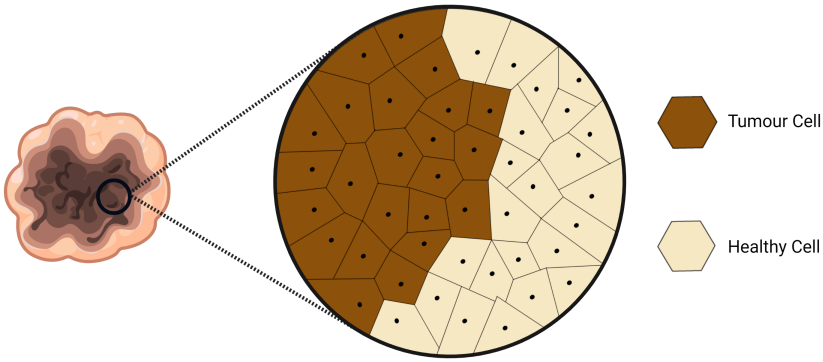

Fig. 2: An illustration of the Voronoi tessellation used to model the cell's shape and its neighbourhood: the boundary of cell i is the line equidistant between the cell's center point (black dot) and one of its connected cell's center points. Cells in the neighbourhood of a particular cell are those sharing a common edge. A Delaunay trangulation is used to determine neighbouring cells connected to cell i.

displacement of the ith point is

$$\boldsymbol{r}_i(t+\Delta t) = \boldsymbol{r}_i(t) + \frac{1}{\mu}\mathbf{F}_i(t)\Delta t = \boldsymbol{r}_i(t) + \lambda \sum_{\forall j} \frac{\boldsymbol{r}_{i,j}(t)}{\|\boldsymbol{r}_{i,j}(t)\|}(s_{i,j}(t) - \|\boldsymbol{r}_{i,j}(t)\|), \quad (5)$$

where μ is the damping constant. This formulation is readily used in other ABM settings, for example see Murray et al. (2012); Murphy et al. (2019); Browning et al. (2019).

2.2.3 Cancer cell proliferation

For modelling cancer cell proliferation, the Euclidean distance, d, between cell i's centre point $\boldsymbol{r}_i(t)$ and the tumour boundary is used as a proxy for the nutrient source. We assume cell i can proliferate when $d < d_{\max}$, where $d_{\max}$ is the maximum radial distance a cell can be from the boundary and still

proliferate. The probability of a cell dividing in a given time step Δt is

$$\mathrm{p}_d = p_0 \left(1 - \frac{d}{d_{\max}}\right), \tag{6}$$

where p_0 is a proliferation constant. In this way, we are defining the probability of proliferation as a function of a cell's distance to the tumour boundary/nutrient source, as has been done similarly in Kansal et al. (2000a); Jiao and Torquato (2011).

Fig 3c shows that when cell i divides it creates two daughter cells i and l, which are placed at a random orientation equidistant from the original cell i's position. Here, the resting spring length between cell i and cell l is computed as

$$s_{i,l} = \frac{s}{p_{\text{age}}}, \tag{7}$$

where s is the mature resting spring length. In this way, we approximate cell growth from a daughter cell size to an adult size which takes p_{age} time steps. This variable spring length is what makes the rest spring length a function of time, i.e. $s_{i,l}(t)$. Once it has been p_{age} time steps since division we have $s_{i,l}(t) = s$. Another factor for cancer cell proliferation is the time taken for a cell to be able to divide into two daughter cells. In the VCBM, the daughter cells take g_{age} time steps to be able to divide again. Note that $p_{\text{age}} < g_{\text{age}}$.

2.2.4 Cell invasion

Due to the invasiveness of cancer, we assume a cancer cell's daughter cell can replace a healthy cell with probability p_{psc}. In our model, this property refers to the probability that invasiveness occurs at the boundary of tumour tissue.

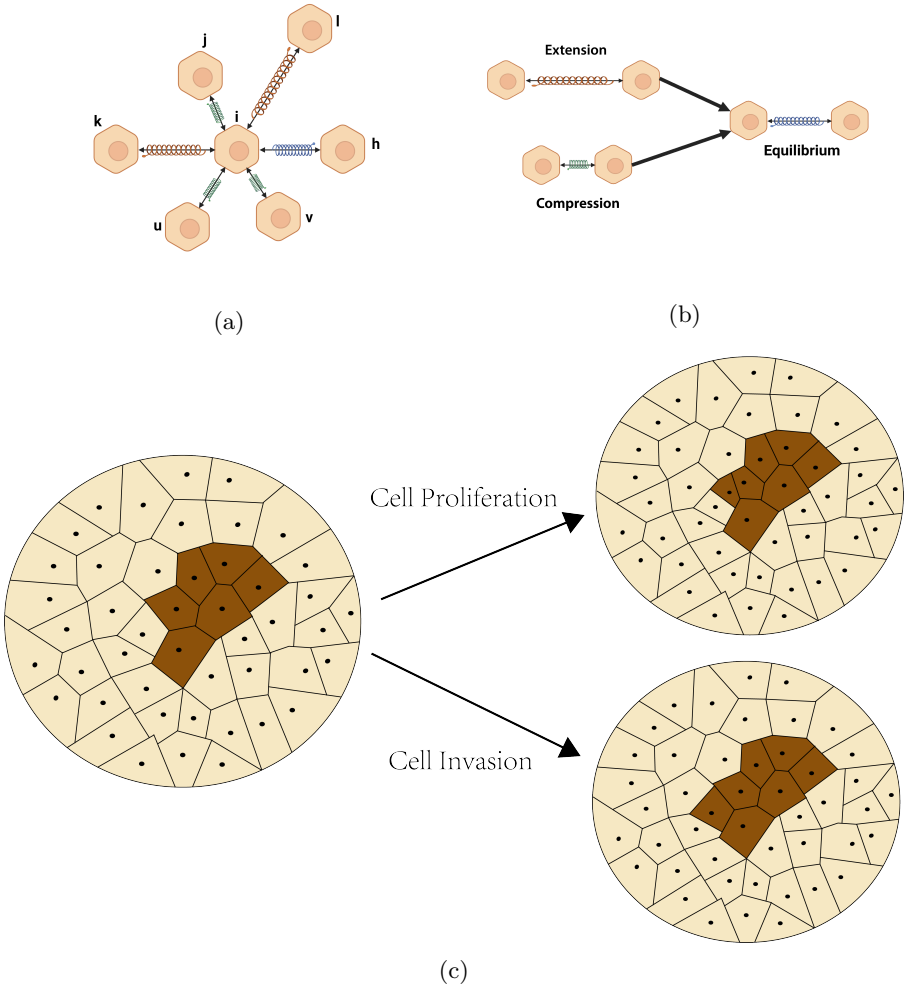

Fig. 3: Cancer cell proliferation: (a) the network of forces interacting at a cell i is governed by Hooke's Law; (b) Hooke's Law is used for cells to reach to equilibrium state; (c) cell proliferation (upper) and cell invasion (lower).

2.3 Bayesian Inference with intractable likelihood

In Bayesian inference, the *posterior distribution* of the parameters $\boldsymbol{\theta}$ is given by

$$\mathrm{P}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{y}|\boldsymbol{\theta})\mathrm{P}(\boldsymbol{\theta}), \tag{8}$$

where the *prior distribution* $\mathrm{P}(\boldsymbol{\theta})$ is updated by the observation data $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)$ with length T via the *likelihood function* $\mathrm{P}(\boldsymbol{y}|\boldsymbol{\theta})$. In this paper, the parameters for the VCBM are $\boldsymbol{\theta} = (p_0, p_{\mathrm{psc}}, d_{\max}, g_{\mathrm{age}})$. However, the

complexity of the biological data simulating processes may render the corresponding likelihood function computationally infeasible (Beaumont, 2019). When the likelihood function is intractable, standard Bayesian inference methods cannot be used to sample the posterior distribution. However, even when the likelihood function is intractable, it often remains feasible to generate simulations from the model.

