Behind the scenes of cellular organization: Quantifying spatial phenotypes of puncta structures with statistical models including random fields

Kyriacos Nicolaou[a,b,*], Josiah B. Passmore[a], Lukas C. Kapitein[a], Bela M. Mulder[a,c], and Florian Berger[a,b]

[a]Cell Biology, Neurobiology and Biophysics, Department of Biology, Faculty of Science, Utrecht University, 3584 CH Utrecht, The Netherlands; [b]Centre for Complex Systems Studies, Utrecht University, 3584 CE Utrecht, The Netherlands; [c]Institute AMOLF, 1098 XG Amsterdam, The Netherlands

ABSTRACT The cellular interior is a spatially complex environment shaped by nontrivial stochastic and biophysical processes. Within this complexity, spatial organizational principles—also called spatial phenotypes—often emerge with functional implications. However, identifying and quantifying these phenotypes in the stochastic intracellular environment is challenging. To overcome this challenge for puncta, we discuss the use of inference of point-process models that link the density of points to other imaged structures and a random field that captures hidden processes. We apply these methods to simulated data and multiplexed immunofluorescence images of Vero E6 cells. Our analysis suggests that peroxisomes are likely to be found near the perinuclear region, overlapping with the endoplasmic reticulum, and located within a distance of $1\,\mu$m to mitochondria. Moreover, the random field captures a hidden variation of the mean density in the order of $15\,\mu$m. This length scale could provide critical information for further developing mechanistic hypotheses and models. By using spatial statistical models including random fields, we add a valuable perspective to cell biology.

New Methods

Open Softwares

SIGNIFICANCE STATEMENT

- The quantification of the remarkable spatial organization within cells from microscopy images is challenging because of spatial variability and unknown—not imaged—effects.
- The authors introduce spatial statistical models based on log-Gaussian Cox processes to quantify spatial phenotypes of puncta. These models can include other imaged components, such as different organelles, prior knowledge about spatial organization, and a Gaussian random field to describe spatial variability. These methods are applied to synthetic data derived from a biophysical model and experimental data of imaged peroxisomes.
- The study quantitatively connects spatial phenotypes via statistical models to a biophysical interpretation.

SEING SPATIAL PHENOTYPES THROUGH THE RANDOMNESS

The interior of cells comprises a spatially complex array of components, including cytoskeletal filaments such as microtubules and actin, various vesicular organelles such as endosomes, peroxisomes, lipid droplets, and polymorphic structures like mitochondria and the endoplasmic reticulum (ER) that constantly undergo membrane remodeling. These components are distributed in distinct ways throughout the cell. Cytoskeletal filaments often have a radial arrangement that spans a significant proportion of the cell. The ER is dense and sheet-like around the nucleus and tubular at the periphery. Mitochondria form dynamic networks that tend to cluster near the nucleus but can also reach out to the cell periphery. Vesicular organelles may be dispersed at seemingly random locations or in proximity to other organelles. This complex spatial landscape emerges from multiple processes such as transport mechanisms, interactions among components, and chemical reaction-diffusion networks, all of which inherently exhibit a random nature manifested in the complex spatial organization of cells.

Behind the complexity and randomness, there are certain spatial organizational principles—which we call *spatial phenotypes*—that are integral to cellular function. For example, the clustering of receptors enhances cooperative activation (Sourjik, 2004), the radial arrangement of microtubules facilitates efficient cargo transport, and the proximity of organelles promotes effective molecular exchange (Elbaz and Schuldiner, 2011; Phillips and Voeltz, 2016). Thus, understanding life at the cellular level requires answering the key questions: What are the spatial phenotypes of the cellular components? How do they emerge? And why does the cell adopt these specific arrangements? Addressing these questions begins with identifying spatial phenotypes from microscopy images, which requires "eing through" the apparent complexity and randomness of the intracellular spatial landscape.

As a first step toward identifying and quantifying spatial phenotypes, we will focus on cellular components that appear point-like in microscopy images, commonly referred to as puncta, such as vesicular organelles, aggregates, and individual molecules. These point-like components, characterized primarily by their locations, are less spatially complex than other cellular components whose spatial organization involves different geometrical features such as shape, orientation, and size. This simplicity makes their quantification more tractable, providing a manageable starting point for identifying and quantifying spatial phenotypes. From a dataset of locations of puncta, which we will refer to as a *point pattern*, we would like to identify spatial phenotypes of these puncta such as clustering, spatial gradients of point densities, and proximity to other structures.

Returning to the "how," "what," and "why" of spatial phenotypes for point-like components, we would like to convey the following conceptual framework, see Figure 1. Generally speaking, biophysical and biochemical processes, genetics, and environmental conditions are the causal factors behind the spatial distribution of point-like components. These complex processes can be represented as a point process—a stochastic mathematical model—whose realizations are point patterns. Assuming that the observed point pattern is generated by such a point process, we can then fit the model to quantify the spatial phenotype, answering the "what." Once a spatial phenotype is identified, it can inform the construction of mechanistic hypotheses on the "how" and "why," which can be further investigated through experiments and biophysical models.

To see the spatial phenotypes behind the apparent randomness, we need statistical methods tailored for point patterns. In cell biology, elements of point-pattern analysis are increasingly utilized to analyze imaged puncta (Nicovich *et al.*, 2017; Levet *et al.*, 2019; Khater *et al.*, 2020; Wu *et al.*, 2020). The most common tools are summary statistics related to the distances between points, indicating whether a pattern is completely random, clustered, or regularly spaced. Although these statistics provide some initial insight, they typically enable only limited interpretations and fail to reveal more of the organizational processes suggested by the observed patterns. A deductive approach to translating a hypothesis for spatial phenotypes into a statistical model is more suitable for quantifying and identifying spatial phenotypes. Such spatial statistical models for point patterns have been broadly applied in other fields, such as geoscience, ecology, and epidemiology, to investigate the spatial correlation of points to other structures, spatial trends, and effective interactions between points (Diggle, 2013; Illian *et al.*, 2013; Raynaud and Nunan, 2014; Baddeley *et al.*, 2015; Velázquez *et al.*, 2016). Despite their relevance in describing cellular point patterns, spatial statistical models have not yet attracted much attention in cell biology.

Toward quantifying spatial phenotypes of puncta using spatial statistical models, it is important to acknowledge that spatial organization is usually influenced by multiple, often hidden, processes and structures. The cellular processes that determine the organization of puncta in cells often leave traces of their spatial characteristics in the point patterns that they generate. Quantifying these hidden spatial characteristics and how they change under perturbation through the point pattern can help us construct mechanistic hypotheses on their origin. To that extent, we need statistical models that can explain spatial heterogeneity through hidden random components. Ideally, these models would be biophysical, including transport mechanisms, physical interactions, and dynamics. However, the complexity and lack of knowledge of the processes that determine the organization make the construction of such models difficult. Therefore, we can instead phenomenologically model the hidden part of the spatial heterogeneity of point patterns using mathematical structures that can capture the collective effect of the processes behind it.

Gaussian random fields (GRFs) are one type of such mathematical structure. They assume random but correlated values throughout space (Rue and Held, 2005). A model that uses GRFs to capture the hidden spatial heterogeneity in the density of points is the log-Gaussian Cox process (LGCP) (Møller *et al.*, 1998). Recent advances in computational statistics have made the inference of LGCPs from point pattern data possible (Rue *et al.*, 2009; Lindgren *et al.*, 2011; Simpson *et al.*, 2016), allowing us to estimate parameters that characterize it, such as its variance and correlation length.
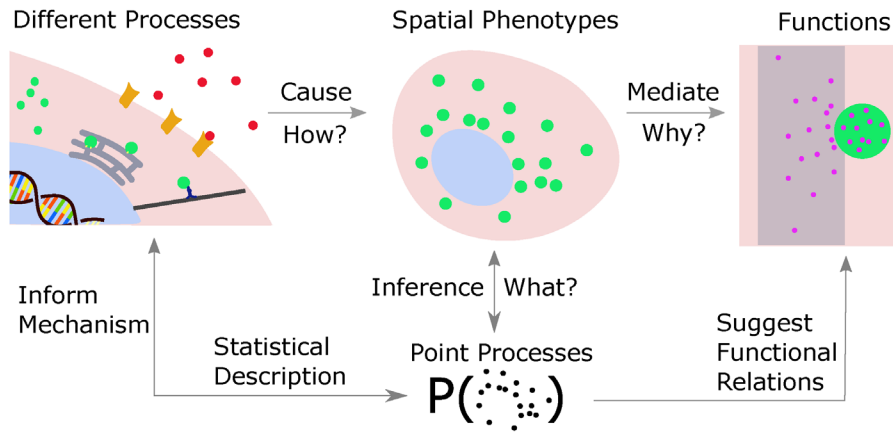
FIGURE 1: Conceptual framework. We relate the "how?," "why?," and "what?" of the spatial organization in cells; Different processes generate spatial phenotypes that, in turn, mediate biological functions. Point processes inferred from the spatial phenotypes or constructed from statistical descriptions can not only quantify the spatial phenotypes but also inform mechanistic hypotheses for the underlying causes. Furthermore, they may suggest functional relations between cellular components.

