Title

Inferring stochastic rates from heterogeneous snapshots of particle positions.

Permalink

https://escholarship.org/uc/item/14g8d18q

Authors

Miles, Christopher
McKinley, Scott
Ding, Fangyuan
et al.

Publication Date

2023-11-08

Peer reviewed

Inferring Stochastic Rates from Heterogeneous Snapshots of Particle Positions

Christopher E. Miles[1], **Scott A. McKinley**[2], **Fangyuan Ding**[4], **Richard B. Lehoucq**[3]

[1]Department of Mathematics, University of California, Irvine, CA, USA

[2]Department of Mathematics, Tulane University, New Orleans, LA, USA

[3]Discrete Math and Optimization, Sandia National Laboratories, Albuquerque, NM, USA

[4]Departments of Biomedical Engineering, Developmental and Cell Biology, University of California, Irvine, Irvine, USA

Abstract

Many imaging techniques for biological systems—like fixation of cells coupled with fluorescence microscopy—provide sharp spatial resolution in reporting locations of individuals at a single moment in time but also destroy the dynamics they intend to capture. These *snapshot observations* contain no information about individual trajectories, but still encode information about movement and demographic dynamics, especially when combined with a well-motivated biophysical model. The relationship between spatially evolving populations and single-moment representations of their collective locations is well-established with partial differential equations (PDEs) and their inverse problems. However, experimental data is commonly a set of locations whose number is insufficient to approximate a continuous-in-space PDE solution. Here, motivated by popular subcellular imaging data of gene expression, we embrace the stochastic nature of the data and investigate the mathematical foundations of parametrically inferring demographic rates from snapshots of particles undergoing birth, diffusion, and death in a nuclear or cellular domain. Toward inference, we rigorously derive a connection between individual particle paths and their presentation as a Poisson spatial process. Using this framework, we investigate the properties of the resulting inverse problem and study factors that affect quality of inference. One pervasive feature of this experimental regime is the presence of cell-to-cell heterogeneity. Rather than being a hindrance, we show that cell-to-cell geometric heterogeneity can *increase* the quality of inference on dynamics for certain parameter regimes. Altogether, the results serve as a basis for more detailed investigations of subcellular spatial patterns of RNA molecules and other stochastically evolving populations that can only be observed for single instants in their time evolution.

✉Christopher E. Miles, chris.miles@uci.edu.

1 Introduction

1.1 Background and Motivation

Advances in microscopy and automated tracking have matured to an age of observing of the spatial evolution of individuals at scales ranging from tissue-scale collectives of cells (Cognet et al. 2014) down to single molecules (Moerner 2007; McLaughlin et al. 2020). Such spatial tracking has led to discoveries in the understanding of viral transmission (Chen et al. 2014), intracellular transport (Osunbayo et al. 2019; Rayens et al. 2023), and many other areas in cell and molecular biology (Manzo and Garcia-Parajo 2015). In these investigations, the dynamics of a population are probed through the tracked motion of individual trajectories. The theory and practice of analyzing these single-particle trajectories is rich and well-developed (Qian et al. 1991; Mellnik et al. 2016; Rehfeldt and Weiss 2023). In contrast, many imaging approaches require the fixation of cells (which requires killing the tissue and cryo-preservation to stabilize agent locations) and therefore provide only a snapshot of the population at one moment of time (Biswas et al. 2021). Without direct access to the evolution of creation, destruction, and motion of this population prior to imaging, what can be inferred about these underlying dynamics?

One very active area of inferring dynamics from snapshots is in the quantification of gene expression from *spatial transcriptomics* imaging of individual RNA molecules (Weinreb et al. 2018; Marx 2021). Although time-lapse imaging and inference of these systems is possible in certain circumstances (Bowles et al. 2022), inferring the dynamics of underlying populations from RNA snapshots is a far wider and mainstream interest. Such spatial transcriptomics have served to be invaluable in recent identifications of cell types (Moses and Pachter 2022) and disease mechanisms (Williams et al. 2022). However, the theory to analyze this data has lagged behind the sophisticated techniques used to harvest it. Spatial imaging techniques (such as smFISH, single-molecule FISH (Shaffer et al. 2013; Omerzu et al. 2019; Ding and Elowitz 2019)) are capable of resolving individual RNA molecules, multiplexed over several genes (Maynard et al. 2020) and many cells(Cai et al. 2006; Eng et al. 2019). This unprecedented single-molecule resolution is largely neglected, with most analysis approaches focusing on RNA counts, binned either per-cell or split into nuclear and cytoplasmic counts (La Manno et al. 2018; Svensson and Pachter 2018; Gorin et al. 2022; Fu et al. 2022). In doing so, the subcellular spatial factors that crucially control gene expression (Heinrich et al. 2017; Cassella and Ephrussi 2022), e.g., geometry-dependent nuclear export (Köhler and Hurt 2007; Rodríguez-Navarro and Hurt 2011), are largely understudied and unincorporated into transcriptomics analyses.

Beyond the motivation for quantification of subcellular patterns of individual RNA molecules, we also emphasize the goals and value of *mechanistic* modeling. A zoo of sophisticated techniques has arisen for the analysis of (spatial) transcriptomics data, but these largely phenomenological or statistical studies (Imbert et al. 2022; Mah et al. 2022) suffer a lack of methodological reproducibility and interpretability (Gorin et al. 2022; Chari and Pachter 2023). The proposed solution to these shortcomings is the use of mechanistic, stochastic models (Lammers et al. 2020; Gorin et al. 2022) as the basis of inferring gene expression dynamics. Such methods have grown in popularity and sophistication in recent

years (Munsky et al. 2015; Gómez-Schiavon et al. 2017; Herbach et al. 2017; Gorin and Pachter 2022; Luo et al. 2022; Bowles et al. 2022; Gupta et al. 2022; Kilic et al. 2023; Vo et al. 2023), but remain primarily focused on non-spatial models. Thus, we have outlined the motivation for subcellular, mechanistic modeling of RNA molecules as a natural next step in expanding the frontier of spatial transcriptomics quantifications. We anticipate such techniques will be useful in advancing understanding of subcellular, spatial factors that modulate gene expression, including post-transcriptional splicing (Ding and Elowitz 2019; Coté et al. 2024; Ding et al. 2022) and nuclear export (Köhler and Hurt 2007; Rodríguez-Navarro and Hurt 2011).

The basis of this work is a mathematical connection between spatial point processes and snapshot observations of stochastic birth-death-diffusion processes. Most crucially, this connection provides statistical information of stochastic particle positions and counts, while circumventing the need for costly spatial stochastic simulation (Lemerle et al. 2005). The connection was suggested by Gardiner in the 1970 s (Gardiner and Chaturvedi 1977; Chaturvedi and Gardiner 1978; Chaturvedi et al. 1977) and recently reintroduced non-rigorously in (Schnoerr et al. 2016) in broad generality of stochastic reaction-diffusion processes. Here, we provide rigorous particle-perspective justification for this connection in the setup of birth, death, and diffusion in a domain. The machinery developed to prove this reveals new insight into generalizations and limitations of the theory. Inference on spatial point processes is a well-established topic in the statistics literature (Baddeley 2001), but focuses largely on phenomenological or purely statistical modeling of the underlying intensities, including in a recent application to the same data considered as our work (Walter et al. 2023). In contrast, the resulting setup here is an intensity derived from the solution of a birth-diffusion-death PDE. The model problem is similar in structure to classical PDE-based "inverse problems" (Aster et al. 2013; Flegg et al. 2020) but the crucial difference is that the data presents as collections of stochastic particle locations rather than the PDE solution itself plus observational noise. Here we embrace the stochastic particle-by-particle nature of the data and study the system using spatial point process inference techniques, which are naturally informed by the solutions of appropriate PDEs.

1.2   Overview of Work

In this work, we investigate a "toy"model motivated by smFISH imaging data of subcellular RNA (Ding and Elowitz 2019) that distills salient features of the intrinsically stochastic (Raj et al. 2006; Raj and Van Oudenaarden 2008), spatial (Heinrich et al. 2017; Cassella and Ephrussi 2022), and heterogeneous nature of the data. In such imaging, see Fig. 1A as an example, individual mRNA can be resolved in both cytoplasm and the nucleus of cells. Panel B shows the model considered throughout the remainder of the text for nuclear mRNA in a domain Ω, with stochastic production at rate κ_+ at a gene site z, decay at rate κ_-, and diffusion with diffusivity D. The nuclear boundary is assumed to be instantaneous export at $\partial\Omega$. To faithfully reflect the cell-to-cell heterogeneity that crucially affects the observed data, we incorporate different nuclear (or cellular) domain sizes and gene site positions, shown through a simplified 1D setup in C. The remainder of this work explores the mathematical nature of the inverse problem on this collection of heterogeneous snapshots.

In Sect. 2 we outline the model setup of snapshot data as an observation of a distribution of particles at time zero in a system in which the initial condition has been pulled back in time to negative infinity. We show that the snapshot data can be described as a nonhomogeneous spatial Poisson point process with an intensity measure that is generally recognized as the *occupation measure* of diffusing particles before exiting the domain or being removed by state-switching (Durrett 1996). The full proof is provided in Sect. 4.1 with a nondimensionalization argument found in Sect. 4.2. This connection is then translated into a likelihood function that can then be used for inference of parameters.

In Sect. 3, we put the theory to practice in the context of our toy model. In this setting, the likelihood function has a fully explicit form, which we report and visualize as a function of the number of cell snapshots that are included in a given experiment. This raises questions concerning the impact on inference from geometric heterogeneity among the cells and, in Sect. 3.1, we calculate the (Fisher) information for experiments run under different heterogeneity assumptions. In Sect. 3.2, we explore how these theoretical results translate into uncertainty in the inference of model parameters. We report on numerical experiments in which we generated synthetic snapshot data from simulations of underlying particle processes, and employed Bayesian methods to quantify joint uncertainty in the inference of particle production and degradation rates. In Sect. 3.3 we report on similarly constructed numerical experiments based on several different cell-heterogeneity assumptions. We find that the statistical information analysis conducted in Sect. 3.2 does faithfully predict Bayesian uncertainty quantifications that result from simulated data. This encourages continued use of these methods for both experimental design and practical statistical inference. Finally, in Sect. 4.4 we use the information-theoretic framework to demonstrate the benefit of using explicit particle locations for inference when compared to spatial binning. Additional numerical implementation details regarding stochastic process simulations and MCMC sampling for Bayesian inference are outlined in Section A.

2   Mathematical and Statistical Methods

2.1   Notation

Throughout this study, we have it in mind that biological cells can be heterogeneous in their geometry, but have similar internal reaction rates and transport dynamics. Let $M \in \mathbb{N}$ be the number of cells in a study, and for each $m \in \{1, 2, \ldots, M\}$, let the compact set $\Omega_m \subset \mathbb{R}^d$ denote the domain of the m th cell. Within each cell there are diffusive particles that emerge at rate κ_+ at locations that are distributed according to a source probability distribution $\varphi_m : \Omega_m \to \mathbb{R}_+$. They then move according to the laws of Brownian motion with uniform diffusivity D, possibly with drift $\alpha : \Omega \to \mathbb{R}^d$. If a particle in the m th cell exits the domain Ω_m, we assume that it cannot return. Furthermore, at rate $\kappa_- \geq 0$, particles undergo a change of state (either they degrade, or engage in a chemical reaction, for example). We will call a particle *active* up until the time it either switches state or exits the domain, and *inactive* thereafter.

In each cell, at time $t = 0$ we observe the locations of the active particles, denoted $\boldsymbol{x}_m = \{x_{1m}, x_{2m} \ldots, x_{n(x_m), m}\}$, where $x_{im} \in \Omega_m$. We write $n(\boldsymbol{x})$ for the number of particles in the collection $\boldsymbol{x}$, and emphasize that the number of particles in each of the M distinct cells Ω_m

will be independent of each other and random. In the foregoing discussion, we emphasize that the boundary condition, the diffusivity constant, and the emergence and switch rates are universal among all cells, but the initial location distribution (often realistically in the cell nucleus) can be different in each different cell.

As we will see later, the three parameters D, κ_+ and κ_- are not mutually identifiable. We therefore introduce the ratios

$$\lambda := \frac{\kappa_+ L_0^2}{D}, \text{ and } \mu := \frac{\kappa_- L_0^2}{D}.$$

(1)

That is, we are nondimensionalizing with respect to the particle diffusivity and the typical cell size L_0 over the population of observations.

2.2 Stochastic Model for Snapshot Observations Within a Single Cell

We now formally introduce the stochastic model that provides the basis for our inference protocol. For the remainder of this section, we will suppress dependence on the cell index m and describe the dynamics within a single focal cell.

In this work, we assume that the emergence (or birth) times of particles are given by a stationary Poisson point process on the negative half of the real line. Let $0 > T_1 > T_2 > \dots$ be an enumeration of points in the negative half-line, starting with the point closest to zero and decreasing from there. The initial locations of particles will be given by the sequence of independent and identically distributed (iid) random variables $\xi_i \overset{iid}{\sim} \varphi$. Let $\left\{ \mathring{B}_i(t) \right\}_{t \geq 0}$ be a sequence of iid standard d-dimensional Brownian motions (with $\mathring{B}_i(0) = 0$), and let $B_i(t) := \mathring{B}_i(t - T_i)$ be the Brownian motions shifted so that their initial time is the associated particle emergence time T_i.

