Investigating Ecological Interactions in the Tumor Microenvironment Using Joint Species Distribution Models for Point Patterns

JOEL ELIASON* AND ARVIND RAO*

Abstract

The tumor microenvironment (TME) is a complex and dynamic ecosystem that involves interactions between different cell types, such as cancer cells, immune cells, and stromal cells. These interactions can promote or inhibit tumor growth and affect response to therapy. Multitype Gibbs point process (MGPP) models are statistical models used to study the spatial distribution and interaction of different types of objects, such as the distribution of cell types in a tissue sample. Such models are potentially useful for investigating the spatial relationships between different cell types in the tumor microenvironment, but so far studies of the TME using cell-resolution imaging have been largely limited to spatial descriptive statistics. However, MGPP models have many advantages over descriptive statistics, such as uncertainty quantification, incorporation of multiple covariates and the ability to make predictions. In this paper, we describe and apply a previously developed MGPP method, the *saturated pairwise interaction Gibbs point process model*, to a publicly available multiplexed imaging dataset obtained from colorectal cancer patients. Importantly, we show how these methods can be used as joint species distribution models (JSDMs) to precisely frame and answer many relevant questions related to the ecology of the tumor microenvironment.

KEYWORDS AND PHRASES: Tumor microenvironment, Spatial point patterns, Joint species distribution models.

1. INTRODUCTION

Similar to natural ecosystems, the TME comprises a dynamic web of relationships, where cancer cells, immune populations, stromal cells, and various molecular cues engage in complex communication and interdependence [1, 2]. Furthermore, just as environmental conditions influence species behavior in ecological systems, the TME's microenvironmental factors, including nutrient availability, oxygen levels, and inflammatory signals, play a pivotal role in shaping cellular behaviors, proliferation, migration, and responses to therapy [3–5]. These interactions exert pivotal influences on tumor progression and therapeutic outcomes [6–9]. This ecosystem-like framework underscores the TME's complexity and underscores the need to employ ecological modeling approaches to gain deeper insights into its intricate dynamics and therapeutic implications [10–14].

The advent of highly multiplexed imaging technologies that allow for the simultaneous measurement of dozens of cell-surface protein biomarkers in tissue samples has allowed for the creation of datasets that have precise spatial coordinates and cell-type annotations for each cell within a tissue [15]. Such multiplexed imaging technologies are of particular relevance to studies of the tumor microenvironment, in which multiplexed images can be used to ascertain the complex interrelationships between cell types and cell states [16, 17]. Such imaging technologies are similar to the remote sensing technologies employed in ecological studies, which allow the localization of various species and environmental covariates in space.

Spatial descriptive statistics are fundamental tools employed in such ecological studies to gain insights into the spatial patterns and relationships within natural systems. These statistics provide a quantitative framework for characterizing the arrangement, dispersion, and clustering of species or individuals across a landscape. Metrics such as nearest neighbor distances, spatial autocorrelation, and Ripley's K-function are commonly used to assess the degree of spatial aggregation or dispersion of species within a given habitat [18–21]. By analyzing these patterns, ecologists can infer crucial information about species interactions, habitat preferences, and the influence of environmental factors on distribution. Such statistics have also been widely adopted by researchers investigating the tumor microenvironment [22–24]. More bespoke methods have also been developed, such as those in which a "neighborhood composition" vector is calculated for each cell [16, 25–27]. Other relatively simple statistics, such as quantifying the amount of immune infiltration based on the Immunoscore metric, in various regions of the TME, are also widely used in clinical models [28, 29].

While spatial descriptive statistics offer essential prelimi-

*Corresponding authors.

nary insights into ecological spatial patterns, more advanced techniques, such as Joint Species Distribution Models (JSDMs), extend this analysis by capturing the complex interactions and dependencies that characterize ecological systems in a more holistic manner [30–34]. In particular, JSDMs are able to model the spatial distributions of multiple species simultaneously, while accounting for the associations between each of these species, as well as their association with environmental variables. Within the context of the TME, these models can enable researchers to identify cell type-cell type or cell type-environment associations while adjusting for other associations, predict cell type distributions, and quantify the uncertainty of model parameters. Such capabilities are typically not available for spatial descriptive statistics.

Point process models are another widely used tool for modeling ecological relationships [21, 35–39]. Such models make explicit the probabilistic data-generating process that is assumed to generate an observed pattern of points, and thus are amenable to uncertainty quantification and producing predictions [18]. A particular class of point process model, called multitype Gibbs point process (MGPP) models, are able to model the pairwise cell type interactions jointly [18, 40–42], along with cell type/environment interactions and thus can be used as JSDMs in the context of the TME ecosystem. Gibbs point process models explicitly model the attraction or inhibition between different points in space; MGPPs generalize this to when there are several types of points in space, such as different cell types [18].

More specifically, modeling interactions via MGPP models can improve on the aforementioned spatial descriptive statistics in a number of ways [18]: a) spatial variables, such as the concentration levels of various inflammatory factors, and non-spatial variables, such as organ of origin or patient demographics, can be easily adjusted for, allowing for the modeling of inhomogeneous distributions of cells; b) the joint modeling of pairwise interactions between cell types can be achieved, as opposed to the case of descriptive statistics, in which case only the marginal distribution of interactions can be estimated; c) uncertainty quantification of interaction effects can be performed; and d) predictions and residuals can be obtained from the model, allowing for the assessment of model fit, as well as the prediction of distributions of cells under new conditions.

Some examples of relevant questions that JSDMs (and thus MGPP models, in this context) can answer, in a general setting and specific to the TME, are found in Table 1. In order to motivate the utility of this model, in this paper we will demonstrate an analysis framework that could be used to answer many of the questions from Table 1. The questions in Table 1 are asked either at the image level (that is, they can be answered after a JSDM model is fit on a particular image) or at the cohort level (they are answered after fitting JSDM models for each image in a cohort, extracting parameters from the fitted models and then using those parameters as inputs for further modeling and analysis). The third column of Table 1 indicates at which level each question can be answered.

2. METHODS

2.1 Dataset

The dataset used to showcase the MGPP method comprises multiplexed images of tumor tissue samples obtained from a publicly available repository for colorectal cancer [16]. The images capture different cell types within the tumor microenvironment, including cancer cells, immune cells, and stromal cells. In the original study, preprocessing was conducted on the images. This involved cell type classification and determination of spatial coordinates for each cell type.

Specifically, this dataset consists of 4 multiplexed images of 56 different cell-surface markers of tumor tissue sections from each of 35 CRC patients, with patients having varying degrees of disease progression and severity. These images were obtained via the PhenoCycler imaging technology (formerly known as CODEX) [17, 43]. The patients are divided into two groups, based on the histopathology of their tumors: those in which tertiary lymphoid structures (TLSs) are visible at the tumor invasive front, and those where diffuse inflammatory infiltration (DII) was present and no TLSs were visible. These patient groups will be subsequently referred to as CLR (Crohn's-like reaction) and DII, with CLR patients generally having better survival prognosis [16]. 16 cell types have been annotated using the 56 different cell-surface markers and in total, the dataset consists of 244,504 annotated cells. In addition, patient metadata is included, including overall/progression-free survival times, tumor grade, and demographic variables such as age and sex. An example image can be seen in Figure 1.

2.2 Multitype Gibbs Point Process (MGPP) Models

Once cell types have been annotated, a multiplex image, such as in Figure 1, can be viewed as a multitype point pattern, or a realization **x** of a particular random variable **X** called a *multitype point process*. Subsequently, a MGPP model can be fit to this multitype point pattern. In the following, we will describe first a general *pairwise interaction* MGPP model, and subsequently the *saturated pairwise interaction* MGPP model.

Mathematically, a pairwise interaction MGPP model characterizes the probability density $f(\mathbf{X})$ of a multitype point process **X** in a domain $D \in \mathcal{R}^2$ in the following way [40], with i and j indexing over the points in **x**:

$$f(\mathbf{X} = \mathbf{x}) = C \exp\left(\left[\sum_{i=1}^{n(\mathbf{x})} \sum_{k=1}^{K} \beta_{k,m_i} X_k(x_i) \right] + \left[\sum_{i<j}^{n(\mathbf{x})} g_{m_i,m_j}(x_i, x_j) \right] \right)$$

(2.1)

Table 1. *Examples of questions that can be asked using MGPP models as JSDMs.*

Example of question in a general setting	Specific to the TME	Image or Cohort level
Are different cell types *positively* or *negatively correlated* in space, after adjusting for the presence of other cell types and spatial covariates?	Are B cells and granulocytes positively or negatively correlated in this patient, after controlling for the effects of other cells?	Image level
How do cell type interactions change across different spatial scales?	At which spatial scales are Tregs repulsive of one another, and at which are they attractive?	Image level
Which cell type or spatial covariate is the best predictor of another cell type at each location in an image?	Which cell type best predicts the locations of tumor cells in the tumor core? How about in stromal regions?	Image level
Given some cell types and spatial covariates in an image, how well can we predict the distribution of a new cell type, using a model fit on other images?	Can we correctly predict the distribution of M2 macrophages in an image where we did not mark for M2 macrophages?	Image level (but may require model fits from other images)
Can the estimated interaction parameters predict patient outcomes?	How well does the interaction between CD8+ T cells and granulocytes predict patient survival?	Cohort level
Can we predict which distributions of cell types are associated with the best tumor responses?	Which distributions of immune cell types are associated with a near absence of tumor cells?	Image level
How do changes in spatial covariates change the distribution of different cell types?	How might changes in oxygenation or angiogenesis affect the distribution of tumor cells?	Image level
Is there evidence of two cell types coexisting in the same metabolic niche? Is the nature of their interaction competitive or mutualistic?	Do immune cells and tumor cells compete for the same metabolic factors? In which areas of the tumor?	Image level
How do cell type interactions compare across images obtained from the same patient?	How variable are the interaction parameters between B cells/CD4+ T cells across images within the same patient among patients with low survival probability?	Cohort level

