Bayesian Nonparametric Learning for Point Processes with Spatial Homogeneity: A Spatial Analysis of NBA Shot Locations

Fan Yin [1] **Jieying Jiao** [2] **Jun Yan** [2] **Guanyu Hu** [3]

Abstract

Basketball shot location data provide valuable summary information regarding players to coaches, sports analysts, fans, statisticians, as well as players themselves. Represented by spatial points, such data are naturally analyzed with spatial point process models. We present a novel nonparametric Bayesian method for learning the underlying intensity surface built upon a combination of Dirichlet process and Markov random field. Our method has the advantage of effectively encouraging local spatial homogeneity when estimating a globally heterogeneous intensity surface. Posterior inferences are performed with an efficient Markov chain Monte Carlo (MCMC) algorithm. Simulation studies show that the inferences are accurate and the method is superior compared to a wide range of competing methods. Application to the shot location data of 20 representative NBA players in the 2017-2018 regular season offers interesting insights about the shooting patterns of these players. A comparison against the competing method shows that the proposed method can effectively incorporate spatial contiguity into the estimation of intensity surfaces.

1. Introduction

Quantitative analytics have been a key driving force for advancing modern professional sports, and there is no exception for professional basketball (Kubatko et al., 2007). In professional basketball, analyses of shooting patterns offer important insights about players' attacking styles and shed light on the evolution of defensive tactics, which has aroused substantial research interests from the statistical community (e.g., Reich et al., 2006; Miller et al., 2014; Franks et al., 2015; Cervone et al., 2016; Jiao et al., 2021a; Hu et al., 2020b). As shot location data are naturally represented by spatial points, developments of novel methods for analyzing spatial point patterns are of fundamental importance.

The literature on spatial point pattern data is voluminous (see, e.g., Illian et al., 2008; Diggle, 2013; Guan, 2006; Guan & Shen, 2010; Baddeley, 2017; Jiao et al., 2021b). The most frequently adopted class of models in empirical research is nonhomogeneous Poisson processes (NHPP), or more generally, Cox processes, including log-Gaussian Cox process (Møller et al., 1998). Such parametric models impose restrictions on the functional forms of underlying process intensity, which can suffer from underfitting of data when there is a misfit between the complexity of the model and the data available. In contrast, nonparametric approaches such as two-way kernel density estimation provide more flexibility compared to parametric modeling as the underfitting can be mitigated by using models with unbounded complexity.

Several important features of the shot location data need to be captured in any realistic nonparametric method. First, near regions are highly likely to have similar intensities. This makes that certain spatial contiguous constraints on the intensity surface desirable. Existing method such as mixture of finite mixtures (MFM) of nonhomogeneous Poisson processes (Taddy & Kottas, 2012; Geng et al., 2021) is lacking in this aspect. There are also rich literature (Blei & Frazier, 2011; Müller et al., 2011; Ghosh et al., 2011; Page et al., 2016; Dahl et al., 2017) discussing spatial constraint prior for regression models however lacking the discussion on intensity estimation of spatial point pattern data. Second, spatial contiguous constraints should not dominate the intensity surface globally (Hu et al., 2020a; Zhao et al., 2020). For some players, there are more than one hot zones in their one-season shooting. Some players will prefer more corner three and top three rather than 45 degrees. Although, the corner is not spatial contiguous with top area, the same intensity value can still belong

[*]Equal contribution [1]Department of Statistics, University of California, Irvine, CA, USA [2]Department of Statistics, University of Connecticut, Storrs, CT, USA [3]Department of Statistics, University of Missouri, Columbia, MO, USA. Correspondence to: Guanyu Hu <guanyu.hu@missouri.edu>.

Proceedings of the 39th International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022. Copyright 2022 by the author(s).

to spatially disconnected regions that are sufficiently similar with respect to intensity values, which is not well accommodated by the penalized method (Li & Sang, 2019). From the application prospect, zone defense is common defense strategy in NBA from 2001. In addition, each players will have floor coverage in court. Knowing the hot regions rather than small hot points will help the team optimize their team defense strategy. For some players, there are more than one hot zones in their one-season shooting. Finally, the extent to which the spatial contiguous affects the intensity surface may differ from player to player, and needs to be learned from the data.

To address these challenges, we consider a spatially constrained Bayesian nonparametric method for point processes to capture the spatial homogeneity of intensity surfaces. Our contributions are three-fold. First, we develop a novel nonparametric Bayesian method for intensity estimation of spatial point processes. Compared to existing methods, the proposed approach is capable of capturing both locally spatially contiguous clusters and globally discontinuous clusters and the number of clusters. Second, an efficient Markov chain Monte Carlo (MCMC) algorithm is designed for our model without complicated reversible jump MCMC. Lastly, we gain important insights about the shooting behaviors of NBA players based on an application to their shot location data.

2. Model Specification

2.1. NHPP

Spatial point process models provide a natural framework for capturing the random behavior of event location data. Let $\mathbf{S} = \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\}$ with $\boldsymbol{s}_i = (x_i, y_i)$, be the set of observed locations in a pre-defined, bounded region $\mathcal{B} \subseteq \mathcal{R}^2$. Let the underlying stochastic mechanism that gives rise to the observed point pattern $\mathbf{S}$ be denoted as spatial point process $\mathbf{Y}$. Process $N_{\mathbf{Y}}(A) = \sum_{i=1}^{N} \mathbb{1}(\boldsymbol{s}_i \in A)$ is a counting process associated with $\mathbf{Y}$, which counts the number of points falling into area $A \subseteq \mathcal{B}$.

The NHPP model assumes conditionally independent event locations given the process intensity $\lambda(\mathbf{s})$. For an NHPP, the number of events in area A, $N_{\mathbf{Y}}(A)$, follows Poisson distribution with rate parameter $\lambda(A) = \int_A \lambda(\mathbf{s})\mathrm{d}\mathbf{s}$. In addition, $N_{\mathbf{Y}}(A_1)$ and $N_{\mathbf{Y}}(A_2)$ are independent if two areas $A_1 \subseteq \mathcal{B}$ and $A_2 \subseteq \mathcal{B}$ are disjoint. Given the observed point pattern $\mathbf{S}$ on fixed region $\mathcal{B}$, the likelihood of the NHPP model is

$$\frac{\prod_{i=1}^{N} \lambda(\mathbf{s}_i)}{\exp(\int_{\mathcal{B}} \lambda(\mathbf{s}) d\mathbf{s})}, \tag{2.1}$$

where $\lambda(\mathbf{s}_i)$ is the intensity function evaluated at location $\mathbf{s}_i$. The NHPP reduces to a homogeneous Poisson process (HPP) when $\lambda(\mathbf{s})$ is constant over the entire study region $\mathcal{B}$, and it is synonymous with *complete spatial randomness* (CSR).

2.2. Nonparametric Bayesian Methods for NHPP

As the CSR assumption over the entire study region rarely holds in real-world problems, and to simplify the potentially overcomplicated problem induced by complete nonhomogeneity on intensity values, Teng et al. (2017) proposed to approximate the intensity function $\lambda(\boldsymbol{s})$ by a piecewise constant function. Specifically, the study region $\mathcal{B}$ is partitioned into n disjoint regions and the intensity over each region is assumed to be constant. Let $A_1, A_2, \ldots, A_n$ be a partition of $\mathcal{B}$, i.e., $\cup_{i=1}^{n} A_i = \mathcal{B}$ and $A_i \cap A_j = \emptyset, \forall i \neq j$. For each region $A_i, i = 1, \ldots, n$, we have $\lambda(\mathbf{s}) = \lambda_i, \forall\, \mathbf{s} \in A_i$. Therefore, the likelihood (2.1) can be written as

$$\prod_{i=1}^{n} f_{\text{poisson}}(N_{\mathbf{Y}}(A_i) | \lambda_i \mu(A_i)), \tag{2.2}$$

where $\mu(A_i) = \int_{A_i} 1 d\boldsymbol{s}$ is the area of region A_i and $f_{\text{poisson}}(\cdot|\lambda)$ is the probability mass function of the Poisson distribution with rate parameter λ. For ease of notation, we use $N(A_i)$ for $N_{\mathbf{Y}}(A_i)$ throughout the remainder of the text.

The heterogeneity in the intensity function across different regions can be naturally represented through a latent clustering structure. The conventional finite mixture modeling framework (McLachlan & Basford, 1988; Bouveyron et al., 2019) assumes that the heterogeneity can be characterized by a discrete set of subpopulations or clusters, such that the points located in the regions belonging to any given subpopulation tend to be produced by similar intensities. The selection of the number of clusters (or components) in finite mixture models are often recasted as statistical model selection problems which can solved using information criteria (Fraley & Raftery, 2002) or cross-validation (Fu & Perry, 2020), among others. Despite its prevalence in empirical research, such model selection procedures for finite mixture model ignore uncertainty in the number of clusters, which may in turn lead to increased erroneous cluster assignments. The Bayesian nonparametric approach provides an alternative to parametric modeling and model selection. The Dirichlet process (Ferguson, 1973) is currently one of the most popular Bayesian nonparametric models and it can be viewed as the limit of the following finite mixture model

$$\begin{aligned} &y_i | Z_i, \beta_{Z_i} \sim F(\beta_{Z_i}) \\ &Z_i | p \sim Discrete(p_1, \cdots, p_K) \\ &\beta_{Z_i} \sim G_0 \quad p \sim Dirichlet_K(\alpha/K, \cdots, \alpha/K) \end{aligned} \tag{2.3}$$

where Z_i stands for the cluster of ith observation, β_{c_i}

means the parameter of c_ith cluster, α stands for the precision parameter, and G_0 is the base measure in Dirichlet process, which is also the prior on cluster specific parameters β's. As $K \to \infty$, the model becomes a Dirichlet process mixture (DPM) model Neal (2000), which can be used to simultaneously estimate the number of clusters and cluster configurations. Combining DPM with NHPP yields the following DPM-NHPP model

$$\begin{aligned} N(A_i) \mid \lambda_1, \ldots, \lambda_n &\sim Poisson(\lambda_i \mu(A_i))\ i = 1, \ldots, n, \\ (\lambda_1, \ldots, \lambda_n) &\sim \prod_{i=1}^{n} G(\lambda_i) \\ G &\sim \mathrm{DP}(\alpha, G_0) \end{aligned} \tag{2.4}$$

where the Dirichlet process (DP) here is parameterized by a base measure $G_0 \equiv \mathrm{Gamma}(a, b)$ and a concentration parameter α. Given $\lambda_1, \ldots, \lambda_n$ drawn from G, a conditional prior can be obtained by integration (Blackwell & MacQueen, 1973)

$$\begin{aligned} \Pr(\lambda_{n+1} | \lambda_1, \ldots, \lambda_n) = &\frac{1}{n+\alpha} \sum_{i=1}^{n} \delta_{\lambda_i}(\lambda_{n+1}) \\ &+ \frac{n}{n+\alpha} G_0(\lambda_{n+1}) \end{aligned} \tag{2.5}$$

where $\delta_{\lambda_i}(\lambda_j) = I(\lambda_i = \lambda_j)$ denote a degenerate distribution concentrated at a single point λ_i.