2.3.1 Approximate Bayesian Computation (ABC)

A method called approximate Bayesian computation (ABC) has been proposed to solve this problem by avoiding evaluation of the likelihood function (Sisson et al., 2018). The idea is to generate approximate samples from the parameter posterior by repeatedly simulating data $\boldsymbol{x}$ from the model for different sets of parameter values and assessing how 'similar' it is with the observed data $\boldsymbol{y}$. Parameter values that generate simulated data that are similar enough to the observed data are retained in the posterior sample. There are two main challenges in the successful implementation of ABC: choosing informative *summary statistics* (Prangle, 2015) and defining a suitable *distance function* (Drovandi and Frazier, 2022), $\rho(\boldsymbol{y}, \boldsymbol{x})$, that measures the 'closeness' of observed data with the simulated data. In our paper, the summary statistics are the observed data itself since the length of tumour growth time series data is short (i.e. T is small), and we select the distance function as

$$\rho(\boldsymbol{y}, \boldsymbol{x}) = \sum_{t=1}^{T} \big(\log(y_t) - \log(x_t)\big)^2. \tag{9}$$

We compare the logarithm of the observed and simulated data since the tumour sizes tend to increase quickly with time. The approximate posterior implied by ABC is given by

$$\mathrm{P}_\epsilon(\boldsymbol{\theta}|\boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{\theta}) \int \mathrm{K}_\epsilon\big(\rho(\boldsymbol{y}, \boldsymbol{x})\big)\mathrm{P}(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x}, \tag{10}$$

where $\mathrm{K}_\epsilon : \mathbb{R}_+ \mapsto \mathbb{R}_+$ is called a *kernel function* with *tolerance* parameter ϵ. Then, the approximate likelihood $\int \mathrm{K}_\epsilon\big(\rho(\boldsymbol{y}, \boldsymbol{x})\big)\mathrm{P}(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x}$ can be estimated unbiasedly by using Monte Carlo via simulating $\boldsymbol{x}$ from the model. Here, the kernel function K_ϵ is a weighting function that assigns higher weight to simulated data $\boldsymbol{x}$ that is closer to observed data $\boldsymbol{y}$. In this paper, we choose the indicator function as the kernel function. That is, $\mathrm{K}_\epsilon\big(\rho(\boldsymbol{y}, \boldsymbol{x})\big) = \mathbb{I}\big(\rho(\boldsymbol{y}, \boldsymbol{x}) < \epsilon\big)$.

Some popular algorithms like the ABC-rejection algorithm (Sisson et al., 2018) and Markov Chain Monte Carlo (MCMC)-ABC (Marjoram et al., 2003; Bortot et al., 2007; Sisson and Fan, 2011) have been used for parameter estimation in the context of biology (Beaumont, 2010; Csilléry et al., 2010; Sunnåker et al., 2013; Beaumont, 2019). Since ABC rejection usually generates proposed

values of $\boldsymbol{\theta}$ from the prior, it can be inefficient if the target posterior differs substantially from the prior. The MCMC-ABC method is often computationally more efficient than ABC-rejection, since it aims to search locally around areas of high posterior probability density. However, MCMC-ABC can also be computationally costly, since the Markov chain can become stuck in low posterior regions, and, due to its serial nature, cannot easily exploit parallel computing architectures. On the other hand, the SMC-ABC algorithm can be more efficient since it can easily harness parallel computing, and it only proposes from the prior in the first iteration, and sequentially improves the proposal distribution. The SMC-ABC replenishment algorithm of Drovandi and Pettitt (2011) is used in a cell biology application in Carr et al. (2021), and we use the same algorithm to calibrate the VCBM. The SMC-ABC replenishment algorithm is explained in more detail in the next section.

2.3.2 SMC-ABC replenishment algorithm

SMC-ABC samples from a sequence of increasingly accurate ABC posteriors based on defining a sequence of non-increasing tolerances $\epsilon_1 \geq \epsilon_2 \geq \cdots \geq \epsilon_T$:

$$\mathrm{P}_{\epsilon_t}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \mathrm{P}(\boldsymbol{\theta}) \int \mathbb{I}\big(\rho(\boldsymbol{y}, \boldsymbol{x}) < \epsilon_t\big) \mathrm{P}(\boldsymbol{x}|\boldsymbol{\theta}) d\boldsymbol{x}, \quad \text{for} \quad t = 1, \ldots, T. \tag{11}$$

The algorithm first draws N independent samples from the prior $\mathrm{P}(\boldsymbol{\theta})$, denoted here as $\{\boldsymbol{\theta}^i\}_{i=1}^N$. Then, for each draw of $\boldsymbol{\theta}^i$ (referred to as a particle), we simulate the stochastic model and compute its corresponding discrepancy ρ^i to produce $\{\boldsymbol{\theta}^i, \rho^i\}_{i=1}^N$. The particle set $\{\boldsymbol{\theta}^i, \rho^i\}_{i=1}^N$ is then sorted by the discrepancy ρ such that $\rho^1 < \rho^2 < \cdots < \rho^N$. We then set the first tolerance threshold as $\epsilon_1 = \rho^N$, i.e. the largest discrepancy value in the particle set. In order to propagate particles through the sequence of target distributions, the next tolerance is defined dynamically as $\epsilon_t = \rho^{N-N_a}$ (where initially $t = 2$) where $N_a = \lfloor Na \rfloor$ and a is a tuning parameter that controls the adaptive selection of discrepancy thresholds and $\lfloor \cdot \rfloor$ is the floor function. Effectively, at each SMC iteration, we discard $a\%$ of the particle set with the largest value of the discrepancy.

After discarding $a\%$ of the particle set, there will only be $N - N_a$ particles remaining. In order to rejuvenate the particle set back to size N, we resample N_a times with equal probability and replacement from the kept set of particles $\{\boldsymbol{\theta}^i, \rho^i\}_{i=1}^{N-N_a}$, where both the parameter and discrepancy values are copied. Although the resampling step ensures that there are N particles in the set, it creates particle duplication.

In order to diversify the particle set, we apply an MCMC kernel to each of the resampled particles. The tuning parameters of the MCMC proposal distribution $q_t(\cdot|\cdot)$ can be obtained from the set of the particles following resampling, which are already distributed according to the ABC target with tolerance ϵ_t. For example, if a multivariate normal random walk proposal is used, its covariance Σ_t can be tuned based on the sample covariance computed from the

Algorithm 1 SMC-ABC replenishment algorithm (Drovandi and Pettitt, 2011)

Input: The observed data $\boldsymbol{y}$, the stochastic model $\mathrm{P}(\boldsymbol{x}|\boldsymbol{\theta})$, distance function $\rho(\cdot,\cdot)$, prior distribution $\mathrm{P}(\boldsymbol{\theta})$, number of particles N, tuning parameters a and c for adaptive selection of discrepancy thresholds and selecting the number of MCMC iterations in the move steps, target tolerance ϵ_T, initial number of trial MCMC iterations S_{init}, minimum acceptable MCMC acceptance rate $p_{\min}$

Output: Samples, $\{\boldsymbol{\theta}^i\}_{i=1}^N$, from the approximate posterior distribution

for $i = 1,\ldots,N$ **do**
  $\boldsymbol{\theta}^i \sim \mathrm{P}(\boldsymbol{\theta})$
  $\boldsymbol{x}^i \sim \mathrm{P}(\boldsymbol{x}|\boldsymbol{\theta}^i)$
  $\rho^i \leftarrow \rho(\boldsymbol{y}, \boldsymbol{x}^i)$
end for
Sort $\{\boldsymbol{\theta}^i\}_{i=1}^N$ by $\{\rho^i\}_{i=1}^N$ such that $\rho^1 \leq \rho^2 \leq \cdots \leq \rho^N$
$N_a \leftarrow \lfloor aN \rfloor, \quad t \leftarrow 2, \quad \epsilon_t \leftarrow \rho^{N-N_a}, \quad \epsilon_1 \leftarrow \rho^N, \quad S_t \leftarrow S_{\text{init}}$
while $\epsilon_{t-1} > \epsilon_T$ **do**
  Compute $\boldsymbol{\Sigma}$ as the sample covariance matrix computed from $\{\boldsymbol{\theta}^i\}_{i=1}^{N-N_a}$
  Generate $\{\boldsymbol{\theta}^i\}_{i=N-N_a+1}^N$ by resampling $\{\boldsymbol{\theta}^i\}_{i=1}^{N-N_a}$
  for $j = N - N_a + 1, \ldots, N$ **do**
    for $k = 1, \ldots, S_t$ **do**
      $p_t^{j,k}, \boldsymbol{\theta}^j, \rho^j \leftarrow \textsc{MCMCMoveStep}(\boldsymbol{\Sigma}, \boldsymbol{\theta}^j, \boldsymbol{y}, \epsilon_t, \rho^j)$
    end for
  end for
  $\tilde{p}_t \leftarrow \sum_{j=N-N_a+1}^N \sum_{k=1}^{S_t} p_t^{j,k} / \big(S_t(N - N_a)\big)$
  $R_t \leftarrow \lceil \log(c) / \big(1 + \log(1 - \tilde{p}_t)\big) \rceil$
  for $j = N - N_a + 1, \ldots, N$ **do**
    for $k = R_t - S_t, \ldots, R_t$ **do**
      $p_t^{j,k}, \boldsymbol{\theta}^j, \rho^j \leftarrow \textsc{MCMCMoveStep}(\boldsymbol{\Sigma}, \boldsymbol{\theta}^j, \boldsymbol{y}, \epsilon_t, \rho^j)$
    end for
  end for
  $\tilde{p}_t^{\text{acc}} \leftarrow \sum_{j=N-N_a+1}^N \sum_{k=1}^{R_t} p_t^{j,k} / \big(R_t(N - N_a)\big)$
  if $\tilde{p}_t^{\text{acc}} < p_{\min}$ **then**
    return $\{\boldsymbol{\theta}^i\}_{i=1}^N$ and terminate the algorithm
  end if
  $S_{t+1} \leftarrow \lceil R_t/2 \rceil$
  Sort $\{\boldsymbol{\theta}^i\}_{i=1}^N$ by new distance $\{\rho^i\}_{i=1}^N$ such that $\rho^1 \leq \rho^2 \leq \cdots \leq \rho^N$
  Update $\epsilon_{t+1} \leftarrow \rho^{N-N_a}, \quad \epsilon_t \leftarrow \rho^N$
  $t \leftarrow t + 1$
end while

