Inference and prediction for stochastic models of biological populations undergoing migration and proliferation

Matthew J. Simpson[1,2] and Michael J. Plank[3]

[1]School of Mathematical Sciences, Queensland University of Technology (QUT), Brisbane, Australia.

[2]ARC Centre of Excellence for the Mathematical Analysis of Cellular Systems, QUT, Brisbane, Australia.

[3]School of Mathematics and Statistics, University of Canterbury, Christchurch, New Zealand.

Abstract

Parameter inference is a critical step in the process of interpreting biological data using mathematical models. Inference provides a means of deriving quantitative, mechanistic insights from sparse, noisy data. While methods for parameter inference, parameter identifiability, and model prediction are well-developed for deterministic continuum models, working with biological applications often requires stochastic modelling approaches to capture inherent variability and randomness that can be prominent in biological measurements and data. Random walk models are especially useful for capturing spatiotemporal processes, such as ecological population dynamics, molecular transport phenomena, and collective behaviour associated with multicellular phenomena. This review focuses on parameter inference, identifiability analysis, and model prediction for a suite of biologically-inspired, stochastic agent-based models relevant to animal dispersal and populations of biological cells. With a particular emphasis on model prediction, we highlight roles for numerical optimisation and automatic differentiation. Open-source Julia code is provided to support scientific reproducibility. We encourage readers to use this code directly or adapt it to suit their interests and applications.

1 Introduction

Mathematical models are an increasingly ubiquitous tool in the life sciences. Frequently, models are built from domain-specific knowledge about the system of interest, and typically contain parameters whose values are *a priori* unknown. Parameter inference refers to statistical methods by which these parameters may be inferred from available data, which are often noisy and incomplete, enabling us to draw conclusions about biological mechanisms or to make predictions about future observations. Whether we are interested in studying disease progression [1], ecological population dynamics [2], cell fate decisions [3], or gene regulation [4], mathematical and computational methods are routinely used to estimate parameters, test hypotheses, and quantify uncertainty [5]. Inference techniques include classical statistical approaches, Bayesian methods, and machine learning, enabling us to integrate diverse data sources, explore hidden patterns, to test and validate mechanistic models.

Continuum models, which describe systems in terms of continuous variables (e.g. density, temperature, displacement) are widely used in classical mathematical modelling applications such as fluid dynamics [6] and material science [7]. Inference for continuum models is well-established across many fields of science and engineering, supported by established foundations in both Bayesian and frequentist methodologies [8–11]. Frequentist approaches, including maximum likelihood estimation and hypothesis testing, have long provided tools for parameter estimation, model validation and model selection [8, 11]. Meanwhile, Bayesian inference incorporates prior knowledge, uncertainty quantification, and performing model selection [9, 10, 12]. The maturity of inference techniques for continuum models has enabled reliable calibration of mathematical models based on ordinary differential equations (ODEs) and partial differential equations (PDEs) across a broad range of applications.

Robust inference in life sciences applications is particularly critical due to the inherent challenges posed by non-standard data collection protocols and limited knowledge of the underlying mechanistic principles, especially when compared to fields like physics and engineering. Biological data are often sparse, noisy, and collected under varying experimental and field conditions. These features can complicate the tasks of parameter estimation, parameter identifiability analysis, and model prediction. Unlike in physics and engineering, where governing equations are typically well-known and able to be derived from first principles, research in life sciences frequently involves phenomenological models that act as coarse approximations of the true biological processes. As a result, careful inference is essential not only for estimating

parameters but also for assessing model identifiability, validating assumptions, appropriately capturing uncertainty, and guiding model refinement.

As data collection protocols and imaging technologies continue to advance, the availability of high-resolution spatiotemporal data is transforming the way biological systems are modelled and analysed [13–19]. Detailed biological datasets capture variability and noise that are often masked in coarse-grained continuum modelling frameworks. This motivates the development and application of stochastic models that more accurately reflect the inherent randomness of biological measurements. In particular, random walk models and related stochastic frameworks have gained prominence for describing a broad range of phenomena including animal movement ecology, multicellular phenomena, molecular diffusion, and intracellular transport [20, 21]. Within the theoretical biology literature, the task of drawing inferences from such models is predominantly approached through Bayesian methods, which are well-suited for handling uncertainty, integrating prior knowledge, and dealing with complex model structures. However, this Bayesian dominance often leads to the undervaluation and under-utilisation of frequentist approaches, which can offer complementary strengths, such as computational efficiency, established theoretical guarantees, and interpretability [22–26].

This review examines the interrelated topics of inference, identifiability, and prediction for stochastic random walk models, and their generalisations. We are particularly interested in applications in population biology, where populations are composed of individuals that undergo some form of random movement and reproduction. Therefore, these models and techniques are well-suited to understanding problems in animal dispersal and collective motion of populations of biological cells. Multi-agent, simulation-based random walk models provide a natural and flexible framework for capturing movement, reproduction and spatial dynamics in these systems, especially where stochasticity plays a central role. We will consider both Bayesian and frequentist approaches to inference, highlighting their respective strengths and limitations in the context of biological data. Special attention will be given to using surrogate models to approximate complex likelihoods, as well as using automatic differentiation to streamline sampling for model prediction. All code required to replicate our results are available as open source Julia code on GitHub, and we encourage readers to use this code directly or adapt it to suit their interests and applications.

2 Motivation

In this Section, we discuss a simple experimental framework that will motivate our high-fidelity, computationally expensive discrete random walk model that aims to describe key features of the experimental mechanisms and observations.

2.1 Experimental observations

The initial experimental motivation for our stochastic simulations is a barrier assay experiment describing the spatial expansion of a circular monolayer of cells, such as the outward circular expansion of a population of fibroblast cells shown in Figure 1(a)–(b) [27]. In this experiment [28, 29], cells are placed into a circular barrier and some time is allowed for individual cells to attach to the cell culture plate. Non-viable cells, and those cells that remain unattached to the plate are removed [27] and the barrier is lifted to initiate the experiment at $t = 0$. The subsequent outward radial expansion of the remaining attached population is imaged, such as the image presented at $t = 72$ hours in Figure 1(b). In this particular experiment, all cells were treated with a chemical inhibitor to prevent proliferation and there was no observed cell death [27]. Images captured during these experiments clearly show the role of stochasticity since random locations of individual cells within the expanding population are visible and can be measured. This kind of experiment motivates us to understand how we can infer macroscopic transport coefficients (e.g., cell diffusivity) and quantify the density of viable cells that have attached to the surface of the cell culture plate at $t = 0$ as the barrier is lifted.

2.2 Discrete model

We first implement a classical non-interacting unbiased random walk model to illustrate some fundamental concepts [20, 30–32]. Later we will extend this fundamental modelling framework to include directional bias, proliferation of agents, and crowding effects. To commence, we consider a lattice-based random walk model on a two-dimensional square lattice with lattice spacing Δ. Each lattice site is indexed in the usual way (i, j), so that each site is associated with location (x_i, y_j), for $i = 1, 2, 3, \ldots, I$ and $j = 1, 2, 3, \ldots, J$. Simulations are performed on a lattice of height H and width W, so that $0 \leq y \leq H$ and $-L \leq x \leq L$, where $W = 2L$. In this case we have $x_i = -L + (i - 1)\Delta$ for $i = 1, 2, 3, \ldots, I$

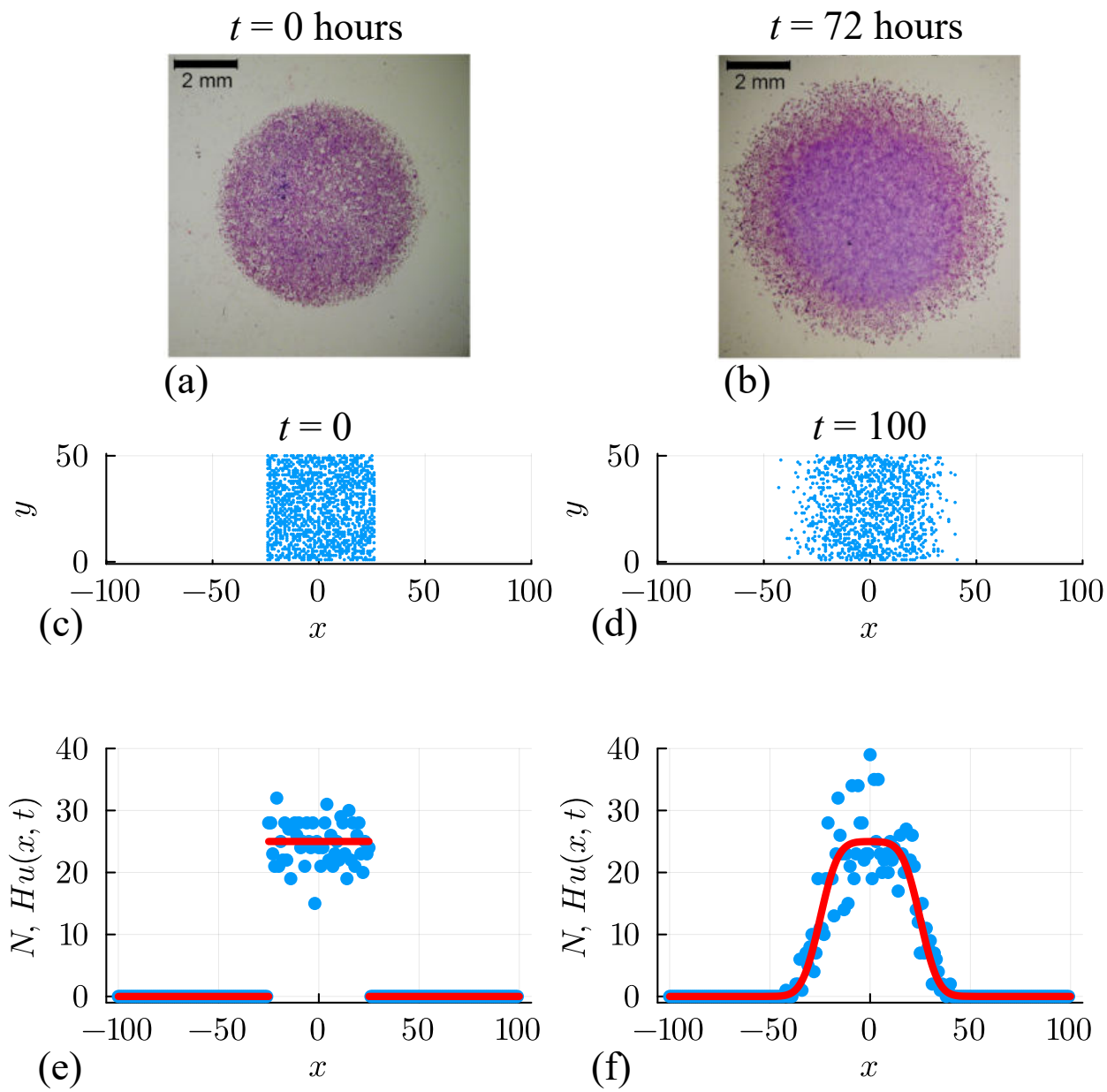

Figure 1: (a)–(b) Barrier assay experiment showing: (a) the initial placement of a population of fibroblast cells within a confined circular region of diameter 6 mm; and, (b) the outward spreading of individual cells after 72 hours [27]. (c)–(d) Unbiased random walk simulation on a square lattice with $\Delta = \tau = 1$, $H = 50$ and $L = 100$. The initial placement of agents in a confined region is shown in (c), and the subsequent outward spreading of agents after 100 time steps is shown in (d). Initially, sites in the region $-25 \leq x \leq 25$ are occupied by, at most, a single agent with probability $U = 0.5$, and then random agent migration takes place with $P = 1$. (e)–(f) Numbers of agents per column at $t = 0$ and $t = 100$, respectively. Noisy count data (blue dots) are superimposed with $Hu(x, t)$ (solid red curve), where $u(x, t)$ is the solution of Equation (2) with $D = P/4$.

and $y_j = (j-1)\Delta$ for $j = 1, 2, 3, \ldots, J$. In this standard model, each lattice site can be occupied by multiple agents, and agents undergo an unbiased random walk within a discrete time stepping framework where each time step has duration τ [20, 30–32]. To evolve the system from time t to time $t+\tau$, we consider a random sequential update method [33] which, for a population of Q agents, involves selecting Q agents independently at random, one at a time, with replacement. Each selected agent moves with probability $P \in [0, 1]$. If an agent attempts to move, the target site is chosen randomly from the four nearest neighbour sites. To maintain generality, we always work with dimensionless simulations by setting $\Delta = \tau = 1$, noting that these simulations can be re-dimensioned to match any dimensional lattice spacing and time step [34]. We will demonstrate the relationship between dimensional experimental data and dimensionless simulations through a Case Study in Section 5.

In this preliminary simulation, we attempt to capture key features of the cell biology experiments in Figure 1, namely we consider a population of cells are initially placed into a confined region, and then we use the discrete model to capture snapshots and data from the spreading population for $t > 0$ [27]. At the beginning of the experiment, $t = 0$, the confined population is shown in Figure 1(a). As the experiment proceeds the population of cells spreads symmetrically outward over time, as illustrated in Figure 1(b). To lay out our analysis as straightforward as possible, we consider simulations performed on a relatively long and narrow lattice with $H = 50$ and $W = 200$ so that $-100 \leq x \leq 100$ and $0 \leq y \leq 50$. Zero flux boundary conditions are applied along all boundaries (i.e., if an agent attempts to move in a direction that would take it off the lattice, that potential move is aborted). Simulations are initialised by randomly occupying each site within the region where $|x| \leq 25$ by a single agent with probability $U = 0.5$, leading to the distribution of agents in Figure 1(c). A single simulation is performed by allowing the system to evolve over 100 time steps with $P = 1$, with the output of the simulation leading to a population that spreads symmetrically in the positive and negative x-directions, as in Figure 1(d). Since there is no variation in the initial macroscopic density in the y-direction, the resulting distribution of agents remains, on average, independent of vertical location for this choice of initial conditions and boundary conditions [34]. This kind of experimental design is both mathematically convenient, and also used in practical cell biology experiments since it is simpler to characterise the spatiotemporal spreading of the population in terms of just one spatial coordinate [35–39]. While it is not immediately obvious from the visualisation in Figure 1(c)-(d), the initialisation of the simulation means that all lattice sites at $t = 0$ either contain zero or, at most, a single agent. In contrast, at subsequent times (e.g. as in Figure 1(d)), some lattice sites

can be occupied by more than one agent since this is a non-interacting random walk model where there is no limit on the number of agents per site and we are not modelling any kind of crowding effect [21].

To quantify the simulations we let $n_{i,j}(t) \in [0, \infty)$ denote the integer number of agents at site (i, j) at time t. Since the average occupancy of lattice sites is independent of vertical location, we quantify the distribution by counting the number of agents per column,

$$N_i(t) = \sum_{j=1}^{J} n_{i,j}(t), \tag{1}$$

so that $N_i(t) \in [0, \infty)$ denotes the integer number of agents in the ith column at time t. For notational simplicity we will sometimes drop the explicit reference to the independent variables by referring to $N_i(t)$ as either N_i or N, as appropriate. This number N is plotted in Figure 1(e)–(f) for the simulation results at $t = 0$ and $t = 100$, respectively. In Figure 1(e) we see that we have $N_i = 0$ for sites where $|x| > 25$, as expected. For sites with $|x| \leq 25$ we have noisy count data where our plot of N_i involves large fluctuations in this region. Since these lattice sites are occupied with probability $U = 0.5$, the expected number of agents is $H \times U = 25$. However, in our example with a lattice of finite height $H = 50$, some columns at $t = 0$ contain as many as $N_i = 36$ agents, whereas others contain as few as $N_i = 17$. This noise is an important feature of real experimental data, which is a key motivation for the use of a stochastic model. To illustrate the importance of these fluctuations we superimpose the expected number of agents per column, $\bar{N}_i$ for $i = 1, 2, 3, \ldots, I$, in Figure 1(e). At $t = 0$ we have $\bar{N}_i = 25$ for $|x| \leq 25$ and $\bar{N}_i = 0$ for $|x| > 25$. Superimposing plots of N_i and $\bar{N}_i$ in this way makes it clear that the data from the stochastic random walk model leads to large fluctuations, and in this review we address the question of how to meaningfully incorporate these large fluctuations using a combination of surrogate models and noise models, as we will now begin to explore.

2.3 Surrogate model

The mean behaviour of the discrete model is related to a PDE that describes the average density of agents as a function of continuous position and time [20, 21, 31, 32],

$$\frac{\partial u}{\partial t} = D \nabla^2 u, \tag{2}$$

where $D = \lim_{\Delta,\tau\to 0} = P\Delta^2/(4\tau)$, such that $u(x, y, t) \geq 0$ denotes the average density of agents at location (x, y) at time t. This continuum limit is valid in the constrained limit as $\Delta \to 0$ and $\tau \to 0$ with the ratio Δ^2/τ held constant [31]. Since our simulations do not involve macroscopic density gradients in the vertical direction, $\partial u/\partial y$ vanishes and this PDE simplifies to

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \tag{3}$$

for the density $u(x, t) \geq 0$, with initial condition

$$u(x, 0) = \begin{cases} U, & |x| \leq h, \\ 0, & |x| > h. \end{cases} \tag{4}$$

Since $u(x, t) \geq 0$ represents the mean agent density, for our simulations the product $H \times u(x, t)$ is the continuous analogue of N_i obtained from the discrete simulations. With our choice of initial conditions, and simulations of an intermediate time scale where agents do not reach the vertical boundaries at $x = \pm L$, we can approximate the solution of Equation (3) with the appropriate infinite domain solution [40]

$$u(x, t) = \frac{U}{2}\left(\operatorname{erf}\left[\frac{h - x}{2\sqrt{Dt}}\right] + \operatorname{erf}\left[\frac{h + x}{2\sqrt{Dt}}\right]\right), \tag{5}$$

where $\operatorname{erf}(x) = (2/\sqrt{\pi})\int_0^x \exp(-z^2)\,\mathrm{d}z$ is the error function [41]. Evaluating this solution with $h = 25$ and $U = 0.5$ to match the initial condition, and $D = 0.25$ to match the jump probability, we plot the quantity $Hu(x, t)$ at $t = 100$ in Figure 1(f) to illustrate that the solution of the PDE model provides a good prediction of the evolution of the average number of agents per column in the discrete simulation. Clearly the PDE solution does not capture the fluctuations present in the discrete simulation data, and much of this review will be devoted to exploring different ways of quantifying and capturing these fluctuations using an appropriate choice of noise model.

3 Inference

We will now work through a number of computational exercises aimed at using discrete count data in Figure 1(f) to infer parameters relevant to the solution of the mathematical model. As we work through various problems, we will introduce different details into the discrete

model and explore the consequences of increasing the complexity of the inference problem. Open access software that can be used to replicate each case we consider is available on GitHub.