Under Dirichlet process mixture model (DPMM), the latent cluster membership variables $\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)$ are distributed according to a Chinese restaurant process (CRP) (Pitman, 1995; Neal, 2000), which is defined through the following conditional distributions or a Pólya urn scheme (Blackwell & MacQueen, 1973)

$$\Pr(Z_i = c | Z_j, j < i; \alpha) \propto \begin{cases} |c|, & c \text{ for existing cluster label} \\ \alpha, & \text{otherwise,} \end{cases} \tag{2.6}$$

where $|c|$ refers to the size of cluster labeled c, and α is the concentration parameter of the underlying Dirichlet process (DP). While CRP allows for simultaneous estimation on the number of clusters and the cluster configuration, it has been proved that CRP can produce extraneous clusters in the posterior leading to inconsistent estimation of the *number of clusters* even with sample size approaching infinity (Miller & Harrison, 2018).

To mitigate the *inconsistency* in estimating the number of clusters caused by CRP, Miller & Harrison (2018) proposed to modify CRP with the Mixture of finite mixtures (MFM) model. An alternative model which will be used as a benchmark for comparison is the MFM of NHPP (MFM-NHPP) (Geng et al., 2021).

2.3. Incorporating Spatial Homogeneity

Spatial events typically obey the so-called first law of geography, "everything is related to everything else, but near things are more related than distant things" (Tobler, 1970). This means spatial smoothness, also known as spatial homogeneity. To incorporate such spatial homogeneity, we impose a Markov random field constraint (Besag et al., 1995; Orbanz & Buhmann, 2008) $M(\lambda_1, \ldots, \lambda_n) := \frac{1}{Z_H} \exp\{-H(\lambda_1, \ldots, \lambda_n)\}$ on $\boldsymbol{\lambda}$ to encourage the intensity parameters in nearby regions to be similar

$$\begin{aligned} G &\sim \mathrm{DP}(\alpha, G_0) \\ (\lambda_1, \ldots, \lambda_n) &\sim M(\lambda_1, \ldots, \lambda_n) \prod_{i=1}^{n} G(\lambda_i) \\ N(A_i) \mid \lambda_1, \ldots, \lambda_n &\sim Poisson(\lambda_i \mu(A_i))\ i = 1, \ldots, n, \end{aligned} \tag{2.7}$$

The cost function $H(\lambda_1, \ldots, \lambda_n) := \sum_{C \in \mathcal{C}} H_C(\lambda_C)$, where $\mathcal{C}$ denotes the set of all cliques, or completely connected subsets in the underlying *neighborhood graph* $\mathcal{N} = (V_\mathcal{N}, E_\mathcal{N}, W_\mathcal{N})$ with vertices $V_\mathcal{N} = (v_1, \ldots, v_n)$ representing n random variables, $E_\mathcal{N}$ denoting a set of edges representing the statistical dependence structure among vertices, and $W_\mathcal{N}$ denoting the edge weights representing the magnitude of the respective dependence.

By the Hammersley—Clifford theorem (Hammersley & Clifford, 1971), the corresponding conditional distributions enjoy the Markov property, i.e., $M(\lambda_i | \boldsymbol{\lambda}_{-i}) = M(\lambda_i | \boldsymbol{\lambda}_{\partial(i)})$, where $\partial(i) := \{j | (i,j) \in E_\mathcal{N}\}$ represents the neighbors of i. In this work, we consider only pairwise interactions by letting

$$H(\lambda_i | \boldsymbol{\lambda}_{-i}) := -\eta \sum_{j \in \partial(i)} I(\lambda_i = \lambda_j) = -\eta \sum_{j \in \partial(i)} I(z_j = z_i) \tag{2.8}$$

where η is a parameter controlling the extent of spatial homogeneity with larger values dictating larger extent of spatial homogeneity. We note that the resulting model defines a valid MRF distribution Π, which can be written as

$$\Pi(\lambda_1, \ldots, \lambda_n) \propto M(\lambda_1, \ldots, \lambda_n) P(\lambda_1, \ldots, \lambda_n) \tag{2.9}$$

and such a constrained model can be shown to only change the finite component of model (2.7) as shown in Theorem 2.1 below. The proof is deferred to the Appendix A.

Theorem 2.1. *Let K^* denote the number of clusters excluding the i-th observation and $n_k^{(-i)}$ denote size of the k-th cluster excluding λ_i, and assume $H(\lambda_i|\boldsymbol{\lambda}_{-i})$ to be a valid cost function for MRF. The conditional distribution of (2.9) takes the following form*

$$\Pi(\lambda_i|\boldsymbol{\lambda}_{-i}) \propto \sum_{k=1}^{K^*} n_k^{(-i)} \frac{1}{Z_H} \exp(-H(\lambda_i|\boldsymbol{\lambda}_{-i}))\delta_{\lambda_k^*}(\lambda_i) + \frac{\alpha}{Z_H} G_0(\lambda_i).$$

Although, there are rich literature discussing constraint based nonparametric Bayesian prior such as the distance dependent Chinese Restaurant Process (Blei & Frazier, 2011), the PPMx prior (Müller et al., 2011), and the Ewens-Pitman attraction distribution (Dahl et al., 2017). Our proposed DPM-MRF prior enjoys attractive properties. First, ddCRP does not hold exchangeability, and its conditional distributions reflect the relationship between observations. However, DPM-MRF is exchangeable since the cohesion function is invariant under permutation (it only depends on the clustering configuration), which by de Finetti's theorem and the conditional distribution directly reflects relationship between the observations with existing clusters. Compared with PPMx prior and EPA distribution, our prior starts from Dirichlet process and incorporate a Markov random field (MRF) structure on partition distribution. Our proposed method inherits the ability of clustering since it provides a full support over the entire space of partitions.

The definition of neighborhood $\partial(i)$ is subject to the nature of the data and the modeler's choice. Common choices include the *rook* contiguity (i.e., the regions which share a border of some length with region i), and the *queen* contiguity (i.e., the regions which share at least a point-length border with region i) (Orbanz & Buhmann, 2008). The MRF-DPM-NHPP model (2.7) reduces to the DPM-NHPP model (2.4) when $\eta = 0$.

3. Bayesian Inference

In this section, we present an efficient MCMC sampling algorithm for our proposed method, post MCMC inference, and model selection criteria for identifying the optimal value of the smoothing parameter.

3.1. A Collapsed Gibbs Sampler

We introduce latent indicator variables $\boldsymbol{Z} = (Z_1, \ldots, Z_n)$ and denote the parameters in (2.7) as $\boldsymbol{\Theta} = \{\boldsymbol{\lambda}, \boldsymbol{Z}\}$. The posterior density of $\boldsymbol{\Theta}$ is

$$\pi(\boldsymbol{\Theta}|\mathbf{S}) \propto L(\boldsymbol{\Theta}|\mathbf{S})\pi(\boldsymbol{\Theta}),$$

where $\pi(\boldsymbol{\Theta})$ is the prior density of $\boldsymbol{\Theta}$, and the likelihood $L(\boldsymbol{\Theta}|\mathbf{S})$ takes the form of (2.1).

We first derive the full conditional distribution of Z_i, which is given by Proposition 3.1.

Proposition 3.1. *Suppose the result of Theorem 2.1 holds. Then, under the model and prior specification (2.7), the full conditional distribution of Z_i, $i = 1, \ldots, n$, is*

$$\Pr(Z_i = c \mid \mathbf{S}, \boldsymbol{Z}_{-i}, \boldsymbol{\lambda}, \boldsymbol{\beta}) \propto \begin{cases} \dfrac{n_c(\boldsymbol{Z}_{-i}) \exp\left(\eta \sum_{j\in\partial(i)} d_{ij} \mathbb{1}(Z_j = c)\right)(\lambda_c \mu(A_i))^{N(A_i)}}{\exp(\lambda_c \mu(A_i))}, \\ \dfrac{\alpha b^a \Gamma(N(A_i)+a)\mu(A_i)^{N(A_i)}}{(b+\mu(A_i))^{N(A_i)+a}\Gamma(a)}, \end{cases} \tag{3.1}$$

for existing labels and new label, respectively, where $\boldsymbol{Z}_{-i}$ is $\boldsymbol{Z}$ with z_i removed, and $\mu(A_i)$ is the area of region A_i.