Consequently, statistical models, including GRFs, have been increasingly applied to elucidate the hidden structures behind point patterns in fields such as ecology and geoscience (Li *et al.*, 2012; D'Angelo *et al.*, 2022).

In the following sections, we briefly introduce point-pattern analysis, discussing spatial Poisson processes and LGCPs, as spatial statistical models that link the density of point-like components to other imaged cellular structures and capture the hidden spatial variability by GRFs. We present these concepts intuitively within the context of cell biology and our framework for connecting point patterns to spatial phenotypes.

First, to showcase how these models capture hidden spatial heterogeneity, we apply them on synthetic data of simulated diffusing particles in a crowded cell. Then, to demonstrate the practical application of these models on experimental data, we fit an LGCP model to point patterns of imaged peroxisomes in Vero E6 cells, a kidney epithelial cell line derived from the African green monkey. This model quantifies the spatial correlation of peroxisomes with two possible partners; the ER, well known to interact with peroxisomes, and the mitochondria, with which the extent of interaction is less clearly understood in mammalian cells (Kumar *et al.*, 2024). In addition, the proximity of peroxisomes to the nucleus is also considered. By including a GRF in the model, we capture residual spatial variability in the peroxisomes' distribution, characterized by a correlation length of about 15 μm, reminiscent of a hidden process that may influence their spatial organization.

FROM COMPLETE SPATIAL RANDOMNESS TO SPATIAL STATISTICAL MODELS

To illustrate the framework shown in Figure 1, let us consider a collection of molecules diffusing inside a cell. Assuming that they follow ideal diffusion without interactions with the cell boundary, these particles are equally likely to be found everywhere within the cell. Here, the physical mechanism behind the distribution of molecules is diffusion. Snapshots of these diffusing particles can be represented with the following point process: We place the points representing the molecules randomly in space and independently of each other. This process is referred to as complete spatial randomness (CSR) or the homogeneous Poisson process and has the following characteristics: 1) the number of points found in any given

region of space follows a Poisson distribution with a mean proportional to its area in two dimensions and to its volume in three dimensions; 2) the mean density of points is the only parameter characterizing CSR. The mean density can be inferred from the data as the number of points in a sample divided by the area or volume of the observation. Therefore, the spatial phenotype here is the CSR quantified by the mean density, a single constant number, which is uniform across space.

The simplicity of CSR and its analytical properties make it the central null model in point-pattern analysis, as it can be used to quantify the degree of randomness in a point pattern. Several widely used statistical tools use CSR as a benchmark to gain statistical insight into the properties of a point pattern. They include the estimation of distributions similar to the nearest-neighbor function to quantify the degree of clustering of point patterns and the estimation of local mean density through quadrat counting and kernel smoothing (Baddeley *et al.*, 2015). These tools can indicate deviations from CSR, such as clustering, regularity, or, in general, heterogeneity of the mean density.

Toward modeling the effect of the highly heterogeneous and dynamic cellular environment on the distribution of puncta, we can consider points that are scattered independently of each other, given a mean density that is nonhomogeneous in space—a process known as the nonhomogeneous Poisson process. In this case, we construct statistical models that include a mean density that depends on other structures, for example, on other simultaneously imaged components, and on hidden spatial effects.

MODELS BASED ON PREDICTIVE MAPS: LINKING THE MEAN DENSITY OF A POINT PATTERN TO IMAGED STRUCTURES

Spatial associations between cellular components are very common and are often related to functional synergies among them. Therefore, images of one structure can be used to predict the presence of another. For point patterns, we can incorporate information from other imaged structures into a model for the mean density. For example, we can directly use the fluorescence intensity of the structures as a predictive map or construct maps of the nearest distance from them to capture proximity. We use the term *predictive maps* to denote that they could potentially have predictive power

for the mean density through spatial correlations, although we remain agnostic about potential causal relationships.

We describe the mean density as the exponential of the weighted sum of such predictive maps because the additivity makes interpretation easier and the exponential ensures that the mean density remains positive. Thus, we write the mean density at position r in the cell as

$$\text{mean density (r)} = \text{Exp}[\beta_0 P_0(r) + \beta_1 P_1(r) + \beta_2 P_2(r) + \cdots].$$

Here, $P_i(r)$ are the predictive maps, of which we set $P_0(r) = 1$ as the zero-bias predictive map, and β_i are their associated weight parameters that quantify the extent of their influence on the mean density. Each component thus scales the mean density by a factor of $e^{\beta_i P_i(r)}$ independently of the other components. Then the coefficient β_i quantifies the spatial association of the puncta to the corresponding map.

To facilitate better interpretation, it is common in spatial statistical modeling to standardize these predictive maps so that their mean is zero and standard deviation is 1. This standardization has two key benefits: 1) it isolates the constant effect within β_0, leaving the other maps to explain spatial fluctuations around that baseline; and 2) the fold-change in density from a standardized predictor value of -1 to $+1$ is simply $e^{2\beta}$. This standardization makes the coefficients β_i directly comparable between different maps.

The physical interpretation of spatial association varies depending on the context and the available information about the system. A positive association might suggest that puncta form on, bind to, or become trapped by a particular imaged structure. Conversely, a negative association with a structure could indicate spatial exclusion or degradation of the puncta component in its presence. It is also important to note that such associations need not arise directly from the imaged structure itself; they may be mediated indirectly through other structures or processes.

This formulation based on the addition of predictive maps does not capture potential interdependencies among them. Interaction terms that represent joint characteristics of multiple predictors can be as simple as the product of two predictive maps. One simple biophysical interpretation of interaction terms is that the presence of a puncta structure may require multiple components to coincide, producing a specific effect that depends on their combined interaction. Because we are introducing these concepts in a cell biology context—and because selecting and interpreting interaction terms can quickly become difficult—we focus here on models without them.

Given a dataset of puncta locations, and available predictive maps, the parameters β_i can then be inferred using, for example, the spatstat package (Baddeley and Turner, 2005). From the inference, we can determine to what extent point-like components spatially correlate with specific cellular structures. Discovering the causal mechanisms behind the spatial correlation, of course, still requires the development of mechanistic hypotheses and perturbation experiments. Analyzing how these perturbations affect the parameters inferred with these models can provide deeper insights into the underlying biological processes that determine organization. Although this approach offers valuable insight, it can only capture part of the complex spatial distribution of point-like components within cells, because it only accounts for spatial associations in relation to the imaged structures. However, a large part of the organization is likely associated with structures and processes not included in the model. Remarkably, in statistical modeling, techniques have been developed to include hidden spatial structures in the model of the mean density, and thus provide a new approach to uncover spatial organizational principles from point patterns.

MODELS BASED ON GRFS: LINKING THE MEAN DENSITY OF A POINT PATTERN TO HIDDEN SPATIAL STRUCTURES

Although predictive maps based on imaged structures provide a starting point for describing the mean density, they barely scratch the surface of the spatial cellular complexity. Numerous other factors that are not observed and hence unknown could play crucial roles in this organization. These hidden factors could potentially contribute to the spatial variability of the observed point pattern and also influence characteristic length scales.