We use the term "time" in the preceding description, but as we show in Sect. 4.2, we can assume that the dynamics have already been rescaled to nondimensional coordinates that result in a process that has diffusivity one, switch rate μ and birth rate λ. In this nondimensional framework, let a drift vector $\alpha: \mathbb{R}^d \to \mathbb{R}^d$ be given, and define $\{X_i(t)\}_{t > T_i}$ to be the solution of the stochastic differential equation (SDE)

$$\begin{aligned} dX_i(t) &= \alpha(X_i(t))dt + \sqrt{2} \ dB_i(t), t > T_i; \\ X_i(T_i) &= \xi_i. \end{aligned}$$

(2)

The partial differential operator associated with this diffusion is defined in terms of its action on twice-differentiable functions $f: \mathbb{R}^d \to \mathbb{R}$,

$$\mathcal{L}_X := \alpha(x) \cdot \nabla f(x) + \Delta f(x)$$

(3)

where Δ is the Laplacian.

Associated with each path is a continuous-time Markov chain (CTMC) $\{J_i(t)\}_{t > T_i}$ that records the state of the particle at time t. We assume that all particles start in the active state $J_i(0) = 1$. At rate μ, the particles switch to the inactive state 0. We label the switch times $\tau_{\text{switch}, i}$ and, due to the Markov assumption, the switch times are exponentially distributed with densities $p_i(t) = \mu e^{-\mu(t - T_i)} 1_{[T_i, \infty)}(t)$. Because particles are assumed to not be able to return to the interior of a cell once they exit, we define the stopping times $\tau_{\text{exit}, i} := \inf\{t \geq T_i : X_i(t) \notin \Omega\}$.

Our model for the "snapshot" data is to define a spatial point process $\boldsymbol{N} : \mathscr{B}(\Omega) \to \mathbb{Z}$ that counts the number of *active* particles in subsets of the domain at time 0. To be precise, for every Borel subset $A \in \mathscr{B}(\Omega)$, we define

$$N\left(A\right) := \sum_{i=1}^{\infty} \delta_A\big((X_i(0 \wedge \tau_{\text{exit}, i}))1_{\{J_i(0 \wedge \tau_{\text{exit}, i}) = 1\}},$$

(4)

where $\delta_A(x) = \int_A \delta(x - y)\mathrm{d}y$ and $t \wedge s := \min(t, s)$. In other words, $N(A)$ counts all particles that have neither exited the system through the boundary $\partial\Omega$, nor switched from active to inactive, and are located in the set A at time $t = 0$. A schematic of this can observation model can be seen in Fig. 2.

We can now state the main mathematical result, which identifies the relationship between collections of particle locations in steady state and the solution of a well-known boundary value problem (BVP).

Theorem 2.1 *Let* $(X_i(t), J(t))_{i \in \mathbb{N}}$ *be the particle/state pairs defined by* (2) *and the foregoing discussion. Let* $\Omega \subset \mathbb{R}^d$ *be an open domain with compact closure that has a boundary with all regular points with respect to the SDE* (2). *Then the associated spatial point process* $\boldsymbol{N}$, *defined by* (4), *giving the locations of active particles at time zero, is a Poisson spatial process with intensity measure u satisfying the BVP*

$$\mathscr{L}^* u(x) - \mu u(x) = -\lambda\varphi(x), \quad x \in \Omega,$$
$$u(x) = 0, \qquad x \in \partial\Omega.$$

(5)

where $\mathscr{L}^* u = -\nabla \cdot (\alpha(x)u(x)) + \Delta u(x)$ *is the adjoint of the operator* (3).

Remark 2.2 This is to say, $\boldsymbol{N}$ satisfies the condition that, given any disjoint sets $A_1, \ldots A_K \subset \mathscr{B}(\Omega)$, the random variables $\{N(A_k)\}_{k=1}^{K}$ are independent with respective distributions

$$N(A_k) \sim \text{Pois}\left(\int_{A_k} u(x) \mathrm{d}x\right).$$

(6)

The proof of this theorem is the content of Sect. 4.1.

2.3 Framework for Inference

The Poisson spatial point process representation of the particle locations organizes the data into a form that readily allows for statistical inference. Suppose that we partition the domain Ω into a fixed collection of sets. We then "bin the data," meaning we record the particle locations in terms of the number of particles in each bin. By Theorem 2.1, the number of particles in each bin is Poisson distributed and independent of number of particles in other bins. This yields a likelihood function that can be used for parameter inference, but depends explicitly on the choice of bins. For example, suppose that the sets $(A_1, A_2, ..., A_K)$ form a partition of cell domain, meaning that they are mutually disjoint and their union is Ω. For a given set of particle locations x, define $n_k(x)$ to be the number of particles observed in the set A_k. Then the likelihood function for the parameter vector $\theta = (\lambda, \mu)$ given the binned data is

$$P_\theta(N(A_1) = n_1(x), ..., N(A_K) = n_K(x)) = \prod_{k=1}^{K} e^{-\int_{A_k} u(x;\theta)\mathrm{d}x} \left(\int_{A_k} u(x;\theta)\mathrm{d}x\right)^{n_k(x)} \frac{1}{n_k(x)!}.$$

(7)

We introduce the notation $u(x;\theta)$ to emphasize the dependence of the intensity measure on θ through the BVP (5).

Now, if we change the partition of the domain, the associated likelihood function for the new binning of the data will be different, leading to different estimates and uncertainty regions for the parameters. Naturally among all choices for which partition to use, we would prefer to use one that yields the "best" statistical inference in some rigorous sense. One standard approach is to consider the likelihood that emerges in a limit in which the partition mesh size (often defined to be the size of the largest bin) is taken to zero. The form of the limiting likelihood is well known for spatial Poisson processes (Baddeley 2001; Daley and Vere-Jones 2003) and can be expressed directly in terms of the particle locations, rather than a summary from binning.

Definition 2.3 (Likelihood function) Let $\Omega \subset \mathbb{R}^d$ satisfy the conditions of Theorem 2.1. For each $\theta = (\lambda, \mu) \in \Theta$, let $u(\cdot;\theta)$ satisfy the BVP (5) for the given parameter vector. Then for any set of particle locations $x = \{x_1, x_2, ..., x_{n(x)}\} \subset \Omega$, we define the likelihood function $L(\cdot;x): \Theta \to \mathbb{R}_+$ as follows:

$$L(\theta;x) := \left(\prod_{i=1}^{n(x)} u(x_i;\theta)\right) e^{-\int_\Omega u(x;\theta)\mathrm{d}x}.$$

While the use of this form of likelihood function is standard and can be found else-where [(for instance, in Jensen and Møller (1991)], in Sect. 4.3 we include a derivation using the Lebesgue Differentiation Theorem in the context of Bayesian methods that we have not explicitly seen in the literature. Moreover, in Sect. 4.4, we use the notion of statistical information to quantify the claim that the likelihood function defined by (8) is "better" than any likelihood function of the form (7) that arises from binning the particle locations.

We note that the above discussion contains a mix of Bayesian and frequentist ideas and we will continue to draw on tools from each school throughout this work. Generally speaking, when we are considering a specific data set and wish to establish point estimates for the parameters and to quantify uncertainty, we will use a Bayesian method. This is because the notion of a highest-density posterior region (HDPR) allows for clean visualization of 2d "joint uncertainty" for the parameters λ and μ, which can then be summarized with a single quantity: its area (Yau and Campbell 2019). Meanwhile, when we wish to make comparisons that are not specific to a given data set, we will use the determinant of the information matrix $\det(\mathcal{I})$. Optimization of this determinant, or *D-optimality* (Atkinson et al. 2007), is a popular choice in experimental design [(including in other studies of gene expression Fox and Munsky (2019)] due to its ability to capture correlations between inferred parameters in the (inverse) area of confidence ellipsoids. However, confidence is a frequentist notion, and we use $\det(\mathcal{I})$ as an analytically tractable proxy for the inverse of the area of the joint uncertainty region in Bayesian inference. The relationship between maximizing $\det(\mathcal{I})$ and minimizing expected HDPR is not precise for finite data, in part because the role of the prior must be considered in the Bayesian setting (Gábor and Banga 2015; Sharp et al. 2022). But, as we show in Sect. 3.3, $\det(\mathcal{I})$ does make useful predictions in the present setting.

3   Inferring Dynamics of the 1D RNA Model

Using the framework established in previous sections, we now explore their consequences and implementation in the context of the 1D model approximating the key features seen in subcellular spatial RNA transcriptomics data. The principal result was the connection between the PDE (5) as the intensity for the corresponding point process. For a single cell, this PDE becomes

$$0 = \Delta u(x) + \lambda\delta(x - z) - \mu u(x) \text{ for } x \in [0, L], u(0) = 0, u(L) = 0,$$

(9)

where (λ, μ) are nondimensional (1) dynamic quantities sought to be inferred, and L is relative to the typical length seen in the population of images.

The solution to (9) can be explicitly computed:

$$u(x) = \frac{\lambda}{\sqrt{\mu}}\operatorname{csch}\left(\sqrt{\mu}L\right)\sinh\left(\sqrt{\mu}\min\{x, z\}\right)\sinh\left(\sqrt{\mu}(L - \max\{x, z\})\right).$$

(10)

The expected number of particles is then, from (6),

$$\mathbb{E}\Big[N\Big] = \int_0^L u\Big(x\Big)\mathrm{d}x = \frac{\lambda}{\mu}\left(1 - \mathrm{sech}\left(\frac{\sqrt{\mu}L}{2}\right)\cosh\left(\frac{1}{2}\sqrt{\mu}\Big(L - 2z\Big)\right)\right)$$

(11)

The expectation in (11) shows an intuitive structure of the product of the pure birth-death mean λ/μ, and a geometry-dependent term (always strictly less than 1) that scales the former.

3.1 Likelihood and Information from Analytical Solution

Equipped with the statistical characterization from Theorem 2.1 and the analytical expression of the intensity for a single snapshot in (10), we can now examine the behavior and emergent lessons arising from this inference setup.

We first explore the behavior of the likelihood (8) as a function of the system parameters and quantity of data. To do so, we perform stochastic simulations of the underlying birth-death-diffusion trajectories and then terminate them at an experimental time T. The particle positions at this observation time are the data used for inference. Further details on the stochastic simulation can be found in Section A. Beginning with the simplest setup, we fix $\lambda = 500, \mu = 10, L = 1$ and $z = 1/2$ and explore how increasing the number of snapshots M shapes the likelihood, the results of which can be seen in Fig. 3A. As M increases, the likelihood first concentrates around a line and then within this line, begins to peak at the true parameter values. To understand this behavior, we plot the $\lambda - \mu$ line of constant $\mathbb{E}[N]$ from (11) superimposed on the likelihood. From this, we see that the likelihood concentrates around the line defining the true number of particles, and then further resolves within this line from their spatial information. In panels B and C of the same figure, the same information is shown as slices of the likelihood with one parameter at a time held to be the true value. From these panels, we see again that sufficiently large M (here, $M \approx 1000$) yields likelihoods that are tightly concentrated around the true values used to generate the synthetic data. In the next section, we will explore more concretely how the spread of the likelihood translates into uncertainty in the inference itself.

Before performing inference on the synthetic data, other useful quantities can be computed directly from the likelihood. For a likelihood $L(\theta; x)$, the corresponding information matrix $\mathscr{I}$ is defined by the elements (Vahid et al. 2020)

$$\mathscr{I}_{ij} := \mathbb{E}_\theta\Big[\big(\partial_{\theta_i}\ \log\ L(\theta; \mathbf{X})\big)\big(\partial_{\theta_j}\ \log\ L(\theta; \mathbf{X})\big)\Big] = -\mathbb{E}_\theta\Big[\partial_{\theta_i}\partial_{\theta_j}\ \log\ L(\theta; \mathbf{X})\Big],$$

(12)

where $\mathbb{E}_\theta$ is the expectation of an observation of the point process $\mathbf{X}$ (a vector with stochastic length) with fixed model parameters θ. The latter equality in (12) holds under regularity conditions straightforward to verify for Poisson point processes (Clark 2022). Our parameters of interest are $\theta = (\lambda, \mu)$ so the information matrix takes the form

$$\mathcal{I}\left(\boldsymbol{\theta}\right) = -\mathbb{E}_\theta \begin{bmatrix} \partial_{\lambda\lambda} \ \log \ L(\lambda,\mu;\mathbf{X}) \ \partial_{\lambda\mu} \ \log \ L(\lambda,\mu;\mathbf{X}) \\ \partial_{\lambda\mu} \ \log \ L(\lambda,\mu;\mathbf{X}) \ \partial_{\mu\mu} \ \log \ L(\lambda,\mu;\mathbf{X}) \end{bmatrix}.$$

(13)

Importantly, the expectation is over both stochastic particle positions and number. To do so, we employ machinery from the theory of spatial point processes, specifically the Campbell-Hardy formula stated in Theorem 3.1.