For the full conditional distribution of λ_k, only data points in the kth component should be considered for simplicity. The full conditional density of λ_k, $k = 1, \ldots, K$, is

$$\begin{aligned} q(\lambda_k \mid \mathbf{S}, \boldsymbol{Z}, \boldsymbol{\lambda}_{-k}) &\propto \frac{\prod_{\ell:\boldsymbol{s}_\ell \in A_j, Z_j = k} \lambda(\boldsymbol{s}_\ell)}{\exp(\int_{\bigcup_{j:Z_j=k} A_j} \lambda(\boldsymbol{s}) \mathrm{d}\boldsymbol{s})} \lambda_k^{a-1} \exp(-b\lambda_k) \\ &= \frac{\prod_{\ell:\boldsymbol{s}_\ell \in A_j, Z_j = k} \lambda_k}{\exp\left(\int_{\bigcup_{j:Z_j=k} A_j} \lambda_k \mathrm{d}\boldsymbol{s}\right)} \lambda_k^{a-1} \exp(-b\lambda_k) \\ &\propto \lambda_k^{N_k+a-1} \exp\left(-\left(b + \sum_{j:Z_j=k} \mu(A_j)\right)\lambda_k\right), \end{aligned} \tag{3.2}$$

which is the kernel of Gamma$\big(N_k + a, b + \sum_{j:Z_j=k} \mu(A_j)\big)$, where $N_k = \sum_{\ell:\boldsymbol{s}_\ell \in A_j, Z_j = k} 1$ is the number of data points in the regions belonging to the kth component. The detailed steps of a Gibbs sampling algorithm using the full conditional distributions from (3.1)–(3.2) is given in Appendix C.

Convergence check for the auxiliary variables $(Z_1, \ldots, Z_n)$ can be done with the help of the Rand Index (RI) (Rand, 1971). The auxiliary variables themselves are nominal labels which cannot be compared from iteration to iteration. The RI is the proportion of concordant pairs between two clustering results with value of 1 indicating the two results are exactly the same. The trajectory of the RI for successive MCMC iterations provides a visual check for convergence. Further, RI values closer to 1 indicate good agreement in the clustering in the MCMC samples.

We carry out posterior inference on the group memberships using Dahl's method (Dahl, 2006) (details in Appendix C). Therefore, the posterior estimates of cluster memberships $Z_1, \ldots, Z_n$ and model parameters $\boldsymbol{\Theta}$ can be based on the draws identified by Dahl's method.

3.2. Selection of Smoothing Parameter

We recast the choice of smoothing parameter $\eta \geqslant 0$ as a model selection problem. In particular, we consider the deviance information criterion (DIC; Spiegelhalter et al. (2002)), logarithm of the Pseudo-marginal likelihood LPML; Gelfand & Dey (1994)) and Bayesian information criterion (BIC; Schwarz (1978)) as candidates.

The DIC for spatial point process can be derived from the standard DIC in a straightforward manner as

$$\mathrm{Dev}(\boldsymbol{\Theta}) = -2\left(\sum_{i=1}^{N} \log \lambda(\boldsymbol{s}_i) - \int_{\mathcal{B}} \lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right),$$
$$\mathrm{DIC} = 2\overline{\mathrm{Dev}}(\boldsymbol{\Theta}) - \mathrm{Dev}(\widehat{\boldsymbol{\Theta}}),$$

where $\overline{\mathrm{Dev}}(\boldsymbol{\Theta})$ is the average deviance evaluated using each posterior sample of Θ, and $\mathrm{Dev}(\widehat{\boldsymbol{\Theta}})$ is the deviance calculated using the point estimation of parameter using Dahl's method.

The LPML for spatial point process can be approximated using the MCMC samples (Hu et al., 2019)

$$\widehat{\mathrm{LPML}} = \sum_{i=1}^{N} \log \widetilde{\lambda}(\boldsymbol{s}_i) - \int_{\mathcal{B}} \overline{\lambda}(\boldsymbol{s})\mathrm{d}\boldsymbol{s},$$
$$\widetilde{\lambda}(\boldsymbol{s})_i = \left(\frac{1}{M}\sum_{t=1}^{M} \lambda(\boldsymbol{s}_i \mid \boldsymbol{\Theta}_t)^{-1}\right)^{-1},$$
$$\overline{\lambda}(\boldsymbol{s}) = \frac{1}{M}\sum_{t=1}^{M} \lambda(\boldsymbol{s} \mid \boldsymbol{\Theta}_t),$$

where $\boldsymbol{\Theta}_t$ denotes the t-th posterior sample of parameters with a total length of M.

The BIC is derived naturally from its general definition

$$\mathrm{BIC}(\boldsymbol{\Theta}) = -2\log L(\boldsymbol{\Theta}) + \widehat{K}\log N,$$
$$\log L(\boldsymbol{\Theta}) = \sum_{i=1}^{N} \log \lambda(\boldsymbol{s}_i) - \int_{\mathcal{B}} \lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s},$$

where $\widehat{K}$ denotes the estimated number of components of the piecewise constant intensity function.

4. Simulation Studies

In this section, we report simulation studies to examine the performance of the MRF-DPM-NHPP model and the proposed Gibbs sampling algorithm. In each setting, we compare the results to that of MFM-NHPP and other methods listed in Table 1 to show that the MRF-DPM-NHPP model can yield better performance.

Table 1: Alternative methods for comparison.

Method	Implementation
CAR prior constrained spatially varying Poisson	`nimble`
Log Gaussian Cox process	`inlabru`
Bayesian additive regression trees	`BayesTree`
Kernel density estimate	`spatstat`
Nonhomogeneous Poisson process (B-spline, order= 3)	`spatstat`

4.1. Design

Consider a study region $\mathcal{B} = [0, 20] \times [0, 20]$ partitioned into $n = 400$ squares of unit area, $\{A_i\}_{i=1}^{n}$. The data generating model was set to be NHPP($\lambda(\boldsymbol{s})$) with a piecewise constant intensity $\lambda(\boldsymbol{s})$ over $\mathcal{B}$. Three settings were considered for $\lambda(\boldsymbol{s})$; see Table 2. The "ground-truth" intensity surfaces of the three settings are displayed in the leftmost column of Figure 1. The first two settings with the different numbers of clusters are similar with the simulation setups in Geng et al. (2021). The third setting contains both spatially contiguous and discontinuous clusters. The point patterns were generated using the `rpoispp()` function from package `spatstat` (Baddeley & Turner, 2005). For each setting, we generated 100 replicates.

The prior distributions were specified as in (2.7), with hyperparameters $a = b = \alpha = 1$. The smoothing parameter $\eta \geqslant 0$ took values on an equally-spaced grid $\eta = \{0, 0.5, \ldots, 7\}$, of which the optimal value is chosen via the model selection criteria introduced in Section 3.2. The neighboring structure was defined based on rook contiguity, and we treat all neighbors equally by letting $d_{ij} = 1, \forall j \in \partial i$. Each MCMC chain was run for a total of 5000 iterations with random starting values, where the first 2000 draws were discarded as burn-in (see Appendix G for traceplots justifying this choice). The remaining 3000 draws were thinned by 3 and stored for posterior inference. We used Dahl's method (Dahl, 2006) to identify the most representative draw from the retained posterior draws as the posterior point estimate.

Table 2: Simulation settings for the piecewise constant intensity function.

	Grid size	$\boldsymbol{\lambda}$	Number of grid boxes
Setting 1	20×20	$(0.2, 4, 12)$	$(90, 211, 99)$
Setting 2	20×20	$(0.2, 1, 4, 8, 16)$	$(80, 80, 80, 80, 80)$
Setting 3	20×20	$(0.2, 4, 10, 20)$	$(90, 145, 66, 99)$

4.2. Results

We evaluate the results of simulation studies on the following aspects, (i) probability of choosing the correct number of clusters, (ii) clustering accuracy quantified by the adjusted Rand index (Hubert & Arabie, 1985), and (iii) estimation accuracy of the intensity surface.

Table 3 (left block) shows the proportion of times the true number of components is identified under different model selection criteria for each simulation setting. Obviously, $\eta = 0$ never recovered the true number of clusters, suggesting that taking spatial contiguity information into account is crucial. For MRF-DPM-NHPP, BIC appears to be better than DIC and LPML as the BIC-selected *optimal* η recovered the true number of clusters more frequently. Although MFM-NHPP seems to be very competitive in terms of identifying the true number of components under setting 1, MRF-DPM-NHPP with smoothing parameter η selected by BIC offers substantially better performance under all other settings. A further investigation revealed that setting $\eta = 0$ always produced overly large numbers of redundant clusters, while DIC and LPML failed more gracefully with wrong numbers of clusters that often fall into the approximate range (A histogram of $\widehat{K}$ is available in the Appendix F).

To assess the clustering performance, we examine the average adjusted RI (calculated using function `arandi` in **R** package `mcclust` (Fritsch, 2012)) over the 100 replicates. Because the "ground-truth" class labels are known in the simulation studies, the adjusted RIs were calculated by comparing the posterior samples with the truth as a measure of clustering accuracy. As shown in Table 3 (right block), MRF-DPM-NHPP with smoothing parameter η selected by BIC yields the highest clustering accuracy. Despite being more capable of identifying the true number of clusters, the clustering accuracy of MFM-NHPP is worse than that of MRF-DPM-NHPP with BIC under setting 1, which suggests that MFM-NHPP might happen to get the number of clusters right by allocating the regions into wrong clusters. For the remainder of this paper, we focus on the results that correspond to optimal η selected by BIC.

We next summarize accuracy in estimating the intensity surfaces. Figure 1 displays the averages of the median, 2.5th percentile, and 97.5th percentile of the estimated intensity surface obtained with the optimal η selected by BIC from the 100 replicates, in comparison with the true surfaces, for the three settings. The median surface agrees with true surface well in all three settings. The 2.5th and 97.5th percentiles of the estimated intensity surfaces over 100 replicates have higher uncertainties occasionally at the boundaries where the true intensities jump, but in general are not far from the true surfaces. The frequency coverage rates of the posterior 95% credible intervals for a representative pixel $(18, 10)$ are 0.92, 0.93 and 0.94 for setting 1, 2 and 3, respectively, which are very close to the nominal level. Figure 2 shows the absolute value of relative bias of element-wise posterior mean estimates under the MFM-DPM-NHPP and other competing methods. The proposed method leads to substantially smaller bias than the competing methods, especially for grids with low true underlying intensity values and/or grids at the boundaries.