3.1 Approximate Bayesian Computation: Estimation and prediction

We now consider the problem of using the discrete data in Figure 1(f) to estimate values of the parameters $\boldsymbol{\theta} = (U, D)^\top$. In other words, we aim to take the noisy output of the discrete simulation, and estimate values of U and D that best explain this noisy data under the random walk model. Throughout this document we will refer to the data using the vector $\boldsymbol{y}^{\text{obs}}$, which in this case is a vector of length 201 so that the ith element is given by N_i as plotted in Figure 1(f). In the first instance we consider a Bayesian setting in which the parameters $\boldsymbol{\theta} = (U, D)^\top$ are treated as random variables, and the uncertainty in $\boldsymbol{\theta}$ is updated using simulation data [42–45]. Simulation data are generated by randomly sampling values of $\boldsymbol{\theta} = (U, D)^\top$ from a uniform prior distribution within the region $0.05 \leq D \leq 0.40$ and $0.30 \leq U \leq 0.70$ since we know in advance that this interval contains the true parameter values used to generate the synthetic data. For the jth random sample $\boldsymbol{\theta}_j$, we initialise the discrete model and perform a discrete simulation over 100 time steps to record the resulting count data $\boldsymbol{N}^{(j)}$, and we compute a discrepancy measure for the jth parameter sample

$$\rho_j\left(\boldsymbol{y}^{\text{obs}}, \boldsymbol{N}^{(j)}\right) = \left\|\boldsymbol{y}^{\text{obs}} - \boldsymbol{N}^{(j)}\right\|, \tag{6}$$

which we record as, $\varepsilon_j = \rho_j$. To implement the algorithm we repeat this process until we obtain 10^5 samples of $\{\boldsymbol{\theta}_j, \varepsilon_j\}_{j=1}^{10^5}$. Samples are sorted in order of increasing ε, and we retain a small proportion of samples so that we focus on parameter combinations that provide a good match to the data. In this case we retain the top 1% of samples as posterior samples, and we checked (results not shown) that the key properties of the posterior distribution are relatively insensitive to this choice of tolerance [46]. We also checked that our ABC rejection results are insensitive to widening the prior distribution of D and U, noting that widening the prior does not have a noticeable impact on the results other than significantly slowing down the computational procedure. We encourage readers to download the Jupyter notebooks from GitHub and explore the consequences of working with different prior distributions and varying the tolerance used to construct the posterior samples.

Results of this elementary algorithm are presented in Figure 2(a), where an unsmoothed histogram of posterior samples is presented. Taking the mean of the posterior samples gives $\boldsymbol{\theta} = (0.489, 0.254)^\top$, which is relatively close to the expected result, $\boldsymbol{\theta} = (0.500, 0.250)^\top$. The shape of the histogram indicates that the posterior samples for U are relatively narrowly distributed about the mean, whereas the posterior samples of D are not. The marginal univariate distributions for D and U, also presented in Figure 2(a), show this more clearly with the posterior distribution of U relatively narrow about the mean, whereas the posterior distribution of D is relatively flat and the credible interval is not very different from the width of the prior distribution. While there are many possibilities for improving the performance of this simple approach (e.g. by exploring different summary statistics or different sampling methods) we will now explore the implications of repeating the process using a surrogate model.

Results in Figure 2(b) summarise the outcome of the same ABC rejection algorithm, using the same data as before, except now we work with the solution of the continuum limit model, Equation (5). As discussed before, the solution of this continuum limit model gives $u(x,t)$, and taking the product $Hu(x,t)$ gives us the expected number of agents in the column at location x under the solution parameterised with $\boldsymbol{\theta} = (U, D)^\top$. The outcome of the jth model simulation is a vector $\bar{\boldsymbol{N}}^{(j)}$, whose ith element is given by $Hu(x_i, t)$, where x_i is the horizontal location of the ith column in the lattice. For each $\bar{\boldsymbol{N}}^{(j)}$ we compute $\varepsilon_j = \rho_j$ and retain 1% of 10^5 random samples with the smallest discrepancy and then generate the histogram of posterior samples as shown in Figure 2(b). The posterior mean is $\boldsymbol{\theta} = (0.514, 0.317)^\top$. Compared to the distributions in Figure 2(a), the posterior samples in Figure 2(b) obtained using the surrogate PDE are much more narrowly distributed about the mean. In this case, however, the credible interval for D does not contain the expected true value, $D = 0.25$. This could be due to some bias in the estimator, or it could simply be the result of the extremely variable data, as previously discussed with reference to the data in Figure 1(f). Since we have obtained reasonably precise (but potentially inaccurate) estimates of U and D, we can solve the process model, Equation (5), for each of the $M = 1000$ retained parameter samples, giving a series of M solutions, $u_m(x,t)$ for $m = 1, 2, 3, \ldots, M$. With these M solutions, we construct a credible interval for N by uniformly discretising the domain, $-100 \leq x \leq 100$ using 201 equally-spaced mesh points. At the ith mesh point we form a credible interval for N by finding the maximum and minimum value of $Hu_m(x_i, t)$, for $m = 1, 2, 3, \ldots, M$ (see Figure 2(c)). While this relatively narrow credible interval for N captures the mean trends in the data reasonably well, it does not capture the fluctuations in the data.

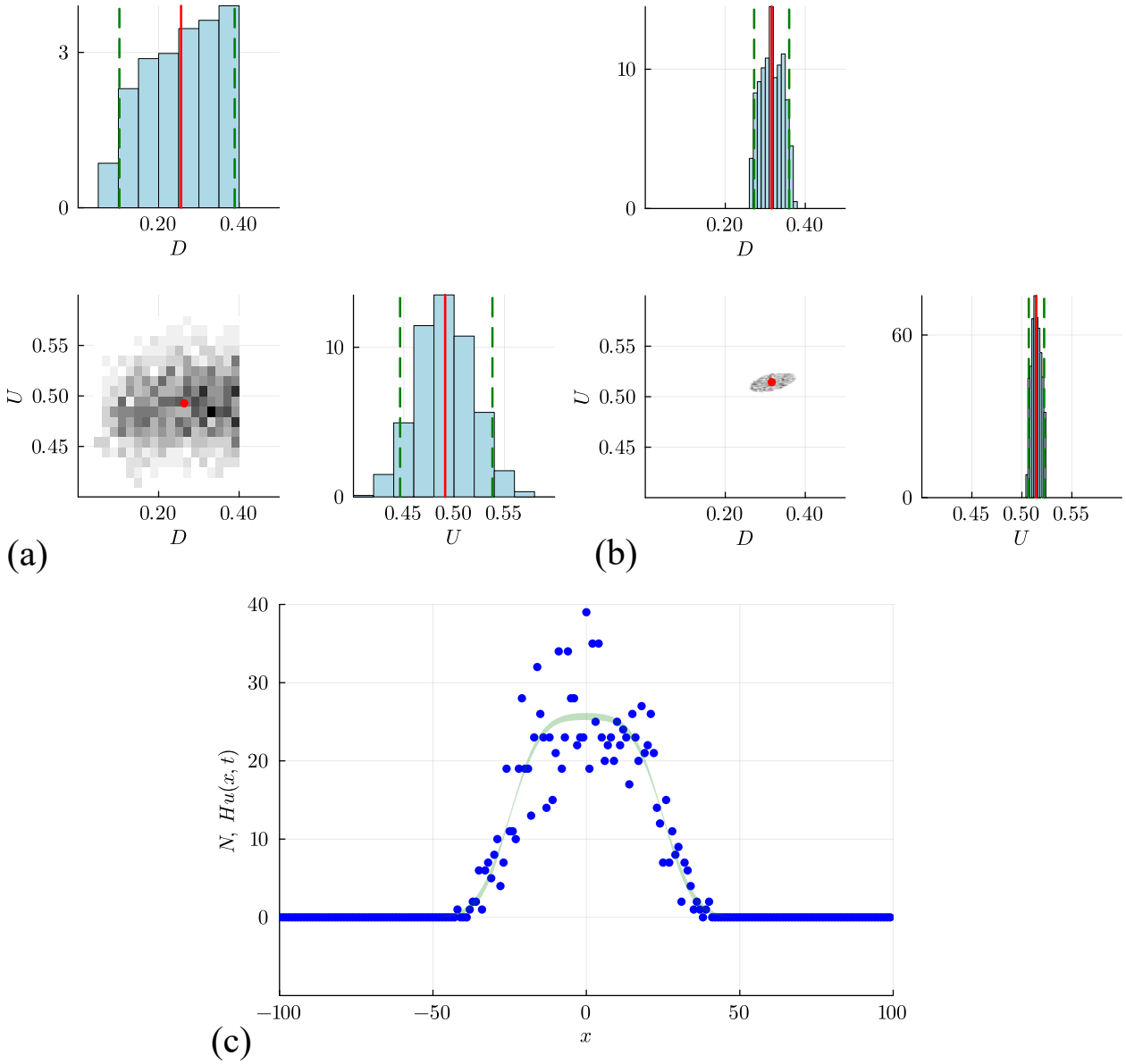

Figure 2: ABC rejection results based on: (a) the discrete simulation model, leading to point estimates of $D = 0.254\,[0.091, 0.394]$ and $U = 0.489\,[0.440, 0.548]$; and, (b) the surrogate PDE model, leading to estimates of $D = 0.317\,[0.269, 0.367]$ and $U = 0.514\,[0.506, 0.523]$. Point estimates correspond to the mean of the posterior samples, and the credible interval is obtained by computing the 0.025 and 0.975 quantiles of the posterior samples. In (c) we superimpose the data, N_i, and the credible interval for N obtained by taking solving the PDE model for each of the $M = 1000$ parameter samples and plotting the point-wise union of these M solutions, $Hu_m(x,t)$ for $m = 1, 2, 3, \ldots, M$, where each solution is evaluated on a uniform discretisation of the domain $-100 \leq x \leq 100$, discretised using 201 mesh points.

Before moving on to a different approach for estimation, we note there are several options available for reporting point estimates from posterior samples, such as those in Figure 2(a)–(b). For simplicity, in this work we report the posterior mean. For example, for the posterior samples in Figure 2(b) the posterior mean for the initial occupancy gives a point estimate of $U = 0.514306$, given here with six decimal places. Alternatively, it could be useful to report the maximum a posteriori (MAP) estimate, giving $U = 0.514445$ for this data. Here, we see that these point estimates are not very different which is consistent with our observation that the posterior distribution is unimodal and approximately symmetric. Therefore, for simplicity, we report the posterior mean, noting that the posterior mean is mathematically related to working with an additive Gaussian noise model [47, 48] that we will now introduce.

3.2 Likelihood-based sampling: Estimation and prediction

We will now turn to a likelihood-based sampling approach for the same inference question of estimating the parameters $\boldsymbol{\theta} = (U, D)^\top$ that best explain the data in Figure 1(f) using the solution of the relevant surrogate model, Equation (5). Unlike the likelihood-free ABC rejection approaches outlined in Section 3.1, here we introduce an observation model, also known as a noise model or measurement error model [49], that relates the solution of the mathematical model at the ith lattice site, $Hu(x_i, t)$, to the observed count data, N_i, by assuming that the data are normally distributed about the solution of the mathematical model with constant variance σ^2. This assumption can be expressed mathematically as

$$N_i \mid \boldsymbol{\theta} = \mathcal{N}(Hu(x_i, t), \sigma^2), \tag{7}$$

at each column of the lattice, $i = 1, 2, 3, \ldots, I$. At this point we have provided no justification to support the standard assumption that the data are normally distributed about the solution of the model [49–53]. For the moment we will proceed under this assumption, and comment on the suitability of this standard approach later. Before proceeding it is worth noting that this assumption introduces another unknown parameter σ^2, and we will estimate this additional parameter from the data. This means that we will work with the expanded vector of parameters $\boldsymbol{\theta} = (U, D, \sigma)^\top$. One advantage of introducing the measurement error model is that we now have an objective way of quantifying how close the solution of the mathematical model is to the data, without the need to specify a distance function like we did in Section 3.1. Here we measure the closeness of the data and the model solution via the probability

density associated with the normal distribution. For a series of measurements at position x_i for $i = 1, 2, 3, \ldots, I$, invoking a standard independence assumption motivates us to take the product of I densities to give the likelihood of observations given the solution of the mathematical model parameterised by $\boldsymbol{\theta}$,

$$\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) = \prod_{i=1}^{I} \phi\left(y_i^{\text{obs}}; Hu(x_i, t), \sigma^2\right), \tag{8}$$

where $\phi(x; \mu, \sigma^2)$ denotes the probability density function of the normal distribution with mean μ and constant variance σ^2 [50, 51]. Under this approach, parameters of the mathematical model that lead to an improved match to the data are associated with a larger value of $\mathcal{L}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}})$. It is convenient to work with the logarithm of the likelihood to avoid working with the product of many extremely small quantities, giving the log-likelihood function

$$\ell(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) = \sum_{i=1}^{I} \log\left[\phi\left(y_i^{\text{obs}}; Hu(x_i, t), \sigma^2\right)\right]. \tag{9}$$

Note that maximising Equation (9) is mathematically identical to minimising the sum of squared differences (or equivalently the mean squared error) for constant σ^2 [48]. This equivalence means that very standard approaches of parameter estimation by minimising a least-squares objective function are equivalent to maximising the log-likelihood function under the assumption of working with an additive Gaussian noise model with constant variance.

Following the approach of Hines et al. [50, 51] we will demonstrate how to take a Bayesian approach to sampling this likelihood function which implicitly relies on Bayes' theorem that we write as

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) \propto p(\boldsymbol{y}^{\text{obs}} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}), \tag{10}$$

where $p(\boldsymbol{y}^{\text{obs}} \mid \boldsymbol{\theta})$ is the likelihood function, $p(\boldsymbol{\theta})$ is the prior and $p(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}})$ is the posterior distribution. Our aim is to estimate the posterior distribution, as it summarises information about the parameter $\boldsymbol{\theta}$ in light of the data $\boldsymbol{y}^{\text{obs}}$ and the prior $p(\boldsymbol{\theta})$. The likelihood function describes information contributed by the data, corresponding to the probabilistic noise model, Equation (7), evaluated at the observed data. We explore the parameter space by sampling the posterior distribution using a Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm [50]. The Markov chain starts at position $\boldsymbol{\theta}^{(0)}$ and a potential move to

$\boldsymbol{\theta}^*$ is accepted with probability

$$\alpha = \min\left[1, \frac{p(\boldsymbol{\theta}^* \mid \boldsymbol{y}^{\text{obs}})}{p(\boldsymbol{\theta}^{(0)} \mid \boldsymbol{y}^{\text{obs}})}\right]. \tag{11}$$

The Markov chain tends to move in the direction of high posterior probability, but also allows movements away from local minima by allowing occasional steps to regions of lower posterior probability [50, 51]. Markov chains produced by this algorithm explore the parameter space in proportion to the posterior probability. Once the MCMC chains have mixed, the sampler has converged to the target posterior distribution and is effectively exploring the full support of that distribution. Under these conditions the MCMC algorithm can be used to generate a finite number of approximately independent, identically distributed samples from the posterior distribution [50, 51]. Proposals in the MCMC algorithm are made by perturbing the current sample point by a draw from the multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Sigma}$. In this example with $\boldsymbol{\theta} = (U, D, \sigma)^\top$ we specify $\boldsymbol{\Sigma} = \text{diag}\,(10^{-4}, 10^{-4}, 10^{-3})^\top$. The entries in the covariance matrix are hyperparameters for the MCMC algorithm, and the choice of these hyperparameters is critical because they can strongly affect the rate of mixing and convergence of the algorithm. In this case with a relatively small number of unknown parameters, we have selected these hyperparameters based on heuristic experience obtained using a number of pilot simulations [50, 51]. For higher dimensional problems it is possible to use adaptive methods to help in the selection of these hyperparameters [54, 55].

With this framework, we generate a Markov chain with 10^4 samples, starting at $\boldsymbol{\theta}^{(0)} = (U, D, \sigma)^\top = (0.1, 0.1, 0.1)^\top$. The MCMC traces in Figure 3 show that samples of $\boldsymbol{\theta}$ rapidly move away from the starting position and appear to settle into a stationary distribution after approximately 500-1000 samples. In this example we have chosen to stop the chain after 10^4 samples as there is no obvious visually detectable drift at this point [50, 51]. The decision regarding when to stop the MCMC chains can be formalised by calculating certain criteria that involve comparing within-chain and between-chain variances [47]. For the purposes of this review we take a standard approach and stop the chain according to a simple visual inspection criteria [50, 51]. For typical applications of MCMC-based parameter estimation, we never know in advance of performing the potentially expensive MCMC simulation whether the chains will eventually mix and converge. If the chains do mix and converge, we never know in advance how many iterations will be required.

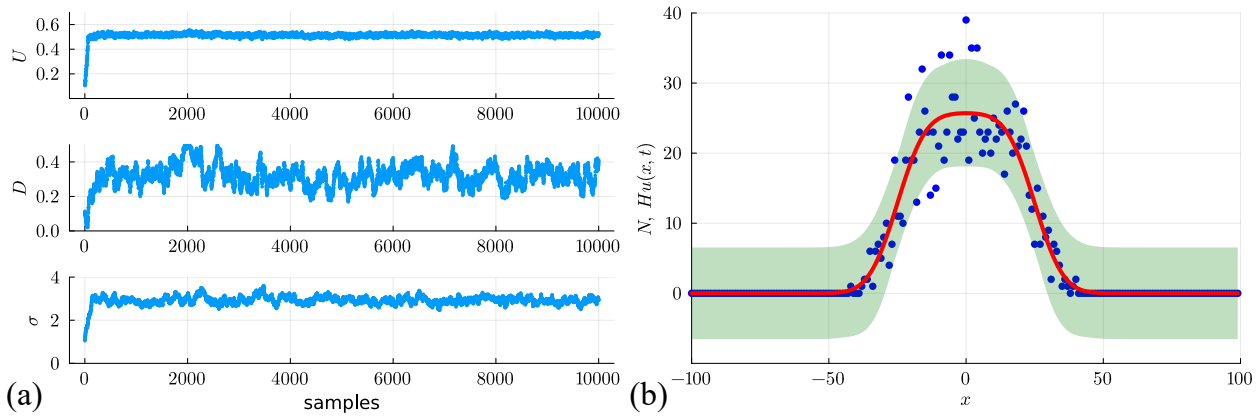

Figure 3: (a) MCMC trace plots for $(U, D, \sigma)^\top$ starting at $\boldsymbol{\theta}^{(0)} = (U, D, \sigma)^\top = (0.1, 0.1, 0.1)^\top$. Proposals are drawn from the multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{\Sigma} = \text{diag}\,(10^{-4}, 10^{-4}, 10^{-3})^\top$. The final 2000 samples are taken as samples from the stationary distribution, and the mean of these samples gives point estimates $(U, D, \sigma)^\top = (0.515, 0.332, 2.959)^\top$. (b) Posterior prediction interval generated using the final $M = 1000$ samples from (a) (green shaded) encloses 94.0% of the stochastic data. This interval is superimposed on the data (blue dots) and the solution of the continuum model $Hu(x, t)$, evaluated at the posterior mean $(U, D, \sigma)^\top = (0.515, 0.332, 2.959)^\top$ (solid red).

To quantify the results of the MCMC chain we take the final 2000 samples from chain and compute various marginal distributions summarised in Figure 4 where we see that each of the univariate posterior marginal distributions are unimodal and relatively narrow about the peak. Accordingly we characterise these distributions by their mean, which gives point estimates of $(U, D, \sigma)^\top = (0.515, 0.332, 2.959)^\top$, confirming that our estimates of U and D are relatively close to the expected result, as well as being consistent with the point estimates and credible intervals obtained using ABC rejection. The MCMC-derived credible intervals are relatively narrow, indicating that our estimates are reasonably precise. Figure 4 also presents histograms illustrating the bivariate marginal distributions between each pair of parameters which gives us a sense of any correlation between parameter pairs. The shapes of the distributions indicate no obvious correlation between D and σ or between U and σ. In contrast there appears to be some degree of positive correlation between D and U, which is consistent with the previous ABC rejection results in Figure 2(b).

