Rapid Parameter Inference for Spatiotemporal Stochastic Biological Models using Neural Posterior Estimation

A Preprint

Tom Kimpson[*1,2], **Jennifer Flegg**[1,2], and **Matthew J. Simpson**[2,3]
[1]School of Mathematics and Statistics, The University of Melbourne, Parkville, VIC 3010, Australia
[2]ARC Centre of Excellence for the Mathematical Analysis of Cellular Systems (MACSYS), The University of Melbourne, Parkville, VIC 3010, Australia
[3]School of Mathematical Sciences, Queensland University of Technology, Brisbane, QLD 4001, Australia

October 27, 2025

Abstract

Parameter inference for stochastic models is essential for quantitative biology, but is hampered by intractable likelihoods. Agent-based random walk models, used widely in ecology and cell biology, exemplify this challenge, often forcing researchers to use likelihood-free methods like Approximate Bayesian Computation (ABC) that introduce biases and tuning difficulties, or else surrogate models which rely on a potentially unfaithful approximation and require an explicit, potentially erroneous, noise model. Here, we overcome these limitations using Neural Posterior Estimation (NPE), a simulation-based, machine learning inference framework that learns the full posterior distribution. Once trained, NPE provides amortized, near-instantaneous posterior estimates. We deploy this framework on a random walk model of a barrier assay experiment describing in vitro cell migration. We demonstrate our approach in two settings: first, using one-dimensional summary statistics (column counts), and second, by augmenting NPE with a novel convolutional neural network (CNN) that enables NPE to learn directly from raw two-dimensional spatial data. In both cases, NPE remains the core inference engine, with the CNN serving as a feature extraction component for spatial data. On synthetic data, our NPE pipeline proves highly effective in both settings, delivering posterior estimates at a substantially reduced computational cost per inference. When applied to 1D summary statistics, the pipeline matches or exceeds the precision of classical methods. Similarly, for 2D spatial data, NPE augmented with CNN performs inference directly on raw images with a precision comparable to that from curated summary statistics. This second capability is critical, as it moves beyond the traditional reliance on lower-dimensional data and opens up new avenues for complex models where informative summary statistics are not obvious or available. We provide an open-source implementation of our pipeline to facilitate its adoption and extension to more complex, spatially-structured biological models. This work provides a practical framework for parameter inference in spatial stochastic models that leverages machine learning to avoid the need for likelihood functions, hand-crafted summary statistics, and approximate noise models.

1 Introduction

Parameter inference is the critical link between mathematical models and biological understanding, enabling quantitative insights from noisy experimental data. Inference methods provide researchers with a powerful toolkit to generate predictions about system behavior, estimate unknown parameters, test competing hypotheses, and rigorously quantify uncertainty. These applications are critical across a wide range of fields, including epidemiology [17, 54], ecology

*Corresponding author: tom.kimpson@unimelb.edu.au

[49, 52], cellular biology [63], gene regulatory networks [35, 66], systems biology [50] phylogenetics [42], and neuroscience [12].

For deterministic continuum models described by ordinary and partial differential equations, inference methodologies are mature and well-established. Both Bayesian [3, 21] and frequentist [13, 9] approaches provide robust foundations for parameter estimation, model validation, and uncertainty quantification. The sophistication of these inference techniques has enabled reliable parameter estimation and prediction across numerous scientific disciplines including climate science [65], cosmology [46], fluid dynamics [10] gravitational wave astrophysics [64].

Spatially explicit, stochastic biological models, however, present fundamental inferential challenges that resist these established approaches. Agent-based models and discrete stochastic processes, while capturing essential biological variability and heterogeneity, typically possess intractable likelihoods that preclude standard statistical methods. Experimental protocols exhibit substantial variation, datasets are frequently sparse and noisy, and the stochastic nature of biological processes introduces complex dependencies that deterministic frameworks cannot adequately represent. These characteristics create a disconnect between the sophisticated stochastic models needed to represent biological reality and the inference tools available to calibrate them from data.

Agent-based models of cell migration and proliferation have become a critical testbed for developing new inference techniques that can preserve the biological realism of stochastic models [8, 55, 57, 53, 67, 36, 33, 41]. These approaches often use experimental data such as scratch and barrier assay experiments to calibrate high-fidelity discrete random-walk models, highlighting applications where stochasticity is visible and decision-relevant [27, 28, 58]. At the same time, these successes have clarified several canonical challenges that have historically limited the broader application of stochastic models [36, 67]. While these approaches demonstrate the potential of sophisticated inference frameworks, they also reveal fundamental limitations that must be addressed to fully realize the promise of stochastic biological modeling. These canonical challenges can be summarized as follows:

1. **Intractable likelihoods:** The likelihood function for most agent-based models is intractable, precluding the use of standard methods like Maximum Likelihood Estimation (MLE) or typical Markov chain Monte Carlo (MCMC) algorithms.
2. **Surrogate-model error:** A common strategy to make inference tractable is to replace the computationally expensive stochastic simulator with a cheap surrogate, typically a deterministic partial differential equation (PDE) describing mean-field behavior. While this enables efficient MCMC or MLE techniques, the resulting inference is conditional on the surrogate being a faithful approximation, an assumption that, if violated, can lead to biased parameter estimates and inaccurate predictions.
3. **Noise-model misspecification:** The use of a deterministic surrogate creates the problem of specifying a noise model to connect its clean output to noisy data. A misspecified choice can lead to non-physical predictions and poorly calibrated uncertainties.
4. **Likelihood-free inefficiency:** Classical likelihood-free alternatives like Approximate Bayesian Computation (ABC) are notoriously computationally expensive and sensitive to the choice of summary statistics, distance metrics, and tolerance thresholds.
5. **Information loss from summary statistics:** Classical methods often require reducing complex, high-dimensional data (e.g., the full 2D spatial arrangement of cells) to lower-dimensional summary statistics (e.g., a 1D cell density profile). This step risks a significant loss of information, and the choice of statistics can heavily bias the inference.

In this paper, we propose that these fundamental challenges can be addressed in a unified manner using Neural Posterior Estimation (NPE), a modern simulation-based inference (SBI) paradigm. NPE leverages flexible neural density estimators to learn the posterior distribution $p(\theta \mid x)$ directly from simulator-generated (θ, x) pairs. This approach offers direct remedies to each of the aforementioned challenges:

1. It is likelihood-free, circumventing intractability.
2. It learns directly from the high-fidelity simulator, eliminating surrogate model bias and removing the PDE as a source of approximation error.
3. It implicitly learns the full, complex, stochastic relationship between parameters and data, obviating the need for an explicit noise model.
4. It provides amortized inference: after an initial, one-time training phase, it can generate posterior estimates for new observations almost instantaneously, overcoming the high per-inference cost of methods like ABC.
5. It can ingest high-dimensional data directly (e.g., raw 2D images), preserving information and avoiding manual selection of summary statistics.

The central thesis of this work is that NPE provides a robust, efficient, and direct framework for parameter inference in stochastic biological models. We demonstrate this by applying a state-of-the-art Sequential NPE (SNPE) workflow to a canonical agent-based random walk model of cell migration and proliferation. This type of model is a well-established tool for studying collective cell behaviour [see e.g. 57] and serves as an ideal testbed for advanced inference methods.

From a modelling perspective, we perform inference directly on the high-fidelity, agent-based random walk model. This simulator-native approach entirely avoids the use of potentially unfaithful surrogate models (e.g., mean-field PDEs), thus eliminating surrogate-model error and preserving the full stochastic realism of the biological system. For infernece, we deploy a modern Neural Posterior Estimation (NPE) framework as the core inference engine. This simulation-based, likelihood-free method directly learns the posterior distribution, overcoming the intractability of the likelihood function. Its amortized nature provides near-instantaneous posterior estimates for new observations after a single upfront training cost, bypassing the high per-inference expense and difficult tuning associated with traditional methods like ABC. A key methodological innovation is our integration of a deep learning architecture, specifically a Convolutional Neural Network (CNN), as a learned feature extractor for the NPE pipeline. This enables the inference framework to learn informative representations automatically and directly from raw, two-dimensional spatial data, removing the need for manual, and potentially information-losing, data curation. We demonstrate how these components form a unified, end-to-end pipeline that represents a new paradigm for inference in spatially-explicit stochastic models. By coupling the CNN feature extractor with the NPE inference engine, we perform Bayesian parameter estimation directly using the raw spatial output of the agent-based model. This synergy bypasses two major traditional bottlenecks: the need for hand-crafted summary statistics and the reliance on approximate surrogate models. Additionally, we release an open-source, reproducible implementation of our pipeline to facilitate its adoption and extension to other complex, spatially-structured biological models. While we use scratch assay experiments as our primary motivating application, the inference framework developed here is applicable to the broader class of spatially-explicit stochastic models commonly used throughout biology and ecology.

The remainder of this paper is structured as follows. Section 2 reviews the canonical stochastic model and establishes a performance baseline by applying classical inference methods. Section 3 details our NPE methodology and demonstrates its performance on one-dimensional summary statistics. Section 4 further enhances this pipeline by embedding a Convolutional Neural Network (CNN) to perform inference directly on raw spatial data. Finally, Section 5 discusses the broader implications of our work for the future of simulation-based inference in biology.

2 Classical Approaches to Inference in Stochastic Biological Models

Before introducing our NPE framework (Section 3), we briefly review the established approaches to inference for stochastic agent-based models. We draw primarily on the comprehensive treatment in [57] (S&P2025hereafter), which provides systematic comparisons across multiple inference paradigms for biological random-walk models. These methods have been the workhorses of parameter inference in mathematical biology, each with distinct strengths and limitations that motivate our proposed alternative. Table 1 provides a comparative overview of these approaches.

Method	**Cost per Inference**	**Requires Surrogate?**	**Requires Summary Stats?**	**Amortized?**
ABC	Very High	No	Yes	No
Surrogate + MCMC	High	Yes	Yes	No
Surrogate + MLE	Medium	Yes	Yes	No
NPE (proposed)	**Low***	**No**	**Optional**	**Yes**

Table 1: Comparison of inference approaches for stochastic biological models. *The low cost per inference for NPE is achieved after an initial, computationally intensive training phase.

2.1 Approximate Bayesian Computation

Approximate Bayesian Computation (ABC) represents the most direct approach to likelihood-free inference for stochastic models [60, 59, 61, 48]. The foundational ABC rejection sampling algorithm approximates the posterior distribution by undertaking the following steps: (i) a parameter vector $\theta^{(j)}$ is drawn from the prior distribution $p(\theta)$; (ii) a dataset $y^{(j)}$ is simulated from the model using $\theta^{(j)}$; (iii) a discrepancy between the summary statistics of the observed data, $s(y^{\text{obs}})$, and simulated data, $s(y^{(j)})$, is calculated using a distance metric $\rho(\cdot, \cdot)$; and (iv) the parameter

$\theta^{(j)}$ is accepted if this discrepancy is less than a tolerance threshold, $\rho(s(y^{\text{obs}}), s(y^{(j)})) \leq \epsilon$. The collection of accepted parameters forms a sample from an approximation to the true posterior, $p(\theta \mid y^{\text{obs}})$ [2].