Here, C indicates a normalizing constant, while the first term in brackets signifies the "first-order" effects on the density of each cell type, that is, the effects from spatial and non-spatial variables. There are K such variables (the $X_k(x_i)$) per cell type. These variables are typically used to model "environmental variables"; in the context of the TME, this could include such variables as the concentration levels of various metabolic or inflammatory factors, or the distance to the nearest blood vessel. The second bracketed term contains the pairwise interaction functions $g_{m_i,m_j}(x_i, x_j)$, which are functions of two points x_i and x_j of types m_i and m_j. These interaction functions are often parameterized with interaction parameters θ_{m_i,m_j}, which typically quantify the direction and magnitude of interaction between points of type m_i and m_j. For example, in the multitype Strauss model, the interaction function $g_{m_i,m_j}(x_i, x_j)$

is written as:

$$g_{m_i,m_j}(x_i, x_j) = \begin{cases} 0, & \text{if } ||x_i - x_j|| > R_{m_i,m_j} \\ \theta_{m_i,m_j}, & \text{if } ||x_i - x_j|| \leq R_{m_i,m_j} \end{cases} \quad (2.2)$$

for user-defined *interaction radii* R_{m_i,m_j}. Here note that the distance function $||x - y||$ between two points $x, y \in \mathbf{x}$ typically denotes the Euclidean distance between the spatial coordinates of the corresponding points. The multitype Strauss model assumes that interactions between cells of types m_i and m_j do not exist if cells are farther than R_{m_i,m_j} apart, and are constant if they are within that radius.

One drawback of many pairwise interaction MGPP models, such as the multitype Strauss model, is that they are only able to model repulsion (and not clustering) between points – that is, that $\theta_{m_i,m_j} < 0$. This has led to the de-
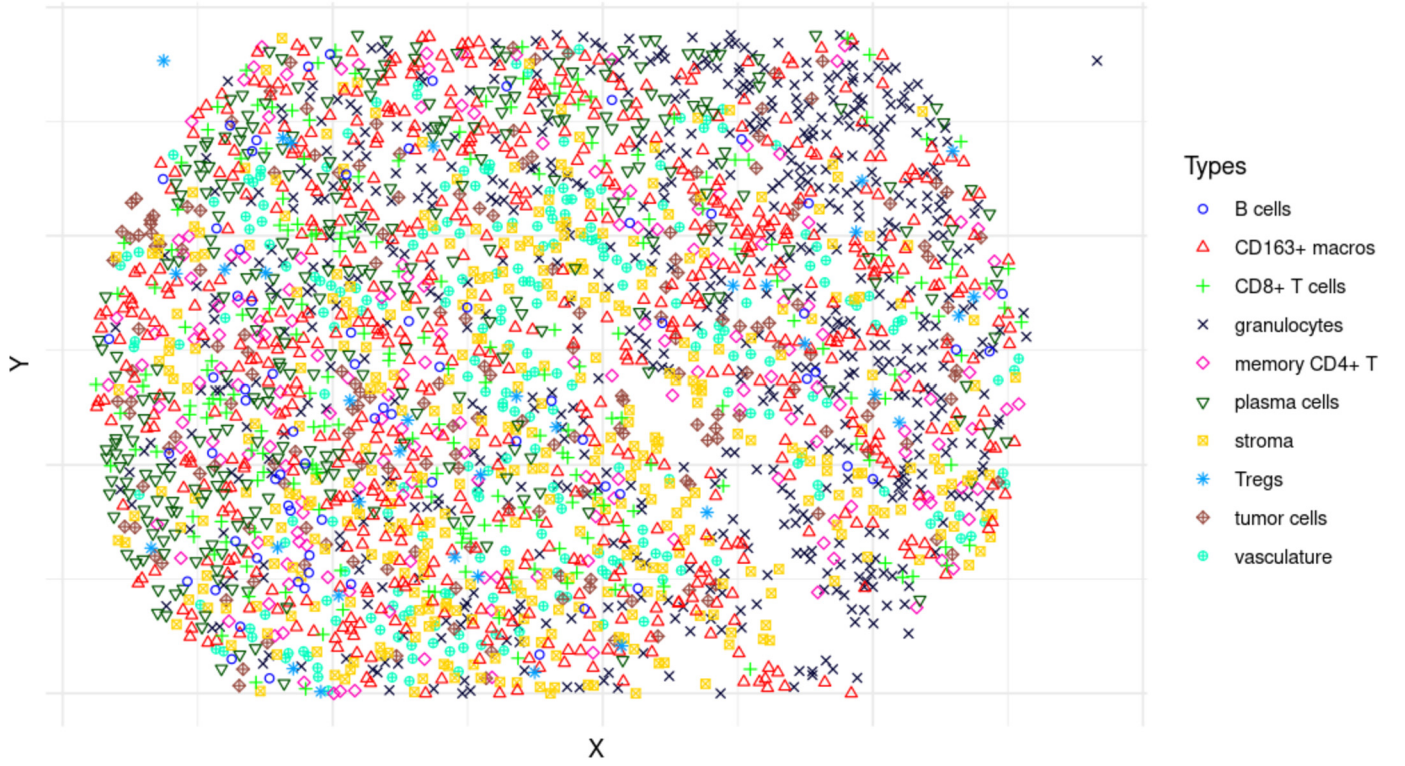
Figure 1: A point pattern representation of a multiplex image of a tissue section of the TME in a CRC patient.

velopment of MGPP models that do allow clustering, such as the *Geyer saturation* model [44] or the *saturated pairwise interaction* MGPP model [42]. For example, in the saturated pairwise interaction Gibbs point process, as in [42] and borrowing some of their notation, we have

$$
\begin{aligned}
f(\mathbf{x}) =& C \exp\left[\sum_{x\in\mathbf{x}}\left(\beta_{0,t(x)} + \sum_{k=1}^{K}\beta_{k,t(x)}X_k(x)\right)\right.\\
&+ \sum_{t_2=1}^{p}\sum_{z\in\mathbf{x}_{t_1}}\alpha_{t_1,t_2}u\left(z,(\mathbf{x}\backslash\{z\})_{t_2}\right)\\
&+ \left.\sum_{t_2=1}^{p}\sum_{z=\in\mathbf{x}_{t_1}}\gamma_{t_1,t_2}v\left(z,(\mathbf{x}\backslash\{z\})_{t_2}\right)\right]
\end{aligned}
\tag{2.3}
$$

with

$$
u(x,\mathbf{x}_{t_2}) = \max_{\eta\in S(\mathbf{x}_{t_2},N)}\sum_{z\in\eta}\phi(||x-z||; R^S_{t(x),t_2})
\tag{2.4}
$$

and v defined similarly as

$$
v(x,\mathbf{x}_{t_2}) = \max_{\eta\in S(\mathbf{x}_{t_2},N)}\sum_{z\in\eta}\psi(||x-z||; R^M_{t(x),t_2})
\tag{2.5}
$$

Here, t_1 and t_2 can take on values in $\{j\}_{j=1}^{p}$ and denote the jth cell type of p cell types, and $t(x)\in\{j\}_{j=1}^{p}$ denotes the type of point $x\in\mathbf{x}$. Note that $t(x_i) = m_i$, as in our