Since our MCMC chain seems to have converged to produce relatively narrow posterior distributions, we can consider generating a posterior predictive interval [9, 10] with the additional information provided through incorporating the additive Gaussian noise model and our estimates of σ. To proceed we take the final $M = 1000$ samples from the Markov

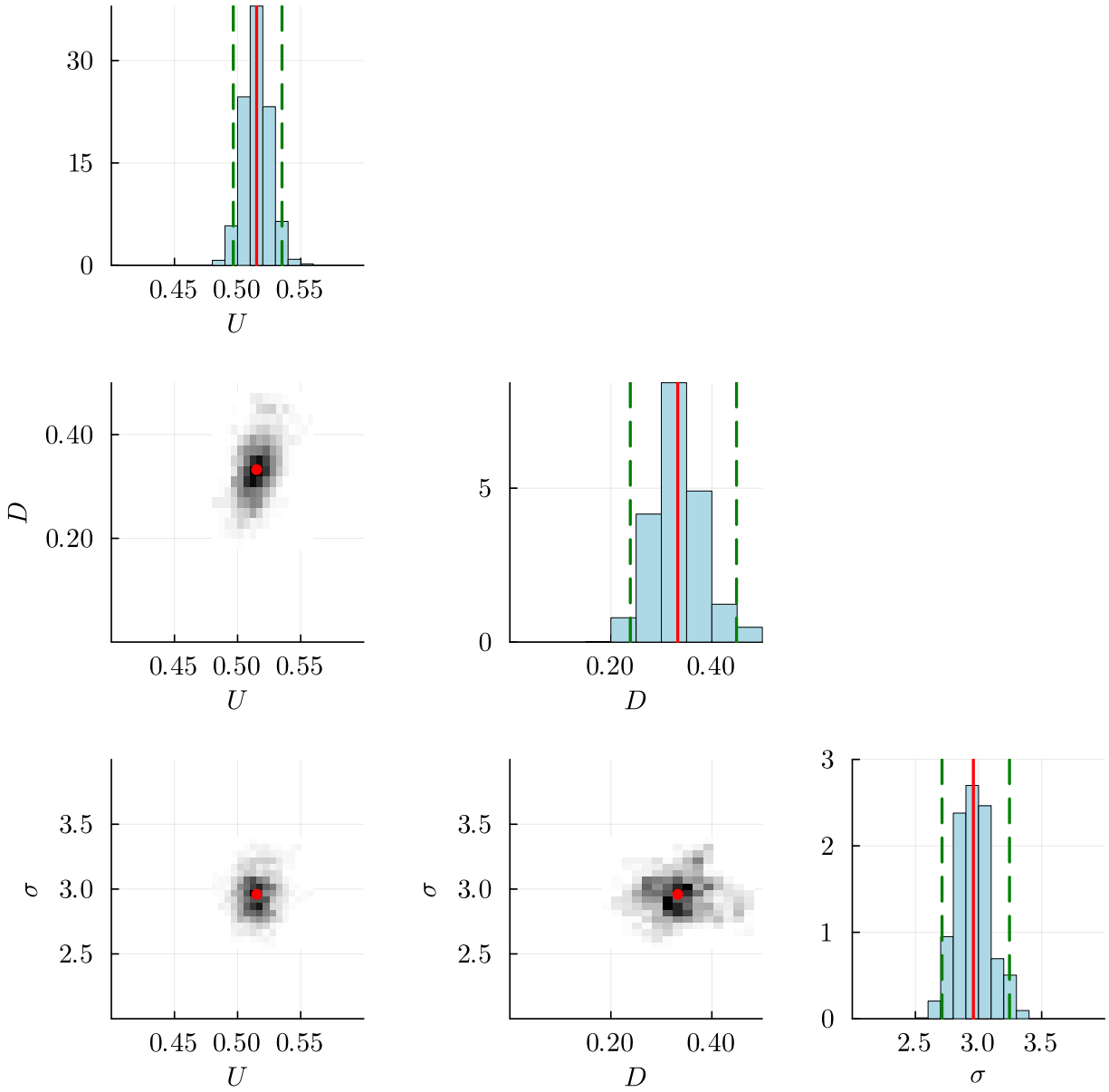

Figure 4: Plot matrix summary constructed from the final 2000 MCMC samples in Figure 3(a). Diagonal entries show histograms of marginal posterior distribution, where each histogram is superimposed with the mean (red vertical lines) and the 0.025 and 0.975 quantiles of the distributions (green dashed vertical lines). These correspond to estimates of $U = 0.515$ $[0.497, 0.535]$; $D = 0.332$ $[0.238, 0.449]$; and, $\sigma = 2.959$ $[2.711, 3.245]$. Off diagonal plots show histograms representing the joint posterior distributions between pairs of parameters with the point estimates superimposed (red dot).

chain in Figure 3, and for each set of parameter values we solve the PDE model, giving M solutions denoted $Hu_m(x,t)$ for $m = 1, 2, 3, \ldots, M$. The quantity $Hu(x,t)$ represents the mean of the measurement error model at position x and time t. With the measurement error model, we no longer need to restrict ourselves to working with the mean of that distribution as we can now incorporate information about the width of the noise model given the estimates of σ from the Markov chain. Therefore, at the ith spatial location, we consider an interval $[Hu_m(x,t)^-, Hu_m(x,t)^+]$, where the lower and upper values correspond to the 0.025 and 0.975 quantiles of the estimated noise distribution, respectively. We construct the posterior predictive interval, as before, except that instead of taking the union of point estimates at each point x_i, we now take the union of the intervals $[Hu_m(x_i,t)^-, Hu_m(x_i,t)^+]$ over $m = 1, \ldots, M$ at each x_i. The resulting prediction interval is given in Figure 3(b) where we also superimpose the data and the solution of the model evaluated at the posterior mean $(U, D, \sigma)^\top = (0.515, 0.332, 2.959)^\top$. The prediction interval takes into account the uncertainty in the model parameters $\boldsymbol{\theta} = (U, D, \sigma)^\top$ as well as the width of the distribution associated with the measurement error model. The solution of the model evaluated at the posterior mean captures key trends in the noisy data, and the prediction interval captures much of the variability of the data. One drawback of working with the additive Gaussian noise model is that the measurement error model has a constant variance, and so the width of the prediction interval is approximately constant for all x. This means that in regions where $u(x,t)$ is sufficiently close to zero the prediction interval leads to a non-physical negative lower bound, $Hu_m(x_i,t)^- < 0$. The fact that the measurement noise model is mis-specified could also bias estimates of the process model parameters D and U [39, 53]. We will describe ways of addressing this issue in Section 3.3.

3.3 Likelihood-based optimisation: Estimation and prediction

We now explore a fourth and final approach to estimate $\boldsymbol{\theta} = (U, D, \sigma)^\top$ from the noisy count data in Figure 1(f). By continuing to work with the additive Gaussian noise model, we have access to the loglikelihood function given by Equation (9) [48]. Unlike the approaches in Sections 3.1–3.2 that rely either on repeated simulations of the stochastic model, or repeated evaluation of the solution of the surrogate model via MCMC sampling, here we take a more direct approach by numerically maximising $\ell(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}})$ to provide point estimates of $\boldsymbol{\theta} = (U, D, \sigma)^\top$. Using a similar approach based on numerical optimisation, we then construct a series of univariate profile loglikelihood functions [48, 56]. This provides a rapid

means of assessing the degree of curvature of the loglikelihood function, which provides key information about inferential precision and parameter identifiability, as we now discuss.

Identifiability is a property which a model must satisfy for precise parameter inference, and identifiability analysis refers to a group of methods that determine how well the model parameters can be estimated given the observed data [48, 57]. Methods of identifiability analysis are typically classified in terms of dealing with *structural* or *practical* identifiability [50]. Structural identifiability analysis determines whether different parameter values generate different probability distributions of the observable variables [58, 59], with the implication that given access to an infinite amount of perfect, noise-free data, it would then be possible to precisely estimate the model parameters. Alternatively, practical identifiability analysis involves working with noisy, sparse data, and exploring the extent to which parameter values can be confidently estimated, given these data. Practical identifiability analysis usually involves fitting a mathematical model to data, and then exploring the extent to which the model fit changes as the parameters vary [48]. A common tool for assessing practical identifiability is the profile likelihood [23, 24, 48, 60].

Instead of repeated sampling of the posterior distribution, our fourth approach to inference and prediction involves applying maximum likelihood estimation (MLE) to obtain the best-fit set of parameters,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left[\ell(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}})\right]. \tag{12}$$

All numerical optimisation computations in this review use the Nelder-Mead algorithm with simple bound constraints [61, 62], and all computations are performed using default stopping criteria within the well-known NLopt family of numerical optimisation routines [63]. Given $\hat{\boldsymbol{\theta}}$, we then work with the normalised loglikelihood function $\bar{\ell}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) = \ell(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) - \ell(\hat{\boldsymbol{\theta}} \mid \boldsymbol{y}^{\text{obs}})$, so that $\bar{\ell}(\hat{\boldsymbol{\theta}} \mid \boldsymbol{y}^{\text{obs}}) = 0$. Working with $\bar{\ell}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}})$ allows us to define asymptotic confidence sets by using Wilks' theorem [64], which states that the likelihood ratio statistic converges in distribution to a χ^2 variable [8, 56, 64]. This result enables us to construct confidence regions and confidence sets, thereby allowing us to quantitatively explore how a noisy set of data can be related to variability in parameter estimates. A key advantage of working with Wilks' theorem is that this theorem holds for arbitrary dimensions and parameter reparameterisations. A limitation of working with Wilks' theorem is that it is an asymptotic result that may not hold when dealing with small sample sizes [8]. In practice, Wilks' theorem allows us to define a likelihood-based threshold, $\bar{\ell}^*$, that defines confidence regions in parameter space $\bar{\ell}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) > \bar{\ell}^*$, and corresponding confidence sets, $\mathcal{C}_\theta = \left\{\boldsymbol{\theta} \in \Theta : \bar{\ell}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) > \bar{\ell}^*\right\}$. We will

explain how to compute values of $\bar{\ell}^*$ later. [8, 56].

To assess parameter identifiability we take a profile likelihood-based approach [22, 56, 60, 65–68] by partitioning $\boldsymbol{\theta}$ into interest parameters ψ, and nuisance parameters ω, so that $\boldsymbol{\theta} = (\psi, \omega)^\mathsf{T}$. For simplicity, here we always take the interest parameter to be a single parameter which allows us to construct various univariate profile likelihood functions. Regardless, for a set of data $\boldsymbol{y}^{\text{obs}}$, the profile loglikelihood for the interest parameter ψ given the partition $(\psi, \omega)^\mathsf{T}$ is

$$\bar{\ell}_{\text{p}}(\psi \mid \boldsymbol{y}^{\text{obs}}) = \max_{\omega|\psi} \bar{\ell}(\psi, \omega \mid \boldsymbol{y}^{\text{obs}}), \tag{13}$$

which implicitly defines a function $\omega^*(\psi)$ of optimal values of ω for each value of ψ, and defines a surface with points $(\psi, \omega^*(\psi))$ in parameter space. The targeting property of the profile likelihood allows us to explore the identifiability of each parameter which is particularly insightful and can be computationally efficient, especially compared to MCMC-based approaches, which normally characterise non-identifiability by poorly mixing MCMC chains [69, 70]. For univariate profiles, $\bar{\ell}_{\text{p}}(\psi \mid \boldsymbol{y}^{\text{obs}})$ is a function of one parameter that can be plotted and visualised in the usual way. For example, considering the current problem with $\boldsymbol{\theta} = (U, D, \sigma)^\top$, we can construct a univariate profile likelihood function for U by setting $\psi = U$ and $\omega = (D, \sigma)^\top$. As for calculating the MLE, profile likelihoods are computed using the Nelder-Mead algorithm to evaluate the necessary numerical optimisation problems [63], and in all cases the iterative optimisation calculations can be initialised using the appropriate components of $\hat{\boldsymbol{\theta}}$, and all profile likelihood functions are computed across an interval of the interest parameter that is uniformly discretised.

For the data in Figure 1(e) and the loglikelihood function given by Equation (9), numerical optimisation gives $\hat{\theta} = (\hat{U}, \hat{D}, \hat{\sigma})^\top = (0.514, 0.314, 2.890)^\top$ within a fraction of a second of computation time. These estimates compare very well those obtained by repeated sampling via the MCMC algorithm in Figure 3. In contrast, however, the optimisation-based estimates can be obtained with several orders of magnitude computational speed-up, and are obtained without the need for specifying hyperparameters for the MCMC parameter proposals. To assess the identifiability of these point estimates, the univariate profile loglikelihood functions for U, D, and σ are given in Figure 5(a). We see that each profile is characterised by a unique maximum at the MLE. The profiles are sufficiently curved that we can construct an approximate confidence interval for each parameter by finding the value of the parameter where the profile loglikelihood intersects with the 95% asymptotic threshold where $\bar{\ell}^* = -1.9207$, as we explain further below. For example, in terms of the initial den-

sity we have $\hat{U} = 0.514$, with an approximate 95% confidence interval of $U \in [0.494, 0.535]$. Just like the MLE calculation, the construction of univariate profile likelihood functions is computationally efficient relative to MCMC sampling [50, 51, 70]. In general, we never know in advance how curved the univariate profile likelihood function will be before we calculate that function. Depending on the degree of curvature of the profile likelihood function, it is possible to conclude that a parameter is practically identifiable, poorly-identified, or one-sided identifiable [71]. This judgement is usually made on a case-by-case basis depending on the width of the approximate confidence interval relative to problem-specific constraints and expectations [72].

Once we have established the normalised loglikelihood function, we define asymptotic confidence sets by identifying a likelihood-based threshold, $\bar{\ell}^* = -\Delta_{q,d}/2$, where $\Delta_{q,d}$ refers to the qth quantile of the χ^2 distribution with d degrees of freedom [8, 64]. The asymptotic confidence set is the set of values of $\boldsymbol{\theta}$ satisfying $\bar{\ell}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) \geq \bar{\ell}^*$. In this work we use a 0.95 quantile threshold for problems with $d = 1, 2, 3$ and 4 degrees of freedom, for which the threshold values are $\bar{\ell}^* = -1.9207, -2.9957, -3.9073$ and -4.7439, respectively. In this context, the number of degrees of freedom corresponds to the number of free parameters [73]. Given $\hat{\boldsymbol{\theta}}$ and confirmation via the profile likelihood function that these estimates are well-identified by the data, we now compute a prediction interval based on these data. We will construct a prediction interval [74] using two methods. First, we generate $M = 1000$ samples of $\boldsymbol{\theta}$ that lie within the approximate 95% confidence region where $\bar{\ell}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) \geq \bar{\ell}^*$. These samples are generated using a simple rejection algorithm where parameters are uniformly randomly sampled within a user-defined region containing $\hat{\boldsymbol{\theta}}$. With these M parameter samples we can solve the mathematical model to give M continuous solution curves $Hu_m(x,t)$ for $m = 1, 2, 3 \ldots, M$. At each spatial location x we compute an interval $[Hu_m(x,t)^-, Hu_m(x,t)^+]$, where the lower and upper values correspond to the 0.025 and 0.975 quantiles of the noise model with mean $Hu(x,t)$, respectively. As before, we take the union of intervals $[Hu_m(x_i,t)^-, Hu_m(x_i,t)^+]$ over $m = 1, 2, 3, \ldots, M$ at each spatial location x_i on a uniform discretisation of the domain, $-100 \leq x \leq 100$ using 201 mesh points to give the green shaded prediction interval in Figure 5(b). As in Section 3.1, we see that the solution evaluated at the MLE captures the main trends and features of the count data, but now we see the value in introducing the noise model as here the prediction interval encloses the stochastic data reasonably well.

Our approach of using uniformly random samples to find M parameter values lying within the region where $\bar{\ell}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) \geq \bar{\ell}^*$ is typically the most computationally-expensive aspect of

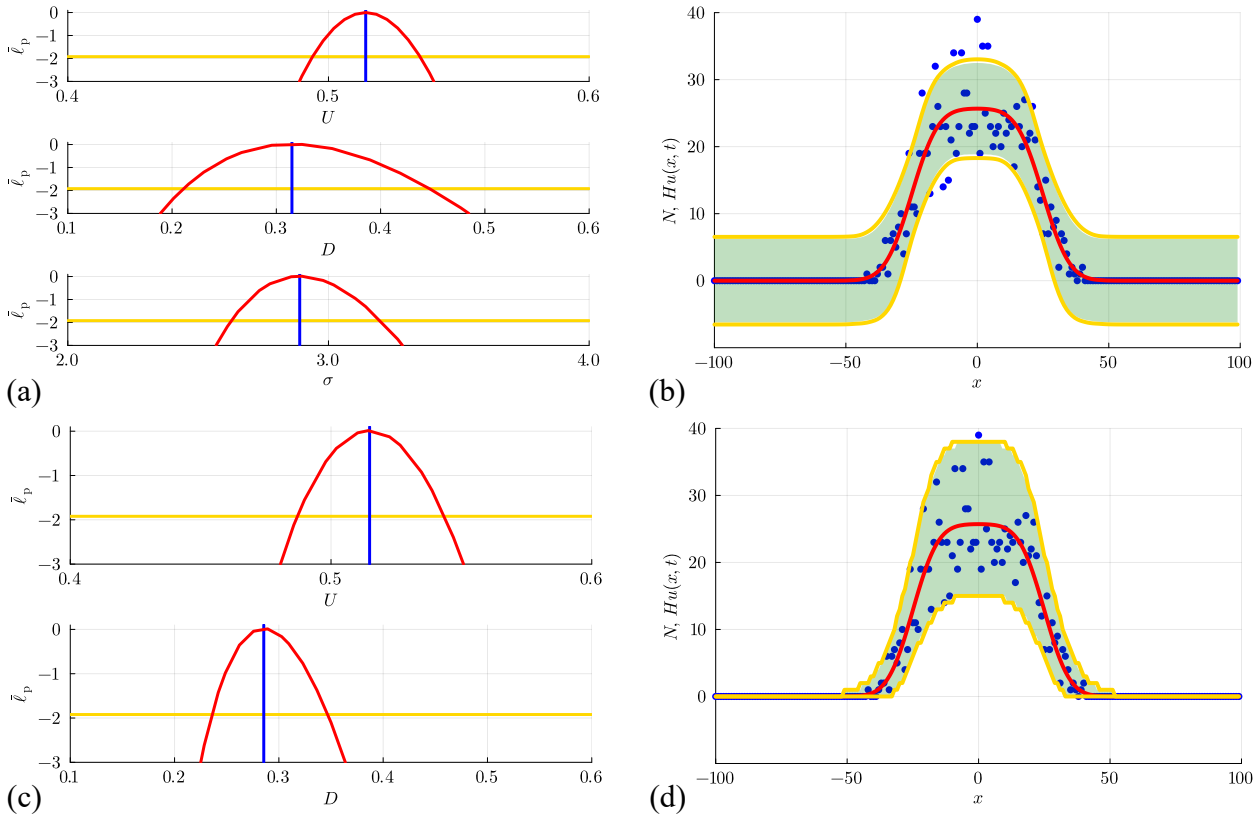

Figure 5: (a)–(b) Estimation, identifiability analysis and prediction under the additive Gaussian noise model. (a) Profiles for U, D and σ with $\hat{\theta} = (\hat{U}, \hat{D}, \hat{\sigma})^\top = (0.514, 0.314, 2.890)^\top$ superimposed (vertical blue lines). Each univariate profile gives an approximate 95% confidence interval: $\hat{U} = 0.514$ $[0.494, 0.535]$; $\hat{D} = 0.314$ $[0.211, 0.446]$; $\hat{\sigma} = 2.890$ $[2.628, 3.198]$ generated by finding the values of the interest parameter where the profile intersects with the 95% threshold (yellow horizontal lines). (b) Prediction interval generated with the additive Gaussian noise model encloses 95.0% of the data. (a)–(b) Estimation, identifiability analysis and prediction under the Poisson noise model. (c) Profiles for U and D with $\hat{\theta} = (\hat{U}, \hat{D})^\top = (0.515, 0.286)^\top$. Each univariate profile gives an approximate 95% confidence interval: $\hat{U} = 0.515$ $[0.487, 0.543]$; $\hat{D} = 0.286$ $[0.236, 0.347]$. (d) Prediction interval generated with the Poisson noise model encloses 99.5% of the data. The prediction intervals in (b) and (d) compare the approach of uniform sampling (solid green region) with the interval generated using Laplace's approximation (solid yellow bounds).

the workflow in Figure 5(a)–(b). The computational overhead of this step depends upon the shape of $\bar{\ell}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}})$, the number of free parameters, and the user-defined size of the region about $\hat{\boldsymbol{\theta}}$ from which random samples are drawn. In our experience, working with this approach requires user-derived expertise in determining the extent of the region about $\hat{\boldsymbol{\theta}}$ from which random samples are drawn. Choosing a region that is too narrow means that the samples do not reach the asymptotic limit given by Wilks' theorem, whereas choosing a region that is too wide leads to very slow sampling.