Algorithm 2 MCMC move step

Input: Parameter value $\boldsymbol{\theta}$, covariance matrix $\boldsymbol{\Sigma}$, observed data $\boldsymbol{y}$, tolerance ϵ_t and discrepancy value ρ

Output: MCMC acceptance probability p_t, parameter value $\boldsymbol{\theta}$ and discrepancy value ρ

```
function MCMCMoveStep(Σ, θ, y, ε_t, ρ)
    θ̃ ~ N(θ, Σ)
    x̃ ~ P(x|θ̃)
    ρ̃ ← ρ(y, x̃)
    p_t = min(1, P(θ̃)/P(θ) I(ρ̃ < ε_t)).
    if U(0,1) < p_t then
        θ ← θ̃, ρ ← ρ̃
    end if
    return p_t, θ, ρ
end function
```

particle set. Defining the current value of a particle we wish to move as $\boldsymbol{\theta}$, we accept a proposal of parameter $\tilde{\boldsymbol{\theta}} \sim \mathcal{N}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ and simulated data $\tilde{\boldsymbol{x}} \sim \mathrm{P}(\boldsymbol{x}|\tilde{\boldsymbol{\theta}})$ according to the probability

$$p_t = \min\left(1, \frac{\mathrm{P}(\tilde{\boldsymbol{\theta}})}{\mathrm{P}(\boldsymbol{\theta})}\mathbb{I}(\rho(\boldsymbol{y}, \tilde{\boldsymbol{x}}) < \epsilon_t)\right). \tag{12}$$

The proposal densities do not appear in the above acceptance probability due to the symmetry of the multivariate normal random walk proposal.

However, the above procedure may reject proposals and we may fail to move a large proportion of the resampled particles. Thus, we propose to apply R_t iterations of the MCMC kernel to each resampled particle, so that there is a high probability of moving each resampled particle at least once. Specifically, we set $R_t = \left\lceil \frac{\log(c)}{\log(1-p_t^{\mathrm{acc}})} \right\rceil$ where c is a tuning parameter of the algorithm and is, theoretically, the probability that a particle is not moved in the R_t iterations. Here, p_t^{acc} is the expected MCMC acceptance probability at the tth SMC iteration. Since p_t^{acc} is unknown, we estimate it from $S_t < R_t$ trial MCMC iterations. Once it is estimated, we compute R_t and perform the remaining $R_t - S_t$ MCMC iterations on each of the resampled particles. For the next iteration we set the number of trial MCMC iterations to $S_{t+1} = \lceil R_t/2 \rceil$.

There are two ways the algorithm can be stopped. One stopping rule is activated if some desired tolerance ϵ_T is reached, i.e. the maximum discrepancy value in the particle set is below ϵ_T. The second stopping rule is activated when the overall MCMC acceptance probability at a given SMC iteration falls below some user-defined p_{acc}, implying that too much computation is required to progress the algorithm further. The overall acceptance rate can be estimated

by

$$\hat{p}_t^{\text{acc}} = \sum_{j=N-N_a+1}^{N} \sum_{k=1}^{R_t} p_t^{j,k} / \big(R_t(N - N_a)\big), \tag{13}$$

where $p_t^{j,k}$ is the MCMC acceptance rate for particle j at MCMC iteration k and is obtained based on (12). For our study, we set the tuning parameters to be $a = 0.5$ and $c = 0.01$. We use the MCMC acceptance rate as the stopping rule of the algorithm and set $p_{\min} = 0.01$.

2.4 Prior knowledge

The parameters p_0 and p_{psc} are constrained to lie between zero and one as they are the *probabilities* of cell proliferation and cell invasion, respectively. A uniformly distributed prior between zero and one is used as it both satisfies the required constraints and is sufficiently "vague" so to not express any strong opinion regarding where the parameter is likely to lie.

Physically, the parameter $d_{\max}$ is bounded between 0 and the maximum observed tumour radius. We select a prior for $d_{\max}$ of $\text{LogNormal}(\log(30), 1)$, since the median of the prior is 30 and the variance is large enough to produce a reasonably vague prior that remains within the parameter constraints. The parameter g_{age} is an integer, but we find that relatively large values of g_{age} are supported by the data, so we treat it as continuous, and round any proposed value of g_{age} to its nearest integer when feeding it into the model simulation. We assign a $\text{LogNormal}(\log(160), 1)$ prior to g_{age} so that the prior median is 160 and has a relatively large variance to create a reasonably vague prior.

3 Results

In this section, we investigate the SMC-ABC algorithm's ability to calibrate the VCBM. First, the SMC-ABC algorithm is applied to synthetic data generated from the model with known parameter values. Then, SMC-ABC is applied to three real tumour growth datasets to explore whether the VCBM can capture the tumour growth pattern for each type of cancer. The code is available upon request.

3.1 Validation with synthetic data

Prior to applying SMC-ABC to the observed experimental data sets, we first perform a preliminary investigation using synthetic datasets generated via simulation from the VCBM. This allows us to verify that the proposed SMC-ABC algorithm is able to produce expected results under simple settings, i.e. exhibit posterior concentration around the true values that were used to generate the data.

We generate three different synthetic time series data from the model, each with a specified, and hence known, parameters $\boldsymbol{\theta}$ (precise values of the parameters are given in Table 1). We apply SMC-ABC to each synthetic time

series dataset and investigate how well it recovers the parameter values that were used to generate the data.

During generation of simulated datasets, the tumour volume is initialized at 200 mm^3, which is similar to the real data considered later. For two of the simulated datasets, we simulate the model for 32 days taking an observation each day, as is the case for the pancreatic cancer data set. For one synthetic dataset, we simulate for 64 days taking an observation every second day, to match the breast cancer dataset.

Table 1: Parameters used in the generation of the three synthetic datasets.

synthetic dataset	p_0	p_{psc}	d_{max}	g_{age}	length (days)
1	0.4	10^{-5}	31	155	32 (32)
2	0.6	10^{-5}	31	171	32 (64)
3	0.91	10^{-5}	20	171	32 (32)

The estimated univariate posterior distributions when applying SMC-ABC to the three synthetic datasets are shown in Fig 4. Tthe true parameter values lie within regions of significant posterior mass, which is an encouraging result as it indicates the calibration is functioning as desired.

The parameter that appears best informed from the data and inference algorithm is g_{age}, followed by d_{max} and p_{psc}. The parameter p_0 does not appear to be informed by the data much, as the posterior remains quite similar to the prior.

We use the posterior samples to approximate the posterior predictive distribution of the tumour volumes. Fig 5 plots the $(0.25, 0.75)$, $(0.1, 0.9)$, and $(0.025, 0.975)$ posterior predictive intervals. For each time point and every synthetic data set, the associated tumour volume falls within at least one of the intervals. Thus, the SMC-ABC algorithm appears effective at calibrating the VCBM.

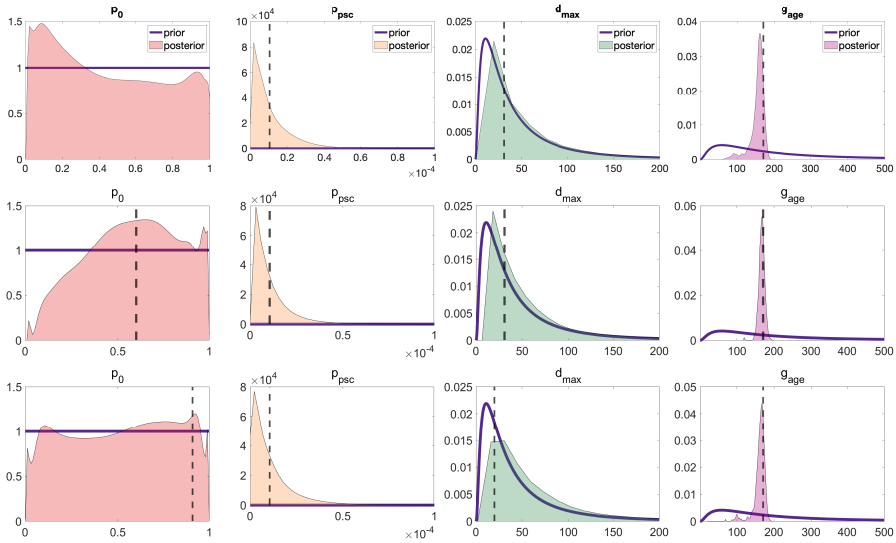

Fig. 4: Marginal prior and (approximated) posterior distributions for parameters $\boldsymbol{\theta} = (p_0, p_{\text{psc}}, d_{\text{max}}, g_{\text{age}})$ for the three synthetic data sets. From top to bottom, each row represents the different synthetic datasets in the order given in Table 1 (the dashed black lines indicate the true data generating parameter values).