Theorem 3.1 *(Campbell-Hardy*, (Baddeley 2001)) *Let N be governed by a general point process on a domain Ω with intensity $u:\Omega \to \mathbb{R}$ so that $\mathbb{E}[N(B)] = \int_B u(x)\mathrm{d}x$. Let $f(x)$ be any measurable function. Then, the expectation of the sum over observations $\mathbf{X} = (X_1, X_2, ..., X_N)$ of the point process is*

$$\mathbb{E}\left[\sum_{i=1}^N f(X_i)\right] = \int_\Omega f\left(x\right)u\left(x\right)\mathrm{d}x.$$

(14)

Through Theorem 3.1, the elements of our information matrix in (13) can be related to the PDE solution u by (10). To do so, the log-likelihood $\ell\left(\boldsymbol{\theta};\boldsymbol{x}\right)$ from (8) is

$$\ell\left(\boldsymbol{\theta};\boldsymbol{x}\right) := \log \ L\left(\boldsymbol{\theta};\boldsymbol{x}\right) = -\int_\Omega u\left(x;\boldsymbol{\theta}\right)\mathrm{d}x + \sum_{i=1}^N \ \log \ u(x_i;\boldsymbol{\theta})$$

and $\partial_{\theta_i\theta_j} \ell = -\int_\Omega \partial_{\theta_i\theta_j} u\left(x;\boldsymbol{\theta}\right)\mathrm{d}x + \sum_{i=1}^N \partial_{\theta_i\theta_j} \ \log \ u(x_i;\boldsymbol{\theta}).$

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) = -\mathbb{E}\left[\partial_{\theta_i\theta_j} \ell\left(\boldsymbol{\theta};\mathbf{X}\right)\right] = -\mathbb{E}\left[-\int_\Omega \partial_{\theta_i\theta_j} u\left(x;\boldsymbol{\theta}\right)\mathrm{d}x\right] - \mathbb{E}\left[\sum_{i=1}^N \partial_{\theta_i\theta_j} \ \log \ u(X_i;\boldsymbol{\theta})\right]$$

$$= \int_\Omega \partial_{\theta_i\theta_j} u\left(x;\boldsymbol{\theta}\right)\mathrm{d}x - \mathbb{E}\left[\sum_{i=1}^N \partial_{\theta_i\theta_j} \ \log \ u(X_i;\boldsymbol{\theta})\right]$$

$$= \int_\Omega \partial_{\theta_i\theta_j} u\left(x;\boldsymbol{\theta}\right)\mathrm{d}x - \int_\Omega \left(\partial_{\theta_i\theta_j} \ \log \ u\left(x;\boldsymbol{\theta}\right)\right)u\left(x;\boldsymbol{\theta}\right)\mathrm{d}x.$$

(15)

The first term is deterministic, and therefore is its own expected value. The second, stochastic sum, takes the form of (14) with $f\left(x\right) = \partial_{\theta_i\theta_j} \ \log \ u\left(x\right)$ and is evaluated with Theorem 3.1. The integrands of (15) can be manipulated to a more tractable form by expanding derivatives

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) = -\int_{\Omega}\big(\partial_{\theta_i\theta_j}\ \log\ u(x;\boldsymbol{\theta})\big)u(x;\boldsymbol{\theta})\mathrm{d}x + \int_{\Omega}\partial_{\theta_i\theta_j}u(x;\boldsymbol{\theta})\mathrm{d}x$$

$$= \int_{\Omega}\big(\partial_{\theta_i}\ \log\ u(x;\boldsymbol{\theta})\big)\big(\partial_{\theta_j}\ \log\ u(x;\boldsymbol{\theta})\big)u(x;\boldsymbol{\theta})\mathrm{d}x\,.$$

$$= \int_{\Omega}\frac{\big(\partial_{\theta_i}u(x;\boldsymbol{\theta})\big)\big(\partial_{\theta_j}u(x;\boldsymbol{\theta})\big)}{u(x;\boldsymbol{\theta})}\mathrm{d}x\,.$$

(16)

If $\lambda > 0$, then $u > 0$, and $\det(\mathcal{I}) > 0$ follows immediately from Cauchy–Schwarz. From this non-degeneracy, structural identifiability of both parameters can be concluded (Browning et al. 2020; Wieland et al. 2021). Moreover, the analytical solution to the PDE allows us to compute the information explicitly. For instance,

$$\mathcal{I}_{11}(\boldsymbol{\theta}) = -\mathbb{E}_{\theta}[\partial_{\lambda\lambda}\ \ell\ (\boldsymbol{\theta};\mathbf{X})] = (\lambda\mu)^{-1}\big[1 - \mathrm{sech}\big(\sqrt{\mu}L/2\big)\cosh\big(\sqrt{\mu}(L-2z)/2\big)\big]$$

(17)

The remaining entries can be computed straightforwardly using Mathematica but have unwieldy expressions not shown here.

In Fig. 4, we employ the analytical expressions for the information matrix to explore the behavior of $\det(\mathcal{I})$ as a function of the underlying parameters. We interpret this quantity's importance through the classical lens of the Cramér-Rao bound (Clark 2022): any unbiased estimator of the parameter's variance is bounded below by the inverse of information. Informally, we interpret larger information setups as "easier" inverse problems. We later verify this interpretation through posterior widths on actual inference setups.

In Fig. 4A and B, we fix $\lambda = 500$, $L = 1$ and vary both source location z and degradation rate μ. For large values of μ, the behavior is intuitive: the information is maximized when the source location is at the center, as this minimizes the overall number of particles exiting the boundary and corresponding information loss. For smaller values of μ, a bifurcation in the behavior occurs and the optimal source location becomes non-trivial: occurring at two symmetric locations somewhere between the cell center and boundary. This non-trivial behavior arises only in the simultaneous inference of parameters, suggesting its origin. We interpret this non-trivial behavior as arising from the trade-off between seeing more particles (sources closer to the center) or seeing a wider range of the spatial gradient (for off-center source locations).

Panels C and D investigate the cell size's impact on the information by fixing z/L and varying L and μ. Intuitively, larger cells (relative to the typical size) yield more information due to less information loss through the absorbing boundary. Interestingly, this effect saturates for sufficiently large values of μ, when the typical length scale of a trajectory is far shorter than the domain size, so no additional information is gained through enlarging the domain.

3.2 Bayesian Inference on Synthetic Data

To deploy Bayesian inference, we must specify prior distributions for λ and μ. For the production rate λ, we take a conjugate prior

$$\lambda \sim \text{Gamma}(\alpha, \beta), \quad p(\lambda) = \frac{\lambda^{\alpha-1} e^{-\beta\lambda} \beta^\alpha}{\Gamma(\alpha)}.$$

(18)

Assuming the total prior is independent for each parameter, $p(\theta) = p(\lambda)p(\mu)$, then for a single observation, we have

$$
\begin{aligned}
p(\theta \mid \mathbf{x}) \propto p(\theta)L(\theta; \mathbf{x}) &= p\left(\mu\right)p\left(\lambda\right)e^{-\lambda\int_\Omega v(x;\mu)\mathrm{d}x}\prod_{i=1}^N \lambda v(x_i; \mu) \\
&= p\left(\mu\right)\prod_{i=1}^N v(x_i;\mu)\frac{\lambda^{\alpha-1}e^{-\beta\lambda}\beta^\alpha}{\Gamma(\alpha)}\lambda^n e^{-\lambda\int_\Omega v(x;\mu)\mathrm{d}x} \\
&= p\left(\mu\right)\prod_{i=1}^N v(x_i;\mu)\frac{\lambda^{(\alpha+n)-1}e^{-\lambda(\beta+\int_\Omega v(x;\mu)\mathrm{d}x)}\beta^\alpha}{\Gamma(\alpha)}.
\end{aligned}
$$

(19)

Here, we have introduced the definition of v as $u(x; \mu, \lambda) = \lambda v(x; \mu)$, noting that (10) is linear in λ and following standard practice of identifying a "scale" parameter in the point process literature (Baddeley 2001). This yields the convenient conditional posterior for λ arising from conjugacy

$$p\left(\lambda \mid \mathbf{x}, \mu\right) = \text{Gamma}\left(\alpha + n(\mathbf{x}), \beta + \int_\Omega v(x; \mu)\mathrm{d}x\right).$$

(20)

For M observations, each with a different domain, this generalizes to

$$p(\lambda \mid \mathbf{x}_1, \ldots, \mathbf{x}_m, \mu) = \text{Gamma}\left(\alpha + \sum_{m=1}^M n(\mathbf{x}_m), \beta + \sum_{m=1}^M \int_{\Omega_m} v(x; \mu)\mathrm{d}x\right).$$

(21)

For μ, we choose a uniform prior $\mu \sim \text{unif}(0, \mu_{\max})$ with large $\mu_{\max}$ for uninformedness. With these choices, of priors, it is straightforward to implement a Gibbs-in-Metropolis sampler for the posterior distribution, with a Gibbs step using the conjugacy for λ and Metropolis step for μ. Further details on the MCMC implementation (including verification of convergence) can be found in Section A.

The results of performing Bayesian inference on synthetic data can be seen in Fig. 5. The data considered here is the same as that in Fig. 3, with fixed parameters and varied number of snapshots. In panel **A**, we depict the 89% HDPR (highest density posterior

region) (Gelman et al. 1995), computed by evaluating $L(\theta; \boldsymbol{x})p(\theta)$ for each posterior sample θ, retaining the top 89%, and then computing the convex hull of these samples. The resulting uncertainty qualitatively mimics the lesson seen from the likelihood. For small values of M, the posterior concentrates around the parameter region that produces an expected number of particles equal to the observed count, and then resolves to be sharply peaked around the true parameters with data from more snapshots. Qualitatively, the same result can be seen in the marginal posterior samples seen in Fig. 5B and C. To more explicitly quantify the dependence of uncertainty on data, we plot the HDPR area as a function of the number of snapshots M in the inset of Fig. 5A, and find a convincing $\sim M^{-1}$ scaling. This scaling can be interpreted as arising from the product (an area) of the uncertainty of each parameter $\sim M^{-1/2}$ arising from the central limit theorem.

3.3 Cell-to-Cell Heterogeneities and their Impact on Inference

In the previous figure, we assumed all cells were identical in shape and source location. However, in experimental setups, there is enormous cell-to-cell heterogeneity. The likelihood results shown in Fig. 3 suggest that this heterogeneity should affect inference quality, as both domain size and source location both modulate the total information. However, those quantities were generated through the computation of the information for a single observation. How do these propagate to a heterogeneous population? To investigate this heterogeneity, we extend the information matrix calculation to now assume that the domain parameters $\boldsymbol{\vartheta} = (z, L)$ follow some probability distribution. The expectation can be extended by the Law of Total Expectation:

$$\mathcal{I}_{ij}(\boldsymbol{\theta}) = -\mathbb{E}_{\theta,\vartheta}\big[\partial_{\theta_i}\partial_{\theta_j}\log\ L(\boldsymbol{\theta},\boldsymbol{\vartheta};\mathbf{X})\big] = -\mathbb{E}_{\vartheta}\big[\mathbb{E}_{\theta}\big[\partial_{\theta_i}\partial_{\theta_j}\log\ L(\boldsymbol{\theta};\mathbf{X})\mid\boldsymbol{\vartheta}\big]\big].$$

(22)

Ultimately, this reduces to integrating the results of previous figures against the probability distributions for z and L. To parameterize the noise, we take $z/L \mid L \sim \text{Beta}(\sigma_z^{-1}, \sigma_z^{-1})$ and $L \sim \Gamma(\sigma_L^{-2}, \sigma_L^2)$. The choice of distribution for source location was chosen so that $\mathbb{E}_{\vartheta}[z \mid L] = L/2$ and increasing σ_z spreads out the source symmetrically around this location, with $\sigma_z = 1$ corresponding to spatially uniform locations. Similarly, L is supported on the non-negative numbers, and the distribution is chosen so that $\mathbb{E}_{\vartheta}[L] = 1$ and increasing σ_L spreads out the distribution of cell sizes in a plausible way.

The impact of cell-to-cell heterogeneity on the information can be seen in Fig. 6. Panel A the information $\det(\mathcal{I})$ as a function of σ_z for various values of μ and fixed $\lambda = 500, L = 1$. For low values of μ, varying source locations *increase* the amount of information in observations, whereas for large μ, they lower the information. This is in alignment with the results seen in Fig. 4 for fixed source locations, where low values of μ corresponded to non-trivial optimal source locations, and large μ have optimal source locations in the center.

A similar effect can be seen for varying L in panel B of Fig.6: for low values of μ, varied cell sizes increase the information, whereas for large μ, they decrease it. This result is somewhat more surprising. For increasing values of σ_L, the median L decreases, meaning that more

mass of the distribution is shifted toward cell sizes *smaller* than the typical size. The fixed-size results say that smaller cells encode less information. However, taken together, the population-heterogeneous results suggest that the information from large cells offsets the lack thereof from small, and creates a net gain in information.

An outstanding curiosity remains of whether these information results are meaningful for actual inference. To investigate this, we repeat the Bayesian inference pipeline from the previous section for various setups of heterogeneities: varied z, varied L, and both varied. For a fixed experiment of $M = 100$ images, synthetic data (with μ chosen in the regime of increase information) is generated and inference is performed. The likelihood is constructed using individual z and L values for each snapshot, assumed known. The HDPR area is computed as a measure of the overall uncertainty in the estimation. This inference experiment is then repeated 1000 times, and the results can be seen in Fig. 6 panel C. The results from estimation support the previous: heterogeneity does indeed *increase* the amount of information in the snapshots, and *lowers* uncertainty in estimation. All three scenarios with random source location, cell size, or both random yield smaller HDPR areas than the static scenario.

4 Support for the Inference Framework.

4.1 Proof of Theorem 2.1

In order to prove that the set of locations of particles is a Poisson spatial process, we need only show that for any finite collection of disjoint sets $A_1, \ldots, A_k$, the number of particles in the sets are independent and Poisson distributed. If we find a function u such that $\int_A u(x)\mathrm{d}x = P(A)$ for all A, then this the intensity measure.