notation in Equation 2.1. Functions ϕ and ψ represent user-defined short- and medium-range *potential functions* which set the shape of short- and medium-range interactions, and are dependent only on the distance between two points. ϕ and ψ are commonly functions such as the exponential or step functions (Figure 2). $R^S_{t_1,t_2}$ and $R^M_{t_1,t_2}$ are user-defined short- and medium-range interaction radii, and the interaction parameters of interest, α_{t_1,t_2} and γ_{t_1,t_2}, are interpreted as the strength and direction of interaction between any two types of points at short and medium ranges. $\mathbf{x}_{t_i}$ denotes the set of points in $\mathbf{x}$ that have type t_i, while $S(\mathbf{x}_{t_i},N)$ is the set of all subsets of points from $\mathbf{x}_{t_i}$ that are of size N – thus, u and v can be interpreted as the sum of the N largest potentials that point x has with points $z\in\mathbf{x}$ of type t_2. Summing only the top N potentials (rather than all of them) is what leads to this being called a *saturated* MGPP model, and the parameter N is called the *saturation parameter* accordingly. Because of its great flexibility in being able to model clustering and repulsion, the saturated pairwise interaction MGPP model will be the one we will be using for the remainder of the paper as the JSDM. It should be noted, furthermore, that currently MGPP models, like the saturated pairwise MGPP model we describe above, are limited to modeling interactions between pairs of points. While higher-order interaction *single-type* Gibbs models do exist, such as the triplet interaction process [44] and the area-interaction process [45], from the authors' perspective,
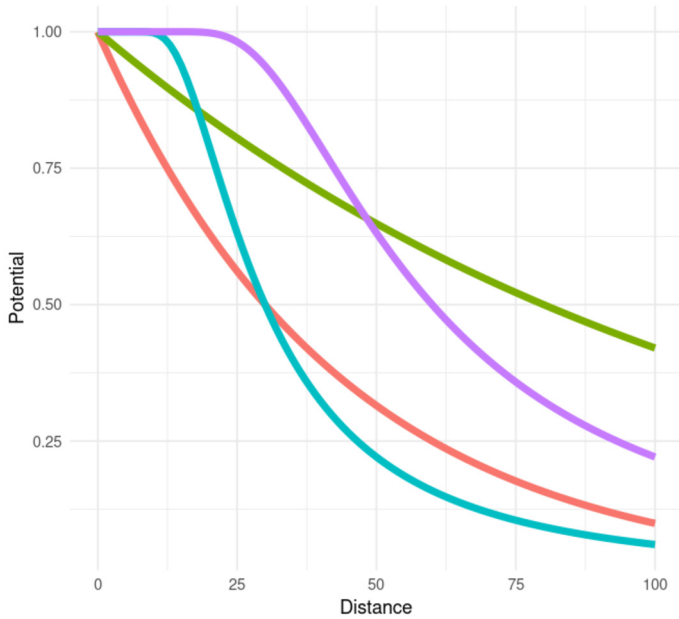
Figure 2: Examples of potential curves commonly used for ϕ and ψ in Equations 2.4 and 2.5.

there has not been any work yet to generalize higher-order interaction processes to *multitype* data.

Once the probability density $f(\mathbf{X})$ of a MGPP model has been defined, estimating the parameters of this model can proceed via standard statistical techniques such as maximum likelihood estimation. However, because this is often quite computationally costly (due to the often intractable normalizing constant C), estimation techniques based on the *conditional intensity* $\lambda(y|\mathbf{x})$ are often preferred. The *intensity* $\lambda(y)$ of a point pattern at a location $y \in D$ is the expected number of points falling in a small window around y, divided by the area of that small window. The *conditional intensity* defined at a spatial location y is defined as $\lambda(y|\mathbf{x}) = \frac{f(\mathbf{x} \cup y)}{f(\mathbf{x} \setminus y)}$, and can be interpreted as the "intensity at y given the rest of the spatial pattern $\mathbf{x}$" [18]. Using the conditional intensity, we can then produce predictions from our MGPP model and subsequently assess model fit. For the specific formulation of the conditional intensity for the saturated pairwise interaction Gibbs point process model, see [42].

In case there are cell types with very low representation, there are a few possible strategies. If an analyst is not interested in interactions involving rare cell types, those cell types can simply be removed from the pattern or lumped into a separate category, as we do in the example analysis below. However, if the analyst is interested in rare cell-type interactions, it is possible to fit a multilevel model across multiple images simultaneously, in order to pool the estimates of interaction strength across multiple images simultaneously. Such a strategy is used to model spatiotemporal

point pattern data in [46, 47], and similar methods could be applied in this case as well.

2.3 An Example Analysis

Here we demonstrate an example analysis to answer a subset of the questions from Table 1. In this example, we jointly model the interactions between sufficiently abundant cell types in each image. Furthermore, we include as a spatial covariate the distance from each point to the nearest blood vessel, as a proxy for angiogenesis and oxygenation found in the TME. Specifically, we set out to answer the following:

1. Which cell types interact with one another, after adjusting for the interactions of other cell types and distance to blood vessels? Which cell types stay away from one another, again, after adjusting for the interactions of other cell types and the distance to blood vessels? What confidence can I have in individual parameter estimates?

   (a) Do these interactions occur at close or far spatial scales?

   (b) Which cell type most strongly uniquely interacts with tumor cells, on average, across the whole image?

2. Which cell types best predict the distribution of tumor cells at each location in space?

3. How much is the distribution of each cell type uniquely predicted by its distance to the nearest blood vessel?

4. What is the difference in interaction strength between each pair of cell types, comparing between the two patient groups (CLR and DII patients)?

5. What levels of intra- and inter-patient heterogeneity are there in how different cell types interact?

6. Are any of these parameter estimates useful clinically for predicting patient outcomes?

7. How much can I trust the results from this model? i.e., how well does it predict spatial distributions of cells?

The code to reproduce the example analysis that follows can be found at https://github.com/jeliason/mgpp_TME/.

2.3.1 Fitting MGPP models

We employed a systematic approach in fitting the MGPP models to the images from the CRC dataset. We focused our analysis on cell types that exhibited a minimum of 20 cells within a given image, ensuring that we focused on sufficiently populated cell populations to enable reliable modeling of interactions. Next, we identified regions devoid of cells and incorporated this information as a spatial covariate within the model. By accounting for these regions, we hoped to avoid any bias in model fitting from assuming all areas in the spatial window were equally likely to contain cells. We further incorporated a spatial covariate of "distance to nearest blood vessel", as a proxy of angiogenesis
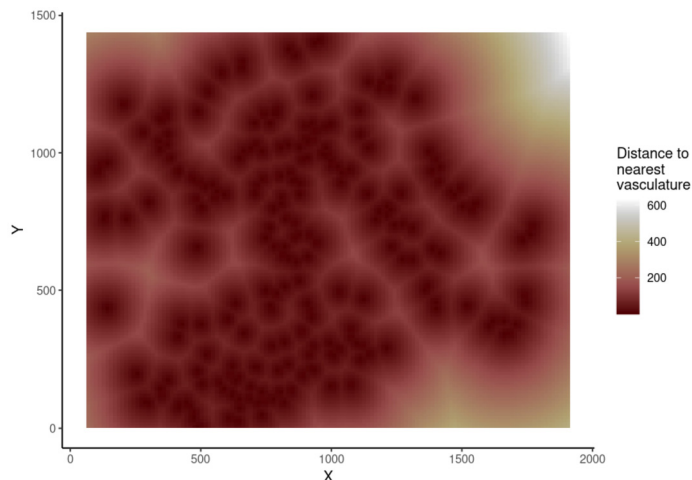
Figure 3: Spatial covariate of "distance to nearest blood vessel", as imaged in spot 59_A.

and oxygenation throughout the TME. An example of this spatial covariate can be seen in Figure 3.

The fitting of the MGPP models was performed using the `ppjsdm` package, an R package tailored for fitting saturated pairwise interaction Gibbs point process models [42]. Specifically, we employed an exponential interaction function to model the pairwise interactions between cell types at both short and medium ranges (that is, the functions ϕ and ψ in Equations 2.4 and 2.5). The interaction radii were set at 30, 70, and 150 microns, reflecting short-, medium- and long-range interaction distances, respectively. A saturation parameter of $N = 5$ was set. Lastly, $N_d = 5,000$ dummy points were included per cell type. The dummy points are related to the inferential procedure for parameter estimation. Briefly, N_d dummy points per cell type are distributed across the image and interactions between dummy points and real points are calculated, according to Equations 2.4 and 2.5. This way of formulating the problem allows one to use fast, optimized inferential procedures developed for logistic regression. Specifically, at each point (dummy or data), the interactions with each cell type at that point are calculated, as well as the associations with spatial covariates. Then, a logistic regression is fitted in which the interactions are the independent variables and the response variable is binary, coded 1 for data points and 0 for dummy points. Increasing N_d will decrease the approximation error of this inferential technique up to a point, and an estimate of the proportion of the standard error of each coefficient that is due to this approximation can additionally be calculated. For more information, the interested reader is referred to [42, 48].

Using 2 cores on a laptop, fitting of the models on these datasets took between 30 seconds–2 minutes per model, while estimating the standard errors for each parameter (a separate step) took 3–10 minutes per model. Memory costs for the fitting of these models are also not extravagant, typically only taking 1–2 GB.

2.3.2 Interpretation of Model Parameters and Predictions

After fitting the models, we extracted the fitted interaction parameters α_{t_1,t_2}, γ_{t_1,t_2} and $\beta_{1,t}$ (corresponding to the effect that distance to nearest blood vessel had on the distribution of each cell type t), along with the respective confidence interval for each parameter. We then plotted these parameters for one image as heatmaps, in order to demonstrate how to answer Questions 1 and 3. We also plotted the exponential potential functions scaled by each interaction parameter, in order to demonstrate how to answer Question 1a. We identified the cell type with the highest magnitude interaction parameter with tumor cells in each image, to answer Question 1b.

Next, we showed how one can answer Question 2 by identifying the cell type that has the largest magnitude contribution to the linear predictor of the conditional intensity of tumor cells at each point in space.

Lastly, we extracted the interaction parameters from all fitted models and plotted heatmaps of the difference in median interaction between the two patient groups, allowing us to answer Question 4.

2.3.3 Heterogeneity in Model Parameter Estimates Within Patient Groups and Across Patients

In order to answer Question 5, we calculated the median absolute deviation (MAD) of each interaction parameter at two levels: pooled within each patient group, as well as across the whole patient cohort, and plotted the MAD estimates as heatmaps. This allows us to identify interactions that are highly variable and likely to be more context-dependent, as well as those that remain more consistent and homogeneous.