Despite its conceptual simplicity, ABC faces significant methodological and computational challenges. A primary issue is the choice of the tolerance ϵ, which defines a fundamental bias-accuracy trade-off [60, 5]. Only in the idealised limit where $\epsilon \to 0$ and the summary statistics are sufficient does the ABC posterior converge to the true posterior [37, 5]. In practice, a smaller ϵ reduces bias but dramatically lowers the acceptance rate, increasing computational cost. Conversely, a larger ϵ increases the acceptance rate at the cost of yielding a more biased and diffuse posterior approximation.

This trade-off is exacerbated by the "curse of dimensionality" [4, 5]. For models with high-dimensional data or parameter spaces, the acceptance rate can be exceptionally low, demanding a vast number of simulations. For instance, S&P2025 generate 10^5 simulations from their computationally expensive random walk model to retain just the top 1%, highlighting this significant computational inefficiency. Furthermore, the method's performance depends critically on the choice of summary statistics. If the statistics are not formally sufficient, crucial information from the data is discarded, which introduces a form of bias that cannot be removed by simply lowering the tolerance ϵ [51, 37]. Selecting informative yet low-dimensional summary statistics in a principled manner remains a difficult and often ad-hoc challenge [20, 5].

2.2 Surrogate Model Approaches

To mitigate the computational burden of methods like ABC, an alternative strategy replaces the stochastic model with a computationally cheaper surrogate model. A common choice is a deterministic partial differential equation (PDE) that captures the system's mean-field behaviour, yielding a tractable likelihood function [19, 1] that enables standard, efficient likelihood-based inference algorithms like Markov chain Monte Carlo (MCMC). However, this computational gain comes at the cost of introducing potential model misspecification bias, as inference is performed on an approximation of the true stochastic process.

The primary limitation of this approach is that the resulting inference is conditional on the fidelity of the surrogate. Any discrepancy between the surrogate and the true stochastic process introduces a systematic bias that cannot be reduced by collecting more data [29, 7]. While the continuum limit accurately describes the mean behaviour of non-interacting random walks, the mean-field PDE neglects crucial stochastic effects in more complex scenarios. For example, in interacting models with crowding effects discussed in S&P2025 (Section 4), the mean-field approximation captures average behavior but cannot represent the bounded nature of agent occupancy or the discrete stochastic fluctuations [6, 1] around this mean. This limitation becomes particularly important when constructing prediction intervals, where the choice of noise model (e.g. Gaussian vs. binomial, see next Section 2.2.2) significantly affects the physical realism of predictions [38, 21]. These fundamental limitations of surrogate-based approaches motivate the need for inference methods that can work directly with the full stochastic simulator.

2.2.1 Inference with Surrogate Likelihoods

A tractable likelihood function from a surrogate model unlocks standard, efficient inference techniques that would be impossible with the original stochastic simulator. These techniques generally fall into two categories: direct optimization or posterior sampling.

The first category, direct optimization, seeks to find the single parameter set that maximizes the likelihood function, yielding the Maximum Likelihood Estimate (MLE). While computationally fast, the MLE provides only a point estimate and does not inherently quantify parameter uncertainty, although indirect (asymptotic) uncertainty quantification can be obtained through likelihood ratio tests and Wilks' theorem. The second category, posterior sampling, uses methods like Markov chain Monte Carlo (MCMC) to approximate the full posterior distribution, offering a complete characterisation of uncertainty. However, MCMC typically requires significant computational effort, careful hyperparameter tuning, and rigorous convergence assessment. A third approach, the Laplace approximation, offers a compromise between the speed of optimization and the detailed uncertainty characterization of MCMC. This method approximates the posterior distribution as a multivariate normal distribution centered at the posterior mode (which equals the MLE, $\hat{\theta}$ under uniform priors). The covariance is given by the inverse of the observed Fisher Information matrix:

$$p(\theta \mid y^{\text{obs}}) \approx \text{MVN}(\hat{\theta}, \mathcal{I}^{-1}(\hat{\theta})), \tag{1}$$

where $\mathcal{I}(\hat{\theta}) = -\nabla\nabla^T \ell(\theta \mid y^{\text{obs}})|_{\theta=\hat{\theta}}$ is the negative Hessian of the log-likelihood evaluated at the mode, often computed using automatic differentiation. While prediction intervals generated via the Laplace approximation can closely match those from more computationally intensive methods (see S&P2025), the method relies on the assumption that the posterior is locally Gaussian. This assumption may be violated for complex models, models with limited

data, or or simply for models with bounded parameters (e.g., $D \geq 0$; see Section 2.3) potentially leading to a poor representation of the true uncertainty.

However, all these likelihood-based approaches require an explicit specification of how the deterministic surrogate output relates to the noisy observational data—a critical modeling choice we examine next.

2.2.2 Noise Model Specification

The use of deterministic surrogates necessitates an explicit choice of how to model the relationship between the clean PDE output and the inherently noisy observations from the stochastic process. This choice profoundly impacts inference results, as the "noise" represents genuine stochastic variability in the agent-based system rather than measurement error.

S&P2025 systematically compare three noise models for relating observed counts to the PDE solution: Gaussian, Poisson, and binomial. The Gaussian noise model assumes observations are normally distributed around the model prediction with constant variance. While mathematically convenient, this leads to unphysical prediction intervals that include negative counts in low-density regions where the pipulation density approaches zero, and fails to respect carrying capacity constraints in high-density regions. The Poisson noise model provides a more natural choice for count data, eliminating the need for an additional variance parameter and ensuring non-negative predictions. The binomial noise model is appropriate for exclusion processes where occupancy is bounded, naturally capturing saturation effects.

S&P2025's extensive comparisons reveal that noise model misspecification can significantly bias parameter estimates and lead to poorly calibrated prediction intervals. Quantitatively, the Gaussian model consistently produces unphysical negative predictions in low-density regions and encloses only 95% of data points, while the Poisson and binomial models better respect the discrete, non-negative nature of count data and achieve coverage rates of 99-100%. Similar conclusions regarding the superiority of binomial measurement models over standard additive Gaussian assumptions for biological count data have been demonstrated in [56].

2.3 Example Results

To illustrate these theoretical considerations in practice, we reproduce a representative example from S&P2025, which focuses on a stochastic agent-based model of a scratch barrier assay. In this model, agents representing biological cells are initially confined to a central region on a two-dimensional square lattice and subsequently spread outwards via an unbiased random walk. The experimental setup, illustrated in Figure 1, mimics a barrier assay where cells are initially placed in a confined region and allowed to spread after barrier removal. The two-dimensional lattice has height $H = 50$ (with $0 \leq y \leq 50$) and width $W = 200$ (with $-100 \leq x \leq 100$), with lattice spacing $\Delta = 1$ and discrete time step $\tau = 1$. Zero-flux boundary conditions are applied at all edges. Initial conditions are set by occupying each site in the central region $|x| \leq 25$ with a single agent with probability $U = 0.5$, as shown in Figure 1(a). At each time step, the system evolves through a random sequential update method where agents move to one of four nearest-neighbor sites with probability $P = 0.7$. After 100 time steps, the population spreads symmetrically as shown in Figure 1(b).

A key feature of this experimental design is the reduction of two-dimensional spatial data to one-dimensional measurements. Since the initial macroscopic density is uniform in the vertical (y) direction and boundary conditions preserve this symmetry, the resulting agent distribution remains independent of vertical location on average. This motivates quantifying the system by counting the number of agents per column:

$$N_i(t) = \sum_{j=1}^{J} n_{i,j}(t), \tag{2}$$

where $n_{i,j}(t)$ denotes the integer number of agents at lattice site (i, j) at time t, and J is the number of lattice sites in the vertical direction. This dimensional reduction from 2D spatial data to 1D column counts is standard in cell biology scratch assays [26], as it simplifies the characterization of spatiotemporal spreading while retaining the stochastic fluctuations inherent to the discrete process, as evident in Figure 1(c)-(d).

For lattice-based random walk models of this type, the mean behaviour of agents in the continuum limit can be described by a PDE of the form [11, 57, 47]:

$$\frac{\partial u}{\partial t} = -\frac{\partial J}{\partial x} + S, \tag{3}$$

where $u(x, t)$ represents mean agent density, J is the macroscopic flux, and S denotes source terms from proliferation. For the simple case considered here (unbiased motion, no proliferation, no crowding), this reduces to the linear diffusion equation:

$$\frac{\partial u}{\partial t} = D\frac{\partial^2 u}{\partial x^2}, \tag{4}$$

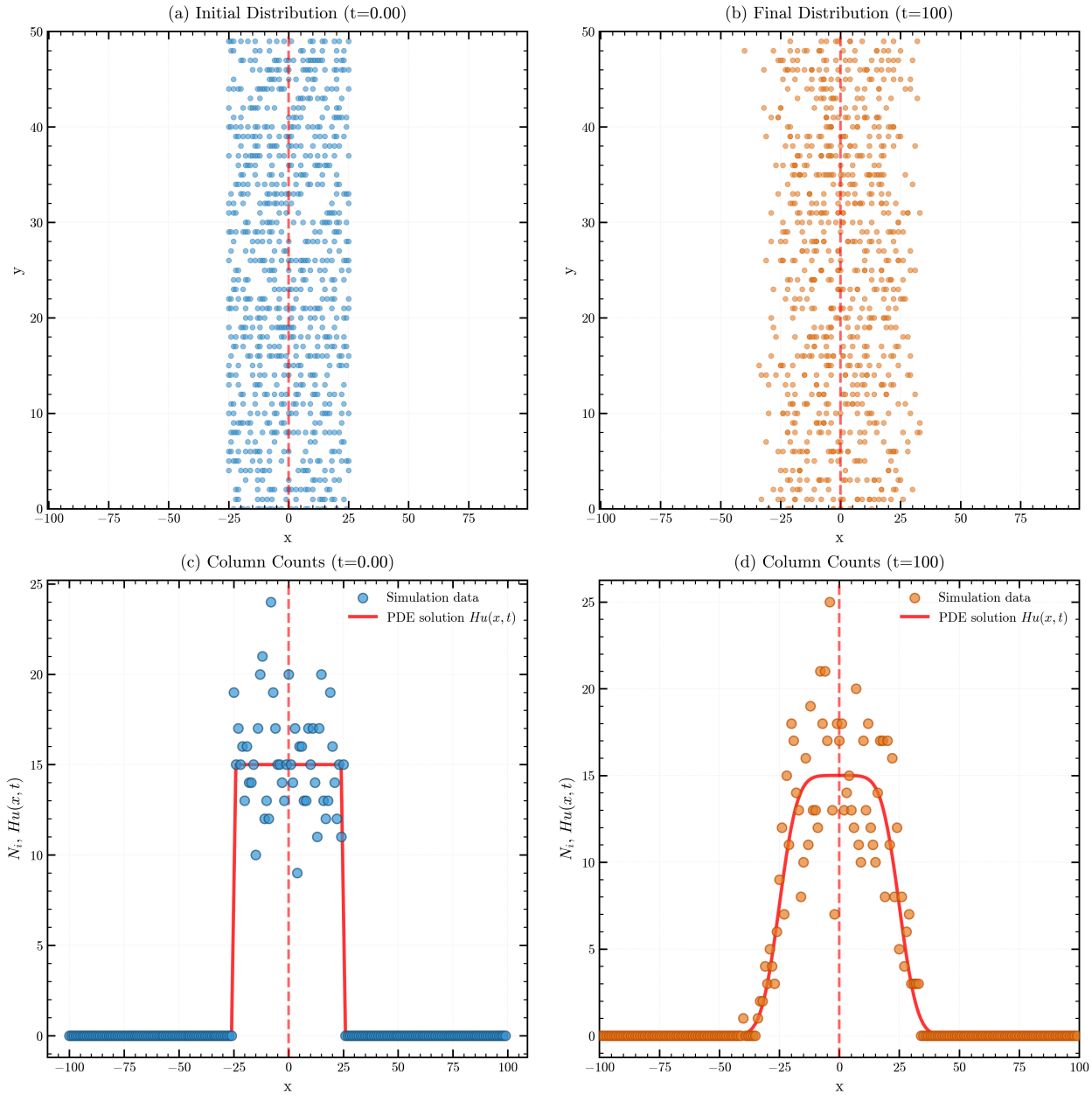


Figure 1: Schematic of the lattice-based random walk model. (a) Initial placement of agents within region $|x| \leq 25$ on a 2D lattice with $H = 50$, $\Delta = \tau = 1$. (b) Agent distribution after 100 time steps showing symmetric outward spreading. (c)-(d) Reduction of 2D data to 1D column counts $N_i(t)$ at $t = 0$ and $t = 100$ respectively. Blue dots show initial stochastic count data; orange dots show final stochastic count data; red curves show the continuum PDE solution $Hu(x,t)$ where $u(x,t)$ solves the diffusion equation with $D = P/4$.

where $D = \lim_{\Delta,\tau\to 0} P\Delta^2/(4\tau)$ is the effective diffusion coefficient. Working with dimensionless units $\Delta = \tau = 1$, this gives $D = P/4 = 0.175$ for our simulation with $P = 0.7$.

The continuous analogue of the column count N_i under the PDE model is $Hu(x_i, t)$, where x_i denotes the horizontal position of the i-th column and H is the lattice height. Figure 1(c)-(d) compares the stochastic count data (blue dots) with this PDE solution (red curves), revealing that while the PDE captures mean trends, substantial stochastic fluctuations remain—a key motivation for carefully modeling observation noise.

While S&P2025 investigates a suite of model variations (e.g., incorporating proliferation, directional bias, and crowding effects) and alternative PDE approximations, we deliberately focus on this simple unbiased, non-interacting case. It provides a clear baseline against which to compare our proposed NPE method, and we refer the reader to the original study for a fuller discussion of more complex scenarios.

We aim to infer two key parameters that govern the model's behaviour: the initial occupation probability, U, and the effective diffusion coefficient, D. For the synthetic data considered here, we use ground truth values $U = 0.3$ and $D = 0.175$.

Figure 2a shows the posterior obtained via ABC. This method operates directly on the stochastic model but relies on one-dimensional summary statistics (column counts), leading to a notably broad posterior. This diffuseness reflects both the inherent stochasticity of the system and the information loss from summarizing the data, showcasing the difficulty in precisely constraining parameters even with a high computational budget. In contrast, Figure 2b displays the Gaussian approximation to the posterior (Laplace approximation) derived from maximizing the likelihood of a deterministic surrogate. This approach is computationally efficient and yields a much tighter posterior. However, this apparent precision is entirely conditional on the fidelity of the surrogate; any mismatch between the surrogate and the true stochastic process introduces a systematic bias not reflected in the narrow posterior. Finally, Figure 2c shows the posterior sampled using MCMC with the surrogate model and an assumed Gaussian noise model. This provides a more flexible characterization of uncertainty. Crucially, this method requires inferring an additional nuisance parameter, σ, which quantifies the noise variance. The corner plot shows the joint posterior over the model parameters and this additional noise term. While comprehensive, this result is still fundamentally limited by the choice of surrogate and the specified noise model.

Collectively, these results highlight the core dilemma: direct methods like ABC are computationally demanding and often imprecise, while surrogate-based methods are faster but introduce hard-to-quantify systematic biases that cannot be reduced by collecting more data.

These persistent challenges across classical approaches motivate our exploration of NPE as a unified framework that can address multiple limitations simultaneously through direct learning from simulator output.

3 Neural Posterior Estimation for Stochastic Biological Models

3.1 Overview and Motivation

NPE [43, 24, 14] offers a fundamentally different approach to inference for simulator-based models. Rather than relying on surrogate approximations or requiring explicit noise models, NPE learns the posterior distribution $p(\theta|x)$ directly from simulations of the stochastic model. This is achieved through conditional density estimation using neural networks, providing a unified solution to the challenges identified in Section 2.

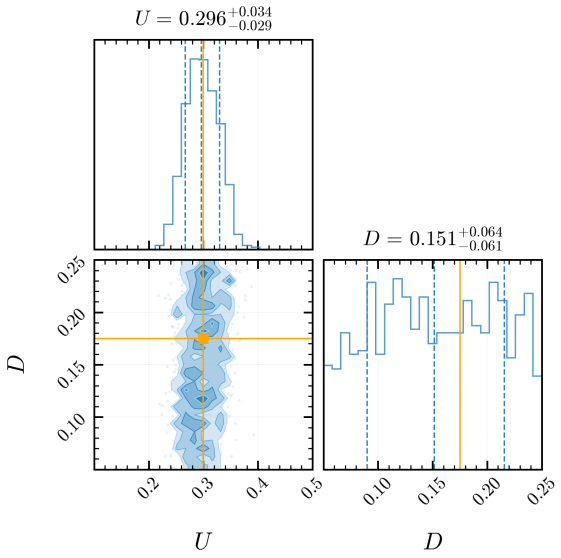
The key insight is that while the likelihood $p(x|\theta)$ may be intractable for complex stochastic simulators, we can still readily generate samples from it by running the simulator. NPE leverages these (parameter, sample) pairs to train a neural network that approximates the inverse mapping — from an observation back to the distribution over parameters that could have produced it. This process is *amortized*; once the network is trained, it can perform posterior inference for any new observation in seconds rather than hours, without requiring additional simulations. For the class of stochastic biological models considered here, this approach directly addresses each limitation of classical methods: it circumvents intractable likelihoods, eliminates surrogate model bias, implicitly captures complex stochastic relationships without explicit noise models, and can process high-dimensional data through automatic feature extraction.

Figure 3 provides a schematic overview of the NPE workflow, illustrating the distinction between the computationally intensive training phase and the rapid inference phase. During training, parameters are sampled from the prior $p(\theta)$ and used to generate simulated data via the stochastic simulator $p(x|\theta)$. These parameter-data pairs $\{(\theta_i, x_i)\}_{i=1}^N$ are then used to train a neural density estimator $q_\phi(\theta|x)$ that learns to approximate the posterior distribution. Once trained, the inference phase is nearly instantaneous: given new observed data x_o, the network directly evaluates the approximate posterior $p(\theta|x_o)$ without requiring additional simulations.

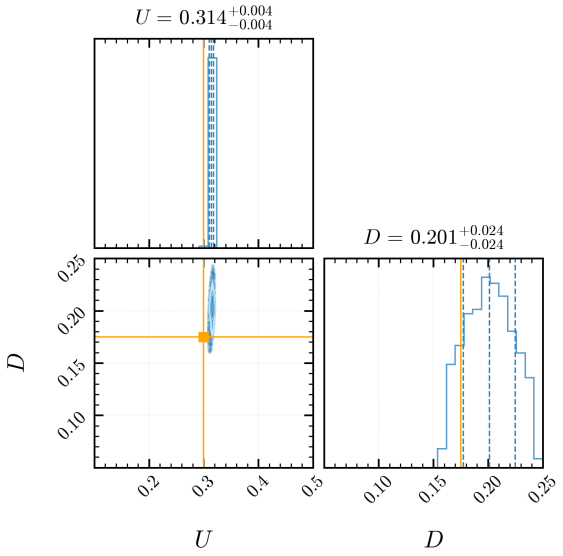
3.2 Mathematical Foundation

3.2.1 Conditional Normalizing Flows

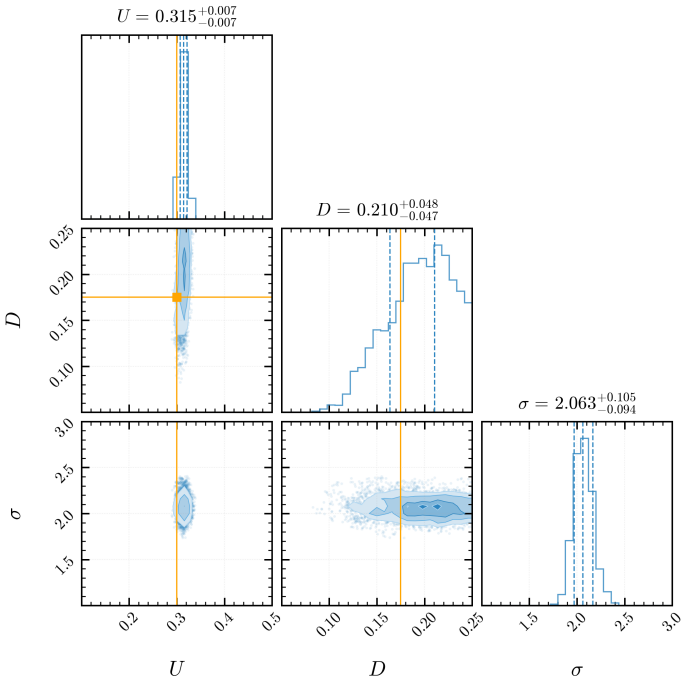
At the core of NPE are normalizing flows, a class of flexible neural architectures that learn complex probability distributions through invertible transformations.Normalizing flows have emerged as a powerful tool for density estimation in high-dimensional spaces [25, 44], offering significant advantages over traditional methods like mixture models or



(a) Approximate posterior from ABC applied directly to the stochastic model. Computational constraints and information loss lead to a diffuse posterior for D



(b) Gaussian approximation of the posterior (Laplace approximation) via MLE on a deterministic surrogate model. This approach is computationally efficient but conditional on the surrogate's fidelity.