Poisson distribution.—To fix notation, let $\rho_A(t)$ be the probability that a particle that is distributed according to φ at time 0 and is in the set A at time $t > 0$. Recall, however, that in the model the particles emerge at a rate λ according to a Poisson arrival process for some interval of time $(-h, 0)$ and we are counting the particles still in the system at time zero. Since this means that the initial times are iid and uniformly distributed over $(-h, 0)$, we can define

$$\rho_A^{(h)} = \frac{1}{h}\int_{-h}^{0}\rho_A\big(-s\big)\mathrm{d}s = \frac{1}{h}\int_{0}^{h}\rho_A\big(t\big)\mathrm{d}t\,.$$

(23)

By Poisson thinning, if the total number of particles that emerge is $\mathrm{Pois}(\lambda h)$, then the distribution of particles in A at time zero is $\mathrm{Pois}\big(\lambda h \rho_A^{(h)}\big)$. Taking the limit as $h \to \infty$, we have that the number of particles in A has a Poisson distribution with mean

$$\lambda \rho_A = \lambda \int_{0}^{\infty}\rho_A(t)\mathrm{d}t\,.$$

(24)

Intensity measure.—The time-dependent PDE that describes the probability density of particles that have not yet been removed by state-switching or by being absorbed at the boundary of a domain is [(p. 230, Cox and Miller (1977)]

$$
\begin{aligned}
\partial_t \rho(x,t) &= L^*\rho(x,t) - \mu\rho(x,t), \quad x \in \Omega, t > 0, \\
\rho(x,t) &= 0, \qquad\qquad\qquad x \in \partial\Omega, t > 0, \\
\rho(x,0) &= \varphi(x), \qquad\qquad\quad x \in \Omega.
\end{aligned}
$$

(25)

Defining $\rho(x) = \int_0^\infty \rho(x,t)\mathrm{d}x$, (p. 163–167, Gardiner (2004)) we integrate (25) to find

$$
\begin{aligned}
\left(\lim_{t \to \infty} \rho\big(x,t\big)\right) - \rho\big(x,0\big) &= L^*\rho(x) - \mu\rho(x), \quad x \in \Omega, \\
\rho(x) &= 0, \qquad\qquad x \in \partial\Omega.
\end{aligned}
$$

(26)

Since particles exit the domain in finite time with probability one, $\rho(x,t) \to 0$ as $t \to \infty$. Imposing the initial condition yields

$$
\begin{aligned}
L^*\rho(x) - \mu\rho(x) &= -\varphi(x), \quad x \in \Omega \\
\rho(x) &= 0, \qquad x \in \partial\Omega.
\end{aligned}
$$

(27)

In order to achieve the desired mean for the Poisson distribution in (24), we define $u(x) = \lambda\rho(x)$ and note that u can be attained by multiplying the initial distribution function by λ. This yields the BVP stated in the theorem and the proof is complete.

□

4.2 Nondimensionalization

In this subsection, we outline the nondimensionalization argument justifying the SDE (2). Suppose that the original SDE satisfies the SDE

$$
\mathrm{d}X(t) = \alpha\big(X(t)\big)\mathrm{d}t + \sqrt{2D}\mathrm{d}W(t)
$$

(28)

with initial condition $X(t_0) = x_0$. The particle can meet one of two fates: a state-switch due to degradation, at rate κ_-, or exit through the boundary of the domain. Define

$$
\tilde{t} := \frac{D}{L_0^2}t, \tilde{x} := \frac{1}{L_0}x, \ \text{ and } \ \tilde{\alpha}\left(\tilde{x}\right) := \frac{L_0}{D}\alpha\big(x\big).
$$

Then we claim that $\widetilde{X}\big(\tilde{t}\big) := \frac{1}{L}X\big(t\big)$ satisfies

$$\mathrm{d}\widetilde{X}(\tilde{t}) = \tilde{\alpha}\big(\widetilde{X}(t)\big)\mathrm{d}\tilde{t} + \sqrt{2}\mathrm{d}\widetilde{B}(\tilde{t})$$

(29)

where $\widetilde{B}(\tilde{t})$ is a standard Brownian motion. The argument follows from the integral form of the SDE:

$$\widetilde{X}(\tilde{t}) - \widetilde{X}(\tilde{t}_0) = \frac{1}{L_0}(X(t) - X(t_0))$$

$$= \int_{t_0}^{t} \frac{1}{L_0}\alpha(X(s))\mathrm{d}s + \sqrt{\frac{2D}{L_0^2}}(B(t) - B(t_0)).$$

For the drift term, under the substitution $\tilde{s} = \frac{D}{L_0^2}s$ we have

$$\int_{t_0}^{t} \frac{1}{L_0}\alpha(X(s))\mathrm{d}s = \int_{\frac{Dt_0}{L_0^2}}^{\frac{Dt}{L_0^2}} \frac{L_0}{D}\alpha\left(X\left(\frac{L_0^2}{D}\tilde{s}\right)\right)\mathrm{d}\tilde{s}$$

$$= \int_{\tilde{t}_0}^{\tilde{t}} \frac{L_0}{D}\alpha\big(L_0\widetilde{X}(\tilde{s})\big)\mathrm{d}\tilde{s}$$

$$= \int_{\tilde{t}_0}^{\tilde{t}} \tilde{\alpha}\big(\widetilde{X}(\tilde{s})\big)\mathrm{d}\tilde{s}$$

Meanwhile, define $\widetilde{B}\left(\tilde{t}\right) = \sqrt{\frac{D}{L_0^2}}B\left(t\right)$. Then $E\big(\widetilde{B}(\tilde{t})\big) = 0$ and

$$\mathrm{Var}\big(\widetilde{B}(\tilde{t})\big) = \frac{D}{L_0^2}\mathrm{Var}(B(t)) = \frac{D}{L_0^2}t = \tilde{t}$$

(30)

meaning that $\widetilde{B}$ is a standard Brownian motion.

In terms of the essential stopping times, we define $\mu = L_0^2/D$ (as in (1)), and so $\tilde{\tau}_{\mathrm{switch}} = \tau_{\mathrm{switch}}D/L_0^2$ is the corresponding nondimensional switch time. Moreover, defining $\widetilde{\Omega} = \{x: L_0x \in \Omega\}$ and $\tilde{\tau}_{\mathrm{exit}} := \inf\big\{\tilde{t} > \tilde{t}_0: \widetilde{X}(\tilde{t}) \notin \widetilde{\Omega}\big\}$. We can readily show path-by-path that $\tilde{\tau}_{\mathrm{exit}} = \tau_{\mathrm{exit}}D/L_0^2$. Analogous to the dimensional case, we define $\tilde{\tau} := \min(\tilde{\tau}_{\mathrm{exit}}, \tilde{\tau}_{\mathrm{switch}})$

Altogether, it follows that, for any $A \in \Omega$, if we let $\widetilde{A} = \{x: L_0x \in \Omega\}$, then

$$P\big(\widetilde{X}(0 \wedge \tilde{\tau}) \in \widetilde{A}\big) = P(X(0 \wedge \tau) \in A)).$$

(31)

Hence, if we wish to solve for the intensity measure of a given dimensional system, we can rescale space by L_0 and the rates as in (1) and use the intensity measure of the associated nondimensionalized system for statistical inference.

4.3 Derivation of the Likelihood Function

Proposition 4.1 *Let* $\{\mathscr{A}_K\}_{K=1}^{\infty}$ *(with* $\mathscr{A}_K = (A_{1,K}, \ldots, A_{K,K})$*) be a sequence of* Ω*-partitions that has the following properties:*

    i.   *the partitions are nested in the sense that if* $K_1 < K_2$*, then every set in* $\mathscr{A}_{K_2}$ *is contained within some member of* $\mathscr{A}_{K_1}$*; and*

    ii.   *the partition mesh sizes,* $(\max_{k \in \{1, \ldots, K\}}\{|A_{k,K}|\})$*, where* $|A|$ *denotes the volume of a set* A*, decrease to zero with* K*.*

Fix $\boldsymbol{x} = (x_1, x_2, \ldots, x_{n(x)})$ *to be a single snapshot of particle locations arising from the dynamics defined for Theorem* 2.1 *and let* $\pi(\theta)$ *be a prior distribution for* $\theta = (\lambda, \mu)$*. Define* $\pi_K(\theta, \boldsymbol{x})$ *be the posterior distribution of* θ *under the prior* π *when the locations are reported through binning in the sets of the partition* A_K*. Then,*

$$\pi(\theta \mid \boldsymbol{x}) := \lim_{K \to \infty} \pi_K(\theta \mid \boldsymbol{x}) = \frac{L(\theta; \boldsymbol{x})\pi(\theta)}{\int_\Theta L(\vartheta; \boldsymbol{x})\pi(\vartheta)\mathrm{d}\vartheta}.$$

(32)

where $L(\theta; \boldsymbol{x})$ *is the likelihood function given in Definition Definition* 2.3.

Remark 4.2 A construction of such a sequence of partitions can be achieved, for example, by starting with any partition of Ω and creating each refinement by randomly selecting one of the partition's sets and splitting in half. For another construction, see the definition of psuedolikelihoods put forward by Jensen and Møller (1991).

Proof We can represent the collection of particle locations as a measure $\delta_x(x) := \sum_{i=1}^{n(x)} \delta_{x_i}(x)$ (Daley and Vere-Jones 2003) and for each member $A_{k,K}$ of a partition $\mathscr{A}_K$, we can define

$$n_{k,K}(\boldsymbol{x}) := \int_{A_{k,K}} \delta_{\boldsymbol{x}}(x)\mathrm{d}x.$$

(33)

We then can write the posterior distribution $\pi_K(\cdot \mid \boldsymbol{x})$ associated with the partition $\mathscr{A}_K$ as follows:

$$\pi_K\left(\theta \mid \boldsymbol{x}\right) = \frac{P_\theta(N(A_{1,K}) = n_{1,K}(\boldsymbol{x}), \ldots, N(A_{K,K}) = n_{K,K}(\boldsymbol{x}))\pi(\theta)}{\int_\Theta P_\vartheta(N(A_{1,K}) = n_{1,K}(\boldsymbol{x}), \ldots, N(A_{K,K}) = n_{K,K}(\boldsymbol{x}))\pi(\vartheta)\mathrm{d}\vartheta}$$

(34)

Here the subscript of the probability measure denotes that we are evaluating the probability with the indicated parameter vector. The full parameter space is denoted Θ, which in our case is $\mathbb{R}_+ \times \mathbb{R}_+$. Since the probability in the above formula is expressed in terms of an event comprised of independent Poisson random variable results, we have the following explicit formula for the likelihood (suppressing dependence on K), which follows from Remark 2.2:

$$P_\theta(N(A_1) = n_1(\boldsymbol{x}), \ldots, N(A_K) = n_K(\boldsymbol{x})) = \prod_{k=1}^{K} e^{-\int_{A_k} u(x;\theta)\mathrm{d}x} \left(\int_{A_k} u(x;\theta)\mathrm{d}x \right)^{n_k(x)} \frac{1}{n_k(\boldsymbol{x})!}.$$

(35)

Now, because the partitions are nested, there exists a K_* such that for all $K > K_*$, all partition sets contain at most one particle. If we write A_{k_i} to be the bin containing the particle location x_i, we have that $n_{k_i}(\boldsymbol{x}) = 1$ for each i, and $n_k(\boldsymbol{x}) = 0$ otherwise. For the purpose of writing the likelihood function, we gather all bins not containing a particle into one set

$$A_0 := \Omega \smallsetminus \left\{ \bigcup_{i=1}^{n(\boldsymbol{x})} A_{k_i} \right\}.$$

For such a partition, (35) has the form

$$\begin{aligned} P_\theta(N(A_1) &= n_1(\boldsymbol{x}), \ldots, N(A_K) = n_K(\boldsymbol{x})) \\ &= P_\theta(N(A_0) = 0, N(A_{k_1}) = 1, \ldots, N(A_{k_{n(x)}}) = 1) \\ &= P_\theta(N(A_0) = 0) \prod_{i=1}^{n(\boldsymbol{x})} P_\theta(N(A_{k_i}) = 1) \\ &= e^{-\int_{A_0} u(x;\theta)\mathrm{d}x} \prod_{i=1}^{n(\boldsymbol{x})} e^{-\int_{A_{k_i}} u(x;\theta)\mathrm{d}x} \left(\int_{A_{k_i}} u(x;\theta)\mathrm{d}x \right) \\ &= e^{-\int_\Omega u(x;\theta)\mathrm{d}x} \prod_{i=1}^{n(\boldsymbol{x})} \int_{A_{k_i}} u(x;\theta)\mathrm{d}x. \end{aligned}$$

(36)

Each term in the final product is going to zero, but by the Lebesgue Differentiation theorem (and observing that for the ith particle location, the sequence of x_i-containing partition sets $A_{k_i,K}$ is nested), we have that

$$\lim_{K \to \infty} \frac{1}{|A_{k_i,K}|} \int_{A_{k_i}} u(x;\theta)\mathrm{d}x = u(x_i).$$

It follows that

$$\lim_{K \to \infty} \frac{P_\theta(N(A_{\epsilon,1}) = n_1(\boldsymbol{x}), \ldots, N(A_{\epsilon,K}) = n_K(\boldsymbol{x}))}{\prod_{i=1}^{n(x)} |A_{k_i,K}|} = e^{-\int_\Omega u(x)\mathrm{d}x} \prod_{i=1}^{n(\boldsymbol{x})} u(x_i).$$

(37)

The right-hand side is precisely $L(\theta;\boldsymbol{x})$, the function appearing in the likelihood definition (8). Multiplying the numerator and denominator of (34) by the rescaling factor $\prod_{i=1}^{n(x)} |A_{k_i,K}|^{-1}$, we obtain the limit of a sequence of posterior distributions $\pi_K(\cdot \mid \boldsymbol{x})$, which can be written

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{x}) := \lim_{K \to \infty} \pi_K(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{L(\boldsymbol{\theta}; \boldsymbol{x})\pi(\boldsymbol{\theta})}{\int_\Theta L(\boldsymbol{\vartheta}; \boldsymbol{x})\pi(\boldsymbol{\vartheta})\mathrm{d}\boldsymbol{\vartheta}}.$$

This justifies the choice of L appearing in Definition 2.3. □

4.4 Spatial Binning Encodes Less Statistical Information

The previous section provides a formal derivation of the point process likelihood (8) by considering increasingly fine-scale partitions of the spatial domain. In this subsection, we motivate this further by showing that the resulting likelihood encodes greater statistical information than any spatial binning of the process, stated as the following proposition.