As a simple illustrative example of hidden spatial structure, let us consider a point process with a spatially dependent mean density that has high values around hidden "activity centers," resulting in point patterns with clusters; these models are generally referred to as Neyman-Scott processes (Baddeley et al., 2015). A special case is a Matérn cluster process that considers circular regions, in which points are placed according to CSR (Matérn, 1960). Alternatively, the mean density of a point pattern can be described as the sum of Gaussian functions whose centers are distributed following CSR—a process known as the Thomas cluster process (Thomas, 1949). These models provide a framework for inferring hidden spatial characteristics, such as length scales and mean densities related to the activity centers directly from point patterns. However, the hidden structure of these models is rather rigid because the activity centers have a fixed shape which makes them inadequate to account for the remarkable spatial complexity of cells.

To effectively describe the combined contribution of different cellular processes that influence the spatial patterns of point-like organelles, we need a more flexible mathematical object. Such an object should first account for spatial fluctuations and, second, follow the principle that proximity implies stronger relatedness—a concept as important in cell biology as it is in geography—in other words, it has random values in space that are spatially correlated. GRFs are such mathematical objects, and their use in uncovering hidden effects behind point patterns is on the rise.

GRFs are continuous random fields whose values at any set of locations are jointly distributed by a multivariate Gaussian distribution with distance-dependent correlation. Thus, a GRF is characterized by its mean and spatial correlation structure, which describes how the values at nearby points depend on each other. The correlation structure is usually imposed through a choice of a covariance function, which is typically a decreasing function of the distance between points and characterized by a single correlation length. This correlation length describes how fast the correlation between two values of the GRF decays as their distance increases. Therefore, our GRF depends on the position r, the mean μ, the standard deviation σ, and the correlation length ρ, combined in the notation GRF(r; μ, σ, ρ). The spatial structures generated by such a GRF depend on the size of the correlation length, see Figure 2. A GRF with zero correlation length, meaning that its values are uncorrelated and drawn independently from a Gaussian distribution, describes spatial white noise. GRFs have been used successfully in spatial modeling in various fields and are discussed in detail in textbooks (Rue and Held, 2005).

GRFs have been implemented as determining factors in the analysis of point patterns by considering the mean density as the exponential of a GRF or equivalently, the logarithm of the mean
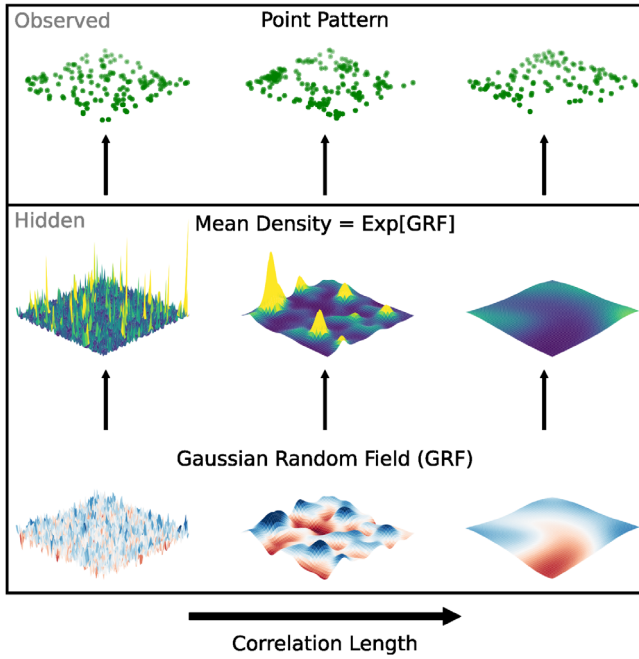
FIGURE 2: LGCP. Each column represents a realization, with the underlying GRF having an increasing correlation length from left to right. The hierarchical structure of the model is illustrated vertically with arrows indicating dependencies. On the top, there are three observed point patterns sampled from the mean densities shown in the middle. The mean densities are derived from the exponential transformation of the respective GRFs, shown at the bottom. Both the mean densities and the GRFs are hidden—not observed.

density as a GRF (Møller *et al.*, 1998). In this case, the mean density is now a random surface. Therefore, a point pattern generated from a Poisson process with such a stochastic mean density is a result of a doubly stochastic process. Such a doubly stochastic Poisson process with a stochastic mean density is called Cox process, and the particular process with a GRF is referred to as the LGCP. For such a LGCP model, we write the mean density as

$$\text{mean density (r)} = \text{Exp}\left[\beta_0 P_0(r) + \text{GRF}(r; \mu = 0, \sigma, \rho)\right],$$

in which the mean μ of the GRF has been absorbed by the coefficient β_0 of the zero-bias predictive map $P_0(r) = 1$, introduced in the previous model.

To intuitively interpret the parameters defining the GRF and thus the inhomogeneity, we first consider the correlation length ρ. This parameter characterizes the size of patches with high and low densities. The magnitude of the density difference between these patches relative to the mean density of points can be described by the fold-change of the mean density between the "hot spots" and "cold spots." Taking a typical value for a hot spot in the GRF, $+\sigma$, and a typical value for a cold spot $-\sigma$, the fold difference in the mean density between them is $e^{2\sigma}$.

If we assume that such a process governs an observed point pattern, we want to quantify the values of the underlying GRF and its characteristics from that point pattern. However, inferring a LGCP presents both conceptual and computational challenges. Despite these difficulties, an increasing number of techniques have been developed to streamline and optimize inference for LGCP models (Taylor *et al.*, 2013; Bachl *et al.*, 2019).

MIXED MODELS BASED ON PREDICTIVE MAPS AND GRFS: LINKING THE MEAN DENSITY OF A POINT PATTERN TO IMAGED AND HIDDEN SPATIAL STRUCTURES

Standard GRFs are spatially stationary and isotropic, meaning that any location in space is statistically equivalent to any other. These properties are unlikely to hold for intracellular organization because there are many well-known organizational trends imposed by subcellular structures and processes. Adequate predictive maps as components of the mean density can capture a large part of this organization. To describe the remaining spatial variation and heterogeneity caused by unobserved, hidden processes, we propose to add a GRF as a complementary component to the mean density.

Such a mixed model for the mean density includes predictive maps and a GRF, and is given by,

$$\text{mean density (r)} = \text{Exp}\left[\beta_0 + \beta_1 P_1(r) + \beta_2 P_2(r)\right.$$
$$\left. + \cdots + \text{GRF}(r; \sigma, \rho)\right].$$

This mixed model consists of several components that contribute to the mean density. Depending on the specific predictors used in the model to capture trends in the data, the free model parameters will acquire different numerical values when fitted to a point pattern. Therefore, the results depend on the combination of components used in the model and how they compete or complement each other. For example, if there is a large-scale spatial trend in the data, the GRF without predictive maps will try to capture that trend with a large correlation length. In contrast, if that trend could be explained by a predictive map, the GRF will capture residual spatial variability. The fitted parameters should thus always be interpreted together for consistent quantitative insights and also carefully compared across different models.

In the next sections, we outline how to fit such models to simulated and experimental data using a Bayesian statistical framework.

BAYESIAN INFERENCE OF MIXED MODELS

Here, we outline the Bayesian inference procedure for modeling a mean density, incorporating predictive maps and a GRF to characterize the spatial phenotype of a point pattern. Following a Bayesian approach, the results are expressed as posterior distributions of the parameters given the observed data. These posterior distributions for the parameters β_i, ρ, σ, as well as for the predicted GRF and the mean density, collectively quantify the spatial phenotype.

To perform the inference, we must first choose appropriate prior distributions for the model parameters. Priors represent our initial beliefs about parameter values. When knowledge is limited, it is common to use broad, uninformative priors. To estimate how broad the priors for β should be, we consider their effect on the mean density which is the exponential of their product with a typical value of the predictive map. For standardized predictive maps, the mean density between a cold spot and a hot spot is given by $e^{2\beta}$, as discussed above. Our aim is to construct a broad distribution for β. Therefore, we consider a few example cases: $\beta = 1$ corresponds to about a $e^2 \approx 7$-fold difference and $\beta = 10$ to a difference of e^{20}, which is an extremely large value. Such a large contribution implies that a predictive map completely reflects the mean density. Thus, selecting priors that cover β values substantially larger than 1 can be considered noninformative, such as a normal distribution with a standard deviation of 100 as used throughout this study.