(c) Posterior distribution from MCMC using the surrogate model with an assumed Gaussian noise model. This method jointly infers the model parameters alongside an additional noise variance parameter, σ.

Figure 2: Comparison of posterior distributions obtained using classical inference methods for a scratch assay model. Each approach demonstrates distinct trade-offs between computational efficiency, precision, and model assumptions. Orange lines indicate the true parameter values.

kernel density estimation when dealing with complex, multimodal distributions typical in biological inference problems. The fundamental idea behind normalizing flows is deceptively simple: construct a complex target distribution by

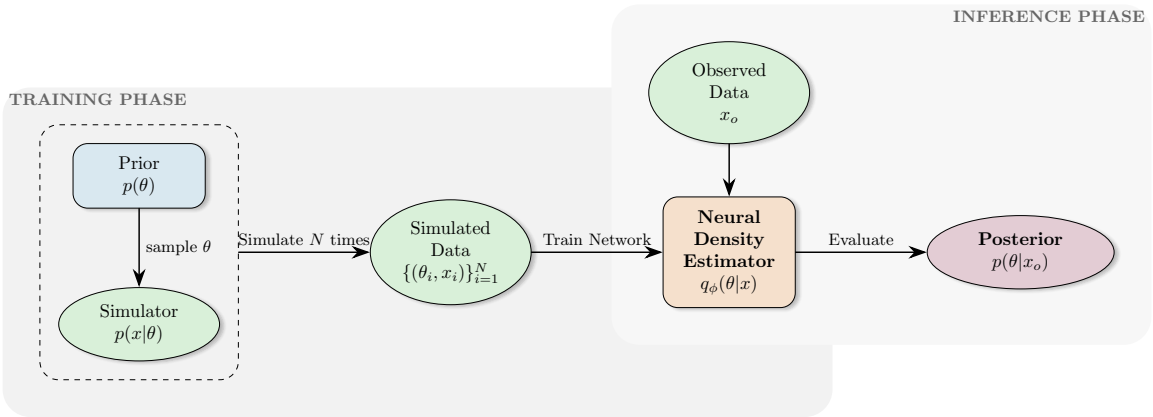


Figure 3: Schematic overview of the Neural Posterior Estimation (NPE) workflow. The training phase (left) involves sampling parameters from the prior, generating simulated data, and training a neural density estimator on the resulting parameter-data pairs. The inference phase (right) uses the trained network to instantly evaluate the posterior for new observations without additional simulations.

transforming samples from a simple base distribution (typically a standard Gaussian) through a series of invertible, differentiable mappings. This approach can be understood as a sophisticated generalization of the familiar inverse transform sampling method, but extended to high dimensions through neural networks [31]. Where traditional change-of-variables techniques might struggle with complex distributions, normalizing flows leverage the representational power of deep learning while maintaining the mathematical rigor needed for proper density estimation.

Formally, let $u \in \mathbb{R}^d$ be a random variable with base density $p_u(u) = \mathcal{N}(0, I)$. The flow defines an invertible transformation $f_\phi : \mathbb{R}^d \to \mathbb{R}^d$ parameterized by neural network weights ϕ. For conditional density estimation, this transformation is conditioned on the observation: $\theta = f_\phi(u; x)$. The invertibility requirement ensures that we can both generate samples (forward pass) and evaluate densities (inverse pass). The key insight is that complex distributions can be learned by composing simple transformations, each of which maintains invertibility and efficient Jacobian computation. This allows the network to learn highly flexible posterior shapes while maintaining computational tractability a crucial feature when inferring parameters in complex biological models where posteriors may exhibit non-Gaussian features such as multimodality, heavy tails, or curved degeneracies. The resulting approximate posterior density follows from the change of variables formula:

$$q_\phi(\theta|x) = p_u(f_\phi^{-1}(\theta; x)) \left|\det J_{f_\phi^{-1}}(\theta; x)\right| \tag{5}$$

where $J_{f_\phi^{-1}}$ denotes the Jacobian of the inverse transformation. Intuitively, the determinant of the Jacobian corrects for the local change in volume caused by the transformation, ensuring the resulting density is properly normalized. To understand this geometrically: when the transformation f_ϕ stretches space in some region, the determinant accounts for this stretching to ensure probability mass is conserved [44]. Modern flow architectures (e.g., coupling flows [16], autoregressive flows [45]) are designed such that this determinant is computationally efficient to calculate—often reducing to a sum or product of scalar terms rather than requiring explicit $d \times d$ matrix determinant computation. For readers familiar with traditional MCMC-based inference in biological models, normalizing flows offer a complementary approach: rather than exploring the posterior through iterative sampling, they learn an explicit parametric approximation that enables fast posterior evaluation and sampling once trained [14]. This amortization — training once on simulated data, then instantly obtaining posteriors for new observations — makes them particularly attractive for scenarios requiring repeated inference, such as model selection, experimental design, or real-time data analysis.

3.2.2 Training Procedure

NPE training requires a dataset of parameter-data pairs generated from the simulator. The training procedure follows three main steps:

1. **Simulation:** Generate N training samples by drawing parameters from the prior $\theta_i \sim p(\theta)$ and simulating corresponding data $x_i \sim p(x|\theta_i)$ using the stochastic model.

2. **Optimization:** Train the normalizing flow by maximizing the log-likelihood of parameters given their corresponding simulated data:

$$\mathcal{L}(\phi) = -\frac{1}{N}\sum_{i=1}^{N} \log q_\phi(\theta_i|x_i) \tag{6}$$

This objective is optimized using stochastic gradient descent (e.g., Adam [30]) with mini-batches of simulated pairs (θ_i, x_i). During training, gradients are computed via backpropagation through the flow transformation $f_\phi(u; x)$, updating the network parameters ϕ until convergence. This learns the transformation that maps from a simple base distribution to the posterior distribution conditioned on the data.

3. **Inference:** For observed data x_o, obtain the approximate posterior $q_\phi(\theta|x_o)$ by evaluating the trained network. Generate posterior samples by drawing $u \sim p_u(u)$ and applying the learned transformation $\theta = f_\phi(u; x_o)$.

The training data requirements scale with problem complexity, but it is difficult in general to provide quantitative recommendation on the number of required simulations to train the model. Typically at least $N = 10^4 - 10^6$ simulations suffice for moderate-dimensional parameter spaces ($d \leq 10$), though higher-dimensional problems may require larger datasets [15]. Crucially, this procedure is *amortized*; the computationally expensive training phase is performed only once, after which posterior inference for any new observation is nearly instantaneous. This represents a significant advantage over methods like ABC, which require fresh simulations for each new dataset. We provide quantitative comparisons in Section 5.3.

3.3 Sequential Neural Posterior Estimation

While the basic NPE algorithm is powerful, it can be inefficient when the prior is much broader than the posterior. Sequential Neural Posterior Estimation (SNPE) [43, 24] addresses this through an iterative refinement process that concentrates computational effort on parameter regions consistent with the observed data.

The sequential algorithm proceeds as follows:

1. **Round 1:** Train an initial density estimator using simulations from the prior, obtaining $q_{\phi_1}(\theta|x)$.
2. **Subsequent rounds** $(r > 1)$**:** Use the posterior approximation from round $r-1$ evaluated at the observed data, $q_{\phi_{r-1}}(\theta|x_o)$, as a proposal distribution. Generate new simulations $\theta \sim q_{\phi_{r-1}}(\theta|x_o)$, $x \sim p(x|\theta)$.
3. **Importance weighting correction:** Train a new density estimator $q_{\phi_r}(\theta|x)$ on the focused simulations, with importance weighting to correct for the proposal distribution bias:

$$\mathcal{L}_r(\phi) = -\frac{1}{N_r}\sum_{i=1}^{N_r} w_i \log q_\phi(\theta_i|x_i) \tag{7}$$

where $w_i = p(\theta_i)/q_{\phi_{r-1}}(\theta_i|x_o)$ corrects for sampling from the proposal rather than the prior.

SNPE is most beneficial when the prior is substantially broader than the posterior, which is common in biological applications where parameter ranges are often set conservatively. The method typically converges within 2-4 rounds, providing substantial efficiency gains over standard NPE while maintaining accuracy. In this work we will exclusively use the sequential version of NPE.

3.4 Implementation Details

All NPE analyses in this work use the `sbi` Python package [62], which provides a well-documented implementation of state-of-the-art simulation-based inference algorithms. We employ neural spline flows [18] as our density estimator architecture, which offer excellent expressivity while maintaining computational efficiency. Additional implementation details can be found in the project Github repository [2]are presented in

3.5 Application to Barrier Assay Data: 1D Column Count Analysis

We demonstrate NPE's effectiveness by applying it to the same synthetic barrier assay data used in Section 2.3. To ensure direct comparison with the classical methods, we first implement NPE using identical 1D column count summary statistics N_i, representing the number of agents in each vertical column of the lattice. In Section 4 we will extend this to operate directly on the 2D spatial data.

[2] https://github.com/tomkimpson/NPE_for_RW_inference

3.5.1 Experimental Setup

Following the protocol of S&P2025, we generate synthetic observations from a stochastic random walk model with true parameters $P = 0.7$ and $U = 0.3$, corresponding to diffusivity $D = P/4 = 0.175$ in the continuum limit. The resulting data consists of a 201-dimensional vector of column counts. We perform inference directly on the stochastic model parameters P and U, then transform to D via the relationship $D = P/4$ for comparison with PDE-based approaches. We use broad uniform priors covering the full parameter space: $P \sim \mathcal{U}(0.0, 1.0)$ and $U \sim \mathcal{U}(0.01, 1.0)$. These priors are determined by the physical constraints of the model: U and P are both probabilities and so must lie in $[0, 1]$. The lower bound of 0.01 for U excludes the degenerate case of zero initial agents while remaining sufficiently permissive. In the absence of strong prior information about parameter values, these uniform priors represent a minimally informative choice that allows the data to dominate the posterior inference. For NPE, the prior serves the dual role of defining the training distribution from which simulation parameters are drawn, and entering into the posterior distribution via Bayes' theorem [21, 22]

3.5.2 Parameter Inference Results

Figure 4 presents the posterior distribution obtained via NPE. NPE recovers a posterior distribution with means $P = 0.698 \pm 0.037$ and $U = 0.301 \pm 0.022$ (mean $\pm$ standard deviation), corresponding to $D = 0.175 \pm 0.009$ after transformation. The results demonstrate excellent agreement with the true parameter values ($P = 0.7$, $U = 0.3$)

This approach yields several key advantages over classical methods. The NPE posterior shows substantially improved precision, than a posterior obtained from a comparable ABC analysis, especially for the diffusion paramter D, all without the need for manual tolerance tuning. Furthermore, because NPE learns directly from the stochastic simulator, it avoids the systematic bias of mean-field approximations and produces a posterior that is not artificially narrow. The resulting credible intervals accurately reflect the natural stochastic variance of the data, a feature often lost when using deterministic surrogate models. Once the initial network training is complete, posterior samples for any new observation can be obtained almost instantly.