Proposition 4.3 *Let $u(x;\theta)$ be the intensity of a Poisson point process on a domain Ω with underlying parameter θ. Let $\{A_j\}_{j=1}^{j=K}$ be any non-overlapping partition of the domain and denote observations of the binned process*

$$N_j = N(A_j) \sim \mathrm{Poisson}\left(\int_{A_j} u(x)\mathrm{d}x\right), \quad j = 1, \ldots, K$$

(38)

with corresponding log-likelihood $\ell_{\mathrm{bin}}(\theta; \boldsymbol{x})$. Let $\mathscr{I}_{\mathrm{bin}}(\theta) = -\mathbb{E}_\theta[\partial_{\theta\theta}\,\ell_{\mathrm{bin}}(\theta; \mathbf{X})]$ be the information for the binned process and $\mathscr{I}_{\mathrm{unbin}}(\theta) = -\mathbb{E}_\theta[\partial_{\theta\theta}\,\log\,L(\theta; \mathbf{X})]$ be the point process information from likelihood (8). Then,

$$\mathscr{I}_{\mathrm{bin}}(\theta) \leq \mathscr{I}_{\mathrm{unbin}}(\theta).$$

Proof The likelihood for the binned process is this process is (35), and the log-likelihood is

$$\ell_{\mathrm{bin}}(\theta; \boldsymbol{x}) := \log\,L_{\mathrm{bin}}(\theta; \boldsymbol{x})$$

(39)

$$= -\int_\Omega u(x;\theta)\mathrm{d}x + \sum_j n_j(\boldsymbol{x})\log\int_{A_j} u(x;\theta)\mathrm{d}x + \sum_j \log\,n_j(\boldsymbol{x})!.$$

(40)

For a parameter θ,

$$\mathcal{I}_{\mathrm{bin}}(\theta) = -\mathbb{E}_\theta[\partial_{\theta\theta}\,\ell_{\mathrm{bin}}(\theta;\boldsymbol{X})]$$

$$= \int_\Omega u_{\theta\theta}(x;\theta)\mathrm{d}x - \sum_j \mathbb{E}[N_j]\,\partial_{\theta\theta}\log\!\left(\int_{A_j} u(x;\theta)\mathrm{d}x\right)$$

$$= \int_\Omega u_{\theta\theta}(x;\theta)\mathrm{d}x - \sum_j \left(\int_{A_j} u(x;\theta)\mathrm{d}x\right)\partial_{\theta\theta}\log\!\left(\int_{A_j} u(x;\theta)\mathrm{d}x\right),$$

$$= \sum_j \frac{\left(\int_{A_j} u_\theta(x;\theta)\mathrm{d}x\right)^2}{\int_{A_j} u(x;\theta)\mathrm{d}x}.$$

(41)

where we have used $\partial_{\theta\theta}n_j(x) = 0$ and $\mathbb{E}_\theta[N_j] = \int_{A_j}u(x;\theta)\mathrm{d}x$. We can compare this directly to the unbinned information obtained via Campbell's formula (16), restated as

$$\mathcal{I}_{\mathrm{unbin}}\!\left(\theta\right) = \int_\Omega u_{\theta\theta}\!\left(x;\theta\right)\mathrm{d}x - \int_\Omega u\,\partial_{\theta\theta}\!\left(x;\theta\right)\,\log\,u\!\left(x;\theta\right)\mathrm{d}x = \int_\Omega \frac{u_\theta(x;\theta)^2}{u(x;\theta)}\mathrm{d}x\,.$$

(42)

To prove $\mathcal{I}_{\mathrm{bin}} \le \mathcal{I}_{\mathrm{unbin}}$, it must be shown that

$$\sum_j \frac{\left(\int_{A_j} u_\theta(x;\theta)\mathrm{d}x\right)^2}{\int_{A_j} u(x;\theta)\mathrm{d}x} \le \int_\Omega \frac{u_\theta(x;\theta)^2}{u(x;\theta)}\mathrm{d}x = \sum_j \int_{A_j} \frac{u_\theta(x;\theta)^2}{u(x;\theta)}\mathrm{d}x,$$

(43)

and noting this must hold for any A_j, this further reduces to showing

$$\frac{\left(\int_{A_j} u_\theta(x;\theta)\mathrm{d}x\right)^2}{\int_{A_j} u(x;\theta)\mathrm{d}x} \le \int_{A_j} \frac{u_\theta(x;\theta)^2}{u(x;\theta)}\mathrm{d}x\,.$$

(44)

Denote the inner product $\langle f, g\rangle_u := \int f(x)g(x)u(x)\mathrm{d}x$. Then the assertion (44) can be written

$$\frac{\langle \partial_\theta\,\log\,u, 1\rangle_u^2}{\langle 1, 1\rangle_u} \le \langle \partial_\theta\log\,u, \partial_\theta\log\,u\rangle_u\,.$$

(45)

In this form (44) is seen to be an expression of the Cauchy–Schwarz inequality, and the claim $\mathcal{I}_{\mathrm{bin}} \le \mathcal{I}_{\mathrm{unbin}}$ is proved. $\square$

Cauchy–Schwarz provides an interpretation of this result. Equality is achieved when $\partial_\theta\,\log\,u \propto$ constant, or when $\partial_\theta u \propto u$. Intuitively, this says that scale parameters (such as λ in the main text) are insensitive to binning, but parameters that modulate the spatial variation of the intensity magnify the information loss from binning.

5 Discussion

In summary, we have established new understanding in the theory and practice of inferring dynamics of spatially stochastic birth-death-diffusion processes in heterogeneous domains from particle snapshot data alone. This pursuit was motivated by technical challenges arising in the study of spatial transcriptomics imaging data, where observable data often consists of positions of a stochastic population of RNA molecules, at a single time (due to cell fixation), in a cell-specific geometry. Using an argument based on the occupation measure of particles, the principal result is a rigorous statistical connection between particle paths and their observation as a Poisson point process with intensity governed by a birth-death-diffusion PDE with appropriate boundaries. This explicit statistical description enables inference that faithfully incorporates often-ignored cell-to-cell heterogeneities. Surprisingly, we find that variation in source location and domain size can increase the accuracy of inference on the underlying dynamics in the regime where particles are long-lived relative to their typical diffusive exit time from the domain. Our study has broader context and future directions in the areas of inverse problems and mathematical modeling of gene expression data.

As an inverse problem, the most important aspect of the work is the formal derivation and justification for the point process likelihood. This non-trivial likelihood must describe a stochastic number of particles and their position, in contrast to previous inference studies with single particle paths (Bernstein and Fricks 2016; Falcao and Coombs 2020) or population positions with fixed number (Chen et al. 2021). Our work clarifies how point processes are the natural resolution of this technical hurdle. Moreover, we show that the point process description encodes more statistical information than spatial binning used in other inference of stochastic reaction-diffusion systems (Dewar et al. 2010). There is future interest in connecting our results to scenarios where particles cannot be tracked individually but temporal fluctuations in spatial bin counts can reveal underlying dynamics, such as the colloidal systems (Mackay et al. 2023) or cell collectives (Simpson et al. 2024). We are far from the first to consider mechanistic models in partnership with spatial point processes in biological systems, e.g., (Plank and Law 2015; Ovaskainen et al. 2014). The distinction of these studies in the ecological literature is subtle but important: in our work, the point process is an emergent statistical description from a birth-death-diffusion process rather than an assumed statistical model for observations. Although the PDE-constraint on the intensity bares resemblance to classical PDE inverse problems (Aster et al. 2013; Flegg et al. 2020), we emphasize the key difference in the observation noise. Rather than observations of the PDE solution (often boundary data) plus additive noise, our inverse problem corresponds to the observations of stochastic particle positions that serve as a proxy for the PDE solution. We provide a preliminary investigation of the statistical properties of this inverse problem, but there are several avenues of promising investigation in discerning similarities and differences in this PDE inverse problem against their classical counterpart.

On the computational side, this connection to other inverse problems suggests there are likely improvements toward more efficient inference. Such improvements will likely be necessary for large-scale inference on now available tissue-scale data (Eng et al. 2019). The primary bottleneck of our approach is each evaluation of the likelihood requires solving the PDE model over all observed domains. The simplicity of the 1D model allowed for

an analytical solution to the PDE and straightforward MCMC sampling. However, these luxuries break down quickly for more realistic investigations. Therefore, one may be able to construct improved samplers seen in other gene model inference (Lin and Buchler 2019; Kilic et al. 2023), or leverage evaluations of the likelihood in approximate techniques such as variational Bayesian methods (Blei et al. 2017; Acerbi 2018). However, these approaches still suffer the bottleneck of each likelihood evaluation requiring the solution to the model over all observed domains. An alternative avenue to circumvent this may be the use of neural network or related approaches. Such techniques have served useful in inverse problems on stochastic chemical reactions (Jiang et al. 2021), classical PDE inverse problems (Raissi et al. 2019), and particle positions (Chen et al. 2021), including with domain generalization (Cao et al. 2023). Therefore, this seems like a promising avenue of study for this problem at the intersection of these topics. Lastly, a variety of investigations remain in adapting the zoo of techniques in the toolbelt for inverse problems to this setup, ranging from model selection (Nardini et al. 2021) to optimal experimental design (Ryan et al. 2016).

The steady-state birth-death-diffusion model was chosen at a deliberate level of complexity to distill the salient features of the data: snapshots of fluctuating populations in a domain. However, there is considerable work that must be done in incorporating known biological complexities before our framework can be used on experimental smFISH data. We acknowledge the context of a vast literature on forward modeling of various non-spatial aspects of stochastic gene expression e.g., (Paulsson 2005; Singh and Soltani 2013; Cao and Grima 2020; Gorin and Pachter 2020; Karamched and Miles 2023). The use of these stochastic models in inference for transcriptomics data is also increasingly prevalent and sophisticated (Herbach et al. 2017; Gómez-Schiavon et al. 2017; Luo et al. 2022). We hope that our investigation serves as a segue for the field to transition into spatially resolved models that can be linked to this already available data. A non-exhaustive list of interesting complexities that may be included in our model is as follows. The production of RNA for many genes is known to be bursty (Jones and Elf 2018; Donovan et al. 2019), with short stochastic intermittent windows of production. A spatial bursting model may be tractable with generating function techniques (Cottrell et al. 2012) that serve fruitful for analytical understanding of non-spatial models (Gorin and Pachter 2020, 2022). On the geometric side, the model must be extended to biological datasets of 2D and 3D FISH imaging (Omerzu et al. 2019). The models can also be extended to multiple compartments (the nucleus, cytoplasm) and serve as a platform for investigating the nature of RNA export through different possible boundary conditions, e.g., Robin boundary conditions arising through stochastic gating of particles (Lawley and Keener 2015) that may approximate the complex dynamics of RNA export (Köhler and Hurt 2007; Azimi et al. 2014). Lastly, the movement model could be extended to include subdiffusive motion seen (Cabal et al. 2006), known to profoundly dictate spatial patterns, especially near boundaries (Holmes 2019).

In summary, our work serves as a basis for more detailed investigations of inferring dynamics from spatial dynamics from snapshots of stochastic particle populations, especially those arising from spatial transcriptomics data. The analytical tractabilityof the model allowed for us to carefully examine the role of cell-to-cell heterogeneities which are surely present in a wide array of biological datasets. This surprising insight of robustness in inference arising from noise contributes a new dimension to the broader theme in

mathematical biology of how noise can sharpen signals (Hänggi 2002). In an era where phenomenological statistical analysis of biological data is commonplace, we hope that our study will inspire others to consider rigorous modeling of biological data to discover other surprises and insights.

Acknowledgements

Data availility

`MATLAB` code for numerical simulations and inference can be found in the Github repository https://github.com/chris-miles/particle_inference_1d. Algorithms are described in Section A.

Appendix A: Numerical Implementation Details

A.1   Stochastic Simulation Algorithm for Synthetic Data

Algorithm 1 describes the generation procedure for synthetic data used in the inference setups of Figs. 5 and 6. Importantly, we simulate trajectories of the particle process, and do inference using the point process theory from Theorem 2.1. The simulation procedure is exact except for a small Δ_t-controlled approximation error in computing whether a Brownian path has exited the domain (Smith and Grima 2019). In all simulations, we take $\Delta_t = 10^{-5}$, several orders of magnitude smaller than any other timescale.