For the parameters of the GRF, a typical choice of prior is a penalized complexity prior designed to penalize very small ρ and large σ, both of which would result in an overly flexible GRF and thus risk overfitting (Simpson *et al.*, 2017; Fuglstad *et al.*, 2019). In practice, these priors can be defined with a likely minimum for ρ and a likely maximum for σ. A reasonable choice for the possible minimum for ρ would be the physical size of the structure corresponding to the puncta. To choose a maximum for σ, we consider what the maximal contribution by the GRF could be. Given the arguments discussed above, the fold difference in the mean density is $e^{2\sigma}$. In this way, a maximum σ can be estimated from the fold difference of cold and hot spots in the dataset.

With the choice of prior distributions for the parameters, Bayesian inference can be carried out using different methods. These methods include Markov chain Monte Carlo (MCMC), variational methods, and the integrated nested Laplace approximation method (INLA) (Teng *et al.*, 2017).

In the following sections, we proceed with two applications in which we present a detailed model construction and interpretation. A thoroughly documented, step-by-step implementation of these analyses is available in our GitHub repository (Nicolaou, 2024), ensuring full reproducibility and deeper insight into the procedures. For both applications, we use the inlabru R package (Bachl *et al.*, 2019), which is based on an INLA method for approximate Bayesian inference of latent Gaussian models and on recent developments in the field of spatial statistical modeling (Rue *et al.*, 2009; Lindgren *et al.*, 2011; Simpson *et al.*, 2016). A key requirement for spatial inference in inlabru is the construction of a two-dimensional mesh covering the extent of the dataset. Because this mesh discretizes the spatial domain, its resolution must be fine enough to capture the relevant spatial details.

A MIXED MODEL CAN QUANTIFY A DENSITY GRADIENT AND HIDDEN CROWDING FROM SIMULATED DATA OF DIFFUSING PARTICLES

To illustrate how a mixed spatial statistical model can reveal key biophysical features in cellular systems, we fit such a model to a synthetic dataset generated by a simple biophysical simulation. Specifically, in this simulation, we describe point particles, representing our puncta, diffusing through a crowded two-dimensional cellular environment, see Figure 3A. Snapshots of the simulation in the steady state provide point patterns that we fit with our mixed model, see Figure 3B.

To construct the simulation, we represent a crowded two-dimensional cellular environment using an annulus: the inner circle corresponding to the nuclear membrane and the outer circle corresponding to the cell membrane. We choose the radius of the outer circle as the unit of length and the radius of the nucleus $1/3$ in this unit. We place disks of radius 0.1 as static, stationary obstacles in the annular domain. The positions of these disks follow a homogeneous Poisson process with an average density of 20 disks per unit area. This construction is also known as a Boolean model of disks (Chiu *et al.*, 2013), which has been widely used to mimic crowding effects in porous media (Horgan and Ball, 1994; Scholz *et al.*, 2015).

Our point particles of interest are produced at a constant rate on the surface of the nucleus and subsequently spontaneously degrade independent of their position. We consider the lifetime of the particles as the time unit, and with this unit, we choose the production rate to be equal to 100. These particles diffuse in the

area of the annular domain that is not occupied by the obstacles with a diffusion coefficient of 0.3 in the introduced units.

For this simple biophysical process, we can immediately deduce some relevant quantities. First, the mean number of particles should be production rate × lifetime = 100, following standard population dynamics. Moreover, given that the particles are produced on the nucleus and diffuse with a mean lifetime of 1, one expects a characteristic decay length of the mean density of particles described by $\sqrt{\text{diffusion coefficient} \times \text{lifetime}} \approx 0.55$. Last, regarding crowding, multiple length scales, such as the obstacle radius, the cluster sizes, and the typical size of the obstacle-free areas, collectively influence local diffusion in this system. We expect that the Gaussian random field component of the statistical model would effectively capture their net effect and suggests the correct order of magnitude of these length scales.

By fitting a mixed model to these simulated point patterns, we can evaluate the model's ability to infer underlying parameters, such as the decay length, and the spatial heterogeneity induced by crowding. Thus, we not only validate our approach with synthetic data but also provide a framework for applying such modeling techniques to experimental datasets from cell biology.

To connect the simulation results to a statistical framework, we construct a statistical model for the mean particle density that incorporates biologically motivated predictive maps. Specifically, we consider the distance from the nucleus as a predictive map, reflecting the assumed gradient of particle density, and a constant zero-bias predictive map, as in all models discussed here. The spatial heterogeneity in the mean density of particles induced by the crowding structures, which are now hidden in the observational data, is then modeled with a Gaussian random field. The dataset used to fit the model only includes the point pattern and the predictive map for the distance, see Figure 3C. No direct information about the crowding structures is provided to the model.

The mean density is thus given by

$$\text{mean density}(r) = \text{Exp}\left[\beta_0 + \beta_d P_d(r) + \text{GRF}(r; \sigma, \rho)\right],$$

in which $P_d(r)$ is the predictive map for the distance from the nucleus at location r, and $\text{GRF}(r, \sigma, \rho)$ is a Gaussian random field characterized with a standard deviation σ and a correlation length ρ. The parameters β_i are the respective weights.

To capture the spatial phenotypes of the point pattern, we perform an aggregated fit that includes nine unique simulated cells, in which we take a snapshot of the particle locations when the simulation has reached a steady state. We choose noninformative priors for the parameters of the statistical model, in which the priors for β_0 and β_1 are normal distributions with a standard deviation of 100. For the GRF component, we choose a likely maximum for the standard deviation of 2, which corresponds to roughly $e^{2 \times 2} \approx 50$-fold difference between hot spots and cold spots in the mean density, and a likely minimum of 0.01 for the correlation length. For the spatial discretization required by the inference, we construct a mesh with a maximum edge length of 0.1.