Another approach involves invoking Laplace's approximation where the shape of the loglikelihood function is approximated with a multivariate normal distribution [9, 10, 75–77]. To implement this approximation we compute the observed Fisher Information at the MLE,

$$\mathcal{I}(\hat{\boldsymbol{\theta}}) = -\nabla\nabla^{\mathsf{T}}\ell\left(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}\right)\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \tag{14}$$

which is a square Hessian matrix, whose elements are the second order partial derivatives of ℓ with respect to the elements of $\boldsymbol{\theta}$. In this study we compute the elements of $\mathcal{I}(\hat{\boldsymbol{\theta}})$ using automatic differentiation [78, 79], which avoids working with finite difference approximations and truncation error [80]. Given our estimates of $\hat{\boldsymbol{\theta}}$ and $\mathcal{I}(\hat{\boldsymbol{\theta}})$, we approximate the shape of the loglikelihood function in the vicinity of the MLE as a multivariate normal distribution, $\text{MVN}(\hat{\boldsymbol{\theta}}, \mathcal{I}^{-1}(\hat{\boldsymbol{\theta}}))$. With this approximation we can draw $M = 1000$ random samples of $\hat{\boldsymbol{\theta}}$ from this multivariate normal distribution. As before, we then calculate the PDE solution $u_m(x,t)$ and construct an interval $[Hu_m(x,t)^-, Hu_m(x,t)^+]$ for each sample $m = 1, \ldots, M$. We then define the prediction interval to be from the 0.025 quantile of $Hu_m(x,t)^-$ to the 0.975 quantile of $Hu_m(x,t)^+$.

It is important to note that drawing samples from a uniform distribution in the region where $\bar{\ell}(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) \geq \bar{\ell}^*$ is different to drawing samples from the multivariate normal distribution given by Laplace's approximation, which has more density close the MLE and some density in the tails of the distribution. In practice we find that the computational effort is significantly lower for the second approach. These two methods give prediction intervals that are very similar, as shown in Figure 5(b). This comparison illustrates that working with Laplace's approximation can give accurate prediction interval estimation with a significantly reduced computational overhead. It should be noted that the intervals defined by both of these methods are technically known as *95%/95% tolerance intervals*, meaning that they are expected to cover 95% of the observation distribution in at least 95% of independent repeats of the same experiment. This is somewhat wider than a 95% prediction interval, which

is expected to contain a single future observation with 95% probability [74]. However, in keeping with much of the literature in this area, we will refer to these as prediction intervals throughout this review.

Before we consider moving on to explore different types of data generated by different random walk models, we briefly re-examine the data in Figure 1(f) using a likelihood-based optimisation approach. In this instance, however, we work with a different noise model to address some of the challenges associated with the commonly-used additive Gaussian noise model underlying results in Figure 5(a)–(b). Given that our data consists of non-negative integer count data, and we express the solution of our continuum limit model as $Hu(x,t)$, which we interpret as the expected number of agents at location x and time t, an alternative and perhaps more natural interpretation of this situation is that the data are drawn from a Poisson distribution whose mean is $Hu(x,t)$, where $u(x,t)$ is the solution of the continuum limit model

$$N_i \mid \boldsymbol{\theta} = \text{Pois}(Hu(x_i,t)), \tag{15}$$

at all locations x_i for $i = 1, 2, 3, \ldots, I$. As we will now explore, there are several advantages, both conceptual and procedural, in working under this Poisson-based noise model. First, by definition, the Poisson distribution has support on non-negative integers which is well-suited to the non-negative integer count data generated by the random walk model or a biological experiment. Second, the Poisson noise model does not involve specifying another parameter which means that here we estimate $\boldsymbol{\theta} = (U, D)^\top$ instead of $\boldsymbol{\theta} = (U, D, \sigma)^\top$ like we did with the additive Gaussian noise model. With this framework numerical optimisation gives $\hat{\theta} = (\hat{U}, \hat{D})^\top = (0.515, 0.286)^\top$, which compares well with the estimates obtained using the additive Gaussian noise model either by sampling or optimisation. To assess the identifiability of these point estimates, the univariate profile loglikelihood functions for U and D are given in Figure 5(c), where we see that each parameter is well-identified by the data as the profiles are characterised by a unique maximum at the MLE, and are sufficiently curved that we can construct an approximate confidence interval for each parameter by estimating the intersection of the univariate profile log-likelihood and the 95% asymptotic threshold [73]. These results are much faster to obtain than the MCMC sampling-based results in Figure 3, as well as being modestly faster to obtain than the optimisation results in Figure 5(a)–(b) for the additive Gaussian noise model since we are dealing with a reduced number of unknown parameters.

Perhaps the most obvious advantage in using the Poisson noise model is clear in terms of the

prediction intervals summarised in Figure 5(d). As before, we obtain the prediction intervals using the same two methods used in Figure 5(b). Both methods give similar results to each other, but working with Laplace's approximation leads to an important computational speed-up. Importantly, regardless of how we sample $\boldsymbol{\theta}$, the Poisson-based prediction intervals are very different from those in Figure 5(b). In the former case we have a prediction interval with constant width that gives the misleading impression that negative count data are possible in regions where $u(x,t)$ is sufficiently close to zero. In contrast, the Poisson-based prediction intervals completely avoids the possibility of negative counts, as the width of the Poisson-based prediction interval increases with $u(x,t)$, and vanishes as $u(x,t) \to 0^+$. This feature gives a much more realistic prediction interval that reflects the fluctuations in the data much more accurately. In this case, the prediction intervals obtained using the additive Gaussian noise model encloses 95.0% of the data, and fails to enclose several data points where the expected counts are high in the centre of the lattice, whereas the Poisson-based prediction interval encloses 99.5% of the data.

In summary, working with the Poisson noise model increases the computational efficiency with which we can estimate parameters, assess their identifiability, and perform model predictions. This approach also increases model fidelity since it leads to predictions that are more realistic and appropriate for count data than making the standard assumption that the data are distributed according to a normal distribution with constant variance. While it is also possible to use the Poisson noise model to repeat the exercise in Figure 3 and use MCMC sampling, we will now move on to a set of new problems by exploring a range of combinations of different random walk models and measurement error models. For the remainder of this review we will focus on the most computationally efficient approaches, namely likelihood-based optimisation approaches for parameter estimation and identifiability analysis, along with invoking Laplace's approximation using automatic differentiation to generate parameter samples and construct prediction intervals.

3.4 Combining different process and noise models

To explore the generality of the approaches considered so far, we will now extend the process model and explore our ability to estimate parameters, understand their identifiability, and make model-based predictions for a range of different models. To keep track of the various combinations of process and noise models we summarise the cases in Table 1. To begin we extend the random walk model used in Sections 3.1–3.3 to now include both biased

motion [20, 81] and agent proliferation [21]. The random walk model is implemented on the same lattice with the same sequential update method. To evolve the system from time t to time $t + \tau$, we consider a random sequential update method where for a population of Q agents, we select Q agents independently at random, one at a time, with replacement. Each selected agent is given the opportunity to move with probability $P \in [0, 1]$. If an agent at site (i, j) is selected to move, the target site is chosen with a directional bias. Lattice sites $(i, j \pm 1)$ are chosen with probability $1/4$, whereas sites $(i \pm 1, j)$ are chosen with probability $(1 \pm \rho)/4$. Here $\rho \in [-1, 1]$ is a constant bias parameter such that setting $\rho = 0$ gives an unbiased random walk, setting $\rho > 0$ means that agents are more likely to step in the positive x-direction, and setting $\rho < 0$ means that agents are more likely to step in the negative x-direction. Once Q potential motility events have been considered, we then select Q agents independently at random, one-at-a-time, with replacement. Each selected agent proliferates with probability $R \in [0, 1]$. If an agent at site (i, j) proliferates we simply place another agent at the same site. Once Q motility events and Q proliferation events have been attempted we update our value of Q and the current time step is completed.

Case	Exclusion	Bias	Growth	Flux	Source	Noise
1: Figure 5(a)-(b)	No	No	No	$J = -D\frac{\partial u}{\partial x}$	$S = 0$	Additive Gaussian
2: Figure 5(c)-(d)	No	No	No	$J = -D\frac{\partial u}{\partial x}$	$S = 0$	Poisson
3: Figure 7(a)-(b)	No	Yes	No	$J = -D\frac{\partial u}{\partial x} + vu$	$S = 0$	Additive Gaussian
4: Figure 7(c)-(d)	No	Yes	No	$J = -D\frac{\partial u}{\partial x} + vu$	$S = 0$	Poisson
5: Figure 9(a)-(b)	No	No	Yes	$J = -D\frac{\partial u}{\partial x}$	$S = ru$	Additive Gaussian
6: Figure 9(c)-(d)	No	No	Yes	$J = -D\frac{\partial u}{\partial x}$	$S = ru$	Poisson
7: Figure 11(a)-(b)	Yes	No	No	$J = -D\frac{\partial u}{\partial x}$	$S = 0$	Additive Gaussian
8: Figure 11(c)-(d)	Yes	No	No	$J = -D\frac{\partial u}{\partial x}$	$S = 0$	binomial
9: Figure 12(a)-(b)	Yes	Yes	No	$J = -D\frac{\partial u}{\partial x} + vu(1-u)$	$S = 0$	Additive Gaussian
10: Figure 12(c)-(d)	Yes	Yes	No	$J = -D\frac{\partial u}{\partial x} + vu(1-u)$	$S = 0$	binomial
11: Figure 13(a)-(b)	Yes	No	Yes	$J = -D\frac{\partial u}{\partial x}$	$S = ru(1-u)$	Additive Gaussian
12: Figure 13(c)-(d)	Yes	No	Yes	$J = -D\frac{\partial u}{\partial x}$	$S = ru(1-u)$	binomial

Table 1: Summary of 12 inference and prediction cases considered in this review.

The mean behaviour of the discrete model is described by a continuous linear PDE that can be written as [20, 21, 31, 32],

$$\frac{\partial u}{\partial t} = D\nabla^2 u - v\frac{\partial u}{\partial x} + ru, \tag{16}$$

where $v = \lim_{\Delta,\tau\to 0} = P\rho\Delta/(2\tau)$ and $r = \lim_{\Delta,\tau\to 0} = R/\tau$, where $u(x, y, t) \geq 0$ denotes the average density of agents at location (x, y) at time t. Again, since our simulations do not involve macroscopic density gradients in the vertical direction the PDE simplifies to

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} - v\frac{\partial u}{\partial x} + ru, \tag{17}$$

for the density $u(x, t) \geq 0$. With our choice of initial conditions, and simulations over an intermediate time scale where agents do not touch the vertical boundaries at $x = \pm L$, we can approximate the solution of Equation (17) with the appropriate infinite domain solution [40]

$$u(x, t) = \frac{U\exp(rt)}{2}\left(\text{erf}\left[\frac{h - (x - vt)}{2\sqrt{Dt}}\right] + \text{erf}\left[\frac{h + [x - vt]}{2\sqrt{Dt}}\right]\right). \tag{18}$$

To illustrate we present a simulation in Figure 6 where we have the same initial condition as before, but the population evolves with $P = 1$, $\rho = 0.5$ and $R = 0$, which is associated with $D = v = 0.25$ and $r = 0$. The drift in the positive x-direction is visually obvious in the distribution of agents in Figure 6(b) as well as being reflected in the count data in Figure 6(d). As for the unbiased simulation in Figure 1, the solution of the mean-field PDE model, Equation (18), provides a good match to the mean count data, but does not capture the stochastic fluctuations generated by the stochastic model.

We first attempt to use data in Figure 6 to estimate parameters, examine parameter identifiability, and construct prediction intervals by making the standard assumption that the count data is normally distributed about $Hu(x, t)$ with constant variance. Numerical optimisation gives $\hat{\theta} = (\hat{U}, \hat{D}, \hat{v}, \hat{\sigma})^\top = (0.508, 0.298, 0.256, 2.488)^\top$, where the estimates of U, D and v that are close to the expected values. Four univariate profile likelihood functions in Figure 7(a) confirm that these parameters are well-identified by these data. The resulting prediction interval in Figure 7(b) encloses 95.0% of the noisy data, but again suffers from the limitation that the additive Gaussian noise model with constant variance leads to prediction intervals that include negative counts in regions where $u(x, t)$ is sufficiently close to zero, and fails to enclose some of the data in regions where $u(x, t)$ is sufficiently large.

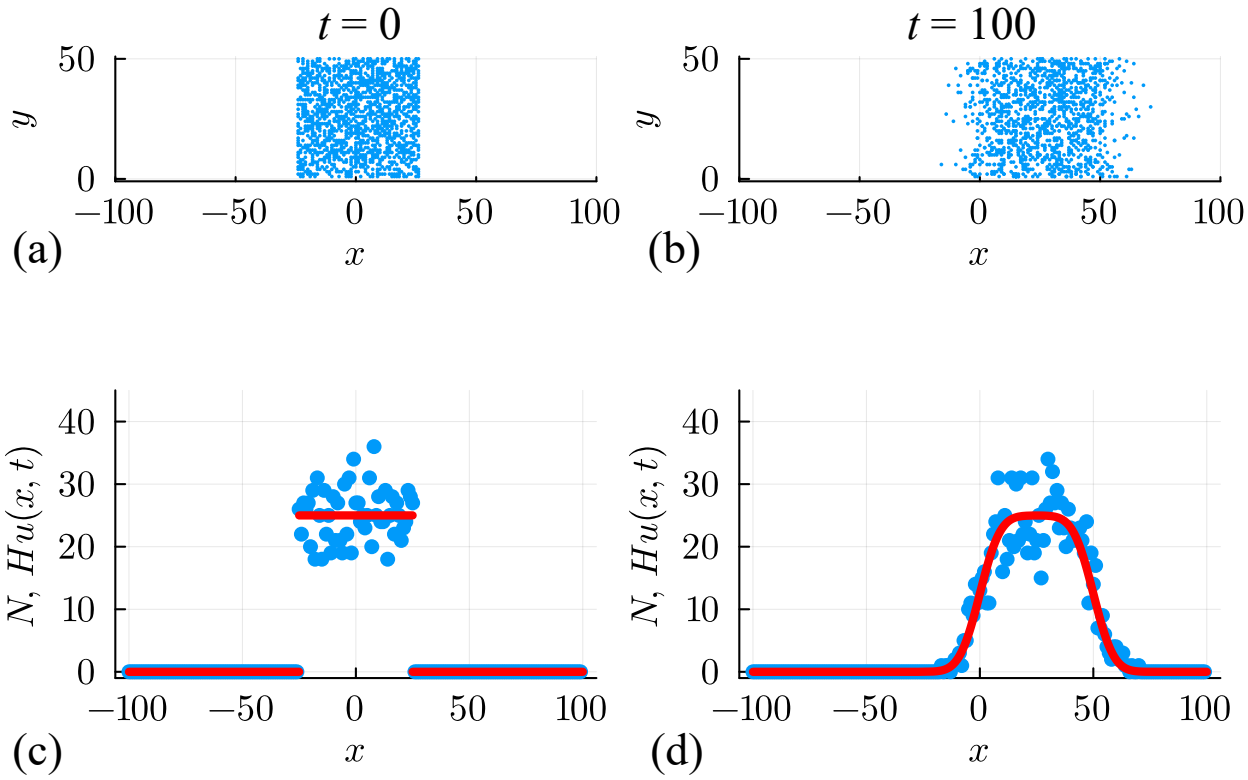

Figure 6: (a)–(b) Biased non-interacting random walk simulation on a square lattice with $\Delta = \tau = 1$, $H = 50$ and $L = 100$. The initial placement of agents is shown in (a), and the outward spreading of agents with a drift in the positive x-direction after 100 time steps is shown in (b). The initial placement of agents is the same as in Figure 1, and random agent migration takes place with $P = 1$ and $\rho = 0.5$. (c)–(d) Numbers of agents per column at $t = 0$ and $t = 100$, respectively. Noisy count data (blue dots) is superimposed with $Hu(x, t)$, where $u(x, t)$ is the solution of Equation (17) with $D = P/4$, $v = P\rho/2$ and $r = 0$ (red curve).

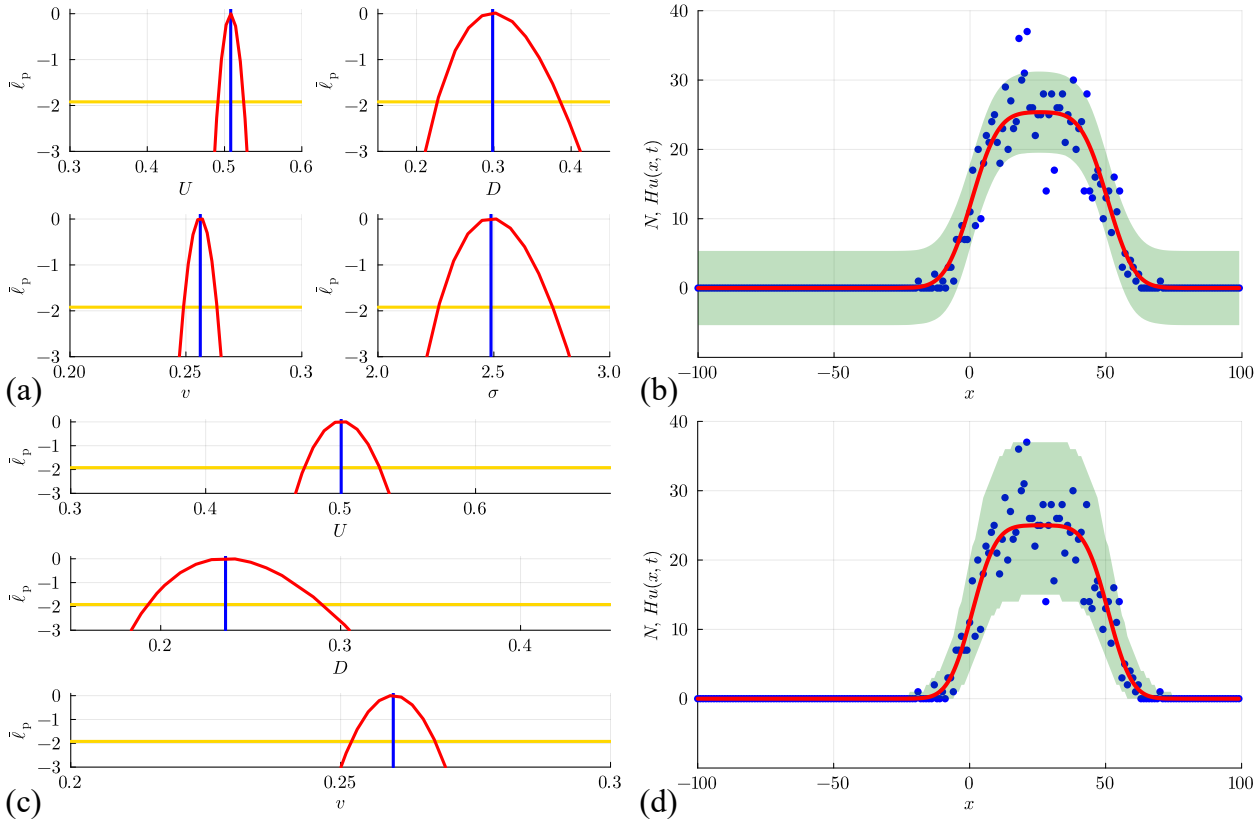

Figure 7: (a)–(b) Estimation, identifiability analysis and prediction under the additive Gaussian noise model. (a) Profiles for U, D, v and σ with $\hat{\theta} = (\hat{U}, \hat{D}, \hat{v}, \hat{\sigma})^\top = (0.508, 0.298, 0.256, 2.488)^\top$ superimposed (vertical blue lines). Each univariate profile gives an approximate 95% confidence interval: $\hat{U} = 0.508$ $[0.491, 0.525]$; $\hat{D} = 0.298$ $[0.227, 0.386]$; $\hat{v} = 0.256$ $[0.249, 0.263]$; $\hat{\sigma} = 2.488$ $[2.262, 2.753]$ generated by finding the values of the interest parameter where the profile intersects with the 95% threshold (yellow horizontal lines). (b) Prediction interval generated with the additive Gaussian noise model encloses 95.0% of the data. (c)–(d) Estimation, identifiability analysis and prediction under the Poisson noise model. (c) Profiles for U, D and v with $\hat{\theta} = (\hat{U}, \hat{D}, \hat{v})^\top = (0.500, 0.236, 0.260)^\top$. Each univariate profile gives an approximate 95% confidence interval: $\hat{U} = 0.500$ $[0.473, 0.529]$; $\hat{D} = 0.236$ $[0.192, 0.290]$; $\hat{v} = 0.260$ $[0.252, 0.267]$. (d) Prediction interval generated with the Poisson noise model encloses 99.0% of the data. Prediction intervals in (b) and (d) are generated using Laplace's approximation.