3.2 Real tumour growth data

In this section, we apply SMC-ABC to the breast, pancreatic and ovarian cancer datasets shown earlier in Fig 1. For each individual dataset, we estimate the posterior distribution (Fig 6), compute the posterior predictive distribution, and compare the latter to the actual data (Fig 7). Summary statistics of the univariate posterior distributions for breast and ovarian cancer dataset are shown in Table 2 and Table 3, respectively.

For all three cancer types, the posterior distributions for p_{psc} and d_{max} are very similar to the prior, indicating that the data is not providing any additional information about these parameters. The posterior distribution for p_0 is also similar to the prior for most datasets, except for some breast and ovarian datasets where there is some preference for smaller values of p_0.

In contrast, for all cancer types, the data is providing substantial information about g_{age}, as indicated by a much more precise posterior distribution compared to the prior. This indicates that g_{age} strongly drives the dynamics of the VCBM, at least in terms of the tumour volume it produces. Furthermore, there is evidence that within the breast and pancreatic cancer, the estimated g_{age} is different between datasets, since in some cases there is limited overlap between the posterior distributions of g_{age}.

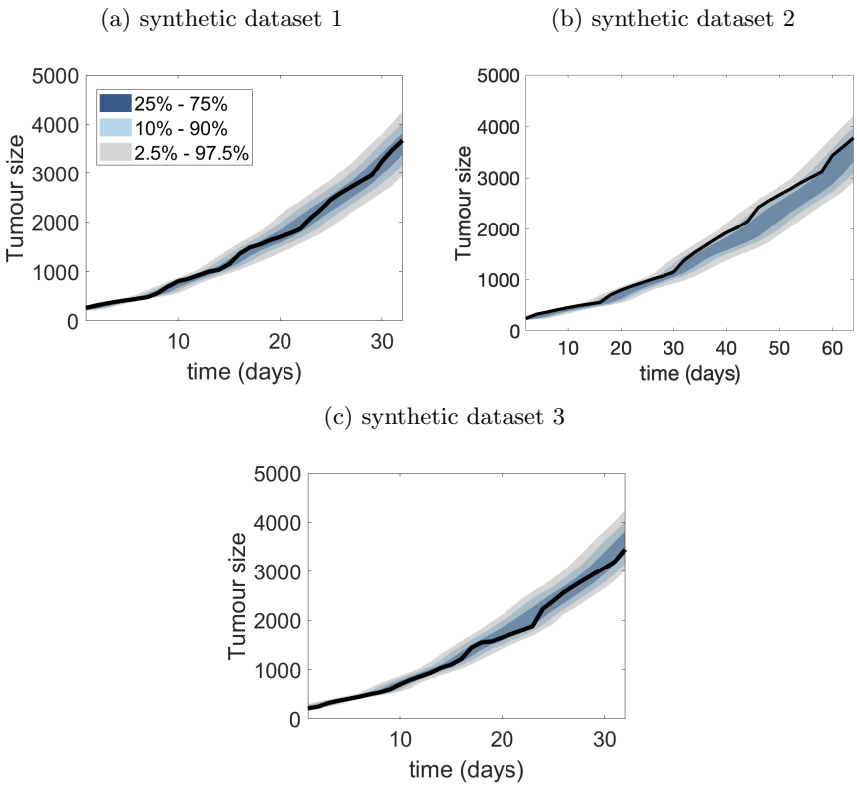

Fig. 5: Posterior predictive intervals obtained via samples from the approximate posterior distributions for the synthetic datasets.

We now examine the ability of the VCBM to accurately model the observed data by analysing the posterior predictive distributions shown in Fig 7. Firstly, it is evident that the VCBM provides a good fit to the three ovarian datasets, as the observed datasets lie comfortably within the prediction intervals (Fig 7c). Based on Fig 7a, the VCBM does a reasonable job at fitting three of the four breast datasets. However, these three datasets show a linear trend in volume size whereas the VCBM predicts exponential growth, and so the model is not as accurately predicting the volume sizes at later times. For the pancreatic datasets, there is a general trend that the VCBM overestimates the volume size at earlier times and underestimates the volume size at later times, particular for mouse 1, 2 and 4. We suggest that the lack of fit of the VCBM could be related to the potentially biphasic nature of tumour growth patterns as exhibited by the pancreatic datasets. We now examine this potential feature in greater detail.

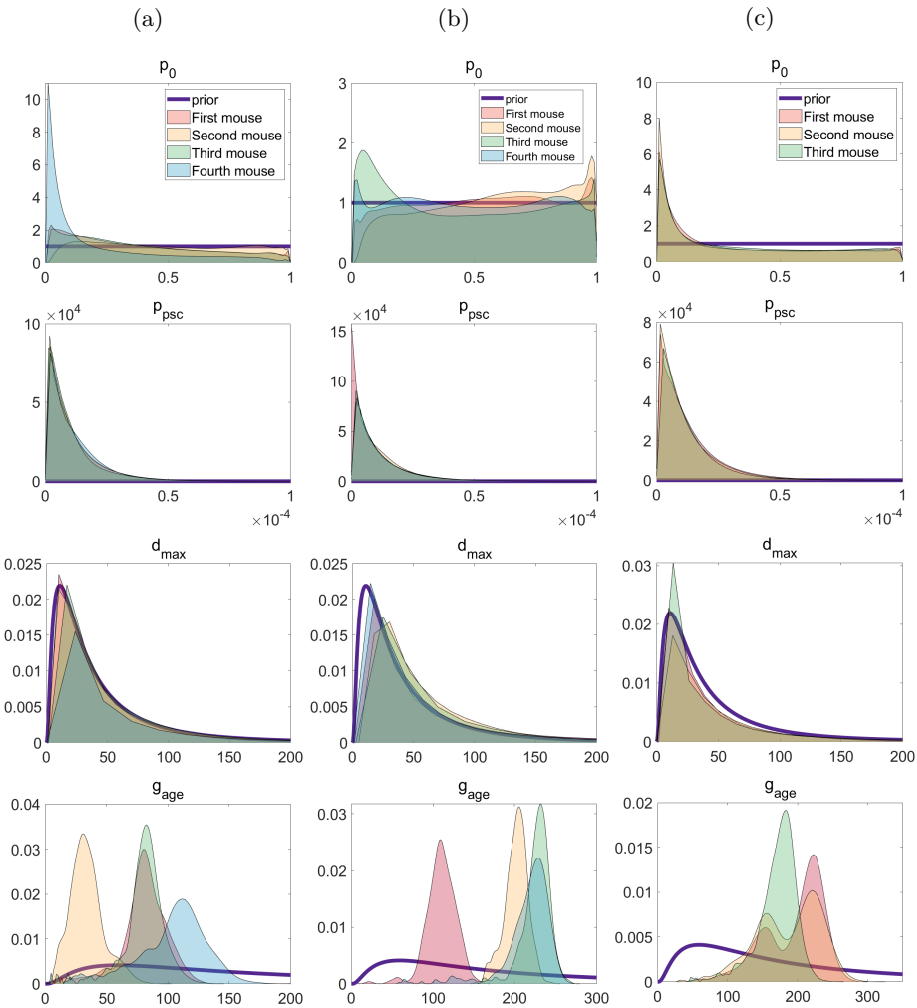

Fig. 6: Marginal prior and mouse-specific posterior distributions for the (a) breast, (b) pancreatic and (c) ovarian tumour growth datasets.

We manually select the tumour growth phase for mouse 1, 2 and 4. Specifically, we define phase 1 for mouse 1, 2 and 4 as 15, 22 and 19 days, respectively, and allocate the rest of the data for phase 2. We allow the parameters $\boldsymbol{\theta}$ to take on different values in these two phases, and hence for each dataset, we estimate 8 parameters as opposed to 4. We use SMC-ABC to estimate the posterior distributions for each phase in three pancreatic datasets, as shown in Fig 8. For all three mice it is evident that the estimate of g_{age} is significantly larger in phase 1 compared to phase 2, demonstrating that the growth rate differs between the two phases (more specifically, faster in phase 2).