From the inference, we obtain marginal posterior distributions for the model parameters that are in good agreement with the underlying biophysical phenomena, see Figure 3D. The mean of the posterior for the zero-bias map coefficient β_0 is ~ 3.3, which corresponds to a mean density of $e^{3.3} \approx 28$. Multiplying this density by the area results in a mean number of 80 particles. The discrepancy to the expected value of 100 is explained by the missing contributions of the other components. By integrating the
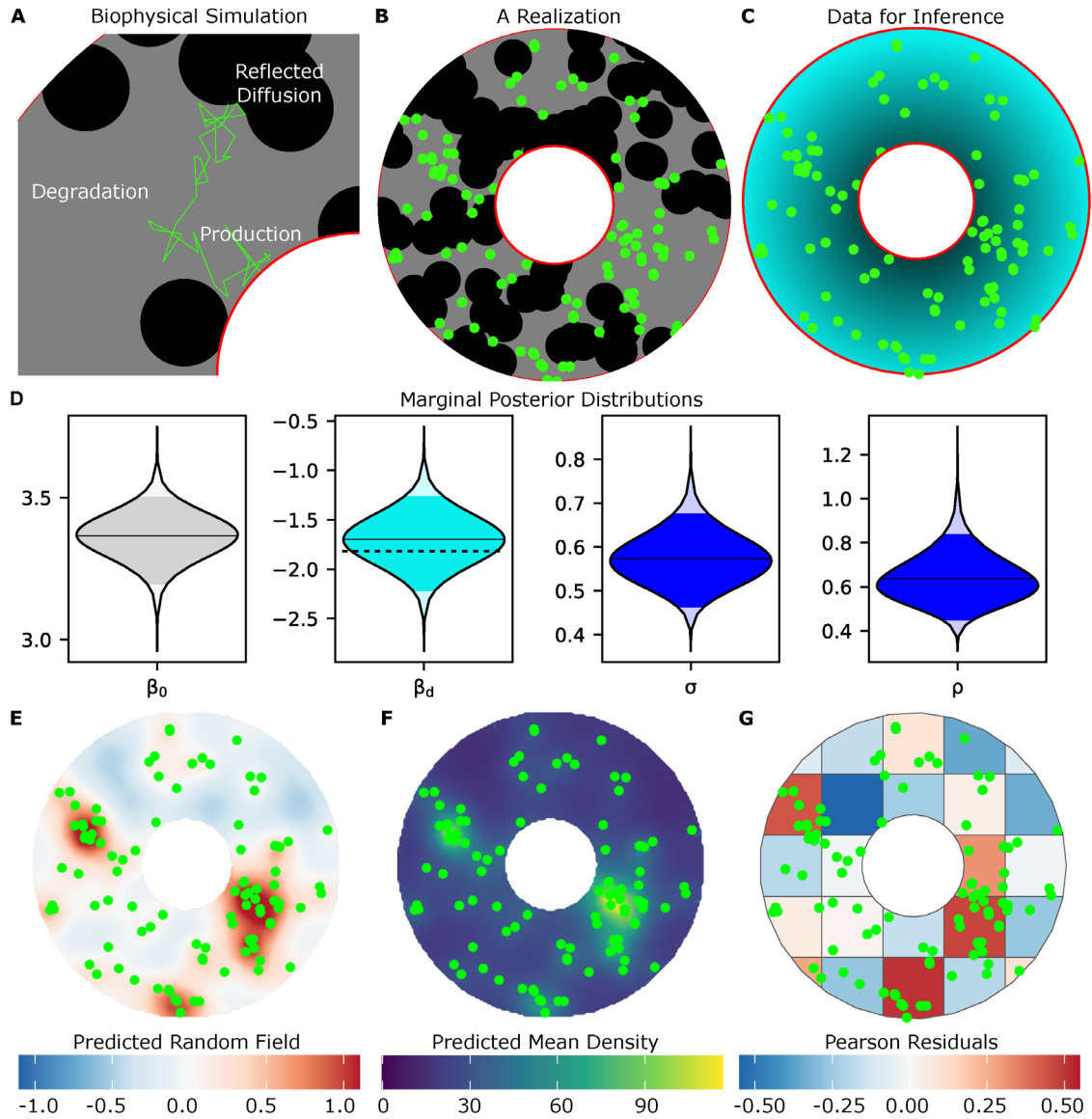
FIGURE 3: A mixed model can quantify the spatial phenotypes of simulated data. (A) Schematic of the biophysical simulation in which particles are produced at the nucleus diffuse through the cell (green trajectory) with obstacles (black disks) and are randomly degraded. (B) A realization of the simulation in the steady state. The particles are shown as green dots and the obstacles are black disks. C: The data used for the inference include the locations of the particles as the puncta (green dots) and the distance from the nucleus (gradient of light blue). (D) The marginal posterior distributions for the model parameter. From left to right: for the weight β_0 of the zero-bias predictive map, for weight β_d of the distance from the nucleus map, for the standard deviation σ of the GRF, and for the correlation length ρ of the GRF. In all distributions, the solid horizontal line indicates the mean, and the light regions are the 0.025 and 0.975 quantiles. In the panel of the marginal posterior distribution for β_d, we indicate the inverse decay length that we theoretically expect with a dashed line. To visualize the GRF (E) and the mean density (F), we drew 100 samples from the respective posterior distributions and determined their mean. We overlaid these averaged samples with the point pattern from the simulation (green dots). The GRF described the heterogeneity of the mean density caused by the obstacles. To quantify the goodness of the fit for the mean density, we determine the Pearson residuals (G). These residuals are between −0.25 and 0.5 well within the standard range of −2 to 2 indicating a good fit.

complete mean density over the cellular domain, we obtain an average of 103 particles with a 95% credible interval from 98 to 105. Furthermore, the mean of the posterior for the distance map coefficient β_d is 1.79 and very close to 1.81, the inverse of the decay length of 0.55. This agreement provides compelling evidence that our statistical model accurately captures what we theoretically expect. The marginal posterior for σ has a mean of around 0.6,

which indicates a $e^{2 \times 0.6} \approx 3$-fold difference in the mean density between hot spots and cold spots. Theoretically, one might expect the mean density in cold spots, which are regions with obstacles, to approach zero. However, given a finite number of particles, the statistical model infers that the density is low in these regions but not exactly zero. The mean of the marginal posterior distribution for the correlation length ρ is around 0.6 and corresponds to a

combined contribution of the multiple characteristic spatial scales introduced by the crowding environment.

In a visualization of the fitted Gaussian random field, we identify that the low-valued patches correspond to areas with many obstacles and the high-valued patches to regions with fewer obstacles, see Figure 3E. The predicted mean density indicates that the model effectively captures the overall spatial variation of the point pattern, see Figure 3F.

To quantitatively assess the predicted mean density, we compute the Pearson residuals, shown in Figure 3G. In the context of Poisson models, these residuals compare the difference between the model-predicted density and an estimate of the density obtained on a coarse mesh of the domain, where the number of particles is divided by the area of each region. Pearson residuals incorporate properties of the Poisson process to regularize the residuals against Poisson variation. Typical residual values between -2 and 2 can be considered acceptable (Baddeley *et al.*, 2015). In our analysis, the Pearson residuals range from -0.25 to 0.5, indicating a very good model fit.

Overall, these results confirm that our model recovers the physical scales of the diffusion process and accurately captures spatial variation in particle density because of the obstacles.

A MIXED MODEL REVEALS THE SPATIAL PHENOTYPE OF PEROXISOMES IN VERO E6 CELLS

For the quantification of the spatial phenotypes of point patterns from experimental data, we apply our method to immunofluorescence images of six Vero E6 cells. In a multiplexed immunofluorescence imaging experiment, we acquired stainings of the nucleus, the ER, mitochondria, the Golgi apparatus, lipid droplets, and peroxisomes, following the iterative indirect immunofluorescence imaging (4i) protocol described in (Kramer *et al.*, 2023), see Figure 4A. We chose peroxisomes as our puncta of interest because of their point-like appearance. After preprocessing the images, we used Bayesian inference to fit a mixed model comprising predictive maps and a GRF to the data.

In the following, we go through the analysis step by step, discussing data preprocessing, model construction, and inference. The analysis notebooks, together with the data, can be found in our GitHub repository (Nicolaou, 2024).

Data preprocessing: Domain definition, location estimation, and construction of predictive maps

First, we define the spatial domain for our statistical model as a two-dimensional slice of the cytoplasm for each cell in the dataset. Therefore, we manually segmented the area of each cell using the fluorescence intensity channel of the ER, as the ER provides a clear boundary of the cell's extent. From these segmented regions, we excluded the nuclei by cropping them out to obtain the final spatial domain for our model, which represents the cytoplasm, Figure 4C. We then estimated the locations of the peroxisomes within the spatial domains by applying the blob detection function of the scikit-image library (Van Der Walt *et al.*, 2014).

Our statistical model includes predictive maps based on other imaged components. As predictive maps, we could directly use the fluorescence intensity of the simultaneously imaged structures, and construct derived maps, such as binary images from them. In addition, we want to construct a predictive map that can account for the dependence of the density of peroxisomes on the distance from the area around the nucleus, known as the perinuclear region, and their scarcity near the cell boundary, see Figure 4C. This organiza-
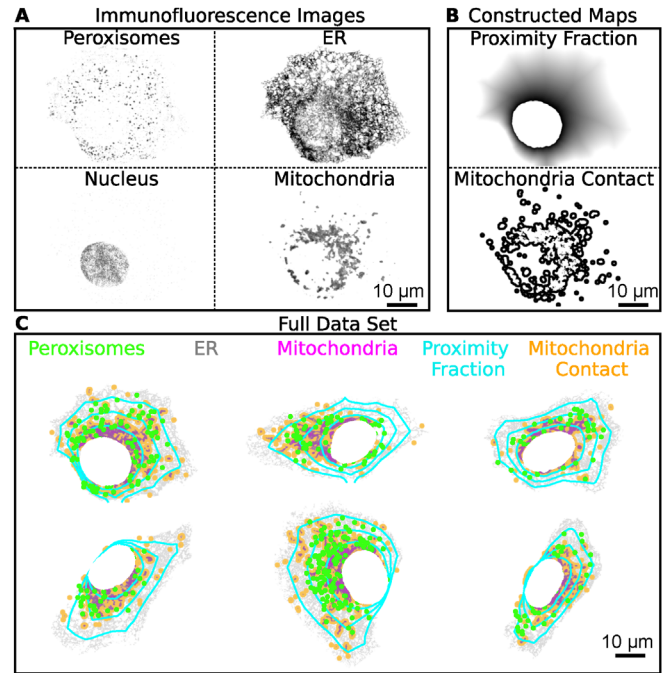

FIGURE 4: Experimental and preprocessed data for the model. (A) Immunofluorescence images of one Vero E6 cell of our dataset, showing the peroxisomes, the nucleus, the ER, and the mitochondria. (B) Constructed predictive maps: the proximity fraction derived from the images of the nucleus and the boundary of the cell; the mitochondria-contact map representing pixels within a distance of $1 \mu m$ from mitochondria. (C) Our full dataset used for the analysis contains six cells with localized peroxisomes (green dots), the ER (gray), mitochondria (magenta), the mitochondria-contact map (orange), and the proximity fraction shown by cyan contour lines at values of 0.25, 0.5, and 0.75 from the periphery to the nucleus, respectively.

tion is a common spatial trend among cellular components, as the perinuclear region is a hot spot for organelle interactions and high metabolic activity. Essentially, on a large length scale, the density of peroxisomes displays a gradient from the nucleus to the boundary. To capture this spatial trend, we construct the following predictive map, which at a location r, combines the nearest distances to the nucleus d_n and the nearest distance to the cell's boundary d_b as $P_{PF}(r) = d_b(r)/(d_b(r) + d_n(r))$. This constructed predictive map represents the proximity fraction and has values from 0 to 1, where $P_{PF}(r_n) \approx 1$ for r_n close to the nucleus and $P_{PF}(r_b) \approx 0$ for r_b close to the cell boundary, see Figure 4, B and C.