Further results in Figure 7(c)–(d) repeat the same estimation, identifiability, and prediction steps under the Poisson noise model, leading to $\hat{\theta} = (\hat{U}, \hat{D}, \hat{v})^\top = (0.500, 0.236, 0.260)^\top$, which is close to the expected result, and each univariate profile likelihood function indicates that all three parameters are practically identifiable. The prediction interval in Figure 7(d) reflects the positive count data more accurately than the results obtained with the additive Gaussian noise model and encloses 99.0% of the data.

An additional simulation in Figure 8 considers the same initial condition as before, except that now agents undergo unbiased movement and proliferation with $P = 1$, $\rho = 0$, and $R = 0.01$, which is associated with $D = 0.25$ and $r = 0.01$. The population growth is clear in the count data in Figure 6(d), where we see that $Hu(x,t) > Hu(x,0) = 50$ near $x = 0$ after 100 time steps. Estimates of $Hu(x,t)$ obtained using Equation (18) provide a good match to the mean count data without capturing stochastic fluctuations. In this case with proliferation we will treat U as a known quantity and aim to estimate D, r and σ, and we will discuss this choice later in Section 6. We estimate parameters, test their identifiability and construct prediction intervals under the standard assumption that count data are normally distributed about $Hu(x,t)$ with constant variance, giving $\hat{\theta} = (\hat{D}, \hat{r}, \hat{\sigma})^\top = (0.303, 0.0103, 4.694)^\top$. Three univariate profile likelihood functions in Figure 9(a) confirm that these parameters are well-identified by this data. The prediction interval in Figure 9(b) encloses 93.5% of the data and suffers from the same limitation as before that the prediction intervals include negative counts in regions where $u(x,t)$ is sufficiently close to zero, and fails to enclose some of the data in regions where $u(x,t)$ is sufficiently large.

Results in Figure 9(c)–(d) repeat the same estimation, identifiability and prediction steps under the Poisson noise model, leading to $\hat{\theta} = (\hat{D}, \hat{r})^\top = (0.2386, 0.0103)^\top$, which is close to the expected result with each univariate profile likelihood function indicating that both parameters are practically identifiable. The prediction interval in Figure 9(d) reflects the positive count data more accurately than the additive Gaussian noise model approach, enclosing 99.0% of the data.

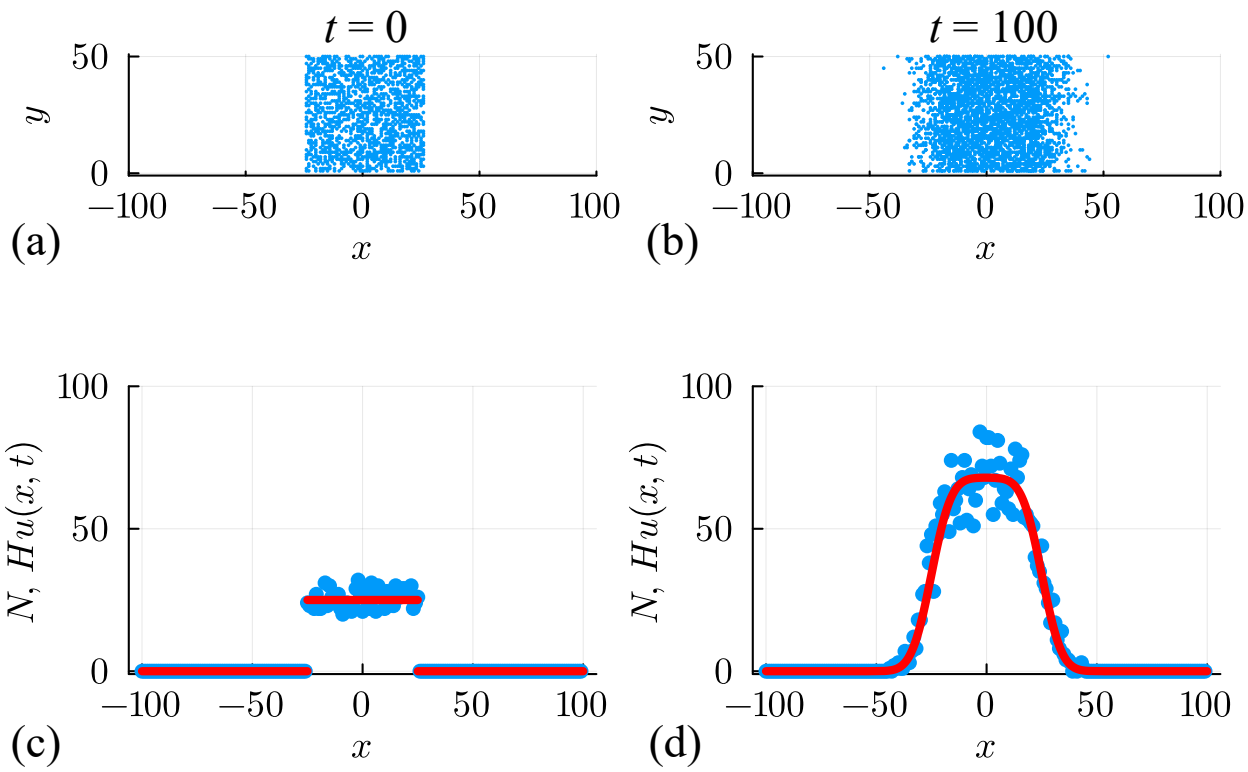

Figure 8: (a)–(b) Unbiased non-interacting random walk simulation with growth on a square lattice with $\Delta = \tau = 1$, $H = 50$ and $L = 100$. The initial placement of agents is shown in (a), and the combined spreading and growth of the population after 100 time steps is shown in (b). The initial placement of agents is the same as in Figure 1, and random agent migration and growth takes place with $P = 1$ and $R = 0.01$. (c)–(d) Numbers of agents per column at $t = 0$ and $t = 100$, respectively. Noisy count data (blue dots) is superimposed with $Hu(x,t)$, where $u(x,t)$ is the solution of Equation (17) with $D = P/4$, $v = 0$ and $r = R$ (red curve).

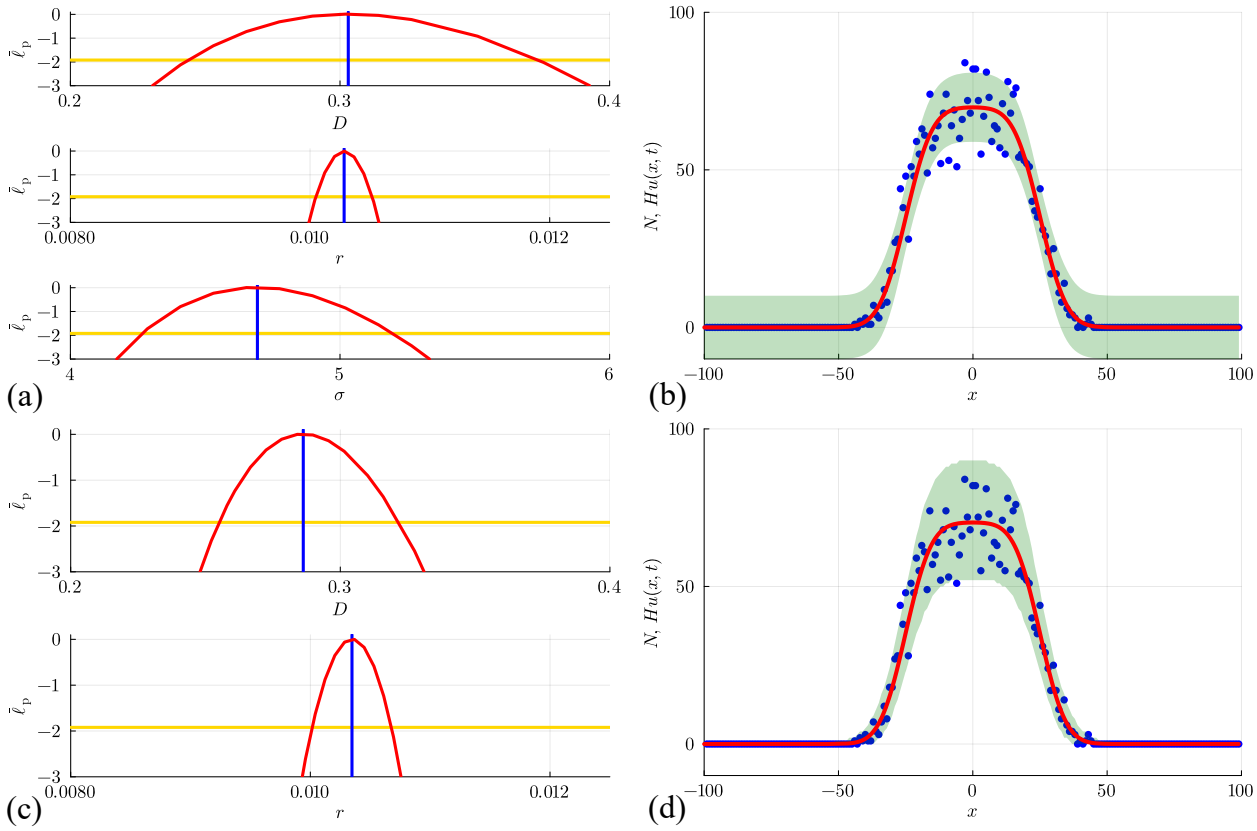

Figure 9: (a)–(b) Estimation, identifiability analysis and prediction under the additive Gaussian noise model. (a) Profiles for D, r and σ with $\hat{\theta} = (\hat{D}, \hat{r}, \hat{\sigma})^\top = (0.303, 0.0103, 4.694)^\top$ superimposed (vertical blue lines). Each univariate profile gives an approximate 95% confidence interval: $\hat{D} = 0.303$ $[0.244, 0.373]$; $\hat{r} = 0.0103$ $[0.0100, 0.0105]$; $\hat{\sigma} = 4.694$ $[4.269, 5.194]$ generated by finding the values of the interest parameter where the profile intersects with the 95% threshold (yellow horizontal lines). (b) Prediction interval generated with the additive Gaussian noise model encloses 93.5% of the data. (c)–(d) Estimation, identifiability analysis and prediction under the Poisson noise model. (c) Profiles for D and r with $\hat{\theta} = (\hat{D}, \hat{r})^\top = (0.286, 0.0103)^\top$. Each univariate profile gives an approximate 95% confidence interval: $\hat{D} = 0.286$ $[0.255, 0.321]$; $\hat{r} = 0.0103$ $[0.0100, 0.0106]$. (d) Prediction interval generated with the Poisson noise model encloses 99.0% of the data. Prediction intervals in (b) and (d) are generated using Laplace's approximation.

4 Interacting random walk models: Crowding effects

All results so far consider a classical non-interacting random walk model, where there is no limit to the number of agents that can occupy a lattice site. We now consider additional results for a very different model where individuals interact in a very simple way that can be taken as a simple model of crowding. These models are closely related to exclusion processes [33], where lattice site occupancy is limited to one agent per site. These models are often used in cell biology applications where cell-to-cell crowding effects are known to be very important [29, 36, 37, 82–89]. Similar interacting random walk models are also used in the context of ecological modelling applications [90–94]. The interacting random walk model is implemented on the same two-dimensional square lattice and we use the same random sequential update method [33]. During the updating algorithm, each selected agent is given the opportunity to move with probability $P \in [0, 1]$, and the target site is chosen with directional bias so that a motile agent at site (i, j) will attempt to step to site $(i, j \pm 1)$ with probability 1/4, and to sites $(i \pm 1, j)$ with probability $(1 \pm \rho)/4$. Any potential motility event that would place an agent on an occupied site is aborted. Once the Q potential motility events have been assessed, we then use the same random sequential update method to simulate agent proliferation. Each selected agent is given the opportunity to proliferate with probability $R \in [0, 1]$. A proliferative agent at site (i, j) will place a daughter agent at sites $(i \pm 1, j)$ and $(i, j \pm 1)$ with equal probability 1/4. Any attempted proliferation event that would place a daughter agent on an occupied site is aborted.

We now consider a suite of discrete simulations using this interacting random walk model on lattice with $H = 50$ and $W = 200$. As before, zero flux boundary conditions are applied along all boundaries, and simulations are initialised by randomly occupying each site within the region where $|x| \leq 25$ with probability $U = 0.5$, leading to the distribution of agents shown in Figure 10(a). A single simulation is performed by allowing the system to evolve over 100 time steps with $P = 1$, $\rho = R = 0$ with the output of the simulation leading to a population that spreads symmetrically in the positive and negative x-directions, as shown in Figure 10(b). This initial and final distribution of agents is then summarised by plotting N_i at $t = 0$ and $t = 100$ in Figure 10(c)–(d), respectively. A key feature of these simulations is that $N_i \leq H$ owing to exclusion. A second simulation with $P = 1$, $\rho = 0.5$ and $R = 0$ is summarised in Figure 10(e)–(h) where we see the impact of the biased motion and exclusion. The spatial distribution of agents at $t = 100$ is asymmetric since motion in the positive x-direction is restricted by other agents in the population. The shape of the resulting distribution is

fundamentally different to the biased simulation for the non-interacting random walk model shown previously in Figure 6. A third simulation with $P = 1$, $\rho = 0$ and $R = 0.01$ is summarised in Figure 10(i)–(l). This shows the impact of the growth and exclusion since the distribution of agents at $t = 100$ is restricted by exclusion and the total number of agents per column can be no greater than $H = 50$ in this case. The shape of the resulting distribution is fundamentally different to the non-interacting simulation with growth shown previously in Figure 8 where in the former case the number of agents per column is unlimited, whereas here exclusion enforces the condition $0 \leq N_i \leq H$.

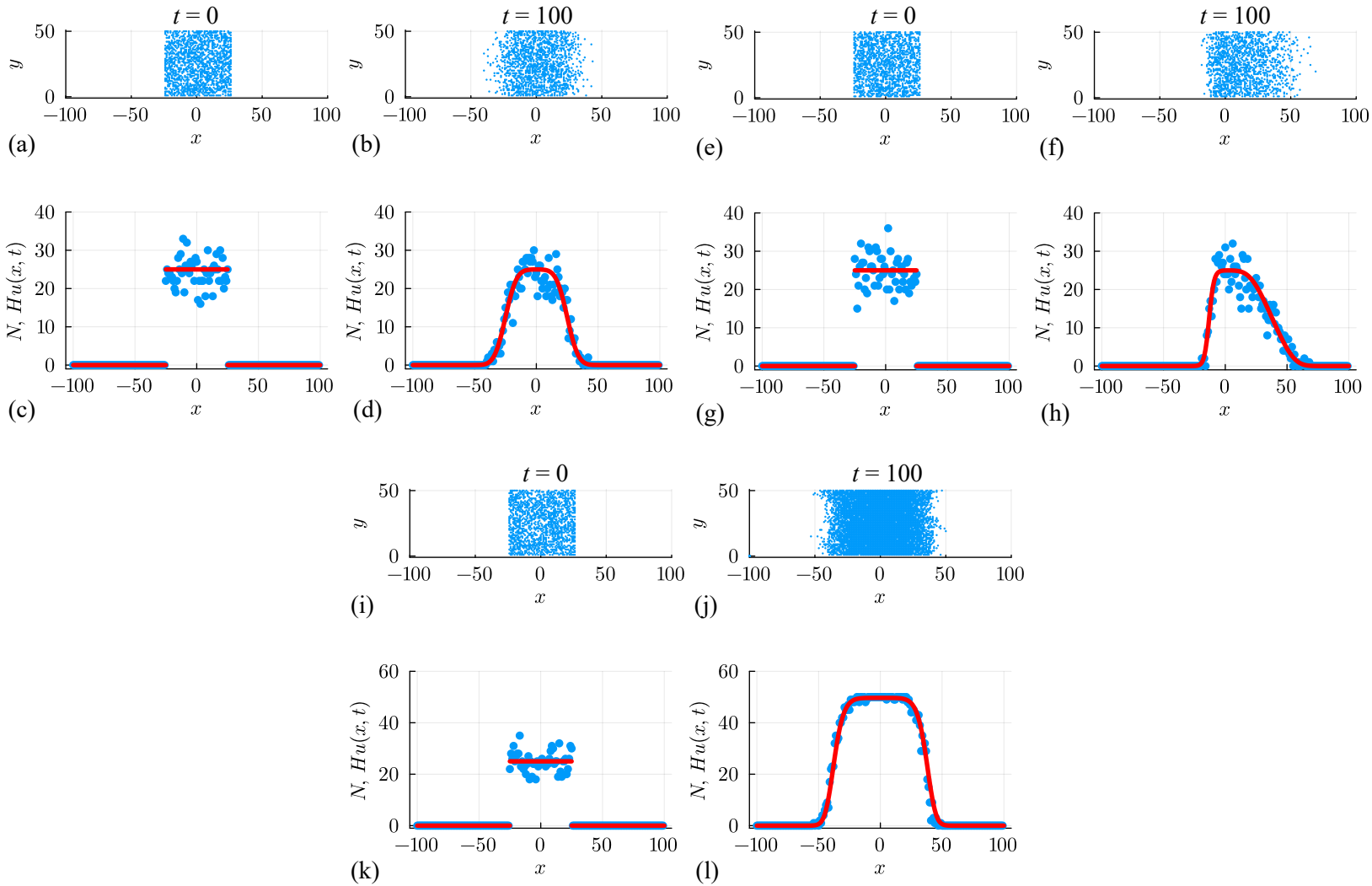

Figure 10: A suite of interacting random walk simulations on a square lattice with $\Delta = \tau = 1$, $H = 50$ and $L = 100$. Simulation results in (a)–(d), (e)–(h) and (i)–(l) correspond to: (i) an unbiased random walk $(P, \rho, R)^\top = (1, 0, 0)^\top$, (ii) a biased random walk $(P, \rho, R)^\top = (1, 0.5, 0)^\top$; and, (iii) an unbiased random walk with growth $(P, \rho, R)^\top = (1, 0, 0.01)^\top$. In each case, the initial placement of agents in (a),(e),(i) involves all sites with $|x| \leq 25$ being occupied by a single agent, at most, with probability $U = 0.5$, and the resulting arrangement of agents in (b),(f),(j) shows the resulting occupancy of sites at $t = 100$. Noisy count data (blue dots) in (c)–(d), (g)–(h), (k)–(l) is superimposed with $Hu(x, t)$, where $u(x, t)$ is the solution of Equation (20) with $D = P/4$, $v = P\rho/2$ and $r = R$ (red curves).