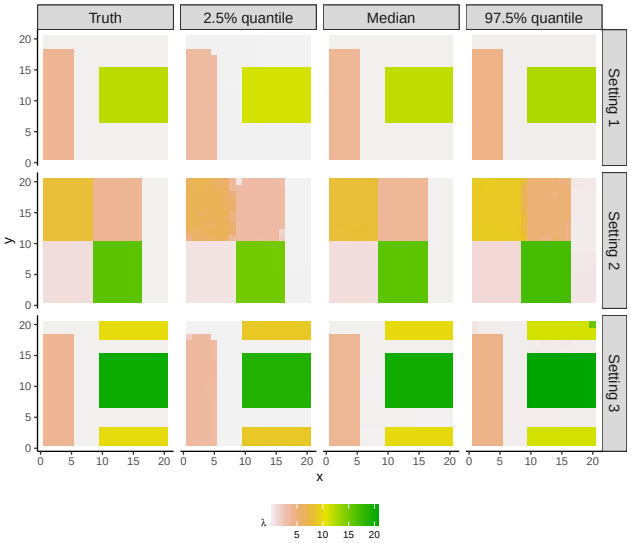

Figure 1: Simulation configurations for intensity surfaces under grid size of 20×20, with fitted intensity surfaces. Element-wise median and quantiles are calculated out of 100 replicates.

In order to show the robustness of our proposed method, we fit kernel density estimate to the Durant's shot location data and use that smooth intensity surface to generate random locations and evaluate the performance of different methods. The simulation results are shown in Table 7 (in Appendix J). We see that the proposed method can yield comparable performance to other methods under this scenario.

In summary, the simulation studies confirm the advantages of the MRF-DPM-NHPP model and the validity of the proposed Gibbs sampling algorithm. The results also suggest that BIC is better than DIC and LPML in selecting the smoothing parameter η in the studied settings. Compared to other methods, the proposed MRF-DPM-NHPP is superior in estimating the intensity surfaces, especially when the transition between different spatially homogeneous regions is not smooth.

Table 3: Proportion of times the true number of cluster is identified, and average adjusted RI across 100 replicates for each simulation setting, under MFM-NHPP, and MRF-DPM-NHPP with $\eta = 0$, optimal η selected by BIC, DIC and LPML.

	Accuracy of $\widehat{K}$					Average adjusted RI				
	MRF-DPM-NHPP				MFM-NHPP	MRF-DPM-NHPP				MFM-NHPP
	$\eta = 0$	BIC	DIC	LPML		$\eta = 0$	BIC	DIC	LPML	
Setting 1	0.00	0.68	0.23	0.26	**0.97**	0.026	**0.940**	0.768	0.770	0.802
Setting 2	0.00	**0.79**	0.59	0.59	0.17	0.045	**0.974**	0.937	0.944	0.402
Setting 3	0.00	**0.70**	0.10	0.11	0.60	0.056	**0.981**	0.833	0.833	0.676

5. Professional Basketball Data Analysis

We applied the MRF-DPM-NHPP model to study the shot data for NBA players in the 2017-2018 NBA regular season (visualized in Appendix D). In particular, we focus on 20 all-star level players that are representative of their positions (Table 5). The study region is a rectangle covering the first 75% of the half court (50 ft $\times$ 35 ft) as the shots made outside this region are often not part of the regular tactics. This rectangle was divided into $50 \times 35 = 1750$ equally-sized grid boxes of 1ft $\times$ 1ft following Miller et al. (2014). For each player, we run parallel MCMC chains with $\eta \in \{0, 0.5, \ldots, 6\}$ for 5000 iterations using random intial values, where the first 2000 were discarded as burn-in and the remainder was thinned by 3 (see Appendix G for traceplots). In this section, we mainly focus on interpreting the results from MRF-DPM-NHPP, and we assess the performance of several alternative methods at the end from a prediction perspective.

Table 5 in Appendix E summarizes the optimal η selected by BIC and the resulting number of clusters. None of the selected $\widehat{\eta}$ lies on the boundary, which assures the validity of candidate values of η. For comparison, the number of clusters from the MFM-NHPP model under the same MCMC setting is also included, and we note that MFM-NHPP leads to higher numbers of clusters for most of the players than that of MFM-DPM-NHPP.

Figure 3 shows the estimated shooting intensity surfaces of selected players under KDE, MFM-NHPP and MRF-DPM-NHPP. Compared to the results of MFM-NHPP, it is clear that the MRF-DPM-NHPP model is capable of capturing distant regions that share similar shooting intensities while preserving the spatial contiguity, which greatly facilitates the interpretability. Taking Karl-Anthony Towns as an example, the estimated shooting intensity surface yielded by MFM-NHPP appears to be too scattered to highlight his preferred shooting regions; the results from the MRF-DPM-NHPP model, however, shows much clearer pattern.

More interesting observations are seen from the estimated shooting intensity surfaces (see Figure 5 in Appendix), and we summarize these observations by the preferred positions of selected players. Among those players with preferred position as center, DeAndre Jordan never makes shots outside the low post, while Dwight Howard seems to have made more shots from the regions between short corner and the restricted area. On the contrary, Joel Embiid and Karl-Anthony Towns are more versatile as attackers in terms of their shot locations — Joel Embiid can attack from low post, high post, top of the key as well as the *point* (i.e., right outside the middle of the arc); Karl-Anthony Towns' shots are mainly initiated either from the low block or outside the arc (right corner and from point to the wing).

The selected power-forward (PF) players show fairly different shooting styles. The shot locations of Kristaps Porziņgis are similar to those of Joel Embiid, and Kristaps Porziņgis seems to be less confined to shooting from low post regions compared to Joel Embiid. Both Giannis Antetokounmpo and LaMarcus Aldridge all make substantial amounts of mid-range shots and seldomly make three-point shots, but it is worth highlighting their differences as Giannis Antetokounmpo appears to be more inclined to make shots from the right while LaMarcus Aldridge's mid-range shots are more spread. Interestingly, the former champion of slam dunk contest, Blake Griffin has higher intensity of shooting outside the arc (in particular, from the right corner, and the regions between the wing and the point).

The selected small-forward (SF) players show versatile shot locations but they differ substantially in their three-point shot locations and the intensity of making shots around restricted area. Speaking about the three-point shots, Kevin Durant prefers shooting around left and right wings, both Paul George and Jimmy Butler prefer shooting around the right corner but the former is clearly more comfortable with launching long-range shots, while LeBron James prefers shooting around the left wing. Compared to the other two SF players, LeBron James have substantially higher intensity of making shots around the restricted area.

The difference in the shooting patterns among backcourt (PG and SG) players is even more sizable. James Harden, Stephen Curry, Damian Lillard and Kyrie Irving all launch

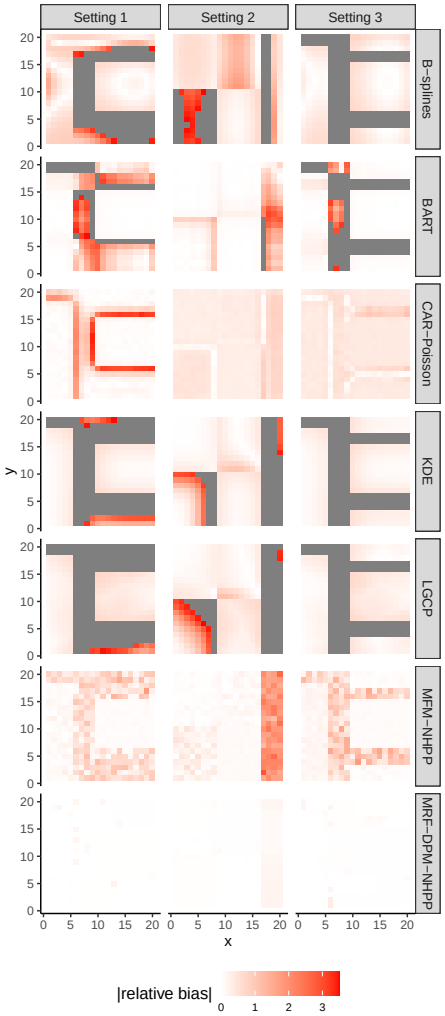

Figure 2: Absolute value of relative bias of element-wise posterior mean estimates for intensity surfaces. Dark grey values correspond to regions with absolute bias beyond 3.5.

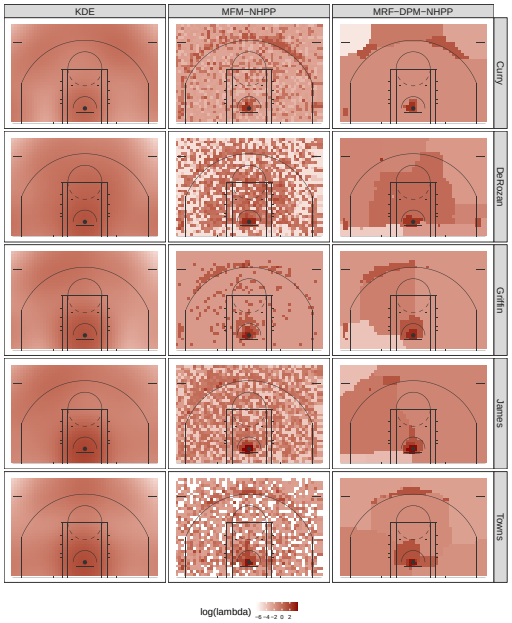

Figure 3: Estimated shooting log-intensity surfaces of selected players (one for each position) based on KDE, MFM-NHPP and MRF-DPM-NHPP. The selected players are Stephen Curry, DeMar DeRozan, Blake Griffin, LeBron James and Karl-Anthony Towns.

considerable amounts of shots within the restricted area and outside the arc, while James Harden makes shots in almost all regions from right wing to left wing outside the arc, Stephen Curry and Kyrie Irving make more shots around left wing rather than right wing, Damian Lillard makes more shots around right wing rather than left wing. Compared to the former three players, Chris Paul, Russell Westbrook, DeMar DeRozan and Klay Thompson make more mid-range shots, but from different angles. Specifically, Russell Westbrook makes shots almost everywhere in the middle, Chris Paul's shots are also mainly located in the middle but slightly biased towards the right, Demar DeRozan's shots are closer to the rim and more spread to the corners, while Klay Thompson's shots are almost evenly distributed across the entire study region.