An organizational principle of peroxisomes mentioned in the literature is their contact with the mitochondria, which has multiple possible functional purposes in mammalian cells (Fransen *et al.*, 2017). In part of our model, we would like to quantify the contact of peroxisomes with mitochondria. Therefore, we constructed a *contact* map—a binary image that identifies all pixels within a distance of $1 \mu m$ to mitochondria, based on a binarized image of their fluorescence signal. The threshold of one micron corresponds to an approximated maximum size of peroxisomes, ensuring that we capture regions of potential contact between these organelles. This mitochondria contact map which we denote by P_{MC} is shown in Figure 4, B and C. The ER is another important component in the organization of peroxisomes, and while it is possible to construct an ER contact map, its extensive spatial coverage and network-like

structure would result in a map that primarily highlights areas devoid of ER presence. Such regions are already represented in the fluorescent ER images, making an additional ER contact map redundant for our analysis.

Model construction and Bayesian inference

Constructing a meaningful statistical model for the mean density is a model selection problem because many different combinations of predictive maps and GRFs can be considered. By adding more components the model becomes more complex with more free model parameters, and thus is more prone to overfitting. For the scope of this study, we include only the images of the ER, P_{ER}, and mitochondria, P_{MT}, from the multiplex immunofluorescence imaging dataset, leaving the Golgi and lipid droplets aside. As constructed maps, we use the mitochondria contact map P_{MC}, and the proximity fraction P_{PF} to capture the presence in the perinuclear region. We add a Gaussian random field to account for spatial variation that is not described by our predictive maps. Taken together, the mean density of our proposed mixed model reads

$$\text{mean density (r)} = \exp[\beta_0 + \beta_{PF}P_{PF}(r) + \beta_{MT}P_{MT}(r) + \beta_{ER}P_{ER}(r)$$

$$+ \beta_{MC}P_{MC}(r) + \text{GRF}(r; \sigma, \rho)].$$

In our model, we consider the mean density of the peroxisome locations across all cells to be a realization of a single underlying process. In this way, we characterize the entire cell population and use information from multiple cells. Therefore, we combine all the predictive maps and peroxisome point patterns for each cell into a single spatial domain, see Figure 4C.

To set up the inference we construct a mesh of the spatial domain, in which the maximum edge length is set to $1\,\mu m$ to capture the spatial characteristics in our predictive maps. We then specify normal distributions with a standard deviation of 100 as noninformative priors for all the β parameters. For the GRF, we use the penalized complexity prior for which we specify a likely minimum ρ of $1\,\mu m$, which is the typical size of peroxisomes, and a likely maximum σ of 2, which corresponds to a 50-fold difference in the mean density of hot spots and cold spots.

In summary, we processed the imaging data, constructed an informative statistical model based on four predictive maps and one GRF, and finally chose priors for the Bayesian inference. Next, we present and discuss the results that we obtained by fitting our model to the experimental data.

RESULTS

The inference performed with INLA yields approximate marginal posterior distributions for the model parameters, see Figure 5A. The marginal posterior distribution for the coefficient β_0 for the constant zero-bias predictive map is distributed around -4, which corresponds to a mean density of $e^{-4} \approx 0.02\,\mu m^{-2}$.

To identify a significant positive or negative contribution to the mean density, we can check whether the confidence interval for the marginal posterior distributions includes zero. In Figure 5A, we see that the posterior distributions for the proximity fraction, the ER, and the contact to mitochondria are positive, while the contribution of mitochondria is not significant.

The marginal posterior distribution for the coefficient β_{PF} of the proximity fraction centers around 0.6. This value implies that the mean density at locations near the nucleus is $e^{0.6P_{PF}(r_n)}/e^{0.6P_{PF}(r_b)} \approx 2$ times larger compared with a location close to the cell boundary.

The coefficient β_{ER} describes the contribution of the respective predictive map to the mean density. The 0.025 quantile of the marginal posterior distribution for β_{ER} is above zero, which indicates a significant positive correlation of the mean density with the intensity of the ER. Indeed, inspection of the fluorescence images often reveals multiple peroxisomes overlapping with the ER, see Figure 4C. Regarding β_{MT}, its contribution to the mean density is uncertain as its posterior marginal includes zero. However, there is a positive correlation to the proximity to the mitochondria which is captured by the mitochondria-contact map, characterized by β_{MC} whose marginal posterior distribution is centered around 0.5, indicating that the mean density is 1.6 times higher within a distance of $1\,\mu m$ from mitochondria.

The mean values of the posteriors for the parameters that describe the GRF indicate a correlation length ρ of $\sim 15\,\mu m$ and a standard deviation σ of around 0.8. The mean of the random field is estimated by sampling 100 times from its posterior and shown in Figure 5B. The field displays hot spots and cold spots with a size of roughly 10 to $20\,\mu m$, fluctuating mostly between -1 and 1, as suggested by the mean of the marginal posterior of the standard deviation. As discussed before, to rationalize these numbers, it helps to consider a mean density that is only determined by the random field or in which the predictive maps are constant. In this case with the posterior mean of $\sigma \approx 0.8$, the GRF can lead to a $e^{2\sigma} = e^{1.6} \approx 5$-fold difference between the intensity values of the hot spots and cold spots.

When interpreting these numerical values, we considered the predictive maps or the GRF independently of other factors influencing the mean density. We acknowledge that this approach is an oversimplification because the numerical values of the parameters are interdependent and should be considered collectively. The combined effect of all predictive maps and the GRF can be quantified by sampling values from the posterior distribution of parameters to construct the mean density. Ensemble averages across these samples showcase the combined contributions of the effects, see Figure 5C. To quantify the goodness of the model, we determine the Pearson residuals. Some of these residuals fall outside of the acceptable range of -2 to 2, indicating that some regions in the model are not captured very well according to this metric, see Figure 5D. Given that we perform an aggregated fit and there is a cell-to-cell variability that we do not explicitly account for in the model, it is not unexpected that the mean density might come with an error on a single-cell level.

In summary, the posterior distributions quantify the spatial phenotype of peroxisomes in our dataset. We observe a positive correlation with proximity to the nucleus, presence of the ER, and contact with mitochondria. Equally important is the heterogeneity in the mean density, captured by the GRF component, which identifies a length scale of $15\,\mu m$.