2.3.4 Predictive Modeling of Patient Outcomes

We employed multivariate Cox proportional hazards models to estimate hazard ratios for interaction effects derived from the fitted MGPP models, in order to answer Question 6. Briefly, the interaction parameters for a given pair of cell types and spatial scale were extracted and used as a predictor of patient survival in a Cox model, while adjusting for the age and sex of each patient, as in Equation 2.6.

$$\lambda(t|\alpha_{ijk}) = \lambda_0(t)\exp(\beta_{0,jk} + \beta_{1,jk}\alpha_{ijk} \\ + \beta_{2,jk}Age_i + \beta_{3,jk}Sex_i) \tag{2.6}$$

Here, α_{ijk} indicates the $\alpha_{j,k}$ parameter from the i-th image. Models were fit for the $\gamma_{j,k}$ parameters in a similar fashion. Pairs of types that appeared in less than 10 images were excluded from this analysis. The resulting hazard ratios serve as potential prognostic biomarkers, indicating the prognostic significance of the identified spatial interactions within the tumor microenvironment. Furthermore, to account for the fact that there are multiple interaction estimates per patient (because there are multiple images, and thus multiple model fits, per patient), we fit these Cox models using a

robust sandwich variance estimator for each fitted parameter that appropriately accounts for the correlation between interaction estimates within patients [49].

2.3.5 Evaluation of Model Fit

Finally, in answering Question 7, we assessed the fit of the MGPP models by employing the Area Under the Curve (AUC) metric to quantify the predictive accuracy that the model had for each cell type within the tumor microenvironment. Briefly, the AUC is a metric that determines how well the predicted conditional intensity is able to partition the spatial window into regions of higher and lower concentrations of points [18]. By examining the AUC for each cell type, we are able to see how well the model predicts the spatial distribution of each cell type. We then calculated the median AUC per cell type, both within each patient group and across patient group, which helped identify cell types that could be reliably predicted by the MGPP models. Furthermore, we calculated the correlation between the AUC values and the abundance of each cell type, to clarify whether cell type abundance improved predictability of that cell type.

3. RESULTS FROM EXAMPLE ANALYSIS

3.1 Interpretation of Model Parameters and Predictions

We extracted the $\beta_{1,t}$ parameters from each of the fitted models, and examples of the fitted parameters from one image are show in Table 2, allowing us to answer Question 3.

Table 2. Coefficients $\beta_{1,t}$ as estimated from spot 59_A.

type	estimate	CI
B cells	−1.4e-04	(−6.2e-03, 5.9e-03)
CD163+ macros	−1.5e-03	(−2.9e-03, −6.1e-05)
CD8+ T cells	9.9e-04	(−1.8e-03, 3.8e-03)
granulocytes	−2.2e-03	(−3.4e-03, −1.1e-03)
memory CD4+ T	−1.9e-03	(−4.7e-03, 9.7e-04)
plasma cells	−1.0e-03	(−2.9e-03, 9.3e-04)
stroma	−6.9e-03	(−9.6e-03, −4.3e-03)
Tregs	3.4e-03	(−5.2e-03, 1.2e-02)
tumor cells	−1.9e-03	(−5.2e-03, 1.3e-03)

The β parameters from this image demonstrate that the intensity of stroma, granulocytes and CD163+ macrophages are negatively associated with distance from vasculature – that is, as the distance from vasculature increases, the less likely it is that stromal cells, granulocytes or CD163+ macrophages will be encountered.

We next extracted the α_{t_1,t_2} and γ_{t_1,t_2} (subsequently α and γ) parameters from each of the fitted models. Examples of the fitted parameters from one image are given in Figure 4. The α parameters from this image demonstrate relatively consistent within-type positive interaction, or clustering, while between-type interactions at this scale tend to be negative, or repulsive. Of course, a significant outlier to this trend is the interaction between B cells and many other cell types, in particular with memory CD4+ T cells. The γ parameters, on the other hand, showcase a quite different story. Here, we again see that within-type interactions
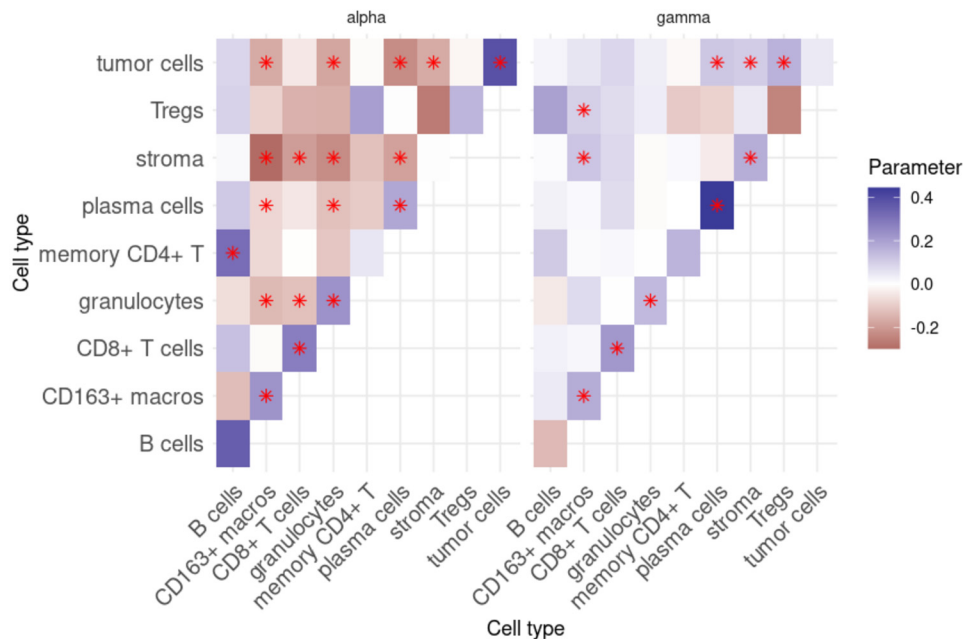

Figure 4: Interaction parameters α_{t_1,t_2} and γ_{t_1,t_2} as estimated from spot 59_A. Stars indicate that the interaction parameter was found to be significant within that image at a significance level of 0.05.
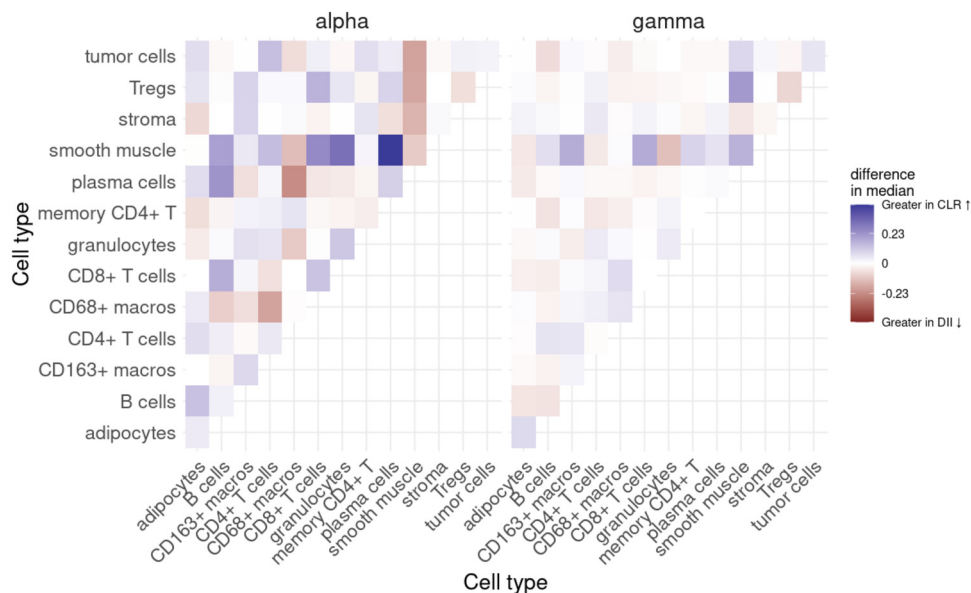
Figure 5: Difference in the median of the short-range interaction parameters α_{t_1,t_2} and γ_{t_1,t_2} between patient groups.

are generally positive (with the notable outlier of Tregs), though now tumor cells and CD163+ macrophages (or M2 macrophages) are generally positively associated with other cell types.

Heatmaps of the interaction parameters, as in Figure 4, allow one to quickly answer questions such as Question 1. From these heatmaps, we can also quickly answer questions like Question 1b: here, we can see that tumor cells seem to be quite strongly positively correlated with each other at short scales, while quite negatively correlated with plasma cells, also at short scales. Interestingly, at larger scales, tumor cells are no longer correlated with each other, and are now positively correlated with plasma cells!

One is likely also to want to compare how these interaction parameters contrast between various patient groups, in order to answer questions like Question 4. Figure 5 demonstrate the difference in the median interaction parameters α and γ from each patient group (averaging across all images within a group). Strikingly, it appears that both the α and γ parameters for smooth muscle generally have the strongest differences in median interaction parameter, with patients from the CLR group generally exhibiting greater attraction between smooth muscle and Tregs, CD8+ T cells and CD163+ macrophages, while smooth muscle and granulocytes exhibit greater repulsion in CLR patients.