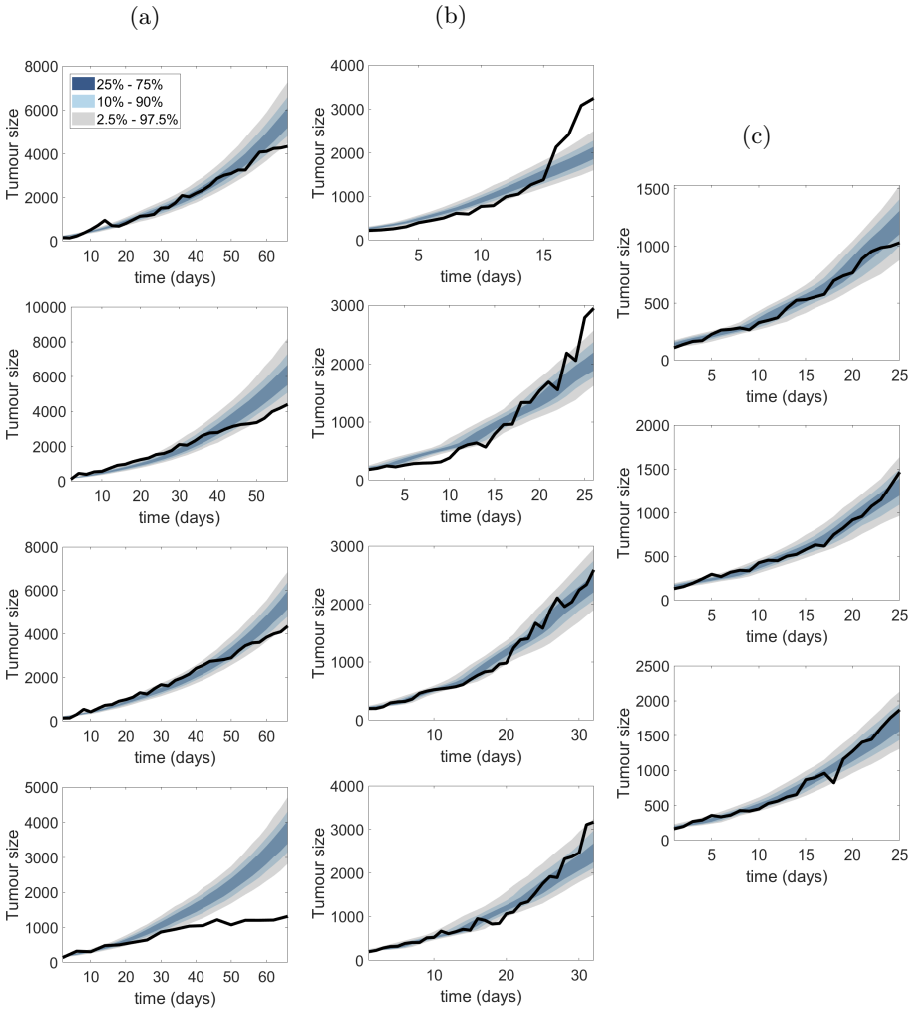

Fig. 7: Posterior predictive intervals for the (a) breast, (b) pancreatic and (c) ovarian tumour growth datasets.

Fig 9 shows that by allowing two separate growth phases that the VCBM provides a much improved fit to the data compared to Fig 7b. Overall, this provides evidence of a possibly biphasic tumour growth pattern for the pancreatic datasets.

Table 2: Posterior means and credible intervals obtained by SMC-ABC for each of the four individual mice in the breast cancer dataset (separate model calibrated for each mouse).

Mouse 1

$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	(25%, 75%)
p_0	0.38	(0.02, 0.95)	(0.15, 0.58)
p_{psc}	1.01×10^{-5}	$(3.01 \times 10^{-7}, 3.37 \times 10^{-5})$	$(2.91 \times 10^{-6}, 1.37 \times 10^{-5})$
$d_{\max}$	49.07	(6.30, 218.55)	(16.26, 58.07)
g_{age}	80.46	(27.94, 114.05)	(73.14, 92.13)

Mouse 2

$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	(25%, 75%)
p_0	0.49	(0.07, 0.96)	(0.25, 0.73)
p_{psc}	1.04×10^{-5}	$(2.91 \times 10^{-7}, 3.89 \times 10^{-5})$	$(2.78 \times 10^{-6}, 1.39 \times 10^{-5})$
$d_{\max}$	50.10	(5.00, 229.44)	(15.64, 60.23)
g_{age}	33.82	(10.68, 64.01)	(24.90, 40.93)

Mouse 3

$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	(25%, 75%)
p_0	0.39	(0.028, 0.94)	(0.16, 0.59)
p_{psc}	9.30×10^{-6}	$(2.45 \times 10^{-7}, 3.59 \times 10^{-5})$	$(2.77 \times 10^{-6}, 3.59 \times 10^{-5})$
$d_{\max}$	50.38	(7.35, 197.02)	(16.98, 62.05)
g_{age}	80.15	(32.65, 109.34)	(74.09, 89.08)

Mouse 4

$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	(25%, 75%)
p_0	0.38	(0.02, 0.95)	(0.15, 0.58)
p_{psc}	1.05×10^{-5}	$(3.88 \times 10^{-7}, 3.80 \times 10^{-5})$	$(3.19 \times 10^{-6}, 1.49 \times 10^{-5})$
$d_{\max}$	42.64	(3.92, 196.70)	(10.34, 47.07)
g_{age}	103.58	(30.51, 146.23)	(88.46, 121.63)

4 Discussion

In this research, we calibrate the off-lattice ABM based on a Voronoi tessellation created by Jenner et al. (2020a) using ovarian cancer, pancreatic cancer and breast cancer tumour growth data sets. The Voronoi Cell-Based model (VCBM) describes the stochastic nature of cancer cell proliferation and has an intractable likelihood function making it challenging to fit to data. The SMC-ABC replenishment algorithm suggested by Drovandi and Pettitt (2011) is suitable for calibrating the VCBM since the likelihood of the VCBM is intractable. While the SMC-ABC replenishment technique has been widely

Table 3: Posterior means and credible intervals obtained by SMC-ABC for each of the three individual mice in the ovarian cancer dataset (separate model calibrated for each mouse).

Mouse 1

$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	(25%, 75%)
p_0	0.35	(0.0096, 0.96)	(0.064, 0.61)
p_{psc}	1.18×10^{-5}	$(2.79 \times 10^{-7}, 4.18 \times 10^{-5})$	$(3.56 \times 10^{-6}, 1.64 \times 10^{-5})$
d_{max}	38.72	(2.85, 187.72)	(9.56, 46.71)
g_{age}	195.62	(101.82, 250.64)	(161.32, 226.59)

Mouse 2

$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	(25%, 75%)
p_0	0.33	(0.0078, 0.96)	(0.048, 0.60)
p_{psc}	1.08×10^{-5}	$(2.84 \times 10^{-7}, 3.74 \times 10^{-5})$	$(3.29 \times 10^{-6}, 1.51 \times 10^{-5})$
d_{max}	40.81	(2.81, 235.99)	(7.00, 46.81)
g_{age}	186.78	(91.47, 247.73)	(153.92, 223.74)

Mouse 3

$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	(25%, 75%)
p_0	0.35	(0.011, 0.95)	(0.063, 0.62)
p_{psc}	1.03×10^{-5}	$(2.03 \times 10^{-7}, 3.65 \times 10^{-5})$	$(2.86 \times 10^{-6}, 1.44 \times 10^{-5})$
d_{max}	37.01	(3.89, 167.17)	(9.74, 41.19)
g_{age}	170.58	(97.78, 211.55)	(159.05, 188.39)

employed Carr et al. (2021); Varghese et al. (2020); Vo et al. (2015); Warne et al. (2019), this is one of the first times that an ABM like this has been calibrated to actual tumour growth data. The outcome of our calibration not only verifies the conclusion of Jenner et al. (2020a) that g_{age} is the most sensitive parameter, but also demonstrates that the model can only represent the monotonic exponential rise of tumour growth and should be further extended to be able to capture biphasic tumour growth.

In our study, we use breast cancer, pancreatic cancer and ovarian cancer tumour growth measurements *in vivo* to examine the robustness and flexibility of the VCBM. It is evident that d_{max} and p_{psc} are not informed by any of the data sets. In contrast, g_{age} provides rich information about the average time a cell takes to proliferate. We also noticed that for data sets with identifiable p_0, these values tend to close to 0. Hence, slowly prolierating tumour cell lines, tend to have identifiable p_0 values whereas, the VCBM is unable to identify faster proliferating tumour cells.