DISCUSSION

In this study, we addressed the problem of identifying and quantifying spatial phenotypes of point-like cellular components as part of an integral approach toward uncovering "how" and "why" spatial phenotypes emerge. For point-like structures, we propose using techniques from point-pattern analysis as they proved to be very promising in other fields such as ecology, epidemiology, and geoscience. Moreover, we advocate for the use of random fields to capture hidden random effects, as demonstrated by analyzing the locations of peroxisomes as a point pattern. Our statistical model that combined four predictive maps and one Gaussian random
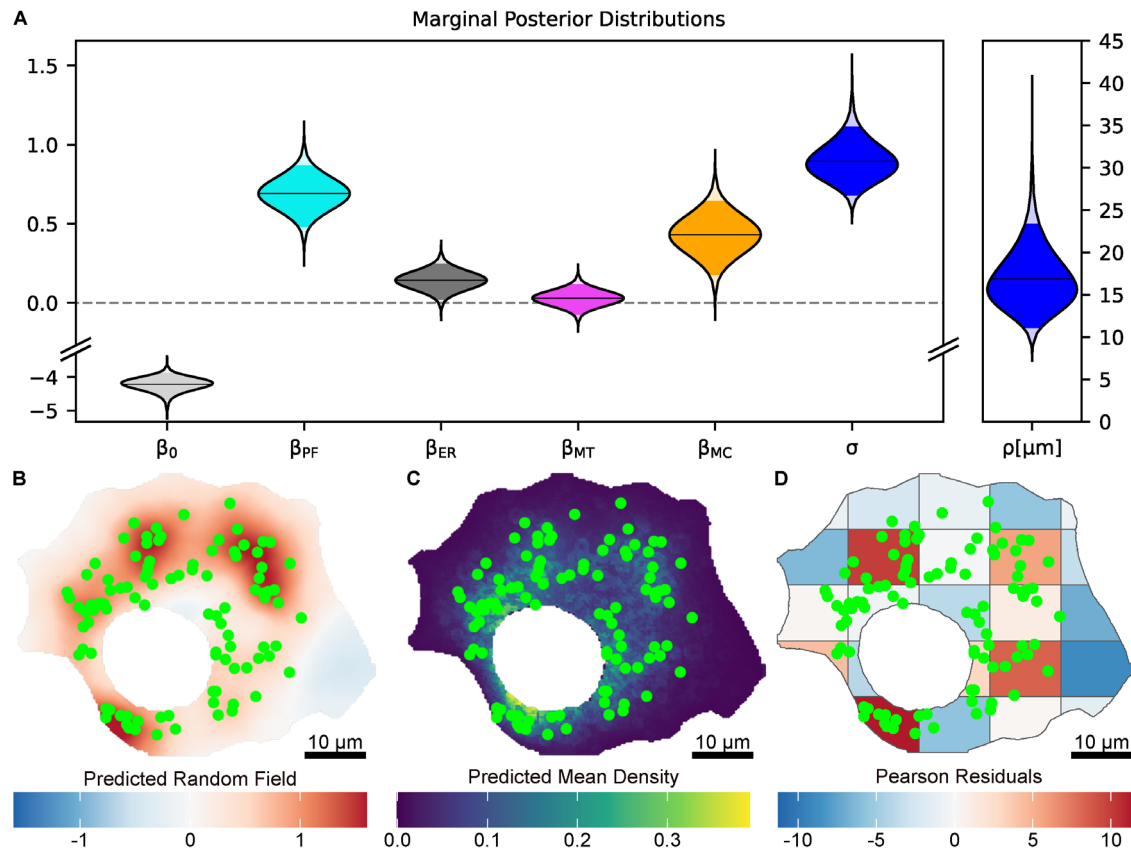
FIGURE 5: Results from Bayesian inference to quantify the spatial phenotypes of peroxisomes in Vero E6 cells. A: The marginal posterior distributions for each model parameter from left to right: the zero-bias predictive map weight β_0, the proximity fraction weight β_{PF}, the ER weight β_{ER}, mitochondria weight β_{MT}, the mitochondria contact map weight β_{MC}, the standard deviation σ of the GRF, and the correlation length ρ of the GRF. In all distributions, the solid horizontal line indicates the mean, and the light regions are the 0.025 and 0.975 quantiles. To illustrate the GRF (B) and the mean density (C), we drew 100 samples from the respective posterior distributions and determined their mean. We overlaid these averaged samples with the point pattern of the peroxisomes (green dots). The GRF described the heterogeneity of the mean density that is not captured by the other components. (D) Pearson residuals of the mean density. Residuals outside the range of −2 to 2 indicate regions where the model may not fit as well.

field uncovered and quantified the following spatial phenotype defined by four features: 1) peroxisomes are likely to be within a distance of $1\,\mu m$ from mitochondria, 2) peroxisomes spatially associate with the ER, 3) there is a higher density of peroxisomes in the region close to the nucleus, and 4) there is a hidden spatial variation in the density in the order of $15\,\mu m$. These results are in line with current literature indicating that peroxisome–ER interaction occurs through direct membrane contact sites, whereby peroxisomes are "wrapped" in ER membrane (Costello *et al.*, 2017), whereas peroxisome-mitochondria contact is considered to be more dynamic and transient (Kumar *et al.*, 2024). The increased presence of peroxisomes close to the nucleus could be a consequence of the higher metabolic activity in the perinuclear region (Wanders *et al.*, 2016).

There are several possible explanations for the hidden spatial variations of $15\,\mu m$. For example, because we have a two-dimensional slice of the cell, intracellular components and structures located in parts of the cell that are not imaged could explain this variation. Additionally, other nonimaged structures, such as the cytoskeleton, could, in principle, also contribute to the organization. Furthermore, time-dependent processes of the past could have influenced the variability of the peroxisomes' locations

at the moment of fixation for imaging. Extending the statistical models discussed here from two-dimensional to three-dimensional and spatiotemporal analogues is nontrivial. Most current methods and packages for LGCPs stem from geostatistics and thus remain oriented toward planar two-dimensional analyses. In contrast, cell biology inherently involves three-dimensional organization and dynamic processes over time, demanding higher-dimensional frameworks. Adapting LGCPs to these settings poses significant computational and theoretical challenges, but doing so could lead to standardized multidimensional inference, potentially providing deeper biological insights.

In our model, we used prior knowledge from cell biology to identify key components involved in the organization of peroxisomes, such as the close proximity to mitochondria. However, in many cases, numerous contributing factors remain unknown and it is necessary to select informative predictive maps from the available data. Often, fluorescence intensities of other imaged structures, binary indicators, and distance maps are available or can be derived. Many different combinations of these predictive maps can be combined into different statistical models. In general, we recommend simplicity in the spirit of Occam's razor, for two reasons: the first one is overfitting, in which too many predictive maps are

used to describe the data, and the inference becomes specific to the dataset. The second is the interpretability of the results, which becomes difficult for models with many parameters. In the absence of prior knowledge regarding which predictive maps are most informative, one can apply the Deviance Information Criterion (DIC) to guide model component selection. The DIC balances model fit against complexity by penalizing the number of effective parameters. In practice, multiple candidate models are fitted, and the one yielding the optimal (typically lowest) DIC value is chosen, thus helping to identify a simple yet explanatory model.

A limitation in the Poisson models is the assumption that the distribution of points is independent of each other given the mean density, implying that there is no interaction between the points. However, in many cases, cellular components organize because of interactions between them. For example, components such as receptors and ion channels often exhibit clustering by interactions, contributing to the nano-organization of synapses, which has important consequences for signaling (Droogers and MacGillavry, 2023). Another class of point processes that incorporate interactions comprises the so-called Gibbs. These models describe spatial configurations with an effective energy function, where the probability of a given configuration depends on the energy. These models are more appropriate, for example, in systems that are known to interact by attraction or exclusion, and they have a direct physical interpretation; for an introduction to these models, see (Baddeley *et al.*, 2015). Ultimately, the choice between a Gibbs model and a Poisson model, such as a LGCP model should be guided by biological insights. Although interactions between puncta are likely always present at short enough lengths, one must consider how significant they are at the spatial scale of interest to justify the added complexity of an interaction-based model.

Standard GRFs are stationary and isotropic processes. Therefore, at first glance, it seems that they must be inadequate to describe the highly nonisotropic organization in cells. However, because we combined them with the predictive maps that account for potential spatial trends, the predicted mean density is nonisotropic. Thus, our model implicitly assumes that the contribution of the GRF to the mean density is stationary and isotropic. Advances in statistics have extended GRFs beyond stationary and isotropic processes by including covariance functions that depend on time, space, and orientation, as well as treating boundary conditions (Fuglstad *et al.*, 2015; Bakka *et al.*, 2019; Lindgren *et al.*, 2024). These refined random fields could potentially improve our descriptions, which is beyond the scope of this study.

Our analysis is based on correlations and therefore a mechanistic interpretation could be problematic. Therefore, we propose combining the quantification of spatial phenotypes with precise experimentation to reveal potential causation. One possible approach is the use of optogenetics to perturb molecular interactions, dynamics, and positioning in space and time (Van Bergeijk *et al.*, 2015; Passmore *et al.*, 2021). Thus, it is possible to change the components under consideration or the components used for the predictive maps. Fitting the same statistical model to imaged cells under different perturbations could reveal potential mechanisms underlying the spatial phenotypes, shedding some light on the complex interplay between organelle dynamics and interactions.