In addition, as we can see from Figure 4 in Appendix D, most of the shots are made either close to the rim or right outside the arc (i.e., 3-point line). The intensity surface of shot locations is not smooth over basketball court. This is in line with the recent trend in the basketball development since it is more efficient for players to pursue higher success rates near the rim or go after higher rewards by making 3-point shots. These patterns confirm that the spatial MRF in our prior will encourage spatial smoothness in the cluster labels to achieve local contiguity, while DPMM in our prior will the globally distant intensity to be clustered together.

Admittedly, the above analysis is far from being exhaustive. We believe, however, that basketball professionals may leverage the proposed method to better understand the shooting patterns of the players and, therefore, design highly targeted offense and defense tactics.

More model assessment results compared with several benchmark methods are given in Appendix I. Based on the results shown in Appendix I, we find that the proposed method is comparable to BART and MFM-NHPP and clearly superior to all other methods. Moreover, we run a simulation study using the fitted intensity of James Harden (Figure 5) under the grid size of 50ft $\times$ 35ft, and the results also confirm the advantage of the proposed method over a wide range of alternative methods.

6. Discussion

The NBA shot location data appear to be modeled by the spatially constrained nonparametric Bayesian model, MRF-DPM-NHPP, reasonably well incorporating local spatial homogeneity. Building upon a combination of Dirichlet process and Markov random field, the proposed method relies on a smoothing parameter η to effectively control the relative contribution of local spatial homogeneity in estimating the globally heterogeneous intensity surface. Statistical inferences are facilitated by a Gibbs sampling algorithm. Selection of the smoothing parameter η is casted as a model selection problem which is handled using standard model selection criteria. Simulation studies show the accuracy of the proposed algorithm and the competitiveness of the model relative to the benchmark MFM-NHPP model (Geng et al., 2021) under several settings in which spatial contiguity is present in the intensity surface. In application to the shot locations of NBA players, the model effectively captures spatial contiguity in shooting intensity surfaces, and provide important insights on their shooting patterns which cannot be obtained from the MFM-NHPP model.

There are several possible directions for further investigations. More sophisticated definition of neighborhood (e.g., higher-order neighborhood, incorporating covariates) than the rook contiguity, which was used in this study and found to be sufficient here, may be useful for more complex data structure. BIC was found to perform well for the purpose of selecting smoothing parameter η, but it is of substantial interest to develop a fully automated procedure that enables the smoothing parameter to be inferred along with the intensity values and the group membership indicators through a single MCMC run. The NBA players shot pattern modeling admits a natural partition for the region of interest. In general settings, however, it is worth investigating how to effectively partition the space such that the piecewise constant assumption is more plausible. As the number of parameters is proportional to the number of grid boxes, developments of more scalable inference algorithms (e.g., variational inference) are critical for finer grid. A player's field goal attempts will vary considerably with respect to several other covariates such time left on shot clock, distance to nearest defender, and score differential. Incorporating non-spatial covariates will help us have better understanding of players' choices. Zero-inflation is a common pattern for some players such as center players attempt very few three-point shots. Considering a zero-inflated model will extend applications of the proposed method. Finally, building a group learning model with pooled data from multiple players merits future research from both methodological and applied perspectives.

Acknowledgements

The authors would like to thank Dr. Yishu Xue for sharing the R code of data visualization.

References

Baddeley, A. Local composite likelihood for spatial point processes. *Spatial Statistics*, 22:261–295, 2017.

Baddeley, A. and Turner, R. spatstat: An r package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(i06), 2005.

Besag, J., Green, P., Higdon, D., and Mengersen, K. Bayesian computation and stochastic systems. *Statistical science*, pp. 3–41, 1995.

Blackwell, D. and MacQueen, J. B. Ferguson distributions via pólya urn schemes. *The Annals of Statistics*, 1(2): 353–355, 1973.

Blei, D. M. and Frazier, P. I. Distance dependent Chinese restaurant processes. *Journal of Machine Learning Research*, 12(Aug):2461–2488, 2011.

Bouveyron, C., Celeux, G., Murphy, T. B., and Raftery, A. E. *Model-Based Clustering and Classification for Data Science: With Applications in R.*, volume 50. Cambridge University Press, 2019.

Cervone, D., D'Amour, A., Bornn, L., and Goldsberry, K. A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association*, 111(514):585–599, 2016.

Dahl, D. B. Model-based clustering for expression data via a dirichlet process mixture model. *Bayesian inference for gene expression and proteomics*, 4:201–218, 2006.

Dahl, D. B., Day, R., and Tsai, J. W. Random partition distribution indexed by pairwise information. *Journal of the American Statistical Association*, 112(518):721–732, 2017.

Diggle, P. J. *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press, 2013.

Ferguson, T. S. A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pp. 209–230, 1973.

Fraley, C. and Raftery, A. E. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631, 2002.

Franks, A., Miller, A., Bornn, L., and Goldsberry, K. Characterizing the spatial structure of defensive skill in professional basketball. *The Annals of Applied Statistics*, 9 (1):94–121, 2015.

Fritsch, A. *mcclust: Process an MCMC Sample of Clusterings*, 2012. URL https://CRAN.R-project.org/package=mcclust. R package version 1.0.

Fu, W. and Perry, P. O. Estimating the number of clusters using cross-validation. *Journal of Computational and Graphical Statistics*, 29(1):162–173, 2020.

Gelfand, A. E. and Dey, D. K. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3): 501–514, 1994.

Geng, J., Shi, W., and Hu, G. Bayesian nonparametric nonhomogeneous poisson process with applications to usgs earthquake data. *Spatial Statistics*, 41:100495, 2021.

Ghosh, S., Ungureanu, A., Sudderth, E., and Blei, D. Spatial distance dependent chinese restaurant processes for image segmentation. *Advances in Neural Information Processing Systems*, 24:1476–1484, 2011.

Guan, Y. A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101(476):1502–1512, 2006.

Guan, Y. and Shen, Y. A weighted estimating equation approach for inhomogeneous spatial point processes. *Biometrika*, 97(4):867–880, 2010.

Hammersley, J. M. and Clifford, P. Markov fields on finite graphs and lattices. *Unpublished manuscript*, 1971.

Hu, G., Huffer, F., and Chen, M.-H. New development of Bayesian variable selection criteria for spatial point process with applications. arXiv e-prints 1910.06870, 2019.

Hu, G., Geng, J., Xue, Y., and Sang, H. Bayesian spatial homogeneity pursuit of functional data: an application to the us income distribution. *arXiv preprint arXiv:2002.06663*, 2020a.

Hu, G., Yang, H.-C., and Xue, Y. Bayesian group learning for shot selection of professional basketball players. *Stat*, 9(1):e324, 2020b. doi: https://doi.org/10.1002/sta4.324. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sta4.324.

Hubert, L. and Arabie, P. Comparing partitions. *Journal of Classification*, 2(1):193–218, 1985.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. *Statistical analysis and modelling of spatial point patterns.*, volume 70. John Wiley & Sons, 2008.

Jiao, J., Hu, G., and Yan, J. A bayesian marked spatial point processes model for basketball shot chart. *Journal of Quantitative Analysis in Sports*, 17(2):77–90, 2021a.

Jiao, J., Hu, G., and Yan, J. Heterogeneity pursuit for spatial point pattern with application to tree locations: A Bayesian semiparametric recourse. *Environmetrics*, pp. e2694, 2021b.

Kubatko, J., Oliver, D., Pelton, K., and Rosenbaum, D. T. A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, 3(3), 2007.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Li, F. and Sang, H. Spatial homogeneity pursuit of regression coefficients for large datasets. *Journal of the American Statistical Association*, 114(527):1050–1062, 2019.

McLachlan, G. J. and Basford, K. E. *Mixture models: Inference and applications to clustering.*, volume 84. M. Dekker New York, 1988.

Miller, A., Bornn, L., Adams, R., and Goldsberry, K. Factorized point process intensities: A spatial analysis of professional basketball. In *International conference on machine learning*, pp. 235–243, 2014.

Miller, J. W. and Harrison, M. T. Mixture models with a prior on the number of components. *Journal of the American Statistical Association*, 113(521):340–356, 2018.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. Log Gaussian Cox processes. *Scandinavian Journal of Statistics*, 25(3):451–482, 1998.

Müller, P., Quintana, F., and Rosner, G. L. A product partition model with regression on covariates. *Journal of Computational and Graphical Statistics*, 20(1):260–278, 2011.

Neal, R. M. Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

Orbanz, P. and Buhmann, J. M. Nonparametric bayesian image segmentation. *International Journal of Computer Vision*, 77(1-3):25–45, 2008.

Page, G. L., Quintana, F. A., et al. Spatial product partition models. *Bayesian Analysis*, 11(1):265–298, 2016.

Pitman, J. Exchangeable and partially exchangeable random partitions. *Probability theory and related fields*, 102 (2):145–158, 1995.

Rand, W. M. Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66(336):846–850, 1971.

Reich, B. J., Hodges, J. S., Carlin, B. P., and Reich, A. M. A spatial analysis of basketball shot chart data. *The American Statistician*, 60(1):3–12, 2006.

Schwarz, G. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639, 2002.

Taddy, M. A. and Kottas, A. Mixture modeling for marked poisson processes. *Bayesian Analysis*, 7(2):335–362, 2012.