3.5.3 Predictive Performance

Beyond parameter estimation, we evaluate NPE's predictive capabilities through posterior predictive sampling. For each posterior sample $\theta^{(j)}$, we generate predicted data $\tilde{x}^{(j)} \sim p(x|\theta^{(j)})$ using the stochastic simulator. The collection of predictions forms empirical prediction intervals that naturally account for both parameter uncertainty and model stochasticity.

The resulting NPE-based predictions are well-calibrated, with 94.8% of true observations falling within the 95% prediction intervals. A key advantage of this simulator-native approach is its physical consistency; the prediction intervals respect the non-negative, discrete nature of the count data without requiring an explicit, and potentially erroneous, noise model. Furthermore, the framework naturally captures heteroscedastic uncertainty, with the prediction intervals appropriately widening in regions of high agent density where stochastic variability is greatest.

4 Learning from Raw Spatial Data

4.1 Motivation for Full Spatial Data

While column counts provide a natural summary of the spreading process, they necessarily discard information about the vertical distribution of agents. In principle, the spatial correlation structure within the 2D configuration contains additional information about the underlying parameters. For instance, clustering patterns, local density fluctuations, and spatial correlations may help distinguish between different combinations of P and U that produce similar column counts, cf. parameter identifiability [3]. Traditional inference methods require hand-crafted summary statistics, making it challenging to leverage such high-dimensional spatial information. NPE, however, can learn directly from raw data through automatic feature extraction. While the simple random walk models considered here are likely well-captured by 1D summary statistics, this approach demonstrates the broader potential of NPE for more complex biological systems where spatial structure is critical. For example, models incorporating agent-agent interactions, complex boundary conditions, spatially heterogeneous environments, or multi-species dynamics would generate spatial patterns that cannot be adequately summarized by simple column counts. In such cases, the ability to learn directly from full spatial data becomes essential for accurate parameter inference. Our framework provides a foundation that can be readily

[3] Although parameter identifiability is not a significant concern for the simple model we consider, additional spatial features are of interest for disentangling parameters in more complex models that incorporate processes like cell-cell adhesion or proliferation. We note that for longer times even this simple model becomes non-identifiable in the diffusion coeccicient D

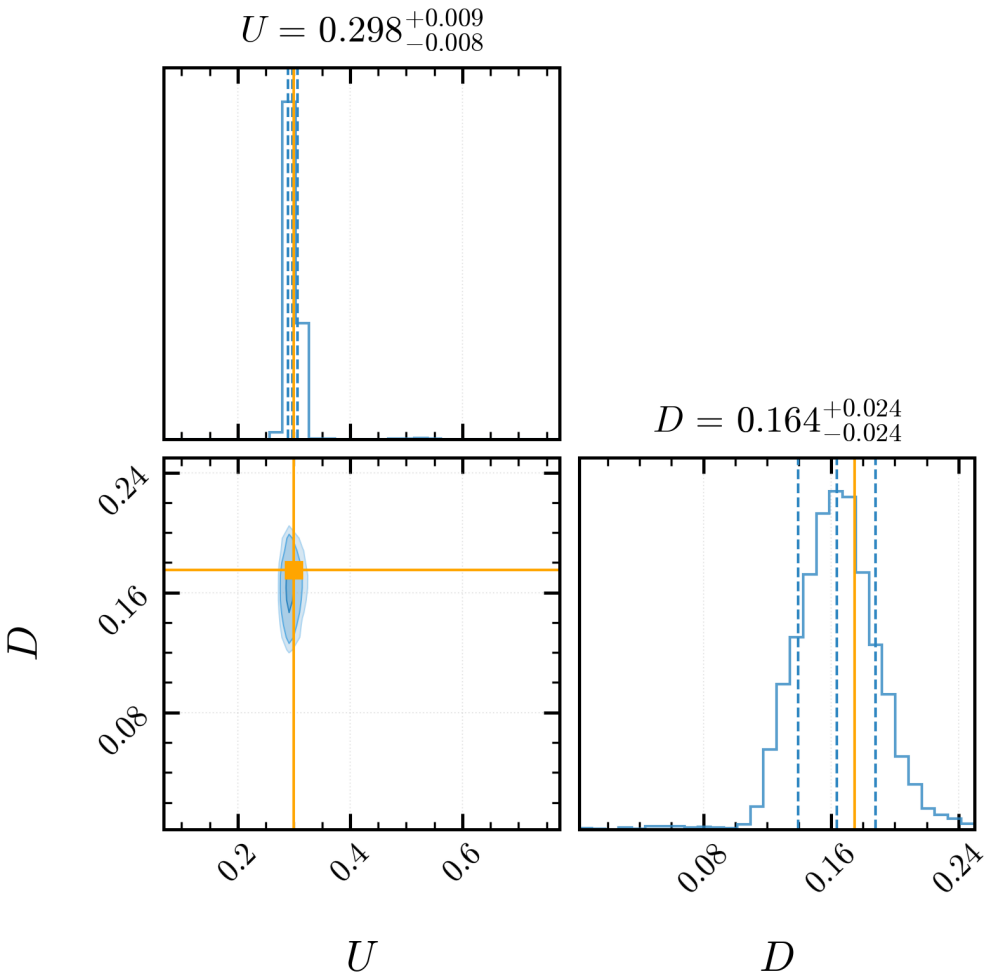


Figure 4: Posterior distribution from NPE using 1D column count data. The contours show 68% and 95% credible regions, with the true parameter values (organge cross) falling well within the high-probability region.

extended to these more sophisticated modeling scenarios without requiring manual specification of appropriate summary statistics.

4.2 Architectural Adaptations for 2D Data

Processing raw 2D spatial data requires a different architecture than the one used for 1D summary statistics. We treat each lattice configuration from the simulator as a binary image, where pixel values indicate the presence or absence of an agent. To learn informative features directly from this high-dimensional input, we replace the simple feed-forward network with a Convolutional Neural Network (CNN) [34, 32, 23]. The CNN acts as a powerful, learnable feature extractor that is embedded within the NPE pipeline. It uses a series of convolutional and pooling layers to progressively downsample the spatial data into a low-dimensional embedding. This process is designed to capture salient spatial features, such as clustering patterns and local agent densities, while preserving translational invariance. The learned embedding is then passed to the remainder of the network to estimate the posterior distribution. This end-to-end training ensures that the features extracted by the CNN are maximally informative for the specific task of parameter inference. A detailed specification of the CNN architecture, training parameters, and implementation details used in this work can be found in Additional implementation details can be found in the project github repository.

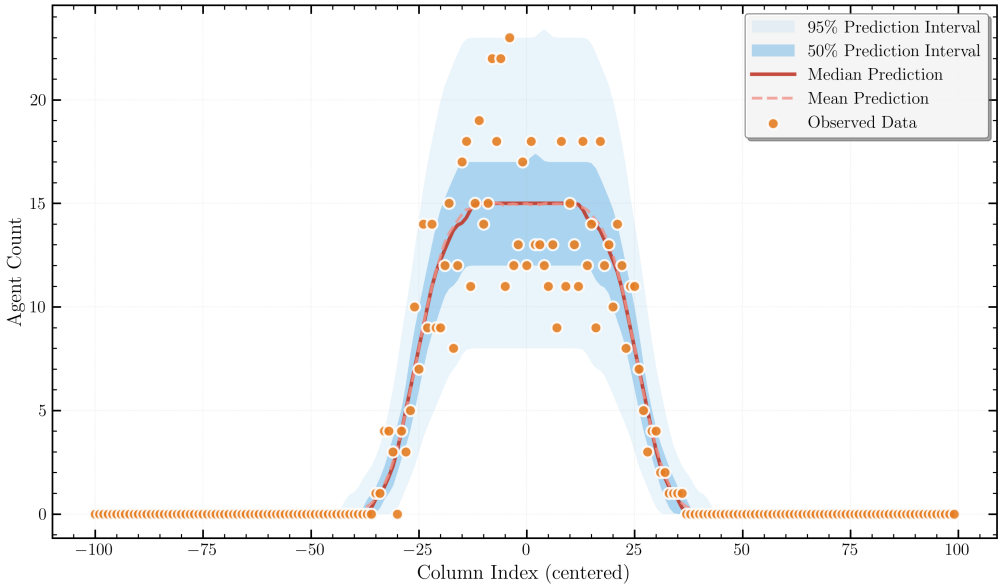


Figure 5: Posterior predictive checks from NPE using 1D column count data. The observed data (orange circles) is overlaid on predictions generated by simulating the model with parameters drawn from the inferred posterior distribution. The median (solid red line) and mean (dashed red line) of the predictions are shown, along with the 50% (dark blue) and 95% (light blue) prediction intervals. For visualization purposes, a Gaussian kernel has been applied to smooth the predictive distributions. The observed data points lie well within the 95% prediction interval, indicating that the calibrated model is consistent with the data

4.3 Parameter Inference from Raw Spatial Data

Having established a baseline with 1D summary statistics, we now apply our NPE pipeline with the embedded CNN to perform inference directly on the raw 2D spatial data. This approach leverages the full spatial configuration of agents, aiming to extract additional information that may be lost when compressing the data into column counts.

4.3.1 Experimental Setup

The experimental setup remains identical to the 1D case, using the same synthetic observation generated with true parameters $P = 0.7$ ($D = 0.7/4$) and $U = 0.3$. However, instead of using the 201-dimensional vector of column counts as input, the network now receives the full 50×200 binary image representing the agent locations on the lattice. We use the same broad uniform priors for P and U as before.

4.3.2 Parameter Inference Results

Figure 6 shows the joint posterior distribution inferred from the 2D spatial data, overlaid with the result from the 1D column count analysis for direct comparison. The CNN-based approach recovers a posterior with means $D = 0.169^{+0.026}_{-0.023}$ and $U = 0.290^{+0.008}_{-0.09}$. As summarised in Table 2, leveraging the full spatial data yields a posterior with a precision that is highly comparable to that obtained using curated 1D summary statistics, cf. Section 3.5.

This result demonstrates that the CNN can automatically learn features from the raw data that are as informative as the carefully chosen summary statistics. This highlights a critical trade-off: the increased computational cost of training the CNN (Table 2) must be weighed against the major advantage of avoiding manual feature engineering. For more complex models where optimal summary statistics are unknown, this direct-from-pixels approach may be advantageous for accurate inference.