Comparing the posterior distributions for g_{age} we see that breast cancer has lower average time for cell division compared to pancreatic cancer and ovarian cancer. We also note that for the pancreatic cancer, predominantly the mice have similar g_{age} posteriors, except for one mouse that has a lower

Table 4: Posterior means and credible intervals for elements of $\boldsymbol{\theta}$ for each phase in the biphasic model (reported for each of three mice).

	Mouse 1: Phase 1[1]		Phase 2[1]	
$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	mean	(2.5%, 97.5%)
p_0	0.41	(0.014, 0.97)	0.64	(0.16, 0.99)
p_{psc}	1.18×10^{-5}	$(3.27 \times 10^{-7}, 3.99 \times 10^{-5})$	1.06×10^{-5}	$(1.83 \times 10^{-7}, 4.13 \times 10^{-5})$
d_{max}	48.96	(4.88, 192.33)	50.52	(4.54, 217.87)
g_{age}	130.36	(67.85, 172.86)	17.22	(3.13, 30.31)
	Mouse 2: Phase 1[2]		Phase 2[2]	
$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	mean	(2.5%, 97.5%)
p_0	0.54	(0.038, 0.98)	0.57	(0.077, 0.97)
p_{psc}	9.97×10^{-6}	$(2.25 \times 10^{-7}, 3.97 \times 10^{-5})$	1.01×10^{-5}	$(2.89 \times 10^{-7}, 3.52 \times 10^{-5})$
d_{max}	68.00	(14.20, 254.90)	47.00	(4.67, 181.02)
g_{age}	216.07	(174.97, 244.00)	30.13	(12.58, 56.55)
	Mouse 3: Phase 1[3]		Phase 2[3]	
$\boldsymbol{\theta}$	mean	(2.5%, 97.5%)	mean	(2.5%, 97.5%)
p_0	0.28	(0.0051, 0.95)	0.51	(0.052, 0.98)
p_{psc}	1.03×10^{-5}	$(3.43 \times 10^{-7}, 4.00 \times 10^{-5})$	9.91×10^{-6}	$(2.76 \times 10^{-7}, 3.73 \times 10^{-5})$
d_{max}	22.35	(2.12, 140.27)	47.59	(4.36, 196.42)
g_{age}	171.87	(75.14, 249.31)	57.11	(33.59, 78.84)

[1]Mouse 1: phase 1 is from day 1 to day 15, phase 2 is from day 15 to day 19.

[2]Mouse 2: phase 1 is from day 1 to day 22, phase 2 is from day 22 to day 26.

[3]Mouse 4: phase 1 is from day 1 to day 19, phase 2 is from day 19 to day 32.

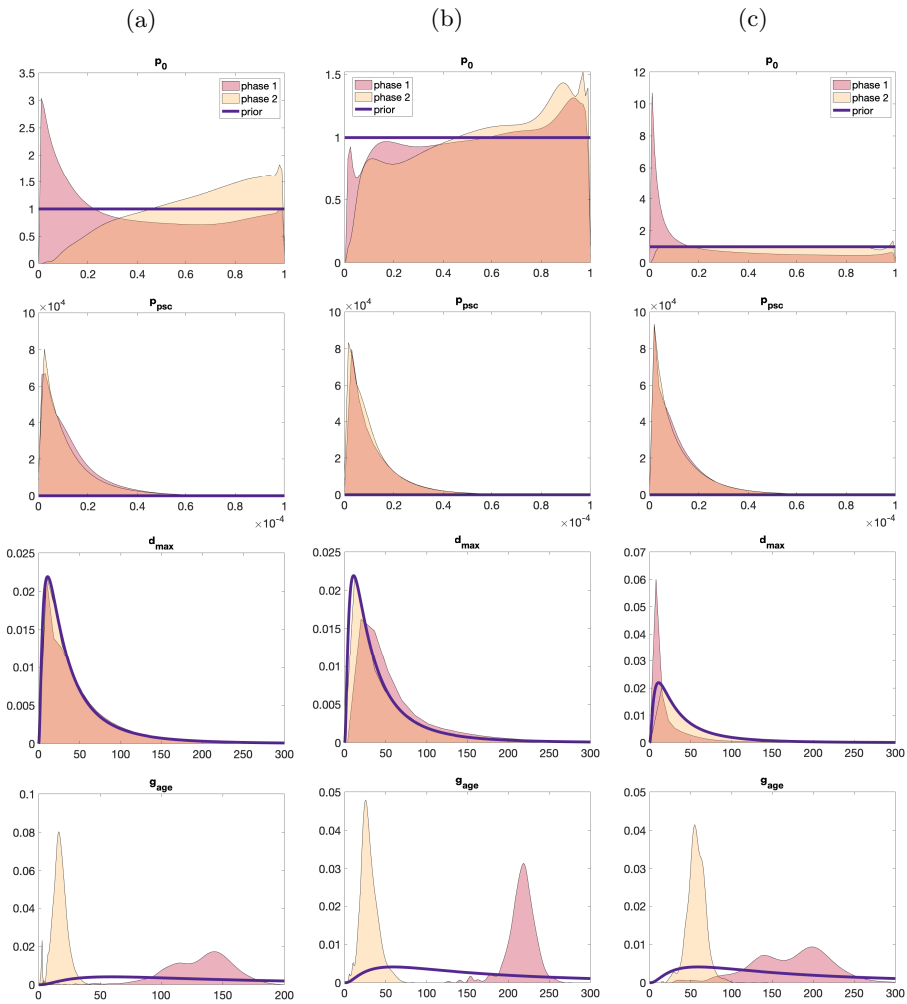

Fig. 8: Marginal prior and posterior distributions for the biphasic tumour growth model for (a) mouse 1, (b) mouse 2 and (c) mouse 3.

average g_{age}. What is also clear, is that there is some bimodality in ovarian cancer g_{age} posteriors, suggesting potentially a subclonal population of ovarian cancer cells with differing times to proliferation.

It is evident that the pancreatic datasets appear to have biphasic tumour growth patterns. To investigate this, we assumed there was some time at which tumour growth could vary based on the deviation of the data from our predicted model simulation. Performing SMC-ABC, we saw that d_{max} and p_{psc} are not able to provide any information as the posterior for both phases are close to the prior. However, the g_{age} posteriors indicate clearly that the average time for cell proliferation is different based on the two phases as shown in

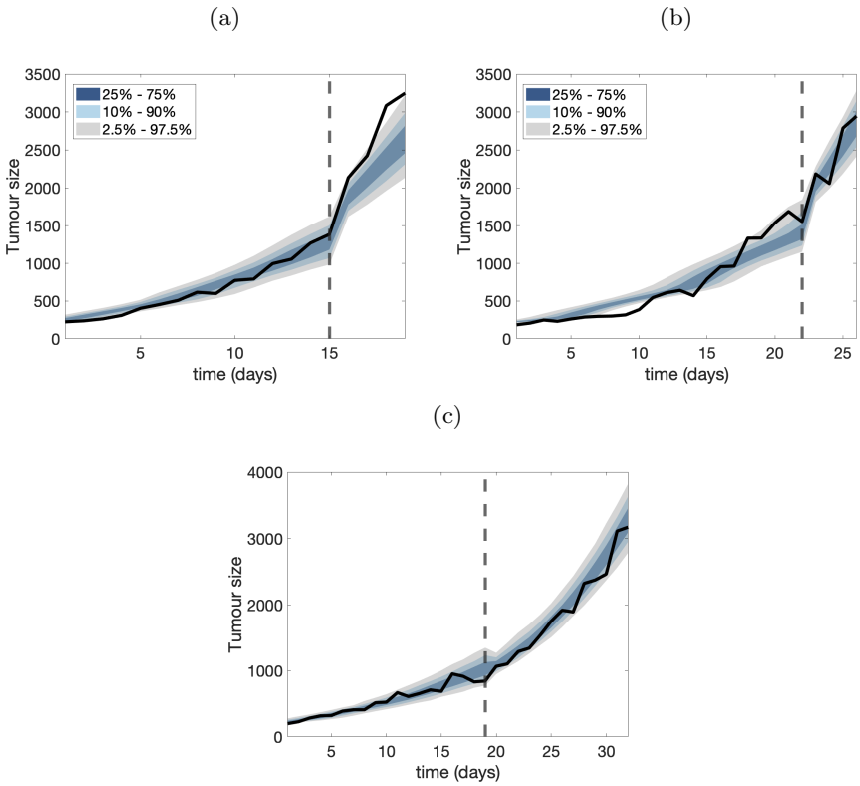

Fig. 9: Posterior predictive distributions for the biphasic tumour growth model fit to the pancreatic tumour datasets. Subfigures (a), (b) and (c) are for mouse 1, 2 and 4, respectively.