Teng, M., Nathoo, F., and Johnson, T. D. Bayesian computation for log-gaussian cox processes: A comparative analysis of methods. *Journal of Statistical Computation and Simulation*, 87(11):2227–2252, 2017.

Tobler, W. R. A computer movie simulating urban growth in the detroit region. *Economic geography*, 46(sup1): 234–240, 1970.

Zhao, P., Yang, H.-C., Dey, D. K., and Hu, G. Bayesian spatial homogeneity pursuit regression for count value data. *arXiv preprint arXiv:2002.06678*, 2020.

A. Proof of Theorem 1

We note that the full conditionals $\Pi(\lambda_i|\boldsymbol{\lambda}_i)$ can be determined up to a constant as the product of the full conditionals of each part

$$\begin{aligned}\Pi(\lambda_i|\boldsymbol{\lambda}_i) &\propto M(\lambda_i|\boldsymbol{\lambda}_{-i})P(\lambda_i|\boldsymbol{\lambda}_{-i}) \\ &\propto M(\lambda_i|\boldsymbol{\lambda}_{-i})\sum_{k=1}^{K^*} n_k^{(-i)}\delta_{\lambda_k^*}(\lambda_i) + M(\lambda_i|\boldsymbol{\lambda}_{-i})\alpha G_0(\lambda_i) \\ &\propto M(\lambda_i|\boldsymbol{\lambda}_{-i})\sum_{k=1}^{K^*} n_k^{(-i)}\delta_{\lambda_k^*}(\lambda_i) + \frac{\alpha}{Z_H}G_0(\lambda_i) \\ &\propto \sum_{k=1}^{K^*} n_k^{(-i)}\frac{1}{Z_H}\exp(-H(\lambda_i|\boldsymbol{\lambda}_{-i}))\delta_{\lambda_k^*}(\lambda_i) + \frac{\alpha}{Z_H}G_0(\lambda_i)\end{aligned} \tag{A.1}$$

As we only consider pairwise interactions $H(\lambda_i|\boldsymbol{\lambda}_{-i}) := -\eta\sum_{j\in\partial(i)} I(\lambda_i = \lambda_j)$, the support of cost function H is the set of existing cluster parameters $\lambda_1^*, \ldots, \lambda_{K^*}^*$. As a result, we have $H(\lambda_i|\boldsymbol{\lambda}_{-i}) = 0$ and hence $M(\lambda_i|\boldsymbol{\lambda}_{-i})$ when λ_i is generated from the base distribution G_0, which allows us to proceed from the second line to the third line. To reach the final line, we simply plug in the definition of $M(\lambda_i|\boldsymbol{\lambda}_{-i})$.

B. Proof of Proposition 1

Following the results of Theorem 1, the full conditional probability that region A_i belongs to an existing component c, i.e., $\exists j \neq i,\ Z_j = c$, can be derived by plugging-in the definition of likelihood and priors

$$\Pr(Z_i = c \mid \mathbf{S}, \boldsymbol{Z}_{-i}, \boldsymbol{\lambda}) \propto \frac{n_c(\boldsymbol{Z}_{-i})\exp\left(\eta\sum_{j\in\partial(i)} d_{ij}\mathbb{1}(Z_j = c)\right)}{n-1+\alpha}\frac{(\lambda_c\mu(A_i))^{N(A_i)}}{\exp(\lambda_c\mu(A_i))}. \tag{B.1}$$

The full conditional probability that A_i belongs to a new component, i.e., $\forall j \neq i,\ Z_j \neq c$, is

$$\begin{aligned}&\Pr(Z_i = c \mid \mathbf{S}, \boldsymbol{Z}_{-i}, \boldsymbol{\lambda}) \\ \propto\ &\frac{\alpha}{n-1+\alpha}\int \frac{(\lambda_c\mu(A_i))^{N(A_i))}}{\exp(\lambda_c\mu(A_i))}\frac{b^a}{\Gamma(a)}\lambda_c^{a-1}e^{-b\lambda_c}\mathrm{d}\lambda_c \\ =\ &\frac{\alpha}{n-1+\alpha}\frac{b^a}{\Gamma(a)}\mu(A_i)^{N(A_i)}\int \lambda_c^{N(A_i)+a-1}e^{-(b+\mu(A_i))\lambda_c}\mathrm{d}\lambda_c \\ =\ &\frac{\alpha b^a\Gamma(N(A_i)+a)\mu(A_i)^{N(A_i)}}{(n-1+\alpha)(b+\mu(A_i))^{N(A_i)+a}\Gamma(a)}\end{aligned} \tag{B.2}$$

C. Gibbs sampling algorithm and Dahl's method

The Dahl's method is given as

1. Define membership matrices $\mathcal{H}^{(l)} = (\mathcal{H}^{(l)}(i,j))_{i,j\in\{1,\ldots,n\}} = (\mathbb{1}(Z_i^{(l)} = Z_j^{(l)}))_{n\times n}$, where $l = 1, \ldots, L$ indexes the number of retained MCMC draws after burn-in, and $\mathbb{1}(\cdot)$ is the indicator function.
2. Calculate the average membership matrix $\overline{\mathcal{H}} = \frac{1}{L}\sum_{l=1}^{L}\mathcal{H}^{(l)}$ where the summation is element-wise.
3. Identify the most *representative* posterior sample as the one that is closest to $\overline{\mathcal{H}}$ with respect to the element-wise Euclidean distance $\sum_{i=1}^{n}\sum_{j=1}^{n}(\mathcal{H}^{(l)}(i,j) - \overline{\mathcal{H}}(i,j))^2$ among the retained $l = 1, \ldots, L$ posterior samples.

Algorithm 1 Collapsed Gibbs sampler for MRF-DPM-NHPP.

Input:

Data: event locations $\mathbf{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N\}$ where $\boldsymbol{s}_i = (x_i, y_i)$, $i = 1, \ldots, N$; regions and their neighbors $\{A_i, \partial(i) : i = 1, \ldots, n\}$.
Prior hyperparameters : a, b, α.
Tuning parameter: η.
Burn-in MCMC sample size: B, post-burn-in MCMC sample size: L.
Initial values: K, $Z_1, \ldots, Z_n$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$, iter = 1.

while iter $\leqslant B + L$ **do**
  Update $\lambda_k|\cdot$, $k = 1, \ldots, K$ as

$$\lambda_k|\cdot \sim \text{Gamma}\left(N_k + a, b + \sum_{j:Z_j=k} \mu(A_j)\right)$$

  where $N_k = \sum_{\ell:\boldsymbol{s}_\ell \in A_j, Z_j=k} 1$ is the number of points belonging to the kth component.
  Update $Z_i|\cdot$, $i = 1, \ldots, n$ following Proposition 3.1.
  iter = iter + 1.
end while

Output: Posterior samples $Z_1^{(l)}, \ldots, Z_n^{(l)}$, $\boldsymbol{\lambda}^{(l)}$, $l = B+1, \ldots, B+L$. =0

D. NBA Shot Location Data

Shot chart data for NBA players from the 2017–2018 regular season were retrieved from the official website of NBA stats.nba.com. The data for each player contain information about all his shots in regular season including game date, opponent team name, game period when each shot was made (4 quarters and a fifth period representing extra time), minutes and seconds left, success indicator (0 represents missed and 1 represents made), action type (like "Cutting dunk shot", "Jump shot", etc.), shot type (2-point or 3-point shot), shot distance, and shot location coordinates. From the data, the half court is positioned on a Cartesian coordinate system centered at the center of rim, with x ranging from -250 to 250 and y ranging from -50 to 420, both with unit of 0.1 foot (ft), as the size of an actual NBA basketball half court is 50 ft $\times$ 47 ft.

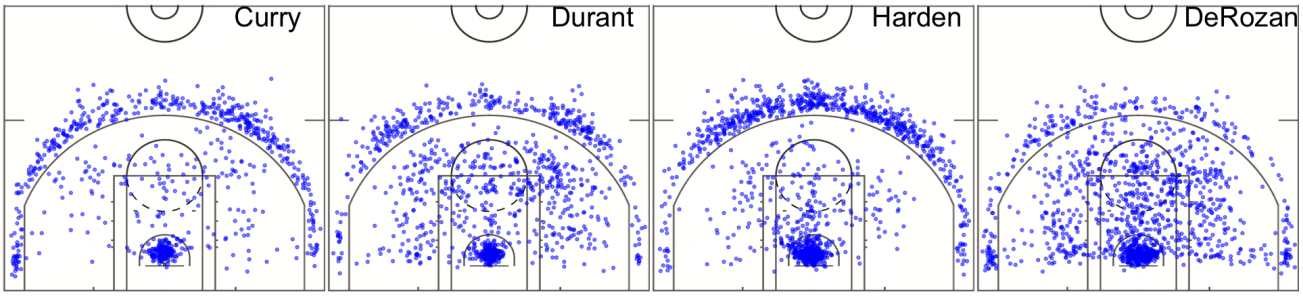

Figure 4: Shot data Display. On half court image, each point represents one shot. From left to right: Stephen Curry, Kevin Durant, James Harden, DeMar DeRozan.

We visualize and summarize the shot data of four key players, Stephen Curry, James Harden, Kevin Durant and DeMar DeRozan. Figure 4 shows their field goal attempts' locations and Table 4 summarizes their other field goal attempts information. As we can see from the plots, most of the shots are made either close to the rim or right outside the arc (i.e., 3-point line). This is in line with the recent trend in the basketball development since it is more efficient for players to pursue higher success rates near the rim or go after higher rewards by making 3-point shots.

Table 4: Shot data summary. Period is for the 1st, 2nd, 3rd, 4th quarter and the extra time. The 2PT shot percentage (%) column gives the percentage of all shots that are 2 PT shots.