4.3.3 Predictive Performance

We assess the predictive performance of the 2D-trained model using the same posterior predictive checking procedure as in Section 3.5.3. As shown in Figures 7, 8, the model generates predictions that are highly consistent with the observed data. The calibrated model accurately captures the spatial characteristics of the agent-based simulation, producing prediction intervals that are well-calibrated and physically consistent with the discrete nature of the data. Importantly, the

framework captures the input-dependent (heteroscedastic) uncertainty—something that standard maximum-likelihood estimators under an i.i.d. noise assumption (as used in much of the mathematical biology literature) cannot represent, often leading to miscalibrated, homoscedastic intervals even when the underlying variability is spatially varying. As in the 1D case, the method remains robust when applied to high-dimensional (2D) spatial inputs.

In the 2D (CNN) results we also note a modest separation between the posterior predictive mean and median, consistent with a mild right-skew in the posterior for D. We attribute the observed skew primarily to structural features of the inference problem—namely a slight U–D trade-off, support constraints ($D \geq 0$ with $D = P/4$), and the nonlinear parameter-to-observable map—together with the fact that the CNN accesses richer spatial cues (e.g., front roughness and local clustering) that are averaged out in 1D summaries and therefore make such asymmetries more apparent. We cannot entirely rule out minor artefacts arising from the higher-dimensional learning task.

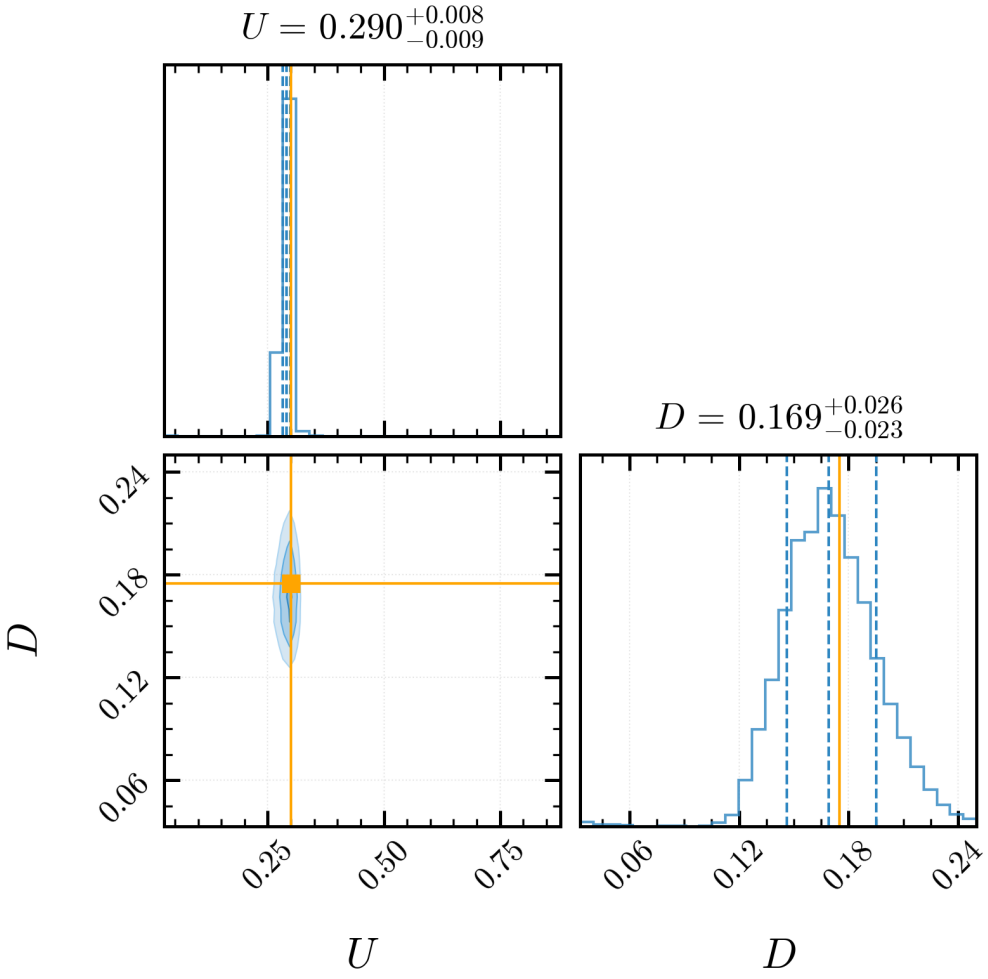


Figure 6: As Figure 4, for posteriors generated from the NPE with CNN embedding networks, using full 2D spatial data.

Data Type	**Posterior for D**	**Posterior for U**
1D Column counts	$0.164\ [0.024, 0.024]$	$0.298\ [0.009, 0.008]$
2D Spatial data	$0.169\ [0.026, 0.023]$	$0.290\ [0.008, 0.009]$

Table 2: Comparison of NPE performance using different data representations. Posteriors are reported as the mean and the 95% credible interval. Inference on the 2D spatial data provides comparable posterior precision to the 1D summary statistics.

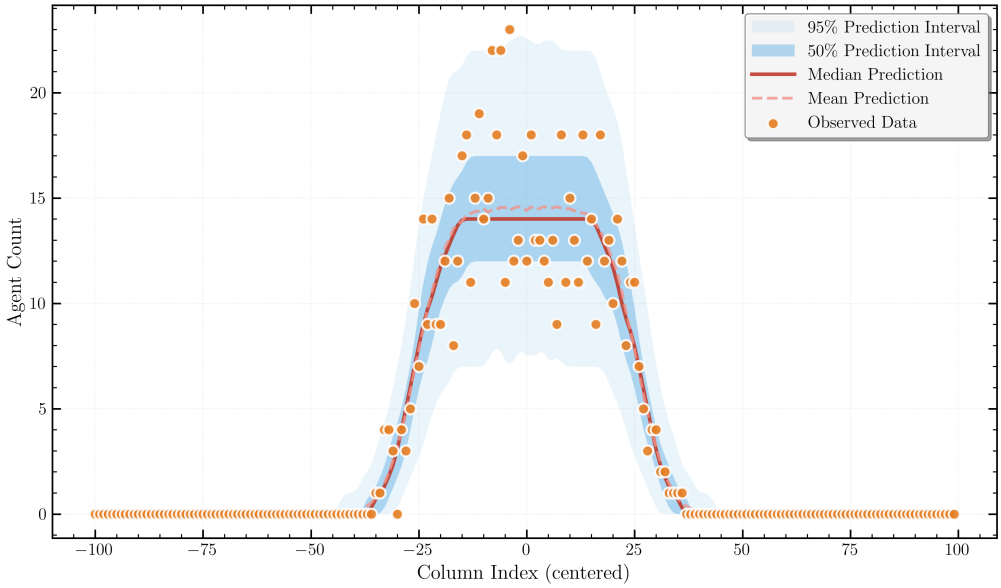


Figure 7: Posterior predictive check for the 2D spatial data model. The format is identical to Figure 5. The predictions are well-calibrated and consistent with the observed data.

5 Discussion

5.1 Advantages and Limitations of Neural Posterior Estimation

NPE offers a robust and flexible framework for parameter inference, directly addressing several challenges that have limited the practical application of stochastic agent-based models. NPE eliminates the need for surrogate models by learning directly from the full stochastic simulator. This direct approach circumvents the systematic bias introduced by mean-field approximations and removes the challenging requirement of specifying explicit noise models that connect deterministic surrogates to stochastic observations. The method's ability to process raw high-dimensional data through automatic feature extraction represents another advantage, avoiding the information loss inherent in manually crafted summary statistics while enabling the analysis of complex spatial patterns that would be difficult to characterize a priori. The amortized nature of NPE provides substantial computational advantages once training is complete. After the initial investment in generating training data and optimizing the neural network, posterior inference for new observations requires only milliseconds of network evaluation rather than hours of simulation-based methods like ABC. This efficiency enables real-time analysis and interactive exploration of parameter space that would be prohibitive with classical approaches. Furthermore, normalizing flows can capture complex, multimodal posterior geometries that arise naturally from model non-identifiabilities without requiring restrictive parametric assumptions about posterior shape.

However, NPE also introduces new challenges and limitations that practitioners must carefully consider. The method's appetite for training data can be substantial, particularly for high-dimensional parameter spaces where the curse of dimensionality becomes prohibitive. Problems with more than 15-20 parameters may require simulation budgets that exceed practical computational limits. Additionally, NPE performance can degrade significantly when test observations fall outside the training distribution, making careful prior specification and robust training data generation critical for reliable inference.

The neural network architecture introduces another layer of complexity, as design choices regarding depth, width, and flow type can dramatically impact performance. Unlike classical methods with well-established theoretical foundations, NPE provides limited formal guarantees about convergence or approximation quality. This black-box nature can make diagnosis of poor performance challenging and requires practitioners to develop intuition about network behavior and training dynamics.

5.2 Spatial information

Our results highlight a critical trade-off between methodological complexity and computational cost. For the simple random walk model, well-chosen summary statistics provide comparable precision to the full spatial analysis at a

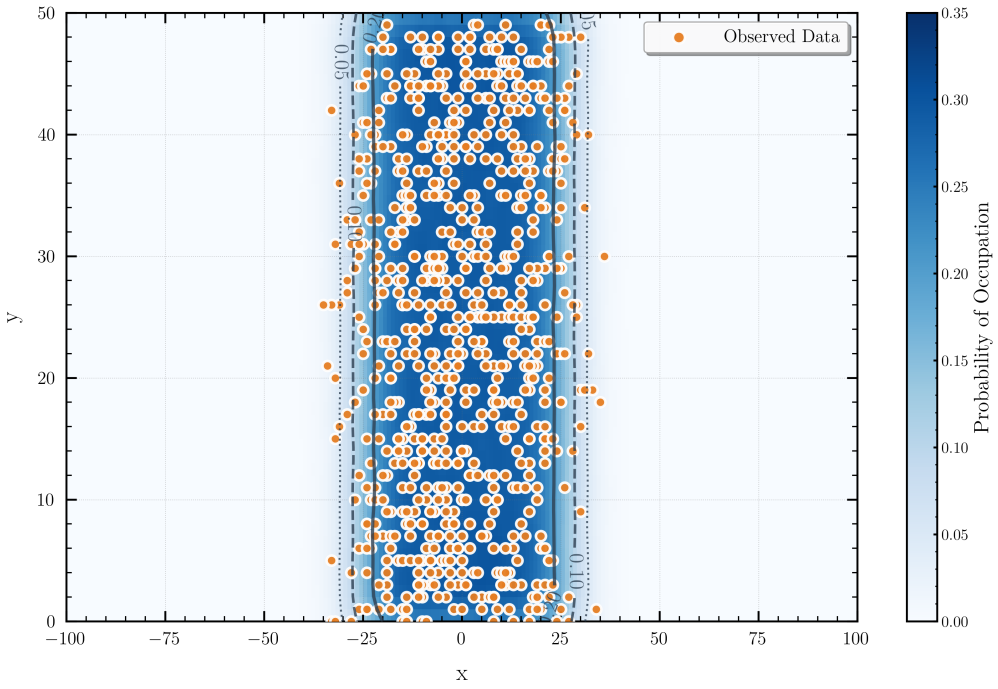


Figure 8: Posterior predictive check for the 2D spatial data model, showing the full observed data points (orange) overlaid on the predicted occupany probability from the NPE model (blue heatmap). The grey contours delineate iso-probability curves at probability levels of 0.05, 0.10, and 0.20 (dotted, dashed, solid, respectively)

fraction of the cost, making them a highly effective and efficient choice. However, this should not obscure the broader potential of spatial inference approaches; the cost-benefit calculation shifts decisively toward spatial approaches for more complex biological systems where crucial information is encoded in patterns that resist simple summarization, or where informative statistics are not known a priori.