Fig 8. Where it is clear that in the first phase, the average time to proliferation is slower than in the second phase. This highlights a shift in tumour growth which may relate to the expansion of the tumour in the subcutaneous tissue or potentially a mutation in the cell cycling. However, to conclude the cause of this shift needs more investigation from both an experimental and modelling point of view.

To reduce computational cost, we implemented the VCBM in only 2-dimensional space. In this way, we approximate tumour growth *in vivo* in 2-dimensions which we feel matched the experimental measurements which are also only in two dimensions. However, having the third dimension measurement of the tumour *in vivo* would provide us with an understanding of the irregularity in tumour volume that we may be missing by only considering 2-dimensional tumour growth. In future work, we hope to use 3-dimensional tumour volume measurements, such as Magnetic resonance imaging (MRI),

to calibrate the 3-dimensional form of the VCBM and improve model accuracy. This would allow us to recapitulate tumour shape irregularities, which we currently assume are negligible in the 2-dimensional VCBM.

As our results suggest, the VCBM is limited to only capturing monophasic tumour growth, while the pancreatic cancer data set suggests that biphasic tumour growth can occur. The next steps for our model are to reformulate the cellular proliferation to be a function of tumour space, so that as the tumour grows we may be able to capture multiphasic growth. One idea for this, would be to consider sub-clonal populations within the tumour that arise stochastically with varying proliferation constants.

While the SMC-ABC is a relatively fast off-line algorithm, it is still computationally expensive as for every proposed parameter value generated in a Bayesian algorithm, a full dataset must be simulated. When simulation time is non-trivial, this then creates a computationally intensive calibration algorithm. On the other hand, online algorithms can iteratively update model parameter estimates as data are introduced sequentially. Such an approach would be useful for our application, since we would only need to simulate data forward one time step at each iteration and future work hopes to investigate this further.

In conclusion, we have applied SMC-ABC to calibrate the VCBM with real breast cancer, pancreatic cancer and ovarian cancer tumour growth datasets. It is evident that g_{age} is the informative parameter for the VCBM and allows it to recapitulate *in vivo* tumour growth data. Unfortunately, p_{psc} and $d_{\max}$ are non-identifiable. We also find the VCBM lacks the ability to extract information from biphasic tumour growth datasets like a pancreatic tumour. We intend to extend the VCBM so that it can capture the biphasic or potentially multiphasic tumour growth pattern and also improve the computational cost of SMC-ABC by moving the algorithm online.

5 Acknowledgement

We thank the computational resources provided by QUT's High Performance Computing and Research Support Group (HPC). We also thank the authors of the previously published experimental data Dr Chae-Ok Yun, Dr Kara Perrow and Dr Samantha Wade for supplying the original published data sets. Xiaoyu Wang and Christopher Drovandi was supported by an Australian Research Council Future Fellowship (FT210100260). Adrianne L. Jenner was supported by the QUT Early Career Researcher Scheme. The project was partly supported by the Centre for Data Science First Byte grant.

6 Statements and Declarations

The authors have no competing interests to declare that are relevant to the content of this article.

References

P. M. Altrock, L. L. Liu, and F. Michor. The mathematics of cancer: integrating quantitative models. *Nature Reviews Cancer*, 15(12):730–745, 2015.

L. An. Modeling human decisions in coupled human and natural systems: Review of agent-based models. *Ecological Modelling*, 229:25–36, 2012.

D. Barbolosi, J. Ciccolini, B. Lacarelle, F. Barlési, and N. André. Computational oncology—mathematical modelling of drug regimens for precision medicine. *Nature reviews Clinical oncology*, 13(4):242–254, 2016.

M. A. Beaumont. Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41:379–406, 2010.

M. A. Beaumont. Approximate Bayesian computation. *Annual Review of Statistics and its Application*, 6:379–403, 2019.

N. Beerenwinkel, R. F. Schwarz, M. Gerstung, and F. Markowetz. Cancer evolution: mathematical models and computational inference. *Systematic biology*, 64(1):e1–e25, 2015.

S. Benzekry, C. Lamont, A. Beheshti, A. Tracz, J. M. Ebos, L. Hlatky, and P. Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS Computational Biology*, 10(8): e1003800, 2014.

P. Bortot, S. G. Coles, and S. A. Sisson. Inference for stereological extremes. *Journal of the American Statistical Association*, 102(477):84–92, 2007.

A. P. Browning, F. G. Woodhouse, and M. J. Simpson. Reversible signal transmission in an active mechanical metamaterial. *Proceedings of the Royal Society A*, 475(2227):20190146, 2019.

M. J. Carr, M. J. Simpson, and C. Drovandi. Estimating parameters of a stochastic cell invasion model with fluorescent cell cycle labelling using approximate Bayesian computation. *Journal of the Royal Society Interface*, 18(182):20210362, 2021.

L. Cheng, K. Yang, Q. Chen, and Z. Liu. Organic stealth nanoparticles for highly effective in vivo near-infrared photothermal therapy of cancer. *ACS nano*, 6(6):5605–5613, 2012.

V. Cristini and J. Lowengrub. *Multiscale modeling of cancer: an integrated experimental and mathematical modeling approach.* Cambridge University Press, 2010.

K. Csilléry, M. G. Blum, O. E. Gaggiotti, and O. François. Approximate Bayesian computation (abc) in practice. *Trends in Ecology & Evolution*, 25 (7):410–418, 2010.

K. Dehingia, H. K. Sarmah, and M. B. Jeelani. A brief review on cancer research and its treatment through mathematical modelling. *Ann Cancer Res Ther*, 29:34–40, 2021.

C. Drovandi and D. T. Frazier. A comparison of likelihood-free methods with and without summary statistics. *Statistics and Computing*, 32(3):1–23, 2022.

C. C. Drovandi and A. N. Pettitt. Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, 67(1):225–233, 2011.

A. G. Fletcher, J. M. Osborne, P. K. Maini, and D. J. Gavaghan. Implementing vertex dynamics models of cell populations in biology within a consistent computational framework. *Progress in Biophysics and Molecular Biology*, 113(2):299–326, 2013.

J. Galle, M. Loeffler, and D. Drasdo. Modeling the effect of deregulated proliferation and apoptosis on the growth dynamics of epithelial cell populations in vitro. *Biophysical Journal*, 88(1):62–75, 2005.

A. Ghaffarizadeh, R. Heiland, S. H. Friedman, S. M. Mumenthaler, and P. Macklin. PhysiCell: An open source physics-based cell simulator for 3-d multicellular systems. *PLoS Computational Biology*, 14(2):e1005991, 2018.

A. Groh and A. K. Louis. Stochastic modelling of biased cell migration and collagen matrix modification. *Journal of Mathematical Biology*, 61(5):617–647, 2010.

I. Irurzun-Arana, C. Rackauckas, T. O. McDonald, and I. F. Trocóniz. Beyond deterministic models in drug discovery and development. *Trends in Pharmacological Sciences*, 41(11):882–895, 2020.

K. Iyer, S. Sankaran, and R. Athale. Stochastic modelling of tumour immune interactions. In *Proceedings of the International Conference on Bioinformatics & Computational Biology (BIOCOMP)*, page 1, 2011.

A. Jenner, W. Kelly, M. Dallaston, R. Araujo, I. Parfitt, D. Steinitz, P. Pooladvand, P. S. Kim, S. J. Wade, and K. L. Vine. Examining the efficacy of localised gemcitabine therapy for the treatment of pancreatic cancer using a hybrid agent-based model. *BioRxiv*, 2022.

A. L. Jenner, F. Frascoli, A. C. Coster, and P. S. Kim. Enhancing oncolytic virotherapy: Observations from a Voronoi Cell-Based model. *Journal of*

Theoretical Biology, 485:110052, 2020a.

A. L. Jenner, F. Frascoli, C.-O. Yun, and P. S. Kim. Optimising hydrogel release profiles for viro-immunotherapy using oncolytic adenovirus expressing il-12 and gm-csf with immature dendritic cells. *Applied Sciences*, 10(8): 2872, 2020b.

Y. Jiao and S. Torquato. Emergent behaviors from a cellular automaton model for invasive tumor growth in heterogeneous microenvironments. *PLoS Computational Biology*, 7(12):e1002314, 2011.

A. Kansal, S. Torquato, G. Harsh Iv, E. Chiocca, and T. Deisboeck. Cellular automaton of idealized brain tumor growth dynamics. *Biosystems*, 55(1-3): 119–127, 2000a.

A. R. Kansal, S. Torquato, G. Harsh Iv, E. Chiocca, and T. Deisboeck. Simulated brain tumor growth dynamics using a three-dimensional cellular automaton. *Journal of Theoretical Biology*, 203(4):367–382, 2000b.