Player	Shot Count	2PT shot percentage (%)	Period percentage (%)
Stephen Curry	753	42.6	(35.0, 20.6, 34.3, 9.7, 0.4)
James Harden	1306	50.2	(28.7, 22.4, 27.9, 20.8, 0.3)
Kevin Durant	1040	66.5	(30.8, 23.8, 30.6, 14.6, 0.3)
DeMar DeRozan	1274	79.9	(29.1, 28.6, 33.3, 17.3, 1.6)

E. More real data results

F. Histograms of $\widehat{K}$ in simulation study

F.1. Simulation study, grid size: 20×20

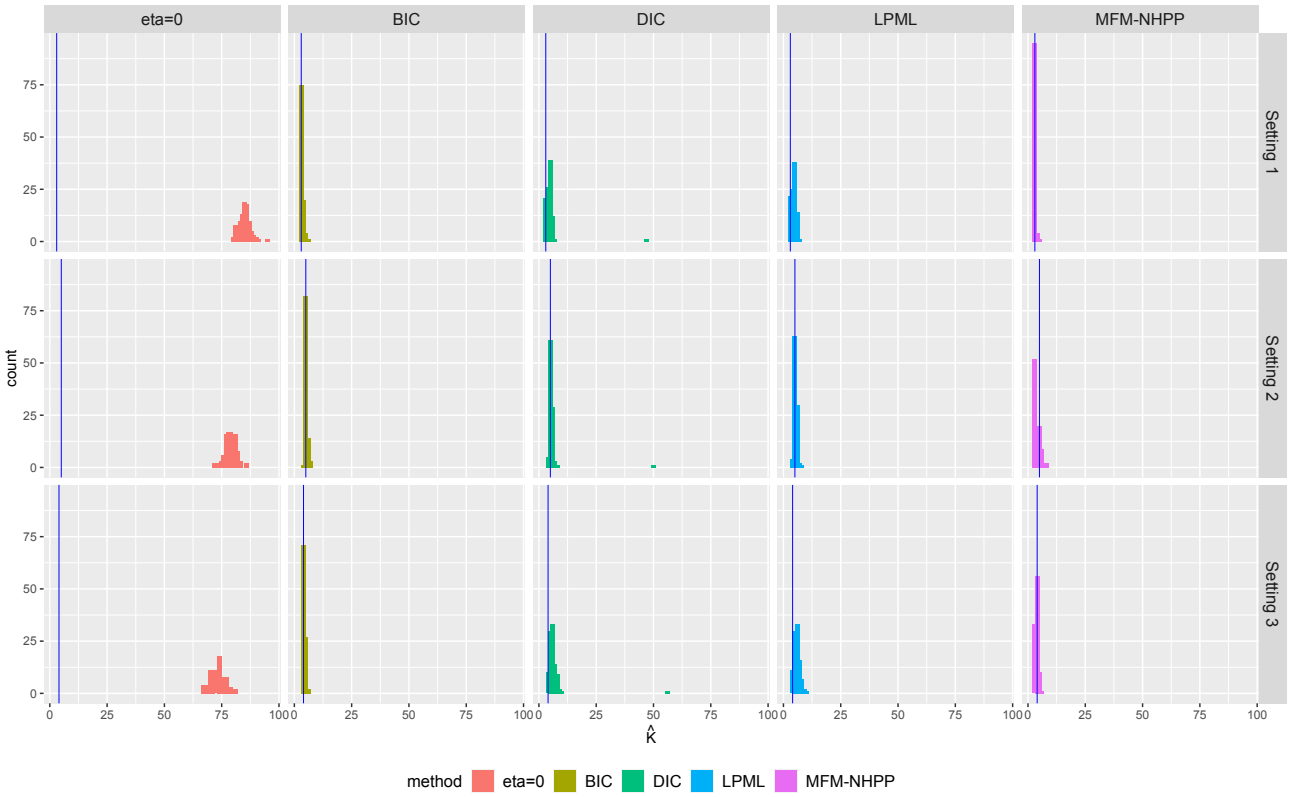

Figure 7: Histograms of $\widehat{K}$ from the 100 replicates for each simulation setting (grid size: 20×20), under MFM-NHPP, and MRF-DPM-NHPP with $\eta = 0$, optimal η selected by BIC, DIC and LPML. The blue vertical line indicates the true number of clusters (K).

Table 5: Basic information (name and the preferred position) of players and the number of clusters given by MRF-DPM-NHPP with the smoothing parameter selected by BIC, and by MFM-NHPP. Player positions: point guard (PG), shooting guard (SG), small forward (SF), power forward (PF), center (C).

		MRF-DPM-NHPP		MFM-NHPP
Player	Position	$\widehat{K}_{\text{BIC}}$	$\widehat{\eta}_{\text{BIC}}$	$\widehat{K}$
DeAndre Jordan	C	7	2.5	5
Joel Embiid	C	7	3.5	10
Karl-Anthony Towns	C	12	3.0	12
Dwight Howard	C	6	3.0	5
Giannis Antetokounmpo	PF	8	3.0	16
Blake Griffin	PF	6	3.0	4
LaMarcus Aldridge	PF	7	4.0	11
Kristaps Porziņģis	PF	6	3.0	11
Stephen Curry	PG	6	3.0	9
Damian Lillard	PG	8	3.0	9
Chris Paul	PG	6	5.0	8
Kyrie Irving	PG	8	3.0	9
Kevin Durant	SF	9	3.5	10
LeBron James	SF	8	3.0	14
Paul George	SF	9	3.0	9
Jimmy Butler	SF	8	3.5	12
James Harden	SG	7	4.0	11
DeMar DeRozan	SG	10	3.0	10
Russell Westbrook	SG	7	3.0	13
Klay Thompson	SG	6	3.0	14

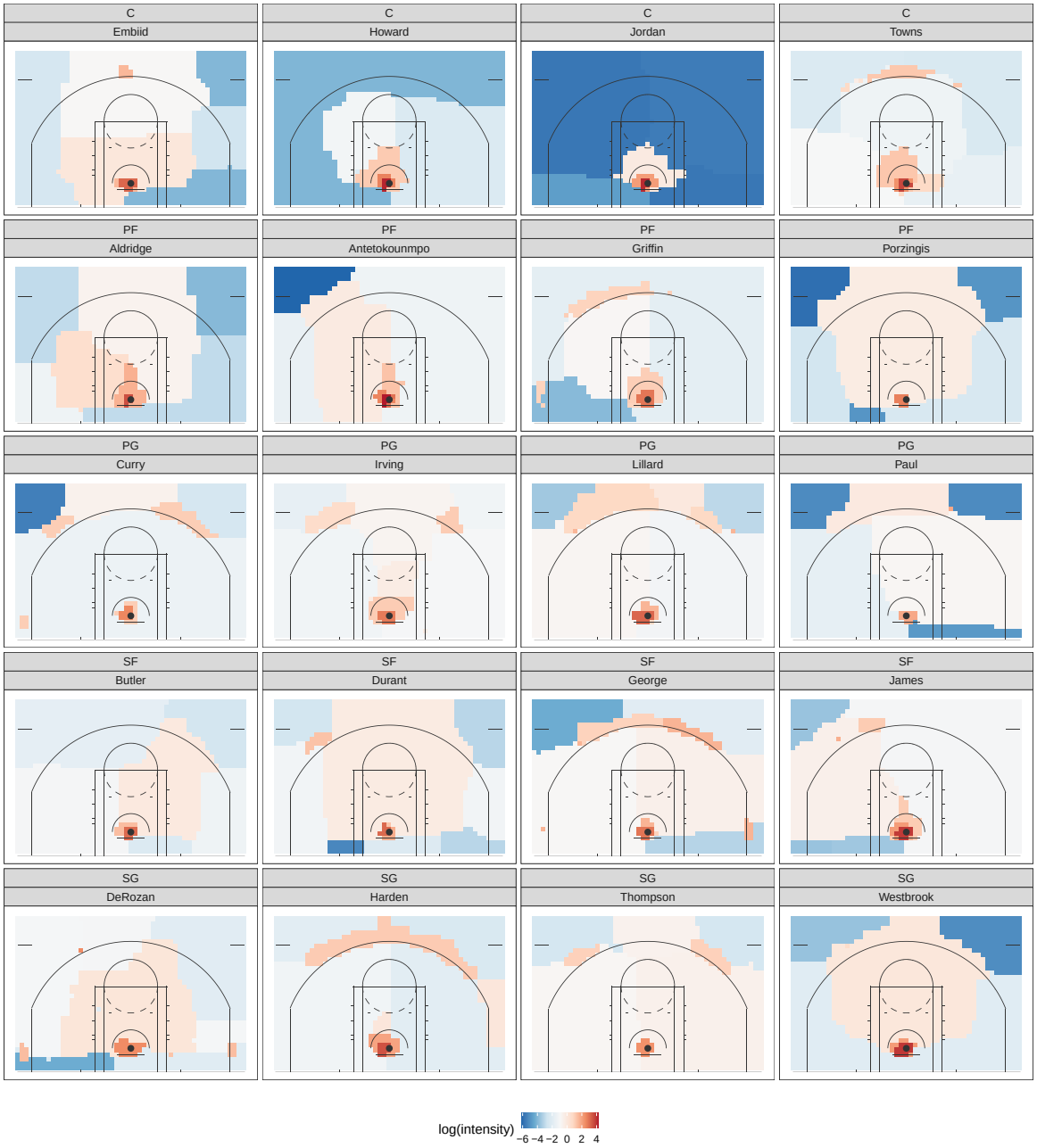

Figure 5: Estimated shooting log-intensity surfaces of selected players based on MRF-DPM-NHPP.