Since the interaction parameters α and γ scale the potential functions used during the model fit (here, the exponential function), in order to answer questions liked Question 1a, it is also useful to view the pairwise potential functions themselves after they have been scaled by the appropriate parameters, as well as the combined effect of the short-range and medium-range scaled potential functions. An example of this can be seen in Figure 6, where the short-range
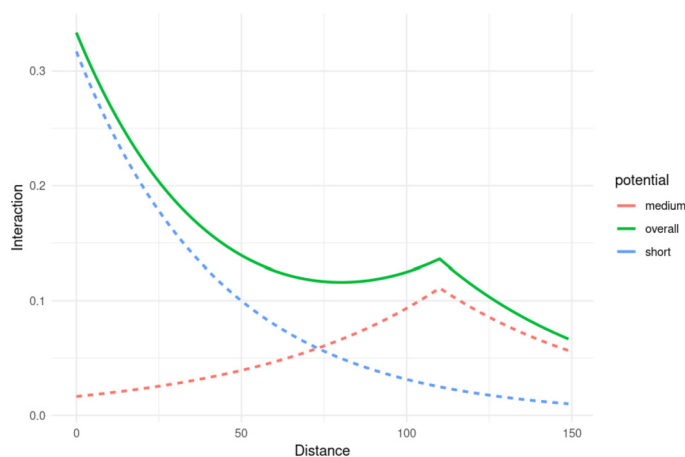

Figure 6: The short-range, medium-range and combined potentials between B cells and memory CD4+ T cells. From model fit on spot 59_A.

potential (scaled by the appropriate α), medium-range potential (scaled by γ), and combined (sum of scaled short- and medium-range potentials) can be seen. The combined (or overall) potential thus gives a more complete picture between two cell types across spatial scales. Here we can see that B cells and memory CD4+ T cells are quite mutually attractive at shorter scales, but much less so at longer spatial scales. Such potentials can also be plotted simultaneously between all cell types of interest, as in Figure 7.

From both of these figures, it is also apparent that there are many significant interactions occurring between these cell types, at multiple scales. These interactions indicate the spatial covariance between pairs of cell types that is unique
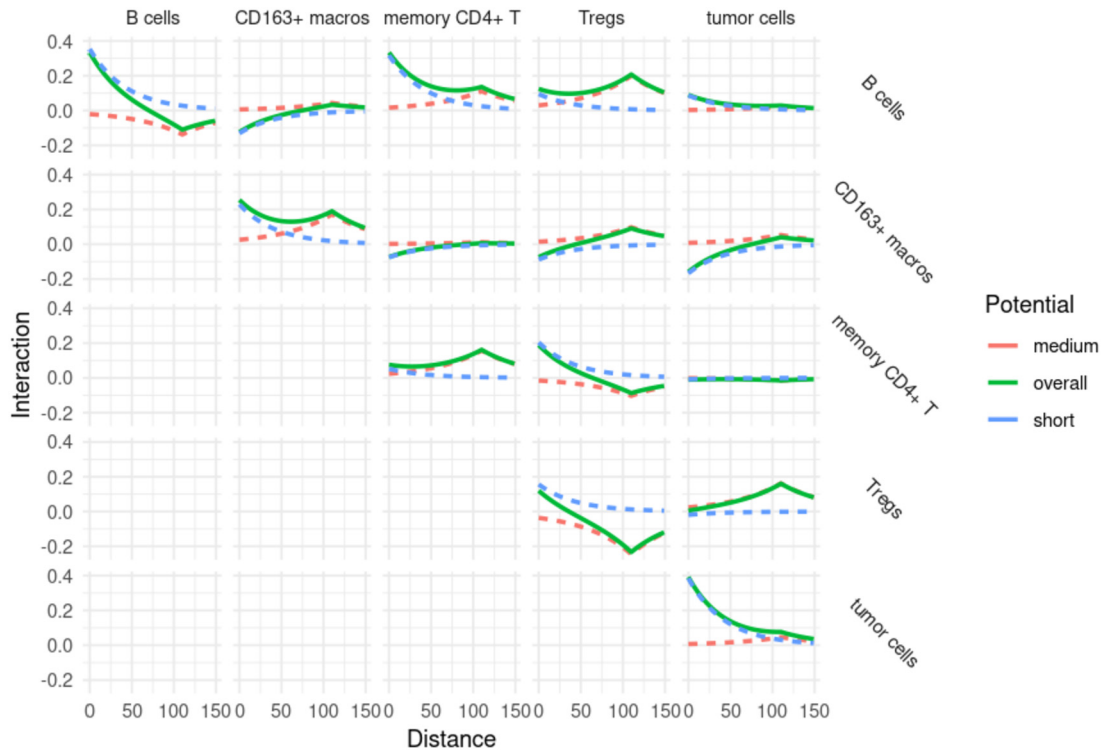
Figure 7: The short-range, medium-range and combined potentials between a subset of cell types. From model fit on spot 59__A.

to that pair. These quantities thus estimate the degree to which one cell type uniquely predicts the concentration of another, after accounting for the contributions to the prediction from all other cell types. From this information, we are thus able to assess which cell type most predicts the intensity of another cell type at each point in space, as in Figure 8. Such information allows one to answer questions

Figure 8: The top contributor to the predicted intensity of tumor cells. From model fit on spot 59__A.

like Question 2. Here we can see that different cell types are correlated more strongly with tumor cells at different points in the image. For example, granulocytes are the best predictors of tumor cells on the right side of the image, whereas CD163+ macrophages and stromal cells are the best predictors in other regions of the image.

3.2 Heterogeneity in Model Parameter Estimates Within Patient Groups and Across Patients

Estimates of the median absolute deviation of the short-range interaction parameters α can be seen in Figures 9 and 10, allowing us to answer Question 5. In Figure 9, we can see that, across the entire patient cohort, within-type interactions for several of the immune cell types (in particular, B cells, CD163+ macros, CD4+ T cells, CD68+ macros and CD8+ T cells) are the most heterogeneous between images, while the interaction between smooth muscle and CD4+ T cells appears to be the most consistent across images.

In Figure 10, it appears that DII patients generally have more homogeneous interactions between images. However, interactions involving smooth muscle in CLR patients appear to be the most homogeneous among CLR patients, sometimes more so than their DII counterpart (for example, the interaction between smooth muscle and stroma). Interestingly, interactions involving smooth muscle also seem
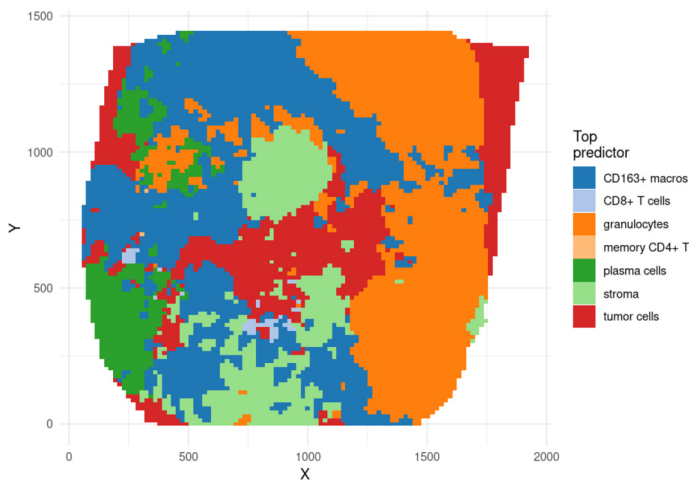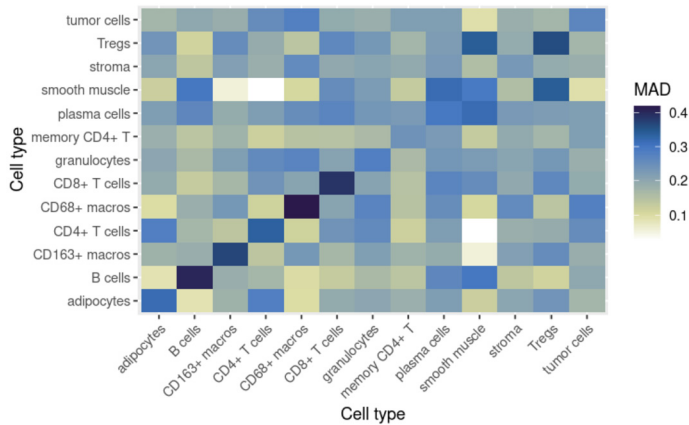
Figure 9: Estimates of the median absolute deviation (MAD) for α parameter estimates, across all patients.

most different between the two patient groups, as detailed in the previous section.

3.3 Predictive Modeling of Patient Outcomes

Next, we turn our attention to answering Question 6. Estimated log-hazard ratios for both short- and medium-range interactions that were found to be significant after FDR adjustment for multiple hypothesis testing are shown in Figure 11. Conspicuously, increased short-range interactions between granulocytes and stroma seemed to be correlated with better patient outcomes, as well as granulocytes and M1 macrophages (CD68+ macros). Furthermore, the short-range interaction between M2 macrophages (CD163+ macrophages) and M1 macrophages seems to be indicative of poor patient outcome.

At longer spatial scales (γ parameters from Figure 11), the Treg and M1 macrophage interaction seems to be highly correlated with poor patient outcome, while the plasma cell – B cell and CD4+ – CD4+ interactions were associated with better patient outcomes.