Algorithm 1 Stochastic simulation to generate particle positions for synthetic data.

```
 1: procedure SIMULATEPARTICLES(κ₊, κ₋, D, z, L, T, Δₜ)
 2:     Nbirths ~ Poisson(κ₊T)
 3:     for n = 1, ..., Nbirths do
 4:         Tbirthₙ ~ unif(0, T)
 5:         τdecayₙ ~ exp(1/κ₋)
 6:         Tdecayₙ ← Tbirthₙ + τdecayₙ
 7:         if Tdecayₙ > T then                               ▷ only simulate if particle is possibly alive T
 8:             x ← z,  tₙ ← Tbirthₙ                          ▷ birth particle at source location
 9:             survivedₙ ← 1
10:             while tₙ ≤ T do
11:                 x ← x + √(2DΔₜ) · randn                   ▷ Brownian step
12:                 tₙ ← tₙ + Δₜ
13:                 if x < 0 or x > L then                    ▷ if exits domain
14:                     survivedₙ ← 0 break                   ▷ instantly absorbed, stop simulating
15:             if survivedₙ then                             ▷ if alive at observation T
16:                 append x                                  ▷ store final position
17:     return x₁, ..., xₙ
```

A.2   Metropolis-in-Gibbs Sampling for Bayesian Inference

See Sect. 3.2 for an explanation of prior choices for λ, μ. The choice of prior for λ provides a (marginal) conjugate posterior that can be computed analytically, so a Gibbs-in-Metropolis MCMC approach is used and described in Algorithm 2.

As is standard practice (Gelman et al. 1995), we do a sequence of warm-up phases to target the range of acceptance probabilities between 0.3 and 0.5, starting with $\sigma = 1$ and rounds of 500 steps. After each warm-up round, if the acceptance frequency is below the target range, $\sigma \to \sigma/2$, and if above $\sigma \to 2\sigma$. Once in the range, the ending values of λ, μ are used as initial

values, and the full algorithm is run with this step size σ for a much greater number of steps. The values from the chains for this burn-in phase are discarded.

Algorithm 2 MCMC Gibbs-in-Metropolis sampling of posterior distribution.

```
1:  procedure POSTERIORSAMPMCMC(X, σ, α_λ, β_λ, μ_max)
2:      init λ_0 ~ Gamma(α_λ, β_λ), μ_0 ~ unif(0, μ_max)
3:      for s = 1 ... , S do
4:          ▷ MH step for μ first                                              ◁
5:          μ_* ← μ + σ · randn
6:          while μ_* < 0 or μ_* > μ_max do μ_* ← μ + σ · randn  ▷ only accept proposals for μ in range of
                                                                     prior
7:          π_* ← ℓ(λ_{s−1}, μ_*, X), π_{s−1} ← ℓ(λ_{s−1}, μ_{s−1}, X)        ▷ proposal & current log-likelihood
8:          p ← min{exp(π_* − π_{s−1}), 1}
9:          if rand < p then μ_s ← μ_*                                         ▷ accept proposal    ◁
10:         else μ_s ← μ_{s−1}                                                 ▷ keep current       ◁
11:         ▷ Gibbs sample λ using Gamma conjugacy                             ◁
12:         α, β ← using prior α_λ, β_λ, data X, and μ_s in (21)   ▷ Gibbs step uses current value of μ
13:         λ_s ~ Gamma(α, β)
14:     return λ_0, ..., λ_S,   μ_0, ..., μ_S
```

To choose the number of MCMC steps, we investigate the Gelman-Rubin diagnostic $\hat{R}$ (Gelman et al. 1995) for the presumed worst-case scenario of $M = 10$ snapshots in the setup of Fig.5. For $M = 10$, we initialize multiple chains and compute $\hat{R}$ as a function of the number of MCMC steps, shown in Fig. 7. Convergence is considered satisfactory for $\hat{R} < 1.1$ (Gelman et al. 1995), which is achieved around 10^3 MCMC steps. We take 10^5 MCMC steps for all setups in the main text and do not monitor convergence further.

References

Aster RC, Borchers B, Thurber CH (2013) Parameter estimation and inverse problems. Elsevier, Cambridge. 10.1016/c2009-0-61134-x

Azimi M, Bulat E, Weis K, Mofrad MRK (2014) An agent-based model for mrna export through the nuclear pore complex. Mol Biol Cell 25(22):3643–3653. 10.1091/mbc.e14-06-1065 [PubMed: 25253717]

Acerbi L (2018). Variational Bayesian Monte Carlo. In: Proceedings of the 32nd international conference on neural information processing systems. NeurIPS'18, pp. 8223–8233. Curran Associates Inc., Red Hook, NY

Atkinson AC, Donev AN, Tobias RD (2007) Optimum experimental designs, with SAS. Oxford University Press, Oxford. 10.1093/oso/9780199296590.001.0001

Baddeley AJ (2001) Likelihoods and pseudolikelihoods for Markov spatial processes. In: Institute of Mathematical statistics lecture notes-monograph series, 21–49. Institute of Mathematical Statistics, Beachwood. 10.1214/lnms/1215090061

Bernstein J, Fricks J (2016) Analysis of single particle diffusion with transient binding using particle filtering. J Theor Biol 401:109–121. 10.1016/j.jtbi.2016.04.013 [PubMed: 27107737]

Bowles JR, Hoppe C, Ashe HL, Rattray M (2022) Scalable inference of transcriptional kinetic parameters from MS2 time series data. Bioinformatics 38(4):1030–1036. 10.1093/bioinformatics/btab765 [PubMed: 34788793]

Blei DM, Kucukelbir A, McAuliffe JD (2017) Variational inference: A review for statisticians. J Am Stat Assoc 112(518):859–877. 10.1080/01621459.2017.1285773

Biswas J, Li W, Singer RH, Coleman RA (2021) Imaging organization of RNA processing within the nucleus. Cold Spring Harb Perspect Biol 13(12):039453. 10.1101/cshperspect.a039453

Browning AP, Warne DJ, Burrage K, Baker RE, Simpson MJ (2020) Identifiability analysis for stochastic differential equation models in systems biology. J R Soc Interface 17(173):20200652. 10.1098/rsif.2020.0652 [PubMed: 33323054]

Cao Z, Chen R, Xu L, Zhou X, Fu X, Zhong W, Grima R (2023) Efficient and scalable prediction of spatiotemporal stochastic gene expression in cells and tissues using graph neural networks. bioRxiv. 10.1101/2023.02.28.530379

Cassella L, Ephrussi A (2022) Subcellular spatial transcriptomics identifies three mechanistically different classes of localizing RNAs. Nat Commun 13(1):6355. 10.1038/s41467-022-34004-2 [PubMed: 36289223]

Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. Nature 440(7082):358–362. 10.1038/nature04599 [PubMed: 16541077]

Chaturvedi S, Gardiner C (1978) The Poisson representation II: two-time correlation functions. J Stat Phys 18(5):501–522. 10.1007/bf01014520

Cao Z, Grima R (2020) Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. Proc Natl Acad Sci 117(9):4682–4692. 10.1073/pnas.1910888117 [PubMed: 32071224]

Chaturvedi S, Gardiner C, Matheson I, Walls D (1977) Stochastic analysis of a chemical reaction with spatial and temporal structures. J Stat Phys 17(6):469–489. 10.1007/bf01014350

Cabal GG, Genovesio A, Rodriguez-Navarro S, Zimmer C, Gadal O, Lesne A, Buc H, Feuerbach-Fournier F, Olivo-Marin J-C, Hurt EC, Nehrbass U (2006) SAGA interacting factors confine sub-diffusion of transcribed genes to the nuclear envelope. Nature 44(7094):770–773. 10.1038/nature04752

Clark DE (2022) A Cramér-Rao bound for point processes. IEEE Trans Inf Theory 68(4):2147–2155. 10.1109/tit.2022.3140374

Cognet L, Leduc C, Lounis B (2014) Advances in live-cell single-particle tracking and dynamic super-resolution imaging. Curr Opin Chem Biol 20:78–85. 10.1016/j.cbpa.2014.04.015 [PubMed: 24875636]

Cox DR, Miller HD (1977) The Theory of Stochastic Processes, vol 134. CRC Press, Boca Raton. 10.1201/9780203719152

Chen A, McKinley SA, Wang S, Shi F, Mucha PJ, Forest MG, Lai SK (2014) Transient antibody-mucin interactions produce a dynamic molecular shield against viral invasion. Biophys J 106(9):2028–2036. 10.1016/j.bpj.2014.02.038 [PubMed: 24806935]

Coté A, O'Farrell A, Dardani I, Dunagin M, Coté C, Wan Y, Bayatpour S, Drexler HL, Alexander KA, Chen F, Wassie AT, Patel R, Pham K, Boyden ES, Berger S, Phillips-Cremins J, Churchman LS, Raj A (2024) Post-transcriptional splicing can occur in a slow-moving zone around the gene. elife. 10.7554/eLife.91357

Chari T, Pachter L (2023) The specious art of single-cell genomics. PLoS Comput Biol 19(8):1011288. 10.1371/journal.pcbi.1011288

Cottrell D, Swain PS, Tupper PF (2012) Stochastic branching-diffusion models for gene expression. Proc Natl Acad Sci 109(25):9699–9704. 10.1073/pnas.1201103109 [PubMed: 22660929]

Chen X, Yang L, Duan J, Karniadakis GE (2021) Solving inverse stochastic problems from discrete particle observations using the Fokker-Planck equation and physics-informed neural networks. SIAM J Sci Comput 43(3):811–830. 10.1137/20M1360153

Ding F, Elowitz MB (2019) Constitutive splicing and economies of scale in gene expression. Nat Struct Mol Biol 26(6):424–432. 10.1038/s41594-019-0226-x [PubMed: 31133700]

Donovan BT, Huynh A, Ball DA, Patel HP, Poirier MG, Larson DR, Ferguson ML, Lenstra TL (2019) Live-cell imaging reveals the interplay between transcription factors, nucleosomes, and bursting. EMBO J 38(12):100809. 10.15252/embj.2018100809

Dewar MA, Kadirkamanathan V, Opper M, Sanguinetti G (2010) Parameter estimation and inference for stochastic reaction-diffusion systems: application to morphogenesis in d. melanogaster. BMC Syst Biol 4(1):1–9. 10.1186/1752-0509-4-21 [PubMed: 20056001]

Ding F, Su CJ, Edmonds KK, Liang G, Elowitz MB (2022) Dynamics and functional roles of splicing factor autoregulation. Cell Rep. 10.1016/j.celrep.2022.110985

Durrett R (1996) Stochastic calculus: a practical introduction, vol 6. CRC Press, Boca Raton

Daley DJ, Vere-Jones D (2003). An Introduction to the Theory of Point Processes: Volume I: elementary theory and methods. Springer, New York. 10.1007/b97277

Eng C-HL, Lawson M, Zhu Q, Dries R, Koulena N, Takei Y, Yun J, Cronin C, Karp C, Yuan G-C, Cai L (2019) Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. Nature 568(7751):235–239. 10.1038/s41586-019-1049-y [PubMed: 30911168]

Falcao RC, Coombs D (2020) Diffusion analysis of single particle trajectories in a Bayesian nonparametrics framework. Phys Biol 17(2):025001. 10.1088/1478-3975/ab64b3 [PubMed: 31860874]

Fox ZR, Munsky B (2019) The finite state projection based fisher information matrix approach to estimate information and optimize single-cell experiments. PLoS Comput Biol 15(1):1006365. 10.1371/journal.pcbi.1006365

Flegg MB, Muñoz MA, Smith-Miles K, Yuen WS, Flegg JA, Carroll JG (2020) Parameter estimation for a point-source diffusion-decay morphogen model. J Math Biol 80(7):2227–2255. 10.1007/s00285-020-01494-x [PubMed: 32335708]

Fu X, Patel HP, Coppola S, Xu L, Cao Z, Lenstra TL, Grima R (2022) Quantifying how post-transcriptional noise and gene copy number variation bias transcriptional parameter inference from MRNA distributions. eLife 11:82493. 10.1101/2021.11.09.467882

Gardiner CW (2004) Handbook of stochastic methods, vol 3. Springer, Berlin

Gábor A, Banga JR (2015) Robust and efficient parameter estimation in dynamic models of biological systems. BMC Syst Biol 9(1):1–25. 10.1186/s12918-015-0219-2 [PubMed: 25582171]

Gardiner CW, Chaturvedi S (1977) The Poisson representation I: a new technique for chemical master equations. J Stat Phys 17(6):429–468. 10.1007/bf01014349

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian data analysis. Chapman and Hall/CRC, New York. 10.1201/9780429258480

Gorin G, Fang M, Chari T, Pachter L (2022) RNA velocity unraveled. PLoS Comput Biol 18(9):1010492. 10.1101/2022.02.12.480214

Gupta A, Martin-Rufino JD, Jones TR, Subramanian V, Qiu X, Grody EI, Bloemendal A, Weng C, Niu S-Y, Min KH, Mehta A, Zhang K, Siraj L, Al' Khafaji A, Sankaran VG, Raychaudhuri S, Cleary B, Grossman S, Lander ES (2022) Inferring gene regulation from stochastic transcriptional variation across single cells at steady state. Proc Natl Acad Sci 119(34):2207392119. 10.1073/pnas.2207392119

Gorin G, Pachter L (2020) Special function methods for bursty models of transcription. Phys Rev E 102:022409. 10.1103/PhysRevE.102.022409 [PubMed: 32942485]