H. Kempf, M. Bleicher, and M. Meyer-Hermann. Spatio-temporal cell dynamics in tumour spheroid irradiation. *The European Physical Journal D*, 60 (1):177–193, 2010.

E. Khain, L. M. Sander, and C. M. Schneider-Mizell. The role of cell-cell adhesion in wound healing. *Journal of Statistical Physics*, 128(1):209–218, 2007.

P.-H. Kim, J.-H. Sohn, J.-W. Choi, Y. Jung, S. W. Kim, S. Haam, and C.-O. Yun. Active targeting and safety profile of peg-modified adenovirus conjugated with herceptin. *Biomaterials*, 32(9):2314–2326, 2011.

J. J. Klowss, A. P. Browning, R. J. Murphy, E. J. Carr, M. J. Plank, G. Gunasingh, N. K. Haass, and M. J. Simpson. A stochastic mathematical model of 4d tumour spheroids with real-time fluorescent cell cycle labelling. *Journal of the Royal Society Interface*, 19(189):20210903, 2022.

B. Lambert, A. L. MacLean, A. G. Fletcher, A. N. Combes, M. H. Little, and H. M. Byrne. Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. *Journal of Mathematical Biology*, 76(7): 1673–1697, 2018.

J. S. Lowengrub, H. B. Frieboes, F. Jin, Y.-L. Chuang, X. Li, P. Macklin, S. M. Wise, and V. Cristini. Nonlinear modelling of cancer: bridging the gap between cells and tumours. *Nonlinearity*, 23(1):R1, 2009.

T. Lundh. Cellular automaton modeling of biological pattern formation: Characterization, applications, and analysis authors: Andreas deutsch and

sabine dormann, birkhäuser, 2005, xxvi, 334 p., 131 illus., hardcover. isbn: 0-8176-4281-, 2007.

P. Macklin, M. E. Edgerton, A. M. Thompson, and V. Cristini. Patient-calibrated agent-based modelling of ductal carcinoma in situ (dcis): from microscopic measurements to macroscopic predictions of clinical progression. *Journal of Theoretical Biology*, 301:122–140, 2012.

C. K. Macnamara. Biomechanical modelling of cancer: Agent-based force-based models of solid tumours within the context of the tumour microenvironment. *Computational and Systems Oncology*, 1(2):e1018, 2021.

P. Marjoram, J. Molitor, V. Plagnol, and S. Tavaré. Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100 (26):15324–15328, 2003.

F. Markowetz. All biology is computational biology. *PLoS Biology*, 15(3): e2002050, 2017.

J. Metzcar, Y. Wang, R. Heiland, and P. Macklin. A review of cell-based computational modeling in cancer biology. *JCO Clinical Cancer Informatics*, 2:1–13, 2019.

R. J. Murphy, P. R. Buenzli, R. Baker, and M. J. Simpson. A one-dimensional individual-based mechanical model of cell movement in heterogeneous tissues and its coarse-grained approximation. *Proceedings of the Royal Society A*, 475(2227):20180838, 2019.

P. J. Murray, C. M. Edwards, M. J. Tindall, and P. K. Maini. Classifying general nonlinear force laws in cell-based models via the continuum limit. *Physical Review E*, 85(2):021921, 2012.

D. Noble. The rise of computational biology. *Nature Reviews Molecular Cell Biology*, 3(6):459–463, 2002.

K.-A. Norton, C. Gong, S. Jamalian, and A. S. Popel. Multiscale agent-based and hybrid modeling of the tumor immune microenvironment. *Processes*, 7 (1):37, 2019.

J. Ozik, N. Collier, J. M. Wozniak, C. Macal, C. Cockrell, S. H. Friedman, A. Ghaffarizadeh, R. Heiland, G. An, and P. Macklin. High-throughput cancer hypothesis testing with an integrated physicell-emews workflow. *BMC bioinformatics*, 19(18):81–97, 2018.

J. Poleszczuk, P. Macklin, and H. Enderling. Agent-based modeling of cancer stem cell driven solid tumor growth. In *Stem Cell Heterogeneity*, pages 335–346. Springer, 2016.

D. Prangle. Summary statistics in approximate Bayesian computation. *arXiv preprint arXiv:1512.05633*, 2015.

S. F. Railsback and V. Grimm. *Agent-based and individual-based modeling: a practical introduction*. Princeton university press, 2019.

H. L. Rocha, I. Godet, F. Kurtoglu, J. Metzcar, K. Konstantinopoulos, S. Bhoyar, D. M. Gilkes, and P. Macklin. A persistent invasive phenotype in post-hypoxic tumor cells is revealed by fate mapping and computational modeling. *iScience*, 24(9):102935, 2021.

R. J. Ross, C. A. Yates, and R. E. Baker. Inference of cell–cell interactions from population density characteristics and cell trajectories on static and growing domains. *Mathematical Biosciences*, 264:108–118, 2015.

R. J. Ross, R. E. Baker, A. Parker, M. Ford, R. Mort, and C. Yates. Using approximate Bayesian computation to quantify cell-cell adhesion parameters in a cell migratory process. *NPJ Systems Biology and Applications*, 3(1): 1–10, 2017.

S. Sahoo, A. Sahoo, and S. Shearer. Stochastic modelling of avascular tumour growth and therapy. *Physica Scripta*, 83(4):045801, 2011.

G. Schaller and M. Meyer-Hermann. Multicellular tumor spheroid in an off-lattice Voronoi-delaunay cell model. *Physical Review E*, 71(5):051910, 2005.

S. A. Sisson and Y. Fan. Likelihood-free MCMC. *Handbook of Markov Chain Monte Carlo*, pages 313–335, 2011.

S. A. Sisson, Y. Fan, and M. Beaumont. *Handbook of approximate Bayesian computation*. CRC Press, 2018.

M. Sunnåker, A. G. Busetto, E. Numminen, J. Corander, M. Foll, and C. Dessimoz. Approximate Bayesian computation. *PLoS Computational Biology*, 9 (1):e1002803, 2013.

S. Tabassum, N. B. Rosli, and M. S. A. B. Mazalan. Mathematical modeling of cancer growth process: a review. In *Journal of Physics: Conference Series*, volume 1366, page 012018. IOP Publishing, 2019.

Y. Tao, Q. Guo, and K. Aihara. A partial differential equation model and its reduction to an ordinary differential equation model for prostate tumor growth under intermittent hormone therapy. *Journal of Mathematical Biology*, 69(4):817–838, 2014.

P. Van Liedekerke, M. Palm, N. Jagiella, and D. Drasdo. Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel

results. *Computational particle mechanics*, 2(4):401–444, 2015.

D. J. VandenHeuvel, C. Drovandi, and M. J. Simpson. Computationally efficient mechanism discovery for cell invasion with uncertainty quantification. *bioRxiv*, 2022.

A. Varghese, C. Drovandi, A. Mira, and K. Mengersen. Estimating a novel stochastic model for within-field disease dynamics of banana bunchy top virus via approximate Bayesian computation. *PLoS Computational Biology*, 16(5):e1007878, 2020.

M. Villasana and A. Radunskaya. A delay differential equation model for tumor growth. *Journal of Mathematical Biology*, 47(3):270–294, 2003.

B. N. Vo, C. C. Drovandi, A. N. Pettitt, and G. J. Pettet. Melanoma cell colony expansion parameters revealed by approximate Bayesian computation. *PLoS Computational Biology*, 11(12):e1004635, 2015.

A. Voss-Böhme. Multi-scale modeling in morphogenesis: a critical analysis of the cellular potts model. 2012.

S. J. Wade. Fabrication and preclinical assessment of drug eluting wet spun fibres for pancreatic cancer treatment. 2019.

S. J. Wade, Z. Sahin, A.-K. Piper, S. Talebian, M. Aghmesheh, J. Foroughi, G. G. Wallace, S. E. Moulton, and K. L. Vine. Dual delivery of gemcitabine and paclitaxel by wet-spun coaxial fibers induces pancreatic ductal adenocarcinoma cell death, reduces tumor volume, and sensitizes cells to radiation. *Advanced Healthcare Materials*, 9(21):2001115, 2020.

Z. Wang, J. D. Butner, R. Kerketta, V. Cristini, and T. S. Deisboeck. Simulating cancer growth with multiscale agent-based modeling. In *Seminars in Cancer Biology*, volume 30, pages 70–78. Elsevier, 2015.

D. J. Warne, R. E. Baker, and M. J. Simpson. Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art. *Journal of the Royal Society Interface*, 16(151):20180943, 2019.

R. A. Weinberg and R. A. Weinberg. *The biology of cancer*. WW Norton & Company, 2006.

R. Yafia. A study of differential equation modeling malignant tumor cells in competition with immune system. *International Journal of Biomathematics*, 4(02):185–206, 2011.