In many realistic scenarios, crucial mechanistic information is encoded in intricate spatial patterns that are lost in simple one-dimensional projections. For example, models incorporating agent-agent interactions through exclusion effects, chemotaxis, or cell-cell adhesion generate complex formations that are lost when compressed into simple one-dimensional projections. Similarly, heterogeneous environments with spatially varying growth rates, physical barriers, or chemical gradients create complex spatial signatures that contain crucial parameter information. Multi-species dynamics involving competition, predation, or cooperation also produce emergent spatial organization that encodes information about underlying mechanisms. Collective behaviors such as swarming, flocking, or tissue morphogenesis involve coordinated movements that are fundamentally spatial phenomena requiring spatial analysis for proper characterization. In these more complex scenarios, CNN approaches would preserve information that would otherwise be discarded by traditional summary statistics.

A critical aspect of applying deep learning models like CNNs in a scientific context is interpretability. While these networks are often treated as "black boxes", it is both possible and important to investigate the features they learn to solve a given inference problem. In principle, techniques such as activation mapping [39] or saliency maps [40] could be used to visualize the spatial features our CNN learns to associate with different parameter values. One might expect the network to develop sensitivity to biologically relevant patterns like the sharpness of the spreading front, local cell density gradients, or the degree of cell clustering. By identifying these learned features, a researcher could validate that the network's reasoning aligns with established biological theory, or discover novel spatial statistics that are informative but not immediately obvious to a human observer. While a detailed feature analysis is beyond the scope of this work, we highlight its importance for future applications. The ability to not only perform inference but also to extract the learned representations that enable it is a key advantage of this approach.

The ability to perform inference with stochastic models on spatiotemporal data has direct implications for experimental design. Traditional protocols often reduce rich data from sources like time-lapse microscopy to simplified measurements. This is a compromise driven by past computational limits, not biological necessity. Our framework removes this barrier, enabling direct analysis that preserves the full spatial information. This should encourage the strategic collection and

retention of rich datasets, which is crucial for improving parameter identifiability in complex systems where spatial structure encodes mechanistic information.

5.3 Computational Scalability and Future Directions

The NPE framework, particularly with the CNN-based spatial analysis, introduces several computational considerations that will influence its applicability to large-scale problems. To provide a practical comparison of these costs, Table 3 summarises representative runtimes for the different inference methods. These figures are illustrative of a typical workflow and are not the result of systematic hyperparameter optimisation. For example, the NPE training time scales roughly linearly with the number of sequential rounds: the reported 1D case used ten rounds, so halving or doubling this number would proportionally reduce or increase the runtime, with an uninvestigated trade-off in posterior accuracy. Similarly, ABC runtime depends directly on the number of simulations generated (here, 10^5). Other hyperparameters (e.g. number of network epochs, batch size, learning rate, width/depth of the embedding network) also have a direct influence on both runtime and accuracy, but were fixed to reasonable defaults rather than tuned for performance. We also note the difference in hardware: the computationally expensive NPE training phase was performed on a GPU, whereas the other methods ran on a standard CPU.

Method	**Model / Data Type**	**Training Time (s)**	**Inference Time (s)**
ABC	Stochastic Simulator (1D Data)	–	$\sim$1200
ABC	Surrogate Model (1D Data)	–	$\sim$2
MCMC	Surrogate Model (1D Data)	–	$\sim$1
NPE	**Stochastic Simulator (1D Data)**	**$\sim$6400**	**$<$ 0.1**
NPE + CNN	**Stochastic Simulator (2D Data)**	**$\sim$9800**	**$<$ 0.1**

Table 3: Representative computational runtimes for various inference methods. NPE requires a significant one-time training cost, after which inference is near-instantaneous. The NPE training was conducted on a GPU, while other methods were run on a CPU.

Two distinct computational strategies emerge from this comparison. Surrogate-based approaches are extremely fast because they entirely bypass repeated stochastic simulations, relying instead on precomputed or deterministic approximations of the model. In contrast, the NPE framework demands a substantial one-time computational investment during training but then performs posterior inference for new observations almost instantaneously. This amortised structure makes NPE particularly attractive for scenarios requiring repeated or real-time inference across many datasets.

The primary cost for any NPE approach is the upfront simulation budget, which can become substantial for models with high-dimensional parameter spaces ($> 15 - 20$ parameters). On top of this, the CNN adds specific challenges; (i) memory requirements scale quadratically with spatial resolution, potentially limiting application to very high-resolution imaging data without careful optimization; (ii) architecture selection for CNN embedding networks requires problem-specific tuning of depth, width, and kernel sizes, adding complexity to the analysis pipeline; (iii) the combined optimization of feature extraction and density estimation can be more challenging than simpler approaches, requiring careful hyperparameter selection and potentially longer training times.

Despite these challenges, several technological trends suggest that spatial inference approaches will become increasingly practical. Advances in GPU computing and specialized hardware for neural network training continue to reduce computational barriers. Improved network architectures and training algorithms may reduce the simulation budgets required for reliable inference. Transfer learning approaches could enable pre-trained spatial feature extractors to be adapted to new biological problems with reduced computational cost.

The framework developed here provides a foundation that can evolve with these technological advances while addressing increasingly complex biological questions. Future extensions might incorporate three-dimensional spatial data, temporal dynamics, or multi-modal observations combining spatial patterns with biochemical measurements. As biological modeling advances toward more realistic representations of complex systems, the ability to learn directly from high-dimensional, spatially structured data will become essential for extracting quantitative insights from biological observations.

6 Conclusion

Parameter inference for spatially explicit stochastic biological models has long been hampered by a fundamental tension: the need for computational tractability versus the desire for biological realism. Classical approaches have required researchers to choose between computationally demanding likelihood-free methods that operate directly on stochastic simulators, or efficient likelihood-based techniques that rely on potentially unfaithful surrogate models and explicit noise specifications. This work demonstrates that Neural Posterior Estimation provides a principled path beyond this dichotomy.

We have shown that NPE, augmented with modern deep learning architectures, addresses each of the canonical challenges that have historically limited stochastic biological modeling. By learning posteriors directly from simulator output, the framework circumvents intractable likelihoods while avoiding the systematic biases introduced by mean-field approximations. The method's implicit learning of complex stochastic relationships eliminates the need for explicit, and potentially erroneous, noise model specification. Once trained, the amortized nature of NPE provides near-instantaneous posterior estimates, overcoming the prohibitive per-inference costs of traditional likelihood-free methods. Perhaps most significantly, by integrating convolutional neural networks as learned feature extractors, the framework can process high-dimensional spatial data directly, removing the need for hand-crafted summary statistics that risk substantial information loss.

Our application to agent-based random walk models demonstrates these advantages in practice. On one-dimensional summary statistics, NPE matches or exceeds the precision of classical methods while requiring no manual tuning of tolerances or distance metrics. When extended to raw two-dimensional spatial data through CNN-based feature extraction, the framework maintains comparable precision without requiring researchers to specify informative summary statistics a priori. This capability becomes critical for complex biological systems where crucial mechanistic information is encoded in spatial patterns that resist simple summarization.

The broader implications extend beyond the specific models considered here. As biological research increasingly demands quantitative frameworks that can accommodate realistic stochasticity, spatial heterogeneity, and multi-scale dynamics, the ability to perform rigorous inference directly on high-fidelity simulators becomes essential. The framework we have developed provides a foundation for this next generation of biological modeling, where machine learning techniques can bridge the gap between model complexity and inferential tractability.

Nevertheless, important challenges remain. The substantial upfront computational cost of generating training data, the difficulty of ensuring robust performance outside the training distribution, and the limited theoretical guarantees compared to classical methods all require careful consideration. The architecture design choices and hyperparameter tuning introduce additional complexity that may present barriers to adoption. Future work should focus on reducing these barriers through improved architectures, more efficient training algorithms, and the development of practical guidelines for applying these methods across diverse biological systems.

We provide an open-source implementation of our complete pipeline[4] to facilitate adoption and enable the community to extend these methods to increasingly complex spatially-structured models. As computational resources continue to expand and neural network architectures become more sophisticated, we anticipate that simulation-based inference frameworks like NPE will become the standard approach for parameter estimation in stochastic biological systems.

[4] github.com/tomkimpson/NPE_for_RW_inference

References

[1] Ruth E. Baker, Christian A. Yates, and Radek Erban. From microscopic to macroscopic descriptions of cell migration on growing domains. *Bulletin of Mathematical Biology*, 72:719–762, 2010.

[2] M. A. Beaumont, W. Zhang, and D. J. Balding. Approximate Bayesian computation in population genetics. *Genetics*, 162(4):2025–2035, 2002.

[3] José M. Bernardo and Adrian F. M. Smith. *Bayesian Theory*. John Wiley & Sons, Chichester, UK, 1994.

[4] M. G. B. Blum. Approximate Bayesian computation: a nonparametric perspective. *Journal of the American Statistical Association*, 105(491):1178–1187, 2010.

[5] M. G. B. Blum, M. A. Nunes, D. Prangle, and S. A. Sisson. A comparative review of approximate Bayesian computation. *Statistical Science*, 28(2):189–208, 2013.

[6] M. Bruna and S. J. Chapman. Excluded-volume effects in the diffusion of hard spheres. *The Journal of Chemical Physics*, 137(20):204116, 2012.

[7] J. Brynjarsdóttir and A. O'Hagan. Learning about physical parameters: the importance of model discrepancy. *Inverse Problems*, 30(11):114007, 2014.

[8] Michael J Carr, Matthew J. Simpson, and Christopher Drovandi. Estimating parameters of a stochastic cell invasion model with fluorescent cell cycle labelling using approximate bayesian computation. *J. R. Soc. Interface.*, 18:20210362, 2021.