We quantify the interacting random walk simulations in a similar way as before by computing N_i, noting that now we have the restriction that $0 \leq N_i \leq H$. The mean-field description can be obtained using the same conservation approach, giving [21]

$$\frac{\partial u}{\partial t} = D\nabla^2 u - v\frac{\partial}{\partial x}\left[u(1-u)\right] + ru(1-u), \tag{19}$$

where $D = \lim_{\Delta,\tau\to 0} = P\Delta^2/(4\tau)$, $v = \lim_{\Delta,\tau\to 0} = P\rho\Delta/(2\tau)$ and $r = \lim_{\Delta,\tau\to 0} = R/\tau$, with $u(x,y,t) \geq 0$ denoting the average density of agents at location (x,y) and time t. For simulations that do not involve macroscopic density gradients in the vertical direction the PDE simplifies to

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2} - v\frac{\partial}{\partial x}\left[u(1-u)\right] + ru(1-u), \tag{20}$$

for the density $u(x,t) \geq 0$. For simulations in Figure 10(a)–(d) with $\rho = R = 0$, this PDE relaxes to the linear diffusion equation and we can work with the same exact solution used previously, Equation (5). For simulations in Figure 10(e)–(l) with $\rho \neq 0$ and/or $R \geq 0$, we solve Equation (20) using a standard method-of-lines numerical approach, with details outlined in the Appendix. Numerical solutions of this nonlinear PDE model are superimposed on distributions of N_i in Figure 10(g)–(h) and Figure 10(k)–(l), where we see that the solution of the nonlinear PDE model captures the mean trends in the simulation data quite well, but without capturing stochastic fluctuations.

Results in Figure 11(a)-(b) deal with the simulation data from Figure 10(a)-(d) for the interacting random walk model without bias or growth. Univariate profiles are superimposed with $\hat{\theta} = (\hat{U}, \hat{D}, \hat{\sigma})^\top = (0.488, 0.258, 1.904)^\top$ are shown, together with the resulting prediction intervals under the standard assumption that $N_i \mid \boldsymbol{\theta} = \mathcal{N}(Hu(x_i,t), \sigma^2)$. In summary, we see that the MLE point estimate is close to the expected result, and the profiles indicate that the parameters are well identified by the data since we have well-defined, relatively narrow confidence intervals. As with previous implementations of the additive Gaussian noise model, the prediction interval encloses the stochastic data reasonably well, but leads to non-physical predictions of negative counts.

For the exclusion process model where each column has $0 \leq N_i \leq H$, it is more natural to assume that the data N_i are drawn from a binomial distribution with H trials and success probability $u(x_i,t)$. For count data from the ith column we have

$$\mathcal{L}\left(\boldsymbol{\theta} \mid y_i^{\text{obs}}\right) \propto u(x_i,t)^{N_i}\left(1-u(x_i,t)\right)^{H-N_i}, \tag{21}$$

which, after invoking a standard independence assumption when considering a group of measurements at different location, x_i for $i = 1, 2, 3, \ldots, I$, gives a loglikelihood function

$$\ell\left(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}\right) \propto \sum_{i=1}^{I} \log\left[u(x_i,t)^{N_i}\left(1-u(x_i,t)\right)^{H-N_i}\right], \tag{22}$$

up to a constant. To evaluate this loglikelihood function we assume the proportionality constant is unity [39]. Working with this binomial loglikelihood function does not introduce any additional parameters. Results in Figure 11(c)-(d) show univariate profiles, superimposed with $\hat{\theta} = (\hat{U}, \hat{D})^{\top} = (0.490, 0.272)^{\top}$, and the resulting prediction intervals. The MLE estimate is close to the expected result, and the profiles indicate that the parameters are well identified with well-defined, relatively narrow confidence intervals. The binomial loglikelihood function leads to a physically meaningful prediction interval that encloses 99.5% of the data, where the width of the prediction interval vanishes as $u(x,t) \to 0^{+}$ which precludes non-physical negative predictions.

To conclude we consider two final sets of results corresponding to discrete simulation data from the interacting random walk model with biased motion without growth in Figure 12, and for unbiased motion with growth in Figure 13. These results correspond to the simulation data associated with the snapshots in Figure 10(e)–(h), and Figure 10(i)–(l), respectively. In both sets of simulations, our count data is restricted to $0 \leq N_i \leq H$ for each $i = 1, 2, 3, \ldots, I$, and the corresponding solution of Equation (20) must be obtained numerically. In both cases the main features of the noisy count data are captured quite well by the numerical solution of Equation (20). Unlike the non-interacting results with growth, shown in Figure 9 where N_i is unbounded above, results in Figure 13 clearly show the impact of having one agent per site since by $t = 100$ the central columns near to $x = 0$ have $N_i = H = 50$ and the solution of the mean-field model appears to have approached a moving front. This is consistent with the observation that setting $v = 0$ in Equation (20) leads to the Fisher-Kolmogorov model which is well-known to have constant speed and constant shape travelling wave solutions [95].

Figures 12(a)–(b) and 13(a)–(b) illustrate the outcome of likelihood-based estimation, identifiability, and prediction analysis under the standard assumption that count data are normally distributed about the solution of the mean-field model, $Hu(x,t)$. While all parameters are identifiable, and prediction intervals enclose most of the stochastic data, both prediction intervals lead to non-physical outcomes. In particular, the prediction interval in Figure 12(b) indicates that negative counts are enclosed by the prediction interval, whereas the prediction

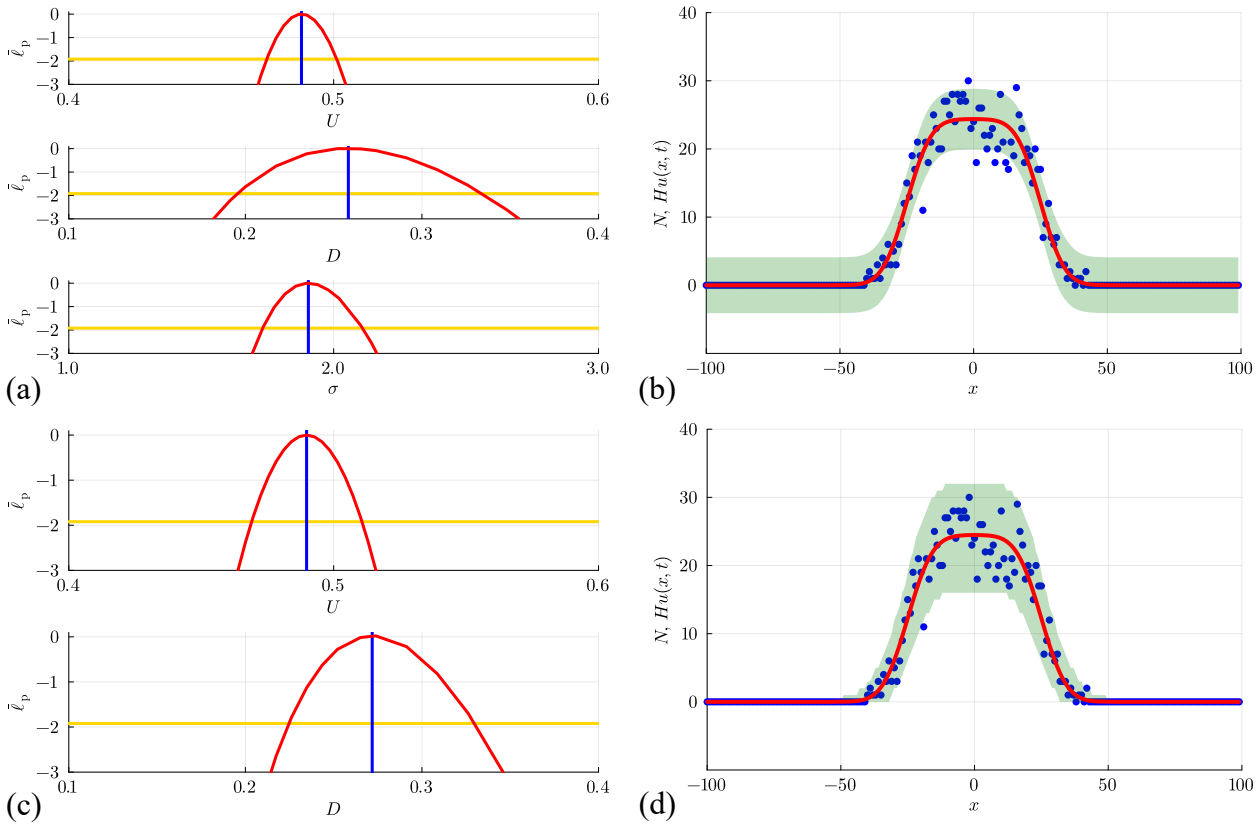

Figure 11: (a)–(b) Estimation, identifiability analysis and prediction under the additive Gaussian noise model. (a) Profiles for U, D and σ with $\hat{\theta} = (\hat{U}, \hat{D}, \hat{\sigma})^\top = (0.488, 0.258, 1.904)^\top$ superimposed (vertical blue lines). Each univariate profile gives an approximate 95% confidence interval: $\hat{U} = 0.488$ $[0.475, 0.501]$; $\hat{D} = 0.258$ $[0.196, 0.334]$; $\hat{\sigma} = 1.904$ $[1.732, 2.107]$ generated by finding the values of the interest parameter where the profile intersects with the 95% threshold (yellow horizontal lines). (b) Prediction interval generated with the additive Gaussian noise model encloses 95.5% of the data. (c)–(d) Estimation, identifiability analysis and prediction under the binomial noise model. (c) Profiles for U and D with $\hat{\theta} = (\hat{U}, \hat{D})^\top = (0.490, 0.272)^\top$. Each univariate profile gives an approximate 95% confidence interval: $\hat{U} = 0.490$ $[0.469, 0.511]$; $\hat{D} = 0.272$ $[0.225, 0.329]$. (d) Prediction interval generated with the binomial noise model encloses 99.5% of the data. Prediction intervals in (b) and (d) are generated using Laplace's approximation.

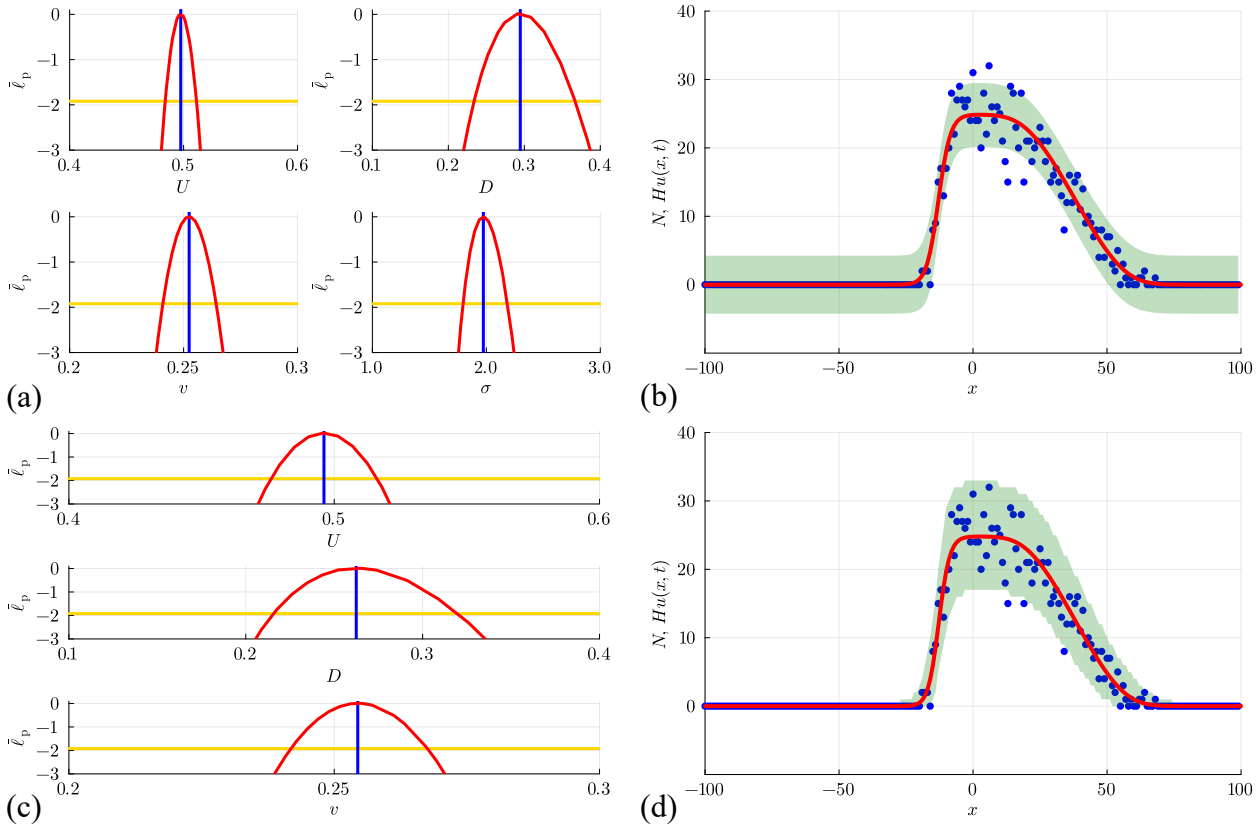

Figure 12: (a)–(b) Estimation, identifiability analysis and prediction under the additive Gaussian noise model. (a) Profiles for U, D, v and σ with $\hat{\theta} = (\hat{U}, \hat{D}, \hat{v}, \hat{\sigma})^{\top} = (0.498, 0.295, 0.252, 1.974)^{\top}$ superimposed (vertical blue lines). Each univariate profile gives an approximate 95% confidence interval: $\hat{U} = 0.498$ $[0.484, 0.511]$; $\hat{D} = 0.295$ $[0.234, 0.367]$; $\hat{v} = 0.252$ $[0.241, 0.264]$; $\hat{\sigma} = 1.974$ $[1.795, 2.184]$ generated by finding the values of the interest parameter where the profile intersects with the 95% threshold (yellow horizontal lines). (b) Prediction interval generated with the additive Gaussian noise model encloses 94.5% of the data. (c)–(d) Estimation, identifiability analysis and prediction under the binomial noise model. (c) Profiles for U, D and v with $\hat{\theta} = (\hat{U}, \hat{D}, \hat{v})^{\top} = (0.496, 0.262, 0.254)^{\top}$. Each univariate profile gives an approximate 95% confidence interval: $\hat{U} = 0.496$ $[0.476, 0.516]$; $\hat{D} = 0.262$ $[0.216, 0.319]$; $\hat{v} = 0.254$ $[0.242, 0.268]$. (d) Prediction interval generated with the binomial noise model encloses 98.0% of the data. Prediction intervals in (b) and (d) are generated using Laplace's approximation.

interval in Figure 13(b) indicates that negative counts, and counts that exceed one agent per site, are enclosed by the prediction interval.

Figures 12(c)–(d) and 13(c)–(d) illustrate the outcome of likelihood-based estimation, identifiability and prediction analysis under the assumption that count data are binomially distributed with mean $Hu(x,t)$. Under this noise model all parameters are identifiable, prediction intervals enclose most of the stochastic data, and both prediction intervals avoid non-physical outcomes.

5 Case study: Scratch assay

We conclude this review with a practical example. Before proceeding it is useful to note that all simulation and inference results so far are non-dimensional in the sense that we have set $\Delta = \tau = 1$ in all discrete simulations. This means that our estimates of U, D, v, and r are also non-dimensional. These non-dimensional simulations and parameter estimates can be re-dimensioned, as appropriate, to match a particular physical application by re-scaling Δ and τ with relevant length and time scales, respectively [34].

In this Section we will deal with a dimensional experimental data set, which means that we will work with dimensional quantities to match the appropriate time and length scales of a particular experiment, in this case a scratch assay [96–98] reported by Jin et al. [38]. Images of the experiment we consider are given in Figure 14(a)–(b) [38]. These images show the monolayer of cells just after a scratch has been made in Figure 14(a). A second image shows how the scratched region becomes recolonised by the impact of combined migration and proliferation in Figure 14(b) after $t = 48$ hours. These experiments use a human prostate cancer cell line, where individual cells have a diameter of approximately 20-25 μm [38]. For the purposes of this Case Study we take the cell diameter to be a constant, $\Delta = 24$ μm. The images in Figure 14(a)–(b) are 1950 μm wide and 1430 μm high. Jin et al. [38] quantified the experiments by placing a series of uniformly-spaced vertical lines across each image and counting the number of cells in each column of width 50 μm. To make this consistent with our simulation approach we make the further minimal assumption that cells are approximately uniformly distributed within each column, which allows us to report count data in Figure 14(c) that is equivalent to counting cells within columns of width 24 μm so that the count data are in a similar format as the count data used in Sections 3.1–3.3. This involves re-scaling the count data reported by Jin et al. [38] by a factor of 24/50, taking care to work