Gorin G, Pachter L (2022) Modeling bursty transcription and splicing with the chemical master equation. Biophys J 121(6):1056–1069. 10.1016/j.bpj.2022.02.004 [PubMed: 35143775]

Gómez-Schiavon M, Chen L-F, West AE, Buchler NE (2017) Bayfish: Bayesian inference of transcription dynamics from population snapshots of single-molecule RNA FISH in single cells. Genome Biol 18:1–12. 10.1186/s13059-017-1297-9 [PubMed: 28077169]

Gorin G, Vastola JJ, Fang M, Pachter L (2022) Interpretable and tractable models of transcriptional noise for the rational design of single-molecule quantification experiments. Nat Commun 13(1):7620. 10.1038/s41467-022-34857-7 [PubMed: 36494337]

Hänggi P (2002) Stochastic resonance in biology how noise can enhance detection of weak signals and help improve biological information processing. ChemPhysChem 3(3):285–290. 10.1002/1439-7641(20020315)3:3<285::AID-CPHC285>3.0.CO;2-A [PubMed: 12503175]

Herbach U, Bonnaffoux A, Espinasse T, Gandrillon O (2017) Inferring gene regulatory networks from single-cell data: a mechanistic approach. BMC Syst Biol 11(1):1–15. 10.1186/s12918-017-0487-0

Heinrich S, Derrer CP, Lari A, Weis K, Montpetit B (2017) Temporal and spatial regulation of mRNA export: single particle RNA-imaging provides new tools and insights. BioEssays 39(2):1600124. 10.1002/bies.201600124

Holmes WR (2019) Subdiffusive dynamics lead to depleted particle densities near cellular borders. Biophys J 116(8):1538–1546. 10.1016/j.bpj.2019.02.021 [PubMed: 30954212]

Imbert A, Ouyang W, Safieddine A, Coleno E, Zimmer C, Bertrand E, Walter T, Mueller F (2022) FISH-quant v2: a scalable and modular tool for smFISH image analysis. RNA 28(6):786–795. 10.1261/rna.079073.121 [PubMed: 35347070]

Jones D, Elf J (2018) Bursting onto the scene? exploring stochastic mRNA production in bacteria. Curr Opin Microbiol 45:124–130. 10.1016/j.mib.2018.04.001 [PubMed: 29705632]

Jiang Q, Fu X, Yan S, Li R, Du W, Cao Z, Qian F, Grima R (2021) Neural network aided approximation and parameter inference of non-Markovian models of gene expression. Nat Commun 12(1):2618. 10.1038/s41467-021-22919-1 [PubMed: 33976195]

Jensen JL, Møller J (1991) Pseudolikelihood for exponential family models of spatial point processes. Ann Appl Probab 1(3):445–461

Köhler A, Hurt E (2007) Exporting RNA from the nucleus to the cytoplasm. Nat Rev Mol Cell Biol 8(10):761–773. 10.1038/nrm2255 [PubMed: 17786152]

Karamched BR, Miles CE (2023) Stochastic switching of delayed feedback suppresses oscillations in genetic regulatory systems. J R Soc Interface 20(203):20230059. 10.1098/rsif.2023.0059 [PubMed: 37376870]

Kilic Z, Schweiger M, Moyer C, Shepherd D, Pressé S (2023) Gene expression model inference from snapshot RNA data using Bayesian non-parametrics. Nat Comput Sci 3(2):174–183. 10.1038/s43588-022-00392-0 [PubMed: 38125199]

Kilic Z, Schweiger M, Moyer C, Pressé S (2023) Monte Carlo samplers for efficient network inference. PLoS Comput Biol 19(7):1011256. 10.1371/journal.pcbi.1011256

Lin YT, Buchler NE (2019) Exact and efficient hybrid monte Carlo algorithm for accelerated Bayesian inference of gene expression models from snapshots of single-cell transcripts. J Chem Phys 151(2):024106 [PubMed: 31301707]

Lemerle C, Di Ventura B, Serrano L (2005) Space as the final frontier in stochastic simulations of biological systems. FEBS Lett 579(8):1789–1794. 10.1016/j.febslet.2005.02.009 [PubMed: 15763553]

Lawley SD, Keener JP (2015) A new derivation of Robin boundary conditions through homogenization of a stochastically switching boundary. SIAM J Appl Dyn Syst 14(4):1845–1867. 10.1137/15M1015182

Lammers NC, Kim YJ, Zhao J, Garcia HG (2020) A matter of time: using dynamics and theory to uncover mechanisms of transcriptional bursting. Curr Opin Cell Biol 67:147–157. 10.1016/j.ceb.2020.08.001 [PubMed: 33242838]

La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg P, Furlan A et al. (2018) RNA velocity of single cells. Nature 560(7719):494–498. 10.1038/s41586-018-0414-6 [PubMed: 30089906]

Luo X, Qin F, Xiao F, Cai G (2022) BISC: accurate inference of transcriptional bursting kinetics from single-cell transcriptomic data. Brief Bioinform. 10.1093/bib/bbac464

Mah CK, Ahmed N, Lopez N, Lam D, Monell A, Kern C, Han Y, Prasad G, Cesnik AJ, Lundberg E et al. (2022) Bento: a toolkit for subcellular analysis of spatial transcriptomics data. BioRxiv. 10.1101/2022.06.10.495510

Marx V (2021) Method of the year: spatially resolved transcriptomics. Nat Methods 18(1):9–14. 10.1038/s41592-020-01033-y [PubMed: 33408395]

Munsky B, Fox Z, Neuert G (2015) Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics. Methods 85:12–21. 10.1016/j.ymeth.2015.06.009 [PubMed: 26079925]

Manzo C, Garcia-Parajo MF (2015) A review of progress in single particle tracking: from methods to biophysical insights. Rep Prog Phys 78(12):124601 [PubMed: 26511974]

McLaughlin GA, Langdon EM, Crutchley JM, Holt LJ, Forest MG, Newby JM, Gladfelter AS (2020) Spatial heterogeneity of the cytosol revealed by machine learning-based 3d particle tracking. Mol Biol Cell 31(14):1498–1511. 10.1091/mbc.E20-03-0210 [PubMed: 32401664]

Mellnik JW, Lysy M, Vasquez PA, Pillai NS, Hill DB, Cribb J, McKinley SA, Forest MG (2016) Maximum likelihood estimation for single particle, passive microrheology data with drift. J Rheol 60(3):379–392. 10.1122/1.4943988

Mackay EKR, Marbach S, Sprinkle B, Thorneywork AL (2023) The Countoscope: measuring self and collective dynamics without trajectories. arXiv:2311.00647

Moerner WE (2007) New directions in single-molecule imaging and analysis. Proc Natl Acad Sci 104(31):12596–12602. 10.1073/pnas.0610081104 [PubMed: 17664434]

Moses L, Pachter L (2022) Museum of spatial transcriptomics. Nat Methods 19(5):534–546. 10.1038/s41592-022-01409-2 [PubMed: 35273392]

Maynard KR, Tippani M, Takahashi Y, Phan BN, Hyde TM, Jaffe AE, Martinowich K (2020) Dotdotdot: an automated approach to quantify multiplex single molecule fluorescent in situ hybridization (smFISH) images in complex tissues. Nucleic Acids Res 48(11):66. 10.1093/nar/gkaa312

Nardini JT, Baker RE, Simpson MJ, Flores KB (2021) Learning differential equation models from stochastic agent-based model simulations. J R Soc Interface 18(176):20200987. 10.1098/rsif.2020.0987 [PubMed: 33726540]

Omerzu M, Fenderico N, De Barbanson B, Sprangers J, De Ridder J, Maurice MM (2019) Three-dimensional analysis of single molecule FISH in human colon organoids. Biology Open 8(8):042812. 10.1242/bio.042812

Ovaskainen O, Finkelshtein D, Kutoviy O, Cornell S, Bolker B, Kondratiev Y (2014) A general mathematical framework for the analysis of spatiotemporal point processes. Theor Ecol 7:101–113. 10.1007/s12080-013-0202-8

Osunbayo O, Miles CE, Doval FF, Reddy BJN, Keener JP, Vershinin MD (2019) Complex nearly immotile behaviour of enzymatically driven cargos. Soft Matter 15(8):1847–1852. 10.1039/c8sm01893f [PubMed: 30698601]

Paulsson J (2005) Models of stochastic gene expression. Phys Life Rev 2(2):157–175. 10.1016/j.plrev.2005.03.003

Plank MJ, Law R (2015) Spatial point processes and moment dynamics in the life sciences: a parsimonious derivation and some extensions. Bull Math Biol 77:586–613. 10.1007/s11538-014-0018-8 [PubMed: 25216969]

Qian H, Sheetz MP, Elson EL (1991) Single particle tracking analysis of diffusion and flow in two-dimensional systems. Biophys J 60(4):910–921. 10.1016/s0006-3495(91)82125-7 [PubMed: 1742458]

Rayens NT, Cook KJ, McKinley SA, Payne CK (2023) Palmitate-mediated disruption of the endoplasmic reticulum decreases intracellular vesicle motility. Biophys J 122(7):1355–1363. 10.1016/j.bpj.2023.03.001 [PubMed: 36869590]

Ryan EG, Drovandi CC, McGree JM, Pettitt AN (2016) A review of modern computational algorithms for Bayesian optimal design. Int Stat Rev 84(1):128–154. 10.1088/0034-4885/78/12/124601

Rodríguez-Navarro S, Hurt E (2011) Linking gene regulation to mRNA production and export. Curr Opin Cell Biol 23(3):302–309. 10.1016/j.ceb.2010.12.002 [PubMed: 21227675]

Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 378:686–707. 10.1016/j.jcp.2018.10.045

Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S (2006) Stochastic mRNA synthesis in mammalian cells. PLoS Biol 4(10):309. 10.1371/journal.pbio.0040309

Raj A, Van Oudenaarden A (2008) Nature, nurture, or chance: stochastic gene expression and its consequences. Cell 135(2):216–226. 10.1016/j.cell.2008.09.050 [PubMed: 18957198]

Rehfeldt F, Weiss M (2023) The random walker's toolbox for analyzing single-particle tracking data. Soft Matter. 10.1039/D3SM00557G

Sharp JA, Browning AP, Burrage K, Simpson MJ (2022) Parameter estimation and uncertainty quantification using information geometry. J R Soc Interface 19(189): 20210940 [PubMed: 35472269]

Smith S, Grima R (2019) Spatial stochastic intracellular kinetics: a review of modelling approaches. Bull Math Biol 81(8):2960–3009. 10.1007/s11538-018-0443-1 [PubMed: 29785521]

Schnoerr D, Grima R, Sanguinetti G (2016) Cox process representation and inference for stochastic reaction-diffusion processes. Nat Commun 7(1):1–11. 10.1038/ncomms11729

Simpson MJ, Murphy RJ, Maclaren OJ (2024) Modelling count data with partial differential equation models in biology. J Theor Biol 580:111732. 10.1016/j.jtbi.2024.111732 [PubMed: 38218530]

Svensson V, Pachter L (2018) RNA velocity: molecular kinetics from single-cell RNA-seq. Mol Cell 72(1):7–9. 10.1016/j.molcel.2018.09.026 [PubMed: 30290149]

Singh A, Soltani M (2013) Quantifying intrinsic and extrinsic variability in stochastic gene expression models. PLoS ONE 8(12):84301. 10.1371/journal.pone.0084301

Shaffer SM, Wu M-T, Levesque MJ, Raj A (2013) Turbo FISH: a method for rapid single molecule RNA FISH. PLoS ONE 8(9):75120. 10.1371/journal.pone.0075120

Vo HD, Forero-Quintero LS, Aguilera LU, Munsky B (2023) Analysis and design of single-cell experiments to harvest fluctuation information while rejecting measurement noise. FrontCell Dev Biol 11:1133994. 10.3389/fcell.2023.1133994

Vahid MR, Hanzon B, Ober RJ (2020) Fisher information matrix for single molecules with stochastic trajectories. SIAM J Imag Sci 13(1):234–264. 10.1137/19M1242562

Wieland F-G, Hauber AL, Rosenblatt M, Tönsing C, Timmer J (2021) On structural and practical identifiability. Curr Opin Syst Biol 25:60–69. 10.1016/j.coisb.2021.03.005

Williams CG, Lee HJ, Asatsuma T, Vento-Tormo R, Haque A (2022) An introduction to spatial transcriptomics for biomedical research. Genome Med 14(1):68. 10.1186/s13073-022-01075-1 [PubMed: 35761361]

Walter FC, Stegle O, Velten B (2023) Fishfactor: a probabilistic factor model for spatial transcriptomics data with subcellular resolution. Bioinformatics 39(5):183

Weinreb C, Wolock S, Tusi BK, Socolovsky M, Klein AM (2018) Fundamental limits on dynamic inference from single-cell snapshots. Proc Natl Acad Sci 115(10):2467–2476 [PubMed: 29453278]

Yau C, Campbell K (2019) Bayesian statistical learning for big data biology. Biophys Rev 11(1):95–102. 10.1007/s12551-019-00499-1 [PubMed: 30729409]