[9] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Press, Belmont, CA, 1990.

[10] Sai Hung Cheung, Todd A. Oliver, Ernesto E. Prudencio, Serge Prudhomme, and Robert D. Moser. Bayesian uncertainty analysis with applications to turbulence modeling. *Reliability Engineering & System Safety*, 96(9):1137–1149, 2011.

[11] Edward A. Codling, Michael J. Plank, and Simon Benhamou. Random walk models in biology. *Journal of the Royal Society Interface*, 5(25):813–834, 2008.

[12] Brandon S Coventry and Edward L Bartlett. Practical bayesian inference in neuroscience: Or how i learned to stop worrying and embrace the distribution. *eNeuro*, 11:ENEURO.0484–23.2024, 2024.

[13] D. R. Cox and D. V. Hinkley. *Theoretical Statistics*. Chapman & Hall, London, 1974.

[14] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

[15] Michael Deistler, Jan Boelts, Peter Steinbach, Guy Moss, Thomas Moreau, Manuel Gloeckler, Pedro L. C. Rodrigues, Julia Linhart, Janne K. Lappalainen, Benjamin Kurt Miller, Pedro J. Gonçalves, Jan-Matthis Lueckmann, Cornelius Schröder, and Jakob H. Macke. Simulation-Based Inference: A Practical Guide. *arXiv e-prints*, page arXiv:2508.12939, August 2025.

[16] Laurent Dinh, Jascha Sohl-Dickstein, and Samy Bengio. Density estimation using Real NVP. *arXiv e-prints*, page arXiv:1605.08803, May 2016.

[17] David B. Dunson. Bayesian perspectives for epidemiological research. *The International Journal of Epidemiology*, 34:1170–1173, 2005.

[18] Conor Durkan, Artur Bekasov, Iain Murray, and George Papamakarios. Neural Spline Flows. *arXiv e-prints*, page arXiv:1906.04032, June 2019.

[19] R. Erban and S. J. Chapman. Stochastic modelling of reaction–diffusion processes in biology. *SIAM Review*, 50(1):37–114, 2009.

[20] P. Fearnhead and D. Prangle. Constructing summary statistics for approximate Bayesian computation: semi-automatic ABC. *Journal of the Royal Statistical Society: Series B*, 74(3):419–474, 2012.

[21] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, London, 1995.

[22] Andrew Gelman, Daniel Simpson, and Michael Betancourt. The Prior Can Often Only Be Understood in the Context of the Likelihood. *Entropy*, 19(10):555, October 2017.

[23] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

[24] David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2404–2414. PMLR, 09–15 Jun 2019.

[25] Danilo Jimenez Rezende and Shakir Mohamed. Variational Inference with Normalizing Flows. *arXiv e-prints*, page arXiv:1505.05770, May 2015.

[26] W. Jin, E. T. Shah, C. J. Penington, S. W. McCue, L. K. Chopin, and M. J. Simpson. Reproducibility of scratch assays is affected by the initial degree of confluence: Experiments, modelling and model selection. *Journal of Theoretical Biology*, 390:136–145, February 2016. Epub 2015-11-29. PMID: 26646767.

[27] Wang Jin, Esha T. Shah, Catherine J. Penington, Scott W. McCue, Lisa K. Chopin, and Matthew J. Simpson. Reproducibility of scratch assays is affected by the initial degree of confluence: experiments, modelling and model selection. *Journal of Theoretical Biology*, 390:136–145, 2016.

[28] Stuart T. Johnston, Matthew J. Simpson, and D. L. S. McElwain. How much information can be obtained from tracking the position of the leading edge in a scratch assay? *Journal of The Royal Society Interface*, 11(95):20140325, 2014.

[29] M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B*, 63(3):425–464, 2001.

[30] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[31] Ivan Kobyzev, Simon J.D. Prince, and Marcus A. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3964–3979, 2021.

[32] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pages 1097–1105, 2012.

[33] John H. Lagergren, John T. Nardini, Ruth E. Baker, Matthew J. Simpson, and Kevin B. Flores. Biologically-informed neural networks guide mechanistic modeling from sparse experimental data. *PLOS Computational Biology*, 16(12):e1008462, 2020.

[34] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[35] Fei Liu, Shao-Wu Zhang, Wei-Feng Guo, Ze-Gang Wei, and Luonan Chen. Inference of gene regulatory network based on local bayesian networks. *PLOS Computational Biology*, 12:e1005024, 2016.

[36] Yihan Liu, David J. Warne, and Matthew J. Simpson. Likelihood-based inference, identifiability, and prediction using count data from lattice-based random walk models. *Physical Review E*, 110(4):044405, 2024.

[37] J.-M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate Bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.

[38] P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman & Hall, London, 2nd edition, 1989.

[39] Anh Pham Thi Minh. Overview of Class Activation Maps for Visualization Explainability. *arXiv e-prints*, page arXiv:2309.14304, September 2023.

[40] T. Nathan Mundhenk, Barry Y. Chen, and Gerald Friedland. Efficient Saliency Maps for Explainable AI. *arXiv e-prints*, page arXiv:1911.11293, November 2019.

[41] John T. Nardini. Forecasting and Predicting Stochastic Agent-Based Model Data with Biologically-Informed Neural Networks. *Bulletin of Mathematical Biology*, 86:130, 2024.

[42] Fabrícia F. Nascimento, Mario dos Reis, and Ziheng Yang. A biologist's guide to bayesian phylogenetic analysis. *Nature Ecology and Evolution*, 1:1446–1454, 2017.

[43] George Papamakarios and Iain Murray. Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. *arXiv e-prints*, page arXiv:1605.06376, May 2016.

[44] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

[45] George Papamakarios, Theo Pavlakou, and Iain Murray. Masked Autoregressive Flow for Density Estimation. *arXiv e-prints*, page arXiv:1705.07057, May 2017.

[46] Planck Collaboration, N. Aghanim, Y. Akrami, et al. Planck 2018 results. vi. cosmological parameters. *Astronomy & Astrophysics*, 641:A6, 2020.

[47] M. J. Plank, M. J. Simpson, and R. E. Baker. Random walk models in the life sciences: including births, deaths and local interactions. *Journal of the Royal Society Interface*, 22:20240422, 2025.

[48] J. K. Pritchard, M. T. Seielstad, A. Perez-Lezaun, and M. W. Feldman. Population growth of human Y chromosomes: a study of Y-chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798, 1999.

[49] Song S. Qian, Kevin Craig, Melissa M. Baustian, and N. N. Rabalais. A bayesian hierarchical modeling approach for analyzing observational data from marine ecological studies. *Marine Pollution Bulletin*, 58:1916–1921, 2009.

[50] Michael Rabinowitz and Swaroop Sridhar. Nested sampling for parameter inference in systems biology. *Journal of the Royal Society Interface*, 10:20130664, 2013.

[51] C. P. Robert, J.-M. Cornuet, J.-M. Marin, and N. S. Pillai. Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences*, 108(37):15112–15117, 2011.

[52] Ori Rosen, Wenxin Jiang, Gary King, and Martin A. Tanner. Bayesian and frequentist inference for ecological inference: the r $\times$ c case. *Statistica Neerlandica*, 55(2):134–156, 2001.

[53] Robert J. H. Ross, Ruth E. Baker, Andrew Parker, Matthew J. Ford, Richard L. Mort, and Christian A. Yates. Using approximate bayesian computation to quantify cell–cell adhesion parameters in a cell migratory process. *npj Systems Biology and Applications*, 3(1):9, 2017.

[54] Daria Semochkina and Cathal D. Walsh. Incorporating additional evidence as prior information to resolve non-identifiability in bayesian disease model calibration. a tutorial. a preprint. *arXiv preprint arXiv:2407.13451v2*, January 2025.

[55] Matthew J. Simpson, Ruth E. Baker, Pascal R. Buenzli, Ruanui Nicholson, and Oliver J. Maclaren. Reliable and efficient parameter estimation using approximate continuum limit descriptions of stochastic models. *Journal of Theoretical Biology*, 549:111201, 2022.

[56] Matthew J. Simpson, Ryan J. Murphy, and Oliver J. Maclaren. Modelling count data with partial differential equation models in biology. *Journal of Theoretical Biology*, 580:111732, 2024.

[57] Matthew J. Simpson and Michael J. Plank. Inference and prediction for stochastic models of biological populations undergoing migration and proliferation. *Journal of the Royal Society Interface*, 22:20250536, 2025.

[58] Matthew J. Simpson, Katrina K. Treloar, Benjamin J. Binder, Parvathi Haridas, Kerry J. Manton, David I. Leavesley, D. L. S. McElwain, and Ruth E. Baker. Quantifying the roles of cell motility and cell proliferation in a circular barrier assay. *Journal of The Royal Society Interface*, 10(82):20130007, 2013.

[59] Scott A. Sisson, Yanan Fan, and Mark Beaumont, editors. *Handbook of Approximate Bayesian Computation*. Chapman and Hall/CRC, New York, 1 edition, 2018. First published 2018, eBook published 3 September 2018.

[60] Mikael Sunnåker, Alberto Giovanni Busetto, Elina Numminen, Jukka Corander, Matthieu Foll, and Christophe Dessimoz. Approximate Bayesian Computation. *PLOS Computational Biology*, 9(1):1–10, 01 2013.

[61] S. Tavare, D. J. Balding, R. C. Griffiths, and P. Donnelly. Inferring coalescence times from dna sequence data. *Genetics*, 145(2):505–518, 1997.

[62] Alvaro Tejero-Cantero, Jan Boelts, Michael Deistler, Jan-Matthis Lueckmann, Conor Durkan, Pedro Gonçalves, David Greenberg, and Jakob Macke. sbi: A toolkit for simulation-based inference. *The Journal of Open Source Software*, 5(52):2505, August 2020.

[63] Catalina A. Vallejos, John C. Marioni, and Sylvia Richardson. BASiCS: Bayesian Analysis of Single-Cell Sequencing Data. *PLoS Computational Biology*, 11:e1004333, 2015.

[64] John Veitch et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the lalinference software library. *Physical Review D*, 91(4):042003, 2015.

[65] Jasper A. Vrugt, Cajo D. ter Braak, Martyn P. Clark, and Eric H. H. van den Eijnden. Computational methods for parameter estimation in climate models. *Bayesian Analysis*, 9:535–575, 2014.

[66] Lingfei Wang, Pieter Audenaert, and Tom Michoel. High-dimensional bayesian network inference from systems genetics data using genetic node ordering. *Frontiers in Genetics*, 10:1196, December 2019.

[67] David J. Warne, Ruth E. Baker, and Matthew J. Simpson. Rapid Bayesian inference for expensive stochastic models. *Journal of Computational and Graphical Statistics*, 31(2):512–528, 2022.