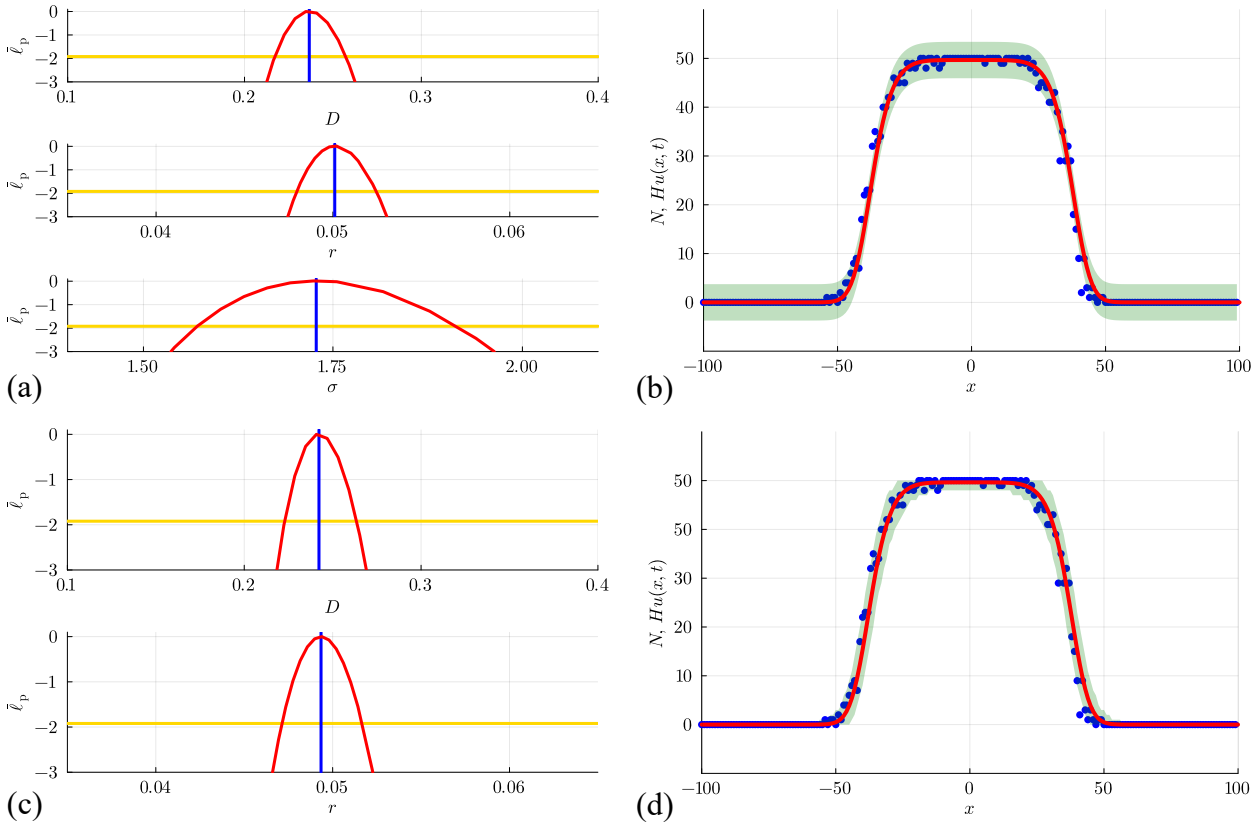

Figure 13: (a)–(b) Estimation, identifiability analysis and prediction under the additive Gaussian noise model. (a) Profiles for D, r and σ with $\hat{\theta} = (\hat{D}, \hat{r}, \hat{\sigma})^\top = (0.237, 0.050, 1.728)^\top$ superimposed (vertical blue lines). Each univariate profile gives an approximate 95% confidence interval: $\hat{D} = 0.237$ $[0.217, 0.257]$; $\hat{r} = 0.050$ $[0.0480, 0.0524]$; $\hat{\sigma} = 1.728$ $[1.571, 1.912]$ generated by finding the values of the interest parameter where the profile intersects with the 95% threshold (yellow horizontal lines). (b) Prediction interval generated with the additive Gaussian noise model encloses 96.0% of the data. (c)–(d) Estimation, identifiability analysis and prediction under the binomial model. (c) Profiles for D and r with $\hat{\theta} = (\hat{D}, \hat{v})^\top = (0.242, 0.0493)^\top$. Each univariate profile gives an approximate 95% confidence interval: $\hat{D} = 0.242$ $[0.223, 0.264]$; $\hat{r} = 0.0493$ $[0.0471, 0.05166]$. (d) Prediction interval generated with the binomial noise model encloses 99.5% of the data. Prediction intervals in (b) and (d) are generated using Laplace's approximation.

with nearest whole numbers to preserve the fact we are working with non-negative integer count data. A simple square packing argument indicates that these columns can hold no more than 60 cells of diameter 24 μm. Together, these minimal assumptions allow us to work with Equation (20) with $v = 0$ since there is no obvious asymmetry in the way that cells recolonise the scratched region of the experiments. This corresponds to working with the Fisher-Kolmogorov model [95, 99] as a surrogate process model, together with a binomial measurement error model to relate the noisy experimental count data to the solution of the surrogate model.

Data in Figure 14(c) show the count data at $t = 0$ and $t = 48$ hours. Following Jin et al. [38] we linearly interpolate data at $t = 0$ to obtain the initial condition, which allows us to solve Equation (20) numerically to model the expected cell density profile at $t = 48$ hours. Using the binomial noise model, we obtain $\hat{\boldsymbol{\theta}} = (\hat{D}, \hat{r})^{\top} = (1287.2, 0.0621)^{\top}$, where the dimensional units of the cell diffusivity is μm^2/hour, and the dimensional units of the cell growth rate is /hour. These parameter values can be simulated using a discrete model with dimensional lattice size $\Delta = 24$ μm (representing one cell diameter). Recall that the macroscopic parameters D and r are related to the microscopic parameters via $D = P\Delta^2/(4\tau)$ and $r = R/\tau$, where P and R are required to be in the interval $[0, 1]$ as they represent probabilities. This imposes an upper limit on the choice of time step τ of 0.112 hours. For example, if we chose $\tau = 0.1$ hours, this would lead to a movement probability per time step of $P = 0.894$ and proliferation probability per time step of $R = 0.00621$. These choices are not unique since we can choose a smaller value of the time step and simply re-scale our values of P and R to match our estimates of D and r.

Using automatic differentiation to calculate the Hessian matrix at the MLE, we invoke Laplace's approximation to obtain the prediction interval in Figure 14(c) which encloses 100% of the data at $t = 48$ hours. As with the synthetic data examples, we see that the binomial noise model leads to bounded prediction intervals that faithfully captures the count data. Univariate profile likelihood functions in Figure 14(d) confirm that both parameters are identified by the data.

This case study provides a quantitative illustration of how we can interpret data from a cell biology experiment in a random walk framework using a surrogate PDE model. The original data presented by Jin et al. [38] involved reporting count data at $t = 0, 12, 24, 36$ and 48 hours. For the sake of simplicity, and to be consistent with the remainder of this review, here we work with data at the final time point only, $t = 48$ hours. It is straightforward,

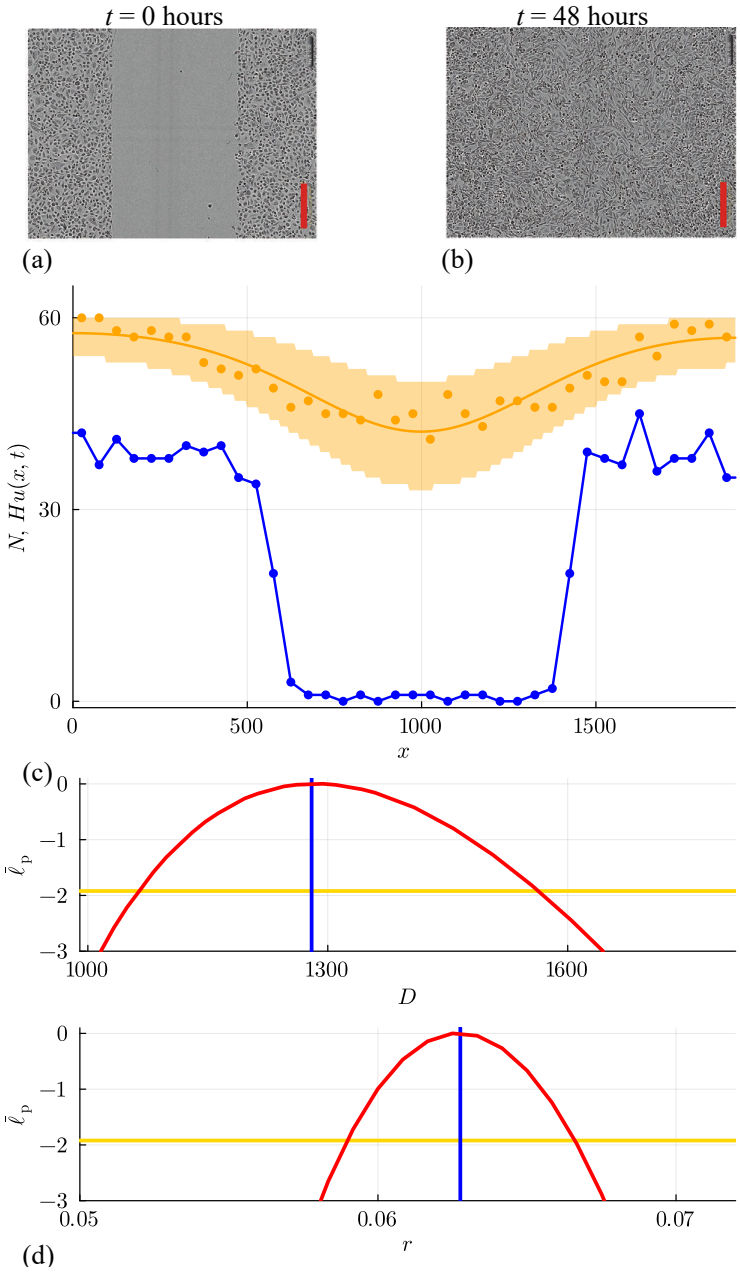

Figure 14: (a)–(b) Images of scratch assays reported by Jin et al. [38] at $t = 0$ and $t = 48$ hours, respectively. The red scale bar in each image corresponds to 300 μm. (c) Count data at $t = 0$ (blue dots) and $t = 48$ hours (orange dots), together with the MLE solution of Equation (20) with $v = 0$ and $\hat{\boldsymbol{\theta}} = (\hat{D}, \hat{r})^\top = (1287.2, 0.0621)^\top$. The prediction interval in (c) for the data at $t = 48$ hours (shaded orange), generated using Laplace's approximation, encloses 100% of the data. (d) Univariate profile likelihood functions for D (upper) and r (lower) superimposed with the MLE (blue vertical lines). Each univariate profile gives an approximate 95% confidence interval: $\hat{D} = 1287.2$ $[1091.0, 1584.1]$ and $\hat{r} = 0.0621$ $[0.05831, 0.0657]$ generated by finding the values of the interest parameter where the profile intersects with the 95% threshold (yellow horizontal lines).

however, to incorporate the intermediate time points into our estimation and we will explain how to achieve this in Section 6.

We conclude the discussion of this Case Study by noting that interpreting this experimental data with a simple interacting random walk model and corresponding mean-field surrogate PDE description has limitations as there are many modelling assumptions that could be explored and relaxed. For the sake of clarity, here we interpret this experiment using data at a single time point which rules out the possibility of identifying time-dependent parameters. This is noteworthy because previous studies that examine this experimental data have used measurements at several time points and concluded that working with time-dependent parameters, $D(t)$ and $r(t)$, captures the experimental data across all time points more accurately than making the standard assumption that D and r are constants [39, 100]. In Section 6 we explain how to deal with data at multiple time points, however here we have not explored the possibility of time-dependent parameters, nor have we addressed more general questions about model selection. Nonetheless, despite working with a parsimonious model this case study illustrates how the ideas developed in Sections 3.1–3.4 can be applied to experimental data to give meaningful and interpretable mechanistic insights. More generally, when dealing with real biological data we are certain to face questions of model selection, and we recommend that parameter identifiability analysis ought to be combined with model selection analysis by prioritising identifiable models [101, 102].

6 Conclusions and future work

In this review, we have surveyed and implemented a range of options for parameter inference, parameter identifiability analysis, and model prediction for a suite of lattice-based random walk models that can be used to study collective motion in biological applications. After an initial exploration of a simple ABC rejection approach relying only on a high-fidelity but computationally expensive stochastic model, we primarily focused on using low-fidelity, computationally inexpensive surrogate PDE models of the form,

$$\frac{\partial u}{\partial t} = -\frac{\partial J}{\partial x} + S, \tag{23}$$

where $u(x,t) \geq 0$ is the macroscopic mean agent density at position x and time t, J is the macroscopic agent flux, and S is the source term representing macroscopic agent prolifera-

tion [20, 21, 30–32].

We have illustrated how solutions of this surrogate PDE model, both exact and numerical, can be used for parameter estimation, identifiability analysis, and model prediction, within both likelihood-based sampling approaches and likelihood-based optimisation approaches. A particular focus of this review is to construct prediction intervals by combining various process models relating to details of the stochastic model, along with various noise models that relate the solution of the PDE to noisy observed data via a specified probability distribution [49]. A recurring result has been to show that the standard choice of working with an additive Gaussian noise model can lead to non-physical prediction intervals. In response, We illustrated how the use of other noise models (such as binomial and Poisson noise) more faithfully reflect the data-generation process, and can more accurately capture variability in the observed data.

In summary, our synthetic data analysis investigated 12 distinct random walk models, both non-interacting and interacting, whose key features are summarised in terms of the associated macroscopic PDE model in Table 1. While the surrogate PDE models in Table 1 describe problems posed in a one-dimensional Cartesian coordinate system, the general methods and workflow can also be applied when considering the same stochastic models and associated PDE descriptions in different coordinate systems. This includes radially symmetric problems and two- or three-dimensional Cartesian problems [103] without the vertical symmetry that we assumed in the synthetic simulation data and in the experimental Case Study presented in Section 5.

To maintain a straightforward style of exposition, we focus on cases where data are collected at a single time point. As mentioned in Section 5, while focusing on a single time point is relevant in many situations, it is also straightforward to extend our workflow to deal with data collected at multiple time points and spatial locations. For example, a modified log-likelihood function

$$\ell(\boldsymbol{\theta} \mid \boldsymbol{y}^{\text{obs}}) = \sum_{i=1}^{I} \sum_{k=1}^{K} \log \left[\phi \left(y_{i,k}^{\text{obs}}; Hu(x_i, t_k), \sigma^2 \right) \right], \tag{24}$$

can be used to deal with data collected at I spatial locations, x_i for $i = 1, 2, 3, \ldots, I$, and at K temporal time points, t_k for $k = 1, 2, 3, \ldots, K$, under a standard independence assumption. The key change here is that now the data vector $\boldsymbol{y}^{\text{obs}}$ is extended to include data collected at the ith spatial location and the kth time point, so that the length of this vector is IK

and individual elements are denoted $y_{i,k}^{\text{obs}}$. This extended form of the log-likelihood is written here for the standard case of working with an additive Gaussian noise model, and similar extensions apply for the other noise models considered in this review.

This review focuses on position-jump random walk models, where the continuum-limit surrogate model takes the form of a parabolic reaction-diffusion or reaction-advection-diffusion PDE [20, 21, 95]. We have not considered velocity-jump random walk models, where individual agents are associated with both a location and a velocity, and the continuum-limit surrogate model takes the form of a system of hyperbolic PDEs in a class known as transport equations [104]. Typically, non-interacting velocity-jump models lead to linear hyperbolic PDEs that are closely related to the telegraph equation [105–107], whereas interacting velocity jump models with crowding effects lead to nonlinear transport PDEs that resemble mathematical models of traffic flow [108]. The same framework and workflow we have outlined of collecting stochastic count data, representing the main trends in the data using a surrogate PDE model, and representing the stochastic data-generating process using an appropriate noise model can also be followed for velocity-jump models.

A key feature of the approaches for inference, identifiability analysis and model prediction in this review is to work with a computationally inexpensive surrogate model. While we have used standard mean-field approaches to arrive at surrogate PDE models [20, 21], there are alternative approaches for deriving appropriate surrogates. For example, methods of equation learning [100, 109–115] aim to *learn* the underlying structure of a differential equation model that gave rise to available observed data. These approaches have been used to generate surrogate ODE and PDE-based models and could be used instead of working with a mean-field PDE model. Other approaches include using a Gaussian process to construct a differential equation model from discrete data [116, 117] or using multidimensional linear regression as a computationally inexpensive option to represent any discrepancy between the solution of the mean-field PDE and averaged data from the discrete model [118]. It is also possible to make use of user-derived expertise to propose phenomenologically motivated differential equation models that can perform very well [119, 120].

Throughout this review, we have focused on identifiable problems where model parameters are reasonably well-identified by the data. A critical step in the workflow of parameter estimation, identifiability analysis, and model prediction is to quantify the extent to which the noisy data constrains our parameter estimates. Identifiability refers to whether a model's parameters can be uniquely determined from observed data, and identifiability analysis involves

evaluating how accurately and reliably these parameters can be estimated given the structure of the model and the data available [57]. Methods of identifiability analysis are typically classified in terms of whether they deal with *structural* or *practical* identifiability [25, 26, 50]. Structural identifiability analysis examines the question of whether different parameter values generate different probability distributions of the observable variables [58, 59, 121]. Structural identifiability is concerned with analysing the structure of the mathematical model, and is often assessed using software that use Lie derivatives to generate a system of input–output equations and examining its solvability properties [122–124]. Practical identifiability analysis concerns the extent to which parameter values can be confidently estimated, given noisy, incomplete data. This usually involves fitting a mathematical model to data, and then exploring the extent to which the quality of the model fit changes as the parameters are varied. Practical identifiability analysis can be performed either locally near a given point, such as near the parameter values that provide the best model fit, or globally over the extended parameter space. A common tool for assessing practical identifiability is the profile likelihood [23, 24, 60], which is the main focus on this work. Our choice to restrict our attention to reasonably well-identified problems explains why we chose to treat the initial density U as a known quantity when exploring random walk models with growth, $R > 0$. Intermediate to long-term data generated with random walk models with growth typically lead to identifiability challenges with the initial density U. This challenge was also circumvented in the case study where we treated the data at $t = 0$ as known and did not attempt to estimate the initial state of the system.

Since we focus on identifiable problems, there are many open questions about how to approach estimation and prediction for random walk models in the face of non-identifiability. Potential approaches involving reparameterisation to deal with structural non-identifiability [122–125], and model simplification or improved experimental design and data collection protocols to deal with practical non-identifiability [70].

In this review we illustrate the use of Gaussian, Poisson, and binomial observation models. However, other distributions may be more appropriate in other applications, for example the negative binomial distribution for overdispersed count data, or the gamma or log-normal distribution for continuous, non-negative data. Switching to a different observation model is straightforward within the workflow we have presented as it simply involves substituting a different likelihood function into the inference method, whether that is profile likelihood or a Bayesian inference method. Throughout, we have assumed that noise in the observations at

different locations is independent, but in the stochastic model (and in reality) fluctuations from the mean field at neighbouring locations will be correlated. Capturing non-independent noise is more challenging, and not covered by the approaches outlined here.

Data Accessibility Julia implementations within Jupyter notebooks for all computations are available on GitHub https://github.com/ProfMJSimpson/RandomWalkInference.

Funding This work is partly supported by the Australian Research Council (DP230100025, CE230100001) and the Marsden Fund (24-UOC-020).

Acknowledgements We thank the MATRIX mathematical research institute for hosting a one-week residential workshop entitled "Parameter identifiability in mathematical biology" (September 2024) where initial work on this project took place. We thank two referees for their supportive comments.

Appendix

We obtain numerical solutions of Equation (20) by discretising the domain $-L \leq x \leq L$ using a uniform mesh, with mesh spacing δ. Each mesh point has position x_i for $i = 1, 2, 3, \ldots, I$ and we write the solution $u(x_i, t)$ as u_i. Discretising spatial derivatives using standard central difference approximations leads to

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = \frac{D}{\delta^2}\left(u_{i-1} - 2u_i + u_{i+1}\right) - \frac{v}{2\delta}\left[u_{i+1}\left(1 - u_{i+1}\right) - u_{i-1}\left(1 - u_{i-1}\right)\right] + ru_i\left(1 - u_i\right), \quad (25)$$

for $i = 2, 3, 4, \ldots, I - 1$, together with boundary conditions $u_1 = u_I = 0$. We integrated this system of ODEs through time using Julia's DifferentialEquations.jl [126] using standard tolerance choices and automatic time stepping with Heun's method. All numerical results presented in this review involve $\delta = 0.5$, and additional results (not shown) indicate that this choice leads to grid-independent results for the problems that we consider.

References

[1] He D, Ionides EL, King AA. Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. Journal of the Royal Society Interface. 2010;7:271-83.

[2] Dennis B, Ponciano JM, Lele SR, Taper ML, Staples DF. Estimating density dependence, process noise, and observation error. Ecological Monographs. 2006;76:323-41.

[3] Perkins TJ, Swain PS. Strategies for cellular decision-making. Molecular Systems Biology. 2009;5:326.

[4] Marku M, Pancaldi V. From time-series transcriptomics to gene regulatory networks: A review on inference methods. PLOS Computational Biology. 2023;19:e1011254.

[5] Smith RC. Uncertainty Quantification: Theory, Implementation, and Applications. vol. 12. SIAM; 2013.