3.4 Evaluation of Model Fit

Lastly, we focus on the issue of model trust, in an effort to answer Question 7. An example of the predicted conditional intensity for plasma cells can be seen in Figure 12. The AUC (0.835) for this cell type seems relatively high, indicating that the model seems to predict the intensity of plasma cells quite well. Additionally, the observed positions of the plasma cells seem to match higher values of the predicted conditional intensity, mirroring the information given by the AUC.

Table 3 shows the median AUC per cell type across all images, as well as the median AUC within each patient group. From this table, we can see that the cell types tend to be similarly predictable on average. Furthermore, there does not seem to be a significant difference in predictability between the patient groups, except, notably, for CD4+ T cells.

Table 3. Median AUC for each cell type, across model fits from all images within each group and across the whole patient cohort.

	AUC		
type	CLR	DII	All
B cells	0.83	0.82	0.82
CD163+ macros	0.72	0.69	0.70
CD4+ T cells	0.81	0.73	0.77
CD68+ macros	0.80	0.76	0.77
CD8+ T cells	0.77	0.74	0.75
NK cells	NA	0.72	0.72
Tregs	0.80	0.77	0.78
adipocytes	0.80	0.78	0.80
generic immune	0.76	0.76	0.76
granulocytes	0.77	0.73	0.75
memory CD4+ T	0.75	0.73	0.74
plasma cells	0.79	0.80	0.79
smooth muscle	0.77	0.75	0.76
stroma	0.72	0.70	0.70
tumor cells	0.72	0.76	0.75

Table 4. Median abundance of each cell type in each patient group.

	abundance	
type	CLR	DII
adipocytes	8.5	4.0
B cells	14.0	17.0
CD163+ macros	172.5	316.5
CD4+ T cells	5.5	2.0
CD68+ macros	9.5	20.0
CD8+ T cells	62.0	109.0
generic immune	12.0	9.0
granulocytes	27.0	123.5
memory CD4+ T	53.0	80.0
NK cells	1.5	1.0
plasma cells	28.0	22.0
smooth muscle	77.0	124.0
stroma	104.0	157.5
Tregs	8.0	21.0
tumor cells	102.5	222.0
vasculature	61.5	75.0

At first, it may seem that this could be simply because of the much larger number of CD4+ T cells in CLR patients than in DII patients. However, as noted in Table 4, CD163+ macrophages are represented almost 2 times more in DII patients than CLR patients, a similar order of difference as found in CD4+ T cells, while maintaining median AUCs that are approximately the same.

To make the aforementioned precise, we measured the correlation between the AUC and median abundance of each cell type. Table 5 shows the correlation between cell type and median abundance, as quantified using AUCs from all model
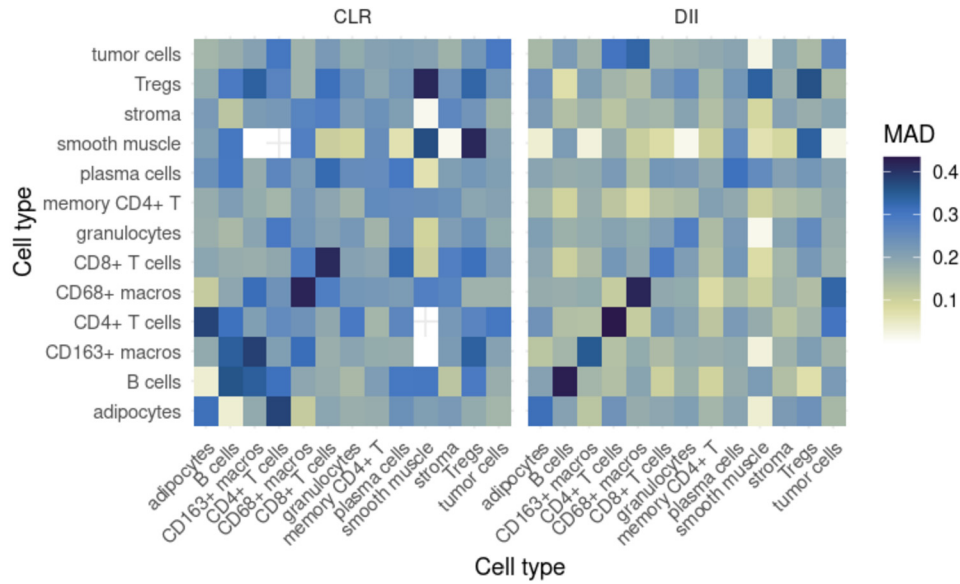
Figure 10: Estimates of the median absolute deviation (MAD) for α parameter estimates, in both patient groups.
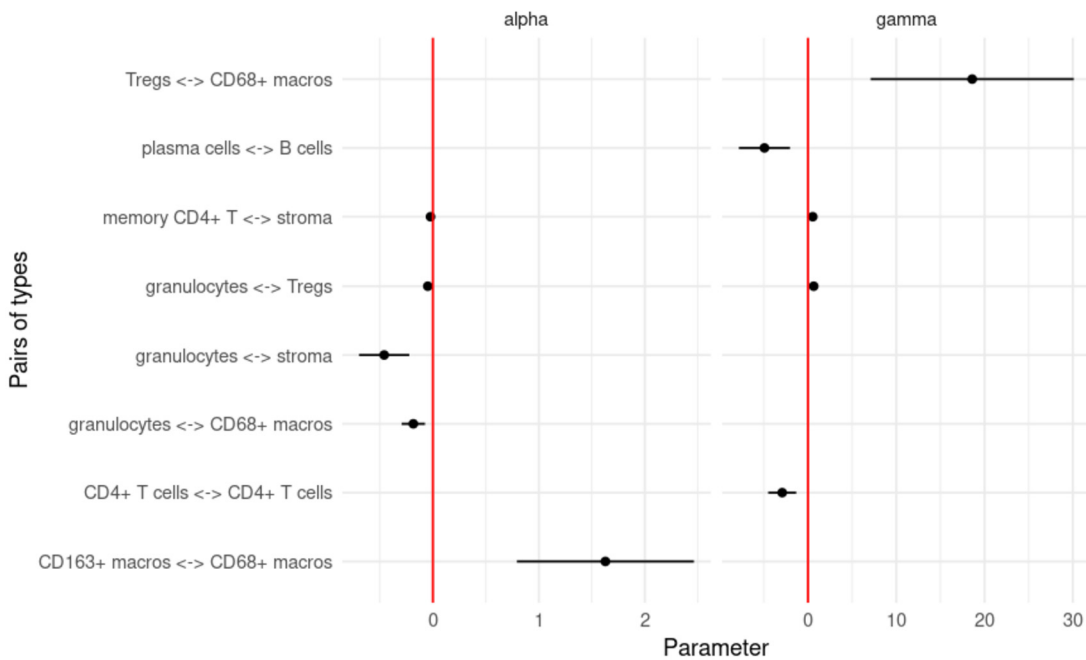

Figure 11: Log-hazard ratios estimated from short- and medium-range interaction parameters that were found to be significant after FDR adjustment.

fits. All correlations can be seen to be negative, some with a quite large magnitude. This indicates that simply having more cells of a given type in an image does not make that cell type necessarily easier to predict using MGPP models. This likely indicates the importance of incorporating higher-order information on cell-type interactions, beyond overall cell-type abundance, in order to correctly predict distributions of cells.

Overall, from Table 3, we can see that the models predict the spatial distribution of cells quite well, indicating that we can likely trust the results from our models.

3.5 Comparison with Previous Analysis

Since the CRC dataset used here comes from a previously published study [16], here we outline a few key points in comparison to the results from that study.
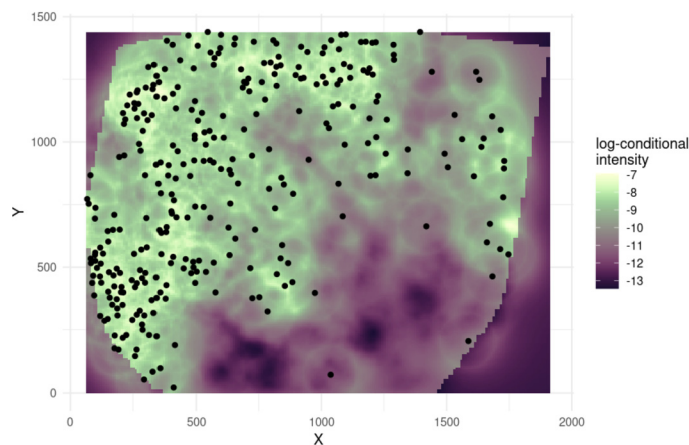
Figure 12: The predicted log-conditional intensity of plasma cells from the model fit for spot 59_A. The AUC for this cell type was 0.835, indicating high predictive ability for plasma cells.

Table 5. Correlation between AUC score and abundance for each cell type.

type	correlation
B cells	−0.18
CD163+ macros	−0.39
CD4+ T cells	−0.30
CD68+ macros	−0.41
CD8+ T cells	−0.49
NK cells	−0.99
Tregs	−0.35
adipocytes	−0.42
generic immune	−0.59
granulocytes	−0.54
memory CD4+ T	−0.29
plasma cells	−0.41
smooth muscle	−0.58
stroma	−0.49
tumor cells	−0.16

1. There is evidence of greater attraction between tumor cells and some subsets of immune cells in DII patients – for example, B cells and M1 macrophages – than CLR patients – similarly, [16] identified the presence of a mixed tumor/immune compartment in DII patients, as opposed to CLR patients, in whom the tumor and immune components were more segregated. Interestingly, though, at short ranges, CD4+ T cells and tumor cells are found to be more attractive in CLR patients (Figure 5).