A: smFISH snapshots

B: nuclear mRNA model

C: 1D snapshots
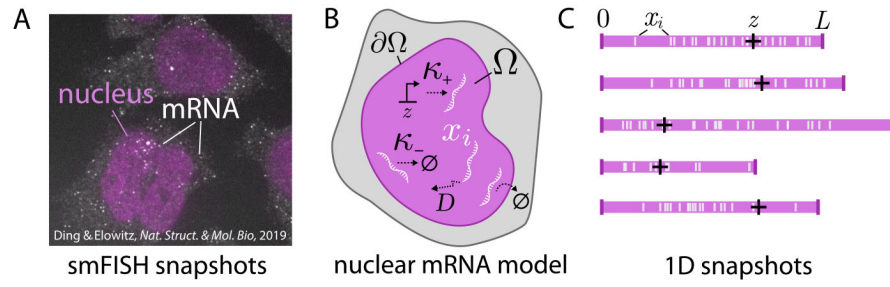
Fig. 1.
Motivation and setup of the work. **A**: Experimental FISH data from (Ding and Elowitz 2019) demonstrating the basis for modeling within this work. Individual mRNAs are resolved in both the nucleus and cytoplasm. **B**: Schematic of the mRNA model in nuclear domain Ω: birth at rate κ_+ at gene site z, decay at rate κ_-, and diffusion with diffusivity D, and instantaneous nuclear export at the boundary $\partial\Omega$ (an absorbing boundary). **C**: 1D simulations of the model demonstrating the heterogeneity in nuclear (or cellular) domain size L and gene site z (black plus)
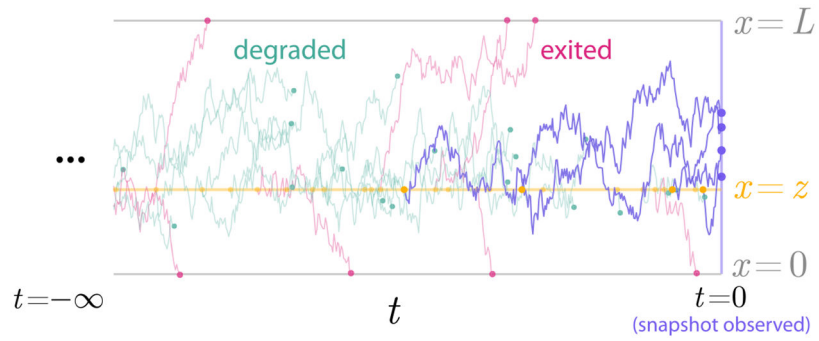
Fig. 2.
Visualization of the stochastic model. Particle populations observed at the snapshot instant ($t = 0$) have long histories that cannot be observed. Here, particles are "born" at a rate λ (yellow circles) in the source location z (yellow line). The particles diffuse until they are either degraded (green circles) or exit the domain (pink circles). The faded trajectories are for particles that left the system before time zero, while the purple trajectories are those that remain active and are observed in the $t = 0$ snapshot. Nondimensional x, t shown with rates $\lambda = 10, \mu = 1.5$ corresponding original simulation parameters $\kappa_+ = 5, \kappa_- = 0.75, D = 0.5, L = 3, L_0 = 1, z = 1$
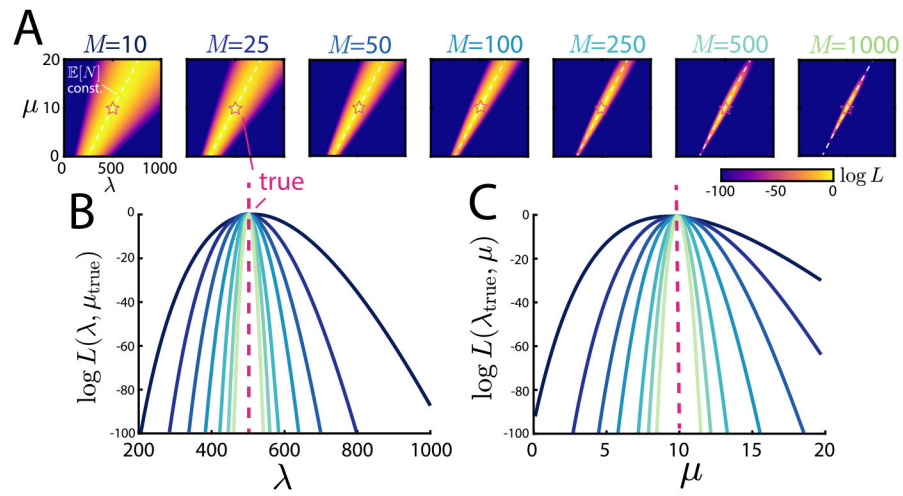
Fig. 3.
Behavior of the likelihood. **A**: $\lambda - \mu$ evaluations of the likelihood $L(\boldsymbol{\theta}; \boldsymbol{x})$ for increasing numbers of snapshots M with fixed $\lambda = 500$, $\mu = 10$, $L = 1$, $z = 0.5$. At low M, the likelihood peaks around the line of constant $\mathbb{E}[N]$, and as M increases, sharply peaks around the true parameter values. **B/C**: Slices of the likelihood shown in panel **A** evaluated at μ_{true} and λ_{true} respectively
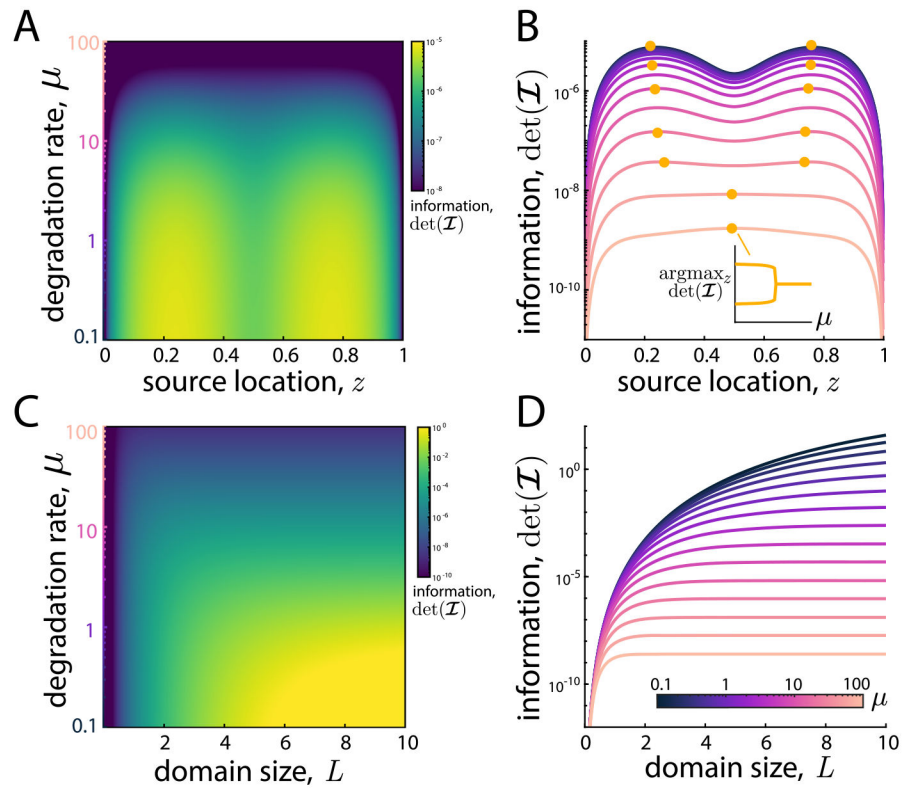
Fig. 4.

Behavior of the information. **A**: Determinant of the information $\det(\mathcal{I})$ in the $z - \mu$ plane with $\lambda = 500$, $L = 1$ fixed. **B**: same information as panel **B**, with $\det(\mathcal{I})$ as a function of z shown for various values of μ, with the maximizing value of z labeled. **B inset**: the-information maximizing source location z as a function of μ. **C**: Determinant of the information $\det(\mathcal{I})$ in the $L - \mu$ plane with $\lambda = 500$, $z = L/2$ fixed. **D**: same information as panel **C**, with $\det(\mathcal{I})$ as a function of L shown for various values of μ
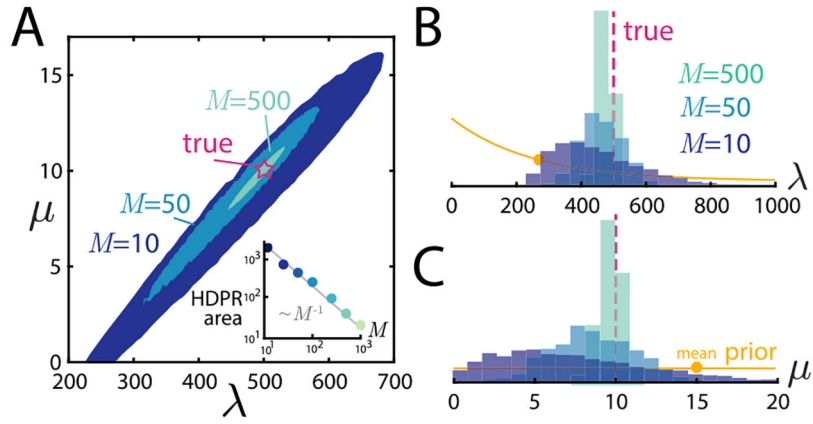
Fig. 5.
Bayesian inference on synthetic data. **A**: Shaded regions show 89% highest-density posterior region (HDPR) for $M = 10, 50, 500$ snapshots and fixed parameters $\lambda = 500, \mu = 10, z = 1/2, L = 1$. As M increases, the posterior regions concentrate around the true parameters. **B,C**: Marginal posterior samples for λ and μ respectively, along with the initial prior distributions for each (with prior means shown with dots on the curves). **A inset**: area of the HDP region as a function of M, showing $\sim M^{-1}$ scaling
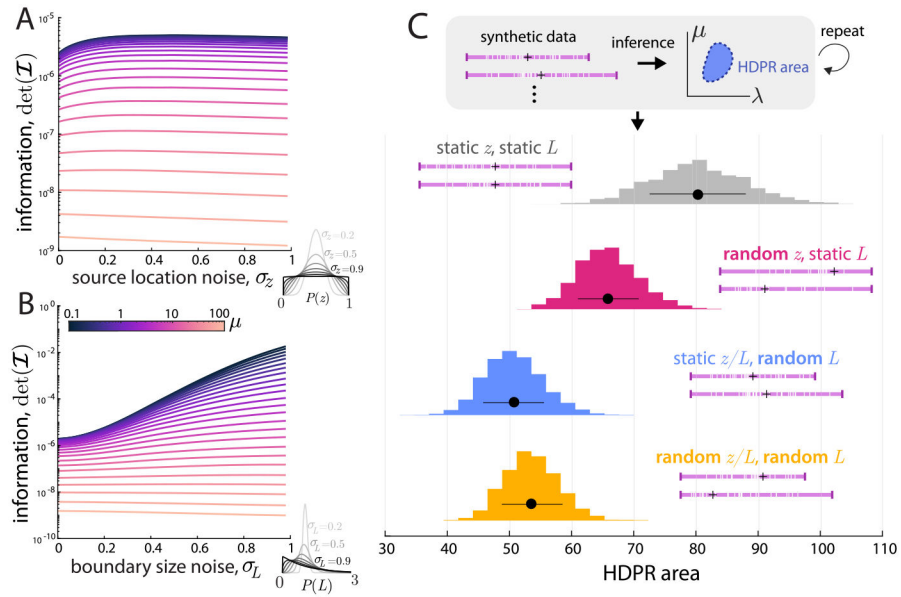
Fig. 6.

Impact of cell-to-cell heterogeneities. **A**: Information with $L = 1$, $\lambda = 500$ fixed and source location varied by a spatially symmetric parameterization with $\sigma_z \to 0$ is a fixed source and $\sigma_z = 1$ is spatially uniform. For low values of μ, spatial heterogeneity increases information, whereas it *decreases* the information for large μ. **B**: Information with $z = L/2$ and varied L parameterized by a distribution such that $\mathbb{E}_\vartheta[L] = 1$ and increasing σ_L spreads out the distribution. Again, lower μ scenarios have increased information from spatial heterogeneity. **C**: Repeated Bayesian inference on synthetic data with $\mu = 1$, varied source location $\sigma_z = 0.75$, cell size $\sigma_L = 0.25$, and both random. All heterogeneous setups have lower uncertainty (smaller HDPR area) than static z, L
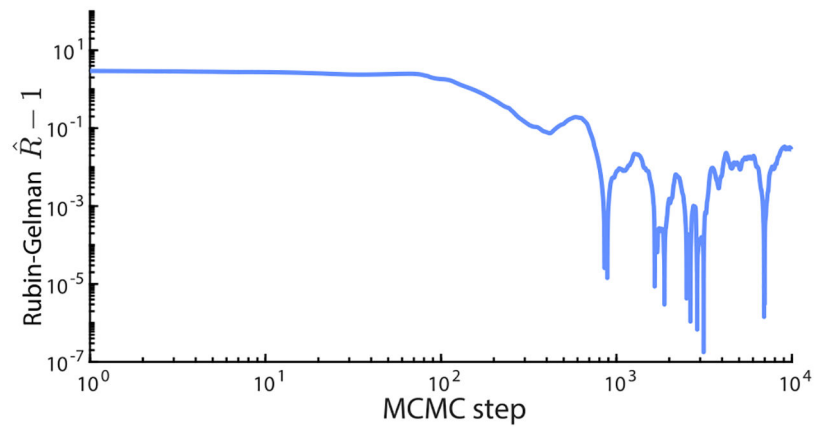
Fig. 7.

Rubin-Gelman $\hat{R} - 1$ demonstrating convergence of MCMC (on the setup in Fig. 5 with $M = 10$) for two chains. Convergence is considered good when $\hat{R} - 1 < 10^{-1}$, achieved in approximately 10^3 MCMC steps