[6] Bear J. Dynamics of Fluids in Porous Media; 1972.

[7] Bird RB, Stewart WE, Lightfoot EN. Transport Phenomena. John Wiley & Sons; 2007.

[8] Casella G, Berger RL. Statistical Inference. Brooks/Cole; 2001.

[9] Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian data analysis. Chapman and Hall; 2014.

[10] Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, et al. Bayesian workflow. arXiv preprint. 2020;2011.01808.

[11] Wasserman L. All of statistics: A concise course in statistical inference. Springer; 2004.

[12] Wilkinson RD. Approximate Bayesian computation (ABC) gives exact results under the assumption of model error. Statistical Applications in Genetics and Molecular Biology. 2013;12:129-41.

[13] Lewis MA, Pacala S. Modeling and analysis of stochastic invasion processes. Journal of Mathematical Biology. 2000;41(5):387-429.

[14] Longo D, Peirce SM, Skalak TC, Davidson L, Marsden M, Dzamba B, et al. Multicellular computer simulation of morphogenesis: blastocoel roof thinning and matrix assembly in Xenopus laevis. Developmental Biology. 2004;271:210-22.

[15] Ciocanel MV, Kreiling JA, Gagnon JA, Mowry KL, Sandstede B. Analysis of active transport by fluorescence recovery after photobleaching. Biophysical Journal. 2017;112:1714-25.

[16] Ciocanel MV, Ding L, Mastromatteo L, Reichheld S, Cabral S, Mowry K, et al. Parameter identifiability in PDE models of fluorescence recovery after photobleaching. Bulletin of Mathematical Biology. 2024;86:36.

[17] Bullara D, Decker YD. Pigment cell movement is not required for generation of Turing patterns in zebrafish skin. Nature Communications. 2015;6:6971.

[18] Mort RL, Ross RJH, Hainey KJ, Harrison OJ, Keighren MA, Landini G, et al. Reconciling diverse mammalian pigmentation patterns with a fundamental mathematical model. Nature Communications. 2016;7:10288.

[19] Volkening A, Sandstede B. Modelling stripe formation in zebrafish: an agent-based approach. Journal of The Royal Society Interface. 2015;12(112):20150812.

[20] Codling EA, Plank MJ, Benhamou S. Random walk models in biology. Journal of the Royal Society Interface. 2008;5:813-34.

[21] Plank MJ, Simpson MJ, Baker RE. Random walk models in the life sciences: including births, deaths and local interactions. Journal of the Royal Society Interface. 2025;22:20240422.

[22] Kreutz C, Raue A, Kaschek D, Timmer J. Profile likelihood in systems biology. The FEBS Journal. 2013;280:2564-71.

[23] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics. 2009;25:1923-9.

[24] Raue A, Kreutz C, Theis FJ, Timmer J. Joining forces of Bayesian and frequentist methodology: a study for inference in the presence of non-identifiability. Philosophical

Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2013;271:20110544.

[25] Wieland FG, Hauber AL, Rosenblatt M, Tönsing C, Timmer J. On structural and practical identifiability. Current Opinion in Systems Biology. 2021;25:60-9.

[26] Heinrich M, Rosenblatt M, Wieland FG, Stigter H, Timmer J. On structural and practical identifiability: Current status and update of results. Current Opinion in Systems Biology. 2025:100546.

[27] Treloar KK, Simpson MJ. Sensitivity of edge detection methods for quantifying cell migration assays. PLOS One. 2013;8:e67389.

[28] Das AM, Eggermont AMM, ten Hagen TLM. A ring barrier–based migration assay to assess cell migration in vitro. Nature Protocols. 2015;10:904-15.

[29] Simpson MJ, Treloar KK, Binder BJ, Haridas P, Manton KJ, Leavesley DI, et al. Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. Journal of the Royal Society Interface. 2013;10:20130007.

[30] Berg HC. Random walks in biology. Princeton University Press; 1983.

[31] Hughes BD. Random walks and random environments. Oxford University Press; 1996.

[32] Redner S. A guide to first passage time processes. Cambridge University Press; 2001.

[33] Chowdhury D, Schadschneider A, Nishinari K. Physics of transport and traffic phenomena in biology: From molecular motors and cells to organisms. Physics of Life Reviews. 2005;2:318 – 352.

[34] Simpson MJ, Landman KA, Hughes BD. Cell invasion with proliferation mechanisms motivated by time-lapse data. Physica A: Statistical Mechanics and its Applications. 2010;389:3779-90.

[35] Gnerucci A, Faraoni P, Sereni E, Ranaldi F. Scratch assay microscopy: A reaction–diffusion equation approach for common instruments and data. Mathematical Biosciences. 2020;330:108482.

[36] Khain E, Sander LM, Schneider-Mizell CM. The role of cell-cell adhesion in wound healing. Journal of Statistical Physics. 2007;128(1-2):209-18.

[37] Khain E, Katakowski M, Hopkins S, Szalad A, Zheng X, Jiang F, et al. Collective behavior of brain tumor cells: the role of hypoxia. Physical Review E. 2011;83(3):031920.

[38] Jin W, Shah ET, Penington CJ, McCue SW, Chopin LK, Simpson MJ. Reproducibility of scratch assays is affected by the initial degree of confluence: Experiments, modelling and model selection. Journal of Theoretical Biology. 2016;390:136-45.

[39] Simpson MJ, Murphy RJ, Maclaren OJ. Modelling count data with partial differential equation models in biology. Journal of Theoretical Biology. 2024;580:111732.

[40] Crank J. The Mathematics of Diffusion. Clarendon Press; 1964.

[41] Abramowitz M, Stegun IA. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. New York: Dover Publications; 1964.

[42] Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C. Approximate Bayesian computation. PLoS Computational Biology. 2013;9(1):e1002803.

[43] Lambert B, MacLean AL, Fletcher AG, Coombes AN, Little MH, Byrne HM. Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. Journal of Mathematical Biology. 2018;76:1673-97.

[44] Schnoerr D, Sanguinetti G, Grima R. Approximation and inference methods for stochastic biochemical kinetics – a tutorial review. Journal of Physics A: Mathematical and Theoretical. 2017;50:093001.

[45] Browning AP, McCue SW, Binny RN, Plank MJ, Shah ET, Simpson MJ. Inferring parameters for a lattice-free model of cell migration and proliferation using experimental data. Journal of Theoretical Biology. 2018;437:251-60.

[46] Browning AP, McCue SW, Simpson MJ. A Bayesian computational approach to explore the optimal duration of a cell proliferation assay. Bulletin of Mathematical Biology. 2017;79:1888-906.

[47] Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Statistical Science. 1992;7:457-72.

[48] Simpson MJ, Baker RE. Parameter identifiability, parameter estimation and model prediction for differential equation models. To appear, SIAM Review. 2025. Available from: `https://arxiv.org/abs/2405.08177`.

[49] Murphy RJ, Maclaren OJ, Simpson MJ. Implementing measurement error models with mechanistic mathematical models in a likelihood-based framework for estimation, identifiability analysis and prediction in the life sciences. Journal of the Royal Society Interface. 2024;21:20230402.

[50] Hines KE, Middendorf TR, Aldrich RW. Determination of parameter identifiability in nonlinear biophysical models: A Bayesian approach. Journal of General Physiology. 2014;143:401.

[51] Hines KE. A primer on Bayesian inference for biophysical systems. Biophysical Journal. 2015;108:2103-13.

[52] Liu Y, Suh K, Maini PK, Cohen DJ, Baker RE. Parameter identifiability and model selection for partial differential equation models of cell invasion. Journal of the Royal Society Interface. 2024;21(212):20230607.

[53] Liu Y, Warne DJ, Simpson MJ. Likelihood-based inference, identifiability, and prediction using count data from lattice-based random walk models. Physical Review E. 2024;110:044405.

[54] Haario H, Saksman E, Tamminen J. An adaptive Metropolis algorithm. Bernoulli. 2001;7:223-42.

[55] Andrieu C, Thoms J. A tutorial on adaptive MCMC. Statistics and Computing. 2008;18:343-73.

[56] Pawitan Y. In all likelihood: statistical modelling and inference using likelihood. Oxford University Press; 2001.

[57] Cobelli C, DiStefano JJ III. Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. American Journal of Physiology. 1980;239:R7-R24.

[58] Meshkat N, Eisenberg M, DiStefano III JJ. An algorithm for finding globally identifiable parameter combinations of nonlinear ODE models using Gröbner bases. Mathematical Biosciences. 2009;222:61-72.

[59] Meshkat N, Sullivant S, Eisenberg M. Identifiability results for several classes of linear compartment models. Bulletin of Mathematical Biology. 2015;77:1620-51.

[60] Fröhlich F, Theis FJ, Hasenauer J. Uncertainty analysis for non-identifiable dynamical systems: Profile likelihoods, bootstrapping and more. International Conference on Computational Methods in Systems Biology. 2014:61-72.

[61] Audet C, Hare W. Derivative-free and blackbox optimization. Springer; 2017.

[62] Nocedal J, Wright SJ. Numerical optimization. Springer; 1999.

[63] Johnson SG. The NLopt module for Julia; 2018. Available from: `https://github.com/JuliaOpt/NLopt.jl`.

[64] Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. The Annals of Mathematical Statistics. 1938;9:60–62.

[65] Campbell DA, Chkrebtii O. Maximum profile likelihood estimation of differential equation parameters through model based smoothing state estimates. Mathematical Biosciences. 2013;246:283-92.

[66] Miles CE. Incorporating spatial diffusion into models of bursty stochastic transcription. Journal of the Royal Society Interface. 2025;22:20240739.

[67] Villaverde AF, Tsiantis N, Banga JR. Full observability and estimation of unknown inputs, states and parameters of nonlinear biological models. Journal of the Royal Society Interface. 2019;16:20190043.

[68] Villaverde AF, Raimúndez E, Hasenauer J, Banga JR. Assessment of prediction uncertainty quantification methods in systems biology. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2023;30:1725-36.

[69] Siekmann I, Sneyd J, Crampin EJ. MCMC can detect nonidentifiable models. Biophysical Journal. 2012;103:2275-86.

[70] Simpson MJ, Baker RE, Vittadello ST, Maclaren OJ. Parameter identifiability analysis for spatiotemporal models of cell invasion. Journal of the Royal Society Interface. 2020;17:20200055.

[71] Simpson MJ, Maclaren OJ. Making predictions using poorly identified mathematical models. Bulletin of Mathematical Biology. 2024;86:80.

[72] Tuncer N, Martcheva M, Ciupe SM. Structural and practical identifiability of within-host models of virus dynamics: review. Current Opinion in Systems Biology. 2025;41:100552.

[73] Royston P. Profile likelihood for estimation and confidence intervals. The Stata Journal. 2007;7:376-87.

[74] Vardeman SB. What about the other intervals? The American Statistician. 1992;46:193-7.

[75] Tierney L, Kadane JB. Accurate approximations for posterior moments and marginal densities. Journal of the American Statistical Association. 1986;81(393):82-6.

[76] Tierney L. Markov chains for exploring posterior distributions. The Annals of Statistics. 1994;22:1701-62.

[77] MacKay DJC. Information theory, inference and learning algorithms. Cambridge University Press; 2003.

[78] Griewank A, Walther A. Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation. SIAM; 2008.

[79] Revels J, Lubin M, Papamarkou T. Forward-mode automatic differentiation in Julia. arXiv. 2016;1607:07892.

[80] Monsalve-Bravo GM, Lawson BAJ, Drovandi C, Burrage K, Brown KS, Baker CM, et al. Analysis of sloppiness in model simulations: unveiling parameter uncertainty when mathematical models are fitted to data. Science Advances. 2022;8:eabm5952.

[81] Codling EA, Bearon RN, Thorn GJ. Diffusion about the mean drift location in a biased random walk. Ecology. 2010;91:3106-13.

[82] Aubert M, Badoual M, Fereol S, Christov C, Grammaticos B. A cellular automaton model for the migration of glioma cells. Physical Biology. 2006;3(2):93.

[83] Bruna M, Chapman SJ. Excluded-volume effects in the diffusion of hard spheres. Physical Review E. 2012;85(1):011103.

[84] Bruna M, Chapman SJ. Diffusion of finite-size particles in confined geometries. Bulletin of Mathematical Biology. 2014;76:947-82.

[85] Dyson L, Maini PK, Baker RE. Macroscopic limits of individual-based models for motile cell populations with volume exclusion. Physical Review E. 2012;86(3):031903.

[86] Plank MJ, Simpson MJ. Models of collective cell behaviour with crowding effects: comparing lattice-based and lattice-free approaches. Journal of the Royal Society Interface. 2012;9:2983-96.

[87] Anguige K, Schmeiser C. A one-dimensional model of cell diffusion and aggregation, incorporating volume filling and cell-to-cell adhesion. Journal of Mathematical Biology. 2009;58:395–427.

[88] Vo BN, Drovandi CC, Pettitt AN, Simpson MJ. Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation. Mathematical Biosciences. 2015;263:133-42.

[89] Vo BN, Drovandi CC, Pettitt AN, Pettet GJ. Melanoma cell colony expansion parameters revealed by approximate Bayesian computation. PLOS Computational Biology. 2015;11:e1004635.

[90] Harada Y, Iwasa Y. Lattice population dynamics for plants with dispersing seeds and vegetative propagation. Researches on Population Ecology. 1994;36:237-49.

[91] Satō K, Iwasa Y. Pair approximations for lattice-based ecological models. In: Dieckmann U, Law R, Metz JAJ, editors. The Geometry of Ecological Interactions: Simplifying Spatial Complexity. Cambridge University Press; 2000. p. 341-58.

[92] Ellner SP. Pair approximation for lattice models with multiple interaction scales. Journal of Theoretical Biology. 2001;210(4):435-47.

[93] Law R, Murrell DJ, Dieckmann U. Population growth in space and time: Spatial logistic equations. Ecology. 2003;84:252-62.

[94] Cantrell RS, Cosner C. Deriving reaction–diffusion models in ecology from interacting particle systems. Journal of Mathematical Biology. 2004;48(2):187-217.

[95] Murray JD. Mathematical biology I: an introduction. Springer; 2002.

[96] Liang CC, Park A, Guan JL. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. Nature Protocols. 2007;2:329-33.

[97] Grada A, Otero-Vinas M, Prieto-Castrillo F, Obagi Z, Falanga V. Research techniques made simple: Analysis of collective cell migration using the wound healing assay. Journal of Investigative Dermatology. 2017;137:e11-6.

[98] Bobadilla AVP, Arévalo J, Sarró E, Byrne HM, Maini PK, Carraro T, et al. *In vitro* cell migration quantification method for scratch assays. Journal of the Royal Society Interface. 2019;16:20180709.

[99] Maini PK, McElwain DLS, Leavesley DI. Traveling wave model to interpret a wound-healing cell migration assay for human peritoneal mesothelial cells. Tissue Engineering. 2004;10(3-4):475-82.

[100] Lagergren JH, Nardini JT, Baker RE, Simpson MJ, Flores KB. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. PLOS Computational Biology. 2020;16:e1008462.

[101] Simpson MJ, Browning AP, Warne DJ, Maclaren OJ, Baker RE. Parameter identifiability and model selection for sigmoid population growth models. Journal of Theoretical Biology. 2022;535:1100998.

[102] Simpson MJ, Walker S, Studerus E, McCue SW, Murphy RJ, Maclaren OJ. Profile-likelihood–based parameter and predictive interval analysis guides model choice for population dynamics. Mathematical Biosciences. 2022;352:108950.

[103] Simpson MJ, Landman KA, Hughes BD. Multi-species simple exclusion processes. Physica A: Statistical Mechanics and its Applications. 2009;388:399-406.

[104] Othmer HG, Dunbar SR, Alt W. Models of dispersal in biological systems. Journal of Mathematical Biology. 1988;26:263-98.

[105] Goldstein S. On diffusion by discontinuous movements, and on the telegraph equation. Quarterly Journal of Mechanics and Applied Mathematics. 1951;4(2):129-56.

[106] Kac M. A stochastic model related to the telegrapher's equation. Rocky Mountain Journal of Mathematics. 1974;4(3):497-509.

[107] Erban R, Othmer HG. From individual to collective behavior in bacterial chemotaxis. SIAM Journal on Applied Mathematics. 2004;65:361-91.

[108] Treloar KK, Simpson MJ, McCue SW. Velocity-jump models with crowding effects. Physical Review E. 2011;84:061920.

[109] Brunton SL, Proctor JL, Kutz JN. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the National Academy of Sciences. 2016;113:3932-7.

[110] Rudy SH, Brunton SL, Proctor JL, Kutz JN. Data-driven discovery of partial differential equations. Science Advances. 2017;3:e1602614.

[111] Schaeffer H, McCalla SG. Sparse model selection via integral terms. Physical Review E. 2017;96:023302.

[112] Lagergren JH, Nardini JT, Lavigne GM, Rutter EM, Flores KB. Learning partial differential equations for biological transport models from noisy spatio-temporal data. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2020;476(2236):20190800.

[113] Nardini JT, Lagergren JH, Hawkins-Daarud A, Curtin L, Morris B, Rutter EM, et al. Learning equations from biological data with limited time samples. Bulletin of Mathematical Biology. 2020;82:119.

[114] Nardini JT, Baker RE, Simpson MJ, Flores KB. Learning differential equation models from stochastic agent-based model simulations. Journal of the Royal Society Interface. 2021;18:20200987.

[115] Nardini JT. Forecasting and predicting stochastic agent-based model data with biologically-informed neural networks. Bulletin of Mathematical Biology. 2024;86:130.

[116] Gramacy RB. Surrogates: Gaussian Process Modeling, Design, and Optimization for the Applied Sciences. Chapman and Hall/CRC; 2020.

[117] VandenHeuval DJ, Drovandi C, Simpson MJ. Computationally efficient mechanism discovery for cell invasion with uncertainty quantification. PLOS Computational Biology. 2022;18:e1010599.

[118] Simpson MJ, Baker RE, Buenzli PR, Nicholson R, Maclaren OJ. Reliable and efficient parameter estimation using approximate continuum limit descriptions of stochastic models. Journal of Theoretical Biology. 2022;549:111201.

[119] Jain HV, Norton KA, Prado BB, Jackson TL. SMoRe ParS: A novel methodology for bridging modeling modalities and experimental data applied to 3D vascular tumor growth. Frontiers in Molecular Biosciences. 2022;9:1056461.

[120] Bergman DR, Norton KA, Jain HV, Jackson T. Connecting agent-based models with high-dimensional parameter spaces to multidimensional data using SMoRe ParS: A surrogate modeling approach. Bulletin of Mathematical Biology. 2024;86:1-30.

[121] Browning AP, Taşcă M, Falcó C, Baker RE. Structural identifiability analysis of linear reaction–advection–diffusion processes in mathematical biology. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2024;480:20230911.

[122] Chiş O, Banga JR, Balsa-Canto E. Structural identifiability of systems biology models: a critical comparison of methods. PLOS ONE. 2011;6:e27755.

[123] Díaz-Seoane S, Rey Barreiro X, Villaverde AF. STRIKE-GOLDD 4.0: user-friendly, efficient analysis of structural identifiability and observability. Bioinformatics. 2022;39:btac748.

[124] Ligon TS, Frölich F, Chiş O, Banga JR, Balsa-Canto E, Hasenauer J. GenSSI 2.0: multi-experimental structural identifiability analysis of SBML models. Bioinformatics. 2018;34:1421-3.

[125] Simpson MJ. Data-informed model reduction for inference and prediction from non-identifiable models. Journal of Theoretical Biology. 2025;611:112155.

[126] Rackauckas C, Nie Q. DifferentialEquations.jl – A performant and feature-rich ecosystem for solving differential equations in Julia. Journal of Open Research Software. 2017;5:15.