2. Schurch et al found much greater contact between a T-cell enriched neighborhood and a macrophage-enriched neighborhood in DII patients. Similarly, we found that M1 macrophages were more attractive with CD4+ T cells and plasma cells in DII patients at short ranges

(Figure 5).

3. The authors further identified that the functional state of a granulocyte-enriched cellular neighborhood was indicative of patient survival in DII patients. Similarly, we also found that granulocyte associations were frequently implicated as the most prognostic biomarkers across patient groups derived from our models (Figure 11).

4. DISCUSSION

In this paper, we apply a novel approach for analyzing ecological interactions within the tumor microenvironment (TME). Specifically, we employ a multitype Gibbs point process model (MGPP) as a joint species distribution model, tailored to the complexities of the TME. This modeling framework was previously introduced as a statistical analysis framework in forest ecology [42]. Our aim in this work is to highlight the significance of this modeling framework and its potential implications for addressing key questions in cancer biology.

Our primary contribution lies in the application of a specific MGPP model to characterize the data-generating process governing cell-cell and cell-environment interactions in the TME. Given the intricate nature of the TME, where diverse cell types interact spatially, the MGPP framework provides a powerful tool for capturing and quantifying these interactions. What sets this modeling framework apart from many others that currently exist is its ability to model relationships between cell types simultaneously, while also being able to quantify the uncertainty in these associations as well as produce predictions. Researchers currently lack such a unified statistical framework capable of jointly modeling interactions between cell types and spatial variables across multiple spatial scales. With a specific MGPP model, we showcase how to bridge this gap, enabling a diverse array of inquiries.

Our emphasis extends beyond model development to model validation and diagnostics. We highlight the utility of our fitted models in generating predictions in order to examine goodness-of-fit and demonstrate the area under the curve (AUC) as a concise summary of model fit. Additionally, we stress the availability of further model diagnostics based on residuals.

Beyond model fitting, our study demonstrates the utility of extracting estimated parameters from single-image models and summarizing them across patient cohorts. This approach can serve as a valuable tool for identifying potential prognostic biomarkers, furthering our understanding of cancer biology and personalized medicine.

While hyperparameter selection is a critical aspect of model development, we do not delve deeply into this topic here. Interested readers can find comprehensive discussions on the selection of hyperparameters, including the scale and shape of potential functions, the saturation parameter and the number of dummy points, in [42].

Lastly, MGPP models offer a unique capability which we also did not address in this paper: the generation of simulations while conditioning on various aspects of the data. This feature has the potential to serve two primary purposes—validating model outputs by comparing them to expected results and simulating new cell types based on different configurations. Such simulations can shed light on the spatial organization of cell types within the TME.

In conclusion, our investigation into cell-type interactions in the tumor microenvironment using multitype Gibbs point process models presents a valuable tool for researchers interested in the complex ecology of the TME. This framework empowers researchers to explore complex questions, validate models, and gain deeper insights into the dynamics of the TME, with potential implications for cancer research and therapy.

CONFLICTS OF INTEREST

REFERENCES

[1] Bożyk, A., Wojas-Krawczyk, K., Krawczyk, P., Milanowski and Tumor, J. (June 2022). Microenvironment—A Short Review of Cellular and Interaction Diversity. *Biology* **11**(6), 929. Publisher: Multidisciplinary Digital Publishing Institute ISSN: 2079-7737. https://www.mdpi.com/2079-7737/11/6/929 (2023).

[2] De Visser, K. E. and Joyce, J. A. (Mar. 13, 2023) The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer Cell* **41**, 374–403. ISSN: 1878-3686.

[3] Boedtkjer, E. and Pedersen, S. F. (Feb. 10, 2020). The Acidic Tumor Microenvironment as a Driver of Cancer. *Annual Review of Physiology* **82**, 103–126 ISSN: 1545-1585.

[4] Petrova, V., Annicchiarico-Petruzzelli, M., Melino, G. and Amelio, I. (Jan. 24, 2018) The hypoxic tumour microenvironment. *Oncogenesis* **7**(1), 1–13. Publisher: Nature Publishing Group, ISSN: 2157-9024. https://www.nature.com/articles/s41389-017-0011-9 (2023).

[5] Li, Y., Zhao, L., Li, Hypoxia, X. -F. and the Tumor Microenvironment. (Aug. 5, 2021). *Technology in Cancer Research & Treatment* **20**, 15330338211036304. ISSN: 1533-0346. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8358492/ (2023).

[6] Tan, K., Naylor and Tumour, M. J. (May 13, 2022). Microenvironment-Immune Cell Interactions Influencing Breast Cancer Heterogeneity and Disease Progression. *Frontiers in Oncology* **12**, 876451. ISSN: 2234-943X. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9138702/ (2023).

[7] Sun, Y. (Sept. 28, 2016). Tumor microenvironment and cancer therapy resistance. *Cancer Letters* **380**, 205–215. ISSN: 0304-3835. https://www.sciencedirect.com/science/article/pii/S0304383515005121 (2023).

[8] Ni, Y. et al. (May 20, 2021). The Role of Tumor-Stroma Interactions in Drug Resistance Within Tumor Microenvironment. *Frontiers in Cell and Developmental Biology* **9**, 637675. ISSN: 2296-634X. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8173135/ (2023).

[9] Li, X. et al. (Oct. 7, 2021). Crosstalk Between the Tumor Microenvironment and Cancer Cells: A Promising Predictive Biomarker for Immune Checkpoint Inhibitors. *Frontiers in Cell and Developmental Biology* **9**, 738373. ISSN: 2296-634X. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8529050/ (2023).

[10] Pienta, K. J., McGregor, N., Axelrod, R. and Axelrod, D. E. (Dec. 2008). Ecological Therapy for Cancer: Defining Tumors Using an Ecosystem Paradigm Suggests New Opportunities for Novel Cancer Treatments. *Translational Oncology* **1**, 158–164. ISSN: 1936-5233. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2582164/ (2021).

[11] Adler, F. R., Amend, S. R., Baratchart, E. and Whelan, C. J. (Feb. 10, 2022). Editorial: From Ecology to Cancer Biology and Back Again. *Frontiers in Ecology and Evolution* **10**, 840375. ISSN: 2296-701X. https://www.frontiersin.org/articles/10.3389/fevo.2022.840375/full (2022).

[12] Somarelli, J. A. (Apr. 28, 2021). The Hallmarks of Cancer as Ecologically Driven Phenotypes. *Frontiers in Ecology and Evolution* **9**, 661583. ISSN: 2296-701X. https://www.frontiersin.org/articles/10.3389/fevo.2021.661583/full (2022).

[13] Thomas, F., Roche, B., Giraudeau, M., Hamede, R. and Ujvari, B. (2020). The interface between ecology, evolution, and cancer: More than ever a relevant research direction for both oncologists and ecologists. *Evolutionary Applications* **13**, 1545–1549. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/eva.13031, ISSN: 1752-4571. https://onlinelibrary.wiley.com/doi/abs/10.1111/eva.13031 (2021).

[14] Reynolds, B. A., Oli, M. W. and Oli, M. K. (2020). Eco-oncology: Applying ecological principles to understand and manage cancer. *Ecology and Evolution* **10**, 8538–8553. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.6590, ISSN: 2045-7758. https://onlinelibrary.wiley.com/doi/abs/10.1002/ece3.6590 (2021).

[15] Van Dam, S., Baars, M. J. D., Vercoulen and Multiplex, Y. (Jan. 2022). Tissue Imaging: Spatial Revelations in the Tumor Microenvironment. *Cancers* **14**(13), 3170. Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 2072-6694. https://www.mdpi.com/2072-6694/14/13/3170 (2023).

[16] Schürch, C. M. et al. (Sept. 3, 2020). Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341–1359.e19. ISSN: 1097-4172.

[17] Black, S. et al. (Aug. 2021). CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nature Protocols* **16**, 3802–3835. ISSN: 1750-2799. https://www.nature.com/articles/s41596-021-00556-8 (2021).

[18] Baddeley, A., Turner and spatstat, R. (Jan. 26, 2005). An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software* **12**, 1–42. ISSN: 1548-7660. https://doi.org/10.18637/jss.v012.i06 (2023).

[19] Fortin, M. -J., Dale, M. R. and Ver Hoef, J. M. (2016). In *Wiley StatsRef: Statistics Reference Online*, 1–13. John Wiley & Sons, Ltd, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118445112.stat07766.pub2, ISBN: 978-1-118-44511-2. https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118445112.stat07766.pub2 (2023).

[20] Ben-Said, M. (Aug. 5, 2021). Spatial point-pattern analysis as a powerful tool in identifying pattern-process relationships in plant ecology: an updated review. *Ecological Processes* **10**, 56. ISSN: 2192-1709. https://doi.org/10.1186/s13717-021-00314-4 (2023).

[21] Velázquez, E., Martínez, I., Getzin, S., Moloney, K. A. and Wiegand, T. (2016). An evaluation of the state of spatial point pattern analysis in ecology. *Ecography.* **39**, 1042–1055. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.

1111/ecog.01579, ISSN: 1600-0587. https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.01579 (2023).