In summary, we propose to include random fields in statistical models to elucidate spatial phenotypes in point-pattern data. This approach seems very promising in the field of cell biology, especially in the context of the intracellular organization of organelles, as we demonstrated with the example of peroxisomes. We believe that such statistical models could serve as a critical link between mechanistic biophysical models and the rich imaging data by modern microscopy. Statistical models that include random fields account for intracellular stochasticity and cell-to-cell variability—a key step toward a better understanding of the principles behind the remarkable spatial organization in living cells.

CODE AND DATA AVAILABILITY

All experimental and synthetic datasets, as well as the code used for preprocessing, simulation, and Bayesian inference, are publicly available in the form of step-by-step notebooks. These notebooks are designed to guide users through each stage of the analysis workflow, facilitating easy reproduction of the results. The entire repository can be accessed at https://github.com/kyriacosn/spatial-phenotypes-of-puncta. Additional details, including example usage instructions and documentation, are provided within the repository.

REFERENCES

Bachl FE, Lindgren F, Borchers DL, Illian JB (2019). inlabru: An R package for Bayesian spatial modelling from ecological survey data. Methods Ecol Evol 10, 760–766.

Baddeley A, Rubak E, Turner R (2015). *Spatial Point Patterns: Methodology and Applications with R*. Boca Raton, Florida: Chapman and Hall/CRC.

Baddeley A, Turner R (2005). spatstat: An *R* package for analyzing spatial point patterns. J Stat Softw 12, 1–42.

Bakka H, Vanhatalo J, Illian JB, Simpson D, Rue H (2019). Non-stationary Gaussian models with physical barriers. Spat Stat 29, 268–288.

Chiu,SN, Stoyan D, Kendall WS, Mecke J (2013). *Stochastic Geometry and its Applications*. Hoboken, New Jersey: Wiley.

Costello JL, Castro IG, Hacker C, Schrader TA, Metz J, Zeuschner D, Azadi AS, Godinho LF, Costina V, Findeisen P, *et al.* (2017). ACBD5 and VAPB mediate membrane associations between peroxisomes and the ER. J Cell Biol 216, 331–342.

D'Angelo N, Siino M, D'Alessandro A, Adelfio G (2022). Local spatial log-Gaussian Cox processes for seismic data. AStA Adv Stat Anal 106, 633–671.

Diggle PJ (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns,*. Hoboken, NJ, USA: Chapman and Hall/CRC.

Droogers WJ, MacGillavry HD (2023). Plasticity of postsynaptic nanostructure. Mol Cell Neurosci 124, 103819.

Elbaz Y, Schuldiner M (2011). Staying in touch: The molecular era of organelle contact sites. Trends Biochem Sci 36, 616–623.

Fransen M, Lismont C, Walton P (2017). The peroxisome-mitochondria connection: How and why? Int J Mol Sci 18, 1126.

Fuglstad G-A, Lindgren F, Simpson D, Rue H (2015). Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. Stat Sin 25, 115–133.

Fuglstad G-A, Simpson D, Lindgren F, Rue H (2019). Constructing priors that penalize the complexity of Gaussian random fields. J Am Stat Assoc 114, 445–452.

Horgan GW, Ball BC (1994). Simulating diffusion in a Boolean model of soil pores. Eur J Soil Sci 45, 483–491.

Illian JB, Martino S, Sørbye SH, Gallego-Fernández JB, Zunzunegui M, Esquivias MP, Travis JMJ (2013). Fitting complex ecological point process models with integrated nested Laplace approximation. Methods Ecol Evol 4, 305–315.

Khater IM, Nabi IR, Hamarneh G (2020). A Review of super-resolution single-molecule localization microscopy cluster analysis and quantification methods. Patterns 1, 100038.

Kramer B, Castillo J, Pelkmans L, Gut G (2023). Iterative indirect immunofluorescence imaging (4i) on adherent cells and tissue sections. Bio Protoc 13, e4712.

Kumar R, Islinger M, Worthy H, Carmichael R, Schrader M (2024). The peroxisome: An update on mysteries 3.0. Histochem Cell Biol 161, 99–132.

Levet F, Julien G, Galland R, Butler C, Beghin A, Chazeau A, Hoess P, Ries J, Giannone G, Sibarita J-B (2019). A tessellation-based colocalization

analysis approach for single-molecule localization microscopy. Nat Commun 10, 2379.

Li Y, Brown P, Gesink DC, Rue H (2012). Log Gaussian Cox processes and spatially aggregated disease incidence data. Stat Methods Med Res 21, 479–507.

Lindgren F, Bakka H, Bolin D, Krainski E, Rue H (2024). A diffusion-based spatio-temporal extension of Gaussian Matérn fields. SORT 48, 3–66.

Lindgren F, Rue H, Lindström J (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. J R Stat Soc Ser B Stat Methodol 73, 423–498.

Matérn B (1960). Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations. Medd Fran Statens Skogsforskningsinstitut 49(5), 1–144.

Møller J, Syversveen AR, Waagepetersen RP (1998). Log Gaussian Cox processes. Scand J Stat 25, 451–482.

Nicolaou K (2024). Spatial phenotypes of puncta. Available at: https://github.com/kyriacosn/spatial-phenotypes-of-puncta. Accessed December 28, 2024.

Nicovich PR, Owen DM, Gaus K (2017). Turning single-molecule localization microscopy into a quantitative bioanalytical tool. Nat Protoc 12, 453–460.

Passmore JB, Nijenhuis W, Kapitein LC (2021). From observing to controlling: Inducible control of organelle dynamics and interactions. Curr Opin Cell Biol 71, 69–76.

Phillips MJ, Voeltz GK (2016). Structure and function of ER membrane contact sites with other organelles. Nat Rev Mol Cell Biol 17, 69–82.

Raynaud X, Nunan N (2014). Spatial ecology of bacteria at the microscale in soil. PLoS One 9, e87217.

Rue H, Held L (2005). *Gaussian Markov Random Fields*. Boca Raton, Florida: Chapman and Hall/CRC.

Rue H, Martino S, Chopin N (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. J R Stat Soc Ser B Stat Methodol 71, 319–392.

Scholz C, Wirner F, Klatt MA, Hirneise D, Schröder-Turk GE, Mecke K, Bechinger C (2015). Direct relations between morphology and transport in Boolean models. Phys Rev E 92, 043023.

Simpson D, Illian JB, Lindgren F, Sørbye SH, Rue H (2016). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. Biometrika 103, 49–70.

Simpson D, Rue H, Riebler A, Martins TG, Sørbye SH (2017). Penalising model component complexity: A principled, practical approach to constructing priors. Stat Sci 32, 1–28.

Sourjik,V (2004). Receptor clustering and signal processing in E. coli chemotaxis. Trends Microbiol 12, 569–576.

Taylor BM, Davies TM, Rowlingson BS, Diggle PJ (2013). lgcp : An R package for inference with spatial and spatio-temporal log-Gaussian Cox processes. J Stat Softw 52, 1–40.

Teng M, Nathoo F, Johnson TD (2017). Bayesian computation for log-Gaussian Cox processes: A comparative analysis of methods. J Stat Comput Simul 87, 2227–2252.

Thomas M (1949). A generalization of Poisson's binomial limit for use in ecology. Biometrika 36, 18.

Van Bergeijk P, Adrian M, Hoogenraad CC, Kapitein LC (2015). Optogenetic control of organelle transport and positioning. Nature 518, 111–114.

Van Der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T (2014). scikit-image: Image processing in Python. PeerJ 2, e453.

Velázquez E, Martínez I, Getzin S, Moloney KA, Wiegand T (2016). An evaluation of the state of spatial point pattern analysis in ecology. Ecography 39, 1042–1055.

Wanders RJA, Waterham HR, Ferdinandusse S (2016). Metabolic interplay between peroxisomes and other subcellular organelles including mitochondria and the endoplasmic reticulum. Front Cell Dev Biol 3, 83.

Wu Y-L, Tschanz A, Krupnik L, Ries J (2020). Quantitative data analysis in single-molecule localization microscopy. Trends Cell Biol 30, 837–851.