[22] CANETE, N. P. et al. (May 26, 2022). spicyR: spatial analysis of in situ cytometry data in R. *Bioinformatics* **38**, 3099–3105. ISSN: 1367-4803. https://doi.org/10.1093/bioinformatics/btac268 (2023).

[23] VU, T. et al. (June 15, 2022). SPF: A spatial and functional data analytic approach to cell imaging data. *PLOS Computational Biology* **18**, e1009486. Publisher: Public Library of Science, ISSN: 1553-7358. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1009486 (2023).

[24] CREED, J. H. et al. (Dec. 7, 2021). spatialTIME and iTIME: R package and Shiny application for visualization and analysis of immunofluorescence data. *Bioinformatics* **37**, 4584–4586. ISSN: 1367-4803. https://doi.org/10.1093/bioinformatics/btab757 (2023).

[25] BEHANOVA, A., KLEMM, A. and WÄHLBY, C. (Jan. 28, 2022). Spatial Statistics for Understanding Tissue Organization. *Frontiers in Physiology* **13**, 832417. ISSN: 1664-042X. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8837270/ (2023).

[26] YUAN, Z. et al. (Nov. 28, 2022). SOTIP is a versatile method for microenvironment modeling with spatial omics data. *Nature Communications* **13**(1), 7330. Publisher: Nature Publishing Group, ISSN: 2041-1723. https://www.nature.com/articles/s41467-022-34867-5 (2023).

[27] TANEVSKI, J., FLORES, R. O. R., GABOR, A., SCHAPIRO, D. and SAEZ-RODRIGUEZ, J. (Apr. 14, 2022). Explainable multiview framework for dissecting spatial relationships from highly multiplexed data. *Genome Biology* **23**, 97. ISSN: 1474-760X. https://doi.org/10.1186/s13059-022-02663-5 (2023).

[28] GALON, J. et al. (Oct. 3, 2012). Cancer classification using the Immunoscore: a worldwide task force. *Journal of Translational Medicine* **10**, 205. ISSN: 1479-5876. https://doi.org/10.1186/1479-5876-10-205 (2023).

[29] ALLAM, M. et al. (Sept. 1, 2022). Spatially variant immune infiltration scoring in human cancer tissues. *npj Precision Oncology* **6**(1), 1–21. Publisher: Nature Publishing Group. ISSN: 2397-768X. https://www.nature.com/articles/s41698-022-00305-4 (2023).

[30] TIKHONOV, G. et al. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution* **11**, 442–447. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13345, ISSN: 2041-210X. https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13345 (2023).

[31] TIKHONOV, G., ABREGO, N., DUNSON, D. and OVASKAINEN, O. (2017). Using joint species distribution models for evaluating how species-to-species associations depend on the environmental context. *Methods in Ecology and Evolution* **8**, 443–452. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12723, ISSN: 2041-210X. https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12723 (2023).

[32] OVASKAINEN, O., ROY, D. B., FOX, R. and ANDERSON, B. J. (2016). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution* **7**, 428–436. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12502, ISSN: 2041-210X. https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12502 (2023).

[33] WILKINSON, D. P., GOLDING, N., GUILLERA-ARROITA, G., TINGLEY, R. and MCCARTHY, M. A. (2021). Defining and evaluating predictions of joint species distribution models. *Methods in Ecology and Evolution* **12**, 394–404. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13518, ISSN: 2041-210X. https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13518 (2023).

[34] CLARK, J. S., GELFAND, A. E., WOODALL, C. W. and ZHU, K. (2014). More than the sum of the parts: forest climate response from joint species distribution models. *Ecological Applications* **24**, 990–999. _eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/13-1015.1, ISSN: 1939-5582. https://onlinelibrary.wiley.com/doi/abs/10.1890/13-1015.1 (2023).

[35] STOYAN, D. and PENTTINEN, A. (2000). Recent Applications of Point Process Methods in Forestry Statistics. *Statistical Science* **15**, 61–78. Publisher: Institute of Mathematical Statistics, ISSN: 0883-4237. https://www.jstor.org/stable/2676677 (2023). https://doi.org/10.1214/ss/1009212674. MR1842237

[36] LAW, R. et al. (2009). Ecological information from spatial patterns of plants: insights from point process theory. *Journal of Ecology* **97**, 616–628. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2745.2009.01510.x, ISSN: 1365-2745. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2745.2009.01510.x (2023).

[37] ILLIAN, J. B. et al. (2013). Fitting complex ecological point process models with integrated nested Laplace approximation. *Methods in Ecology and Evolution* **4**, 305–315. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210x.12017, ISSN: 2041-210X. https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210x.12017 (2023).

[38] ILLIAN, J. B., MØLLER, J. and WAAGEPETERSEN, R. P. (Sept. 1, 2009). Hierarchical spatial point process analysis for a plant community with high biodiversity. *Environmental and Ecological Statistics* **16**, 389–405. ISSN: 1573-3009. https://doi.org/10.1007/s10651-007-0070-8 (2023). https://doi.org/10.1007/s10651-007-0070-8. MR2749847

[39] RENNER, I. W. et al. (2015). Point process models for presence-only analysis. *Methods in Ecology and Evolution* **6**, 366–379. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12352, ISSN: 2041-210X. https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.12352 (2023).

[40] BADDELEY, A., RUBAK, E. and TURNER, R. (Nov. 11, 2015). *Spatial Point Patterns: Methodology and Applications with R*. CRC Press. Google-Books-ID: rGbmCgAAQBAJ. 830 pp. ISBN: 978-1-4822-1021-7.

[41] RAJALA, T., MURRELL, D. J., OLHEDE and DETECTING, S. C. (Nov. 1, 2018). Multivariate Interactions in Spatial Point Patterns with Gibbs Models and Variable Selection. *Journal of the Royal Statistical Society Series C: Applied Statistics* **67**, 1237–1273. ISSN: 0035-9254, 1467-9876. https://academic.oup.com/jrsssc/article/67/5/1237/7058395 (2023). https://doi.org/10.1111/rssc.12281. MR3873707

[42] FLINT, I., GOLDING, N., VESK, P., WANG, Y. and XIA, A. (Nov. 1, 2022). The Saturated Pairwise Interaction Gibbs Point Process as a Joint Species Distribution Model. *Journal of the Royal Statistical Society Series C: Applied Statistics* **71**, 1721–1752. ISSN: 0035-9254. https://doi.org/10.1111/rssc.12596 (2023). https://doi.org/10.1111/rssc.12596. MR4511129

[43] GOLTSEV, Y. et al. (Aug. 9, 2018). Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968–981.e15. ISSN: 0092-8674. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6086938/ (2023).

[44] GEYER, C. (1999). (O. Barndorff-Nielsen, W. Kendall and M. Van Lieshout, eds.) In *Stochastic Geometry: Likelihood and Computation*, chapter 3. *Monographs on Statistics and Applied Probability* **80**, 79–140. Chapman and Hall/CRC, ISBN: 978-0-203-73827–978-0-203-6. https://doi.org/10.1007/978-1-4899-3210-5. MR1317097

[45] BADDELEY, A. J. and VAN LIESHOUT, M. N. M. (Dec. 1, 1995). Area-interaction point processes. *Annals of the Institute of Statistical Mathematics* **47**, 601–619. ISSN: 1572-9052. https://doi.org/10.1007/BF01856536 (2024). https://doi.org/10.1007/BF01856536. MR1370279

[46] ILLIAN, J. B. and HENDRICHSEN, D. K. (2010). Gibbs point process models with mixed effects. *Environmetrics* **21**, 341–353. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.1008, ISSN: 1099-095X. https://onlinelibrary.wiley.com/doi/abs/10.1002/env.1008 (2024). https://doi.org/10.1002/env.1008. MR2842247

[47] KING, R., ILLIAN, J. B., KING, S. E., NIGHTINGALE, G. F. and HENDRICHSEN, D. K. (Dec. 1, 2012). A Bayesian Ap-

proach to Fitting Gibbs Processes with Temporal Random Effects. *Journal of Agricultural, Biological, and Environmental Statistics* **17**. 601–622. ISSN: 1537-2693. https://doi.org/10.1007/s13253-012-0111-0 (2024). https://doi.org/10.1007/s13253-012-0111-0. MR3041887

[48] BADDELEY, A., COEURJOLLY, J. -F., RUBAK, E. and WAAGEPETERSEN, R. (2014). Logistic regression for spatial Gibbs point processes. *Biometrika* **101**, 377–392. Publisher: [Oxford University Press, Biometrika Trust], ISSN: 0006-3444. https://www.jstor.org/stable/43305620 (2024). https://doi.org/10.1093/biomet/ast060. MR3215354

[49] SPIEKERMAN, C. F., LIN and MARGINAL, D. Y. (1998). Regression Models for Multivariate Failure Time Data. *Journal of the American Statistical Association* **93**, 1164–1175. Publisher: [American Statistical Association, Taylor & Francis, Ltd.], ISSN: 0162-1459. https://www.jstor.org/stable/2669859 (2023). https://doi.org/10.2307/2669859. MR1649210

Joel Eliason. Department of Computational Medicine and Bioinformatics, University of Michigan, USA. E-mail address: joelne@umich.edu

Arvind Rao. Department of Computational Medicine and Bioinformatics, Department of Radiation Oncology and Department of Biostatistics, University of Michigan, USA. E-mail address: ukarvind@med.umich.edu