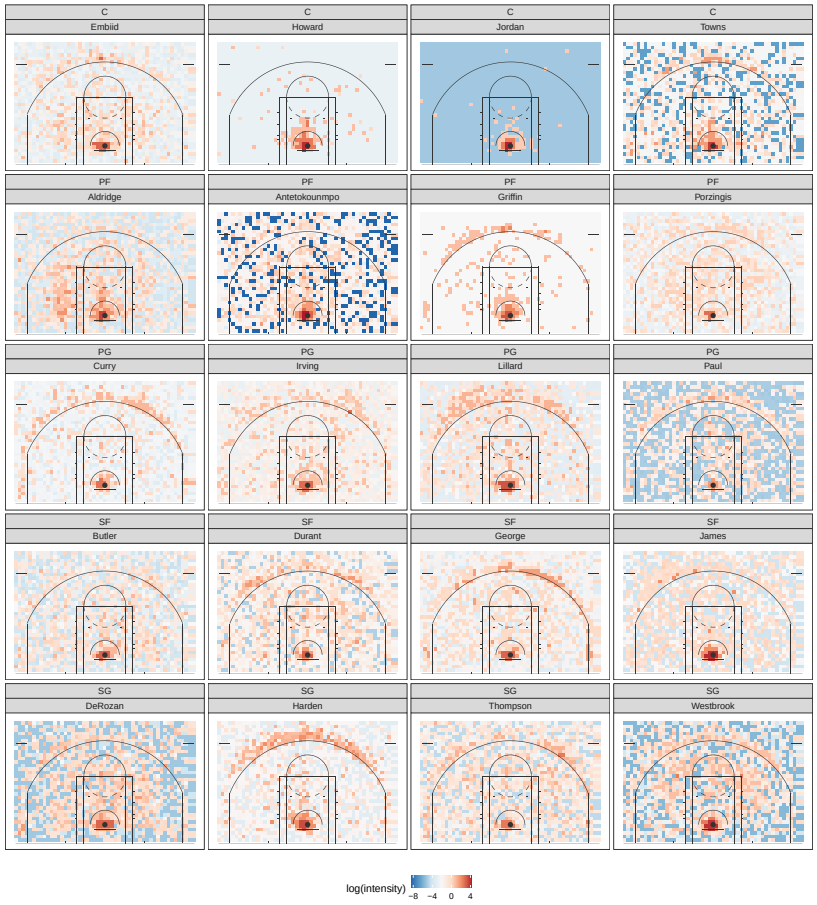

Figure 6: Estimated shooting log-intensity surfaces of selected players based on MFM-NHPP.

F.2. Simulation study, grid size: 50×35

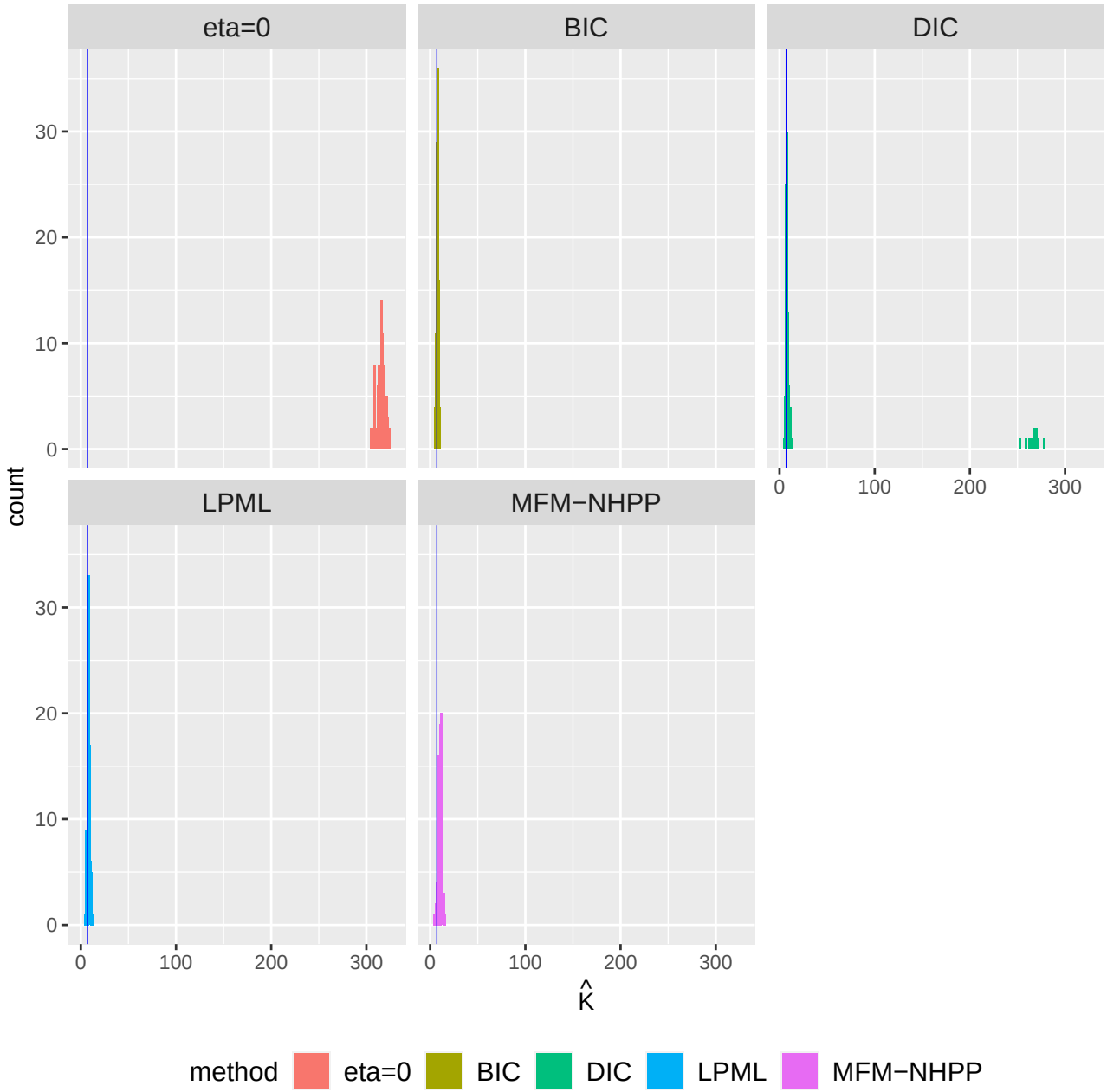

Figure 8: Histograms of $\widehat{K}$ from the 100 replicates for simulation setting ((grid size: 50×35)), under MFM-NHPP, and MRF-DPM-NHPP with $\eta = 0$, optimal η selected by BIC, DIC and LPML. The blue vertical line indicates the true number of clusters (K).

G. Traceplots of RI between ground truth and each partition in simulation study

G.1. Simulation study, grid size: 20×20

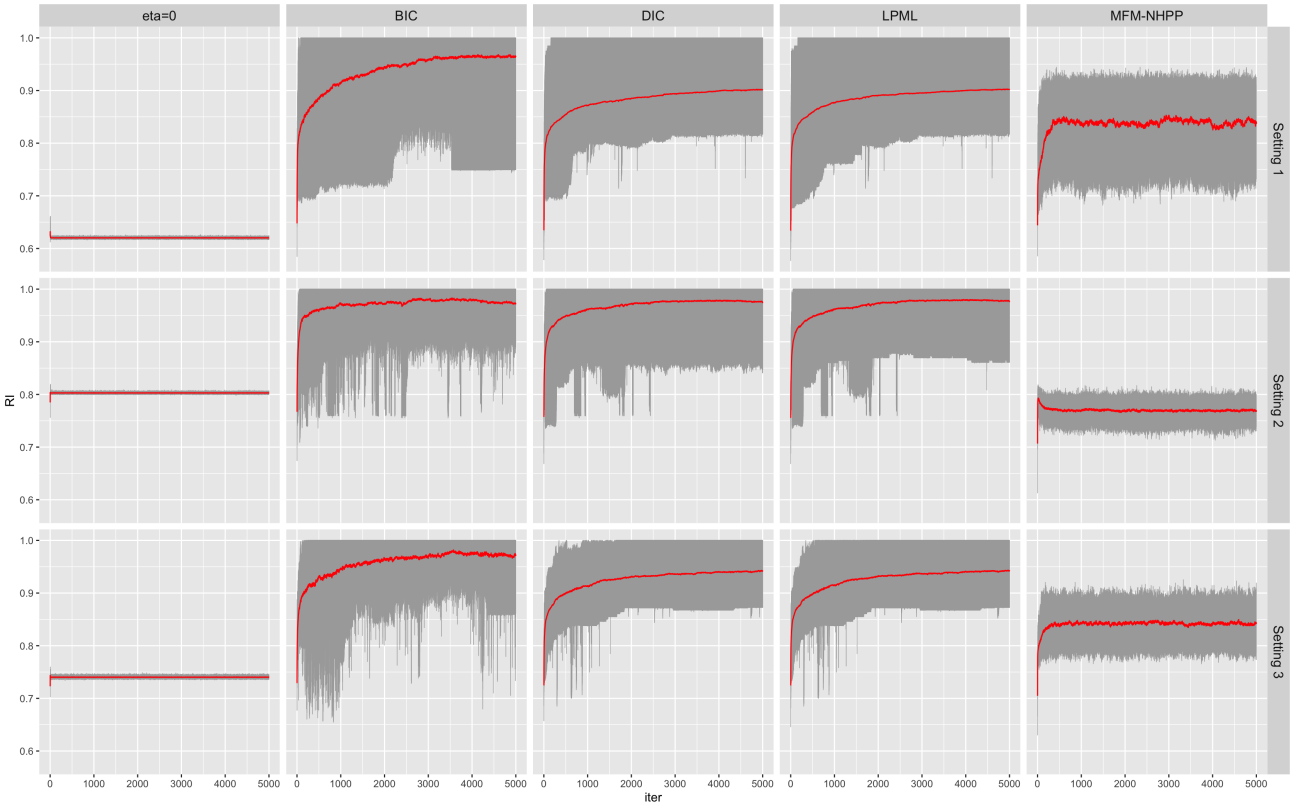

Figure 9: Overlaid trace plots of RI between the truth and the partition obtained at each iteration from 100 replicates for each simulated setting under grid size 20×20. The thick lines are the average traces over the 100 replicates.

G.2. Simulation study, grid size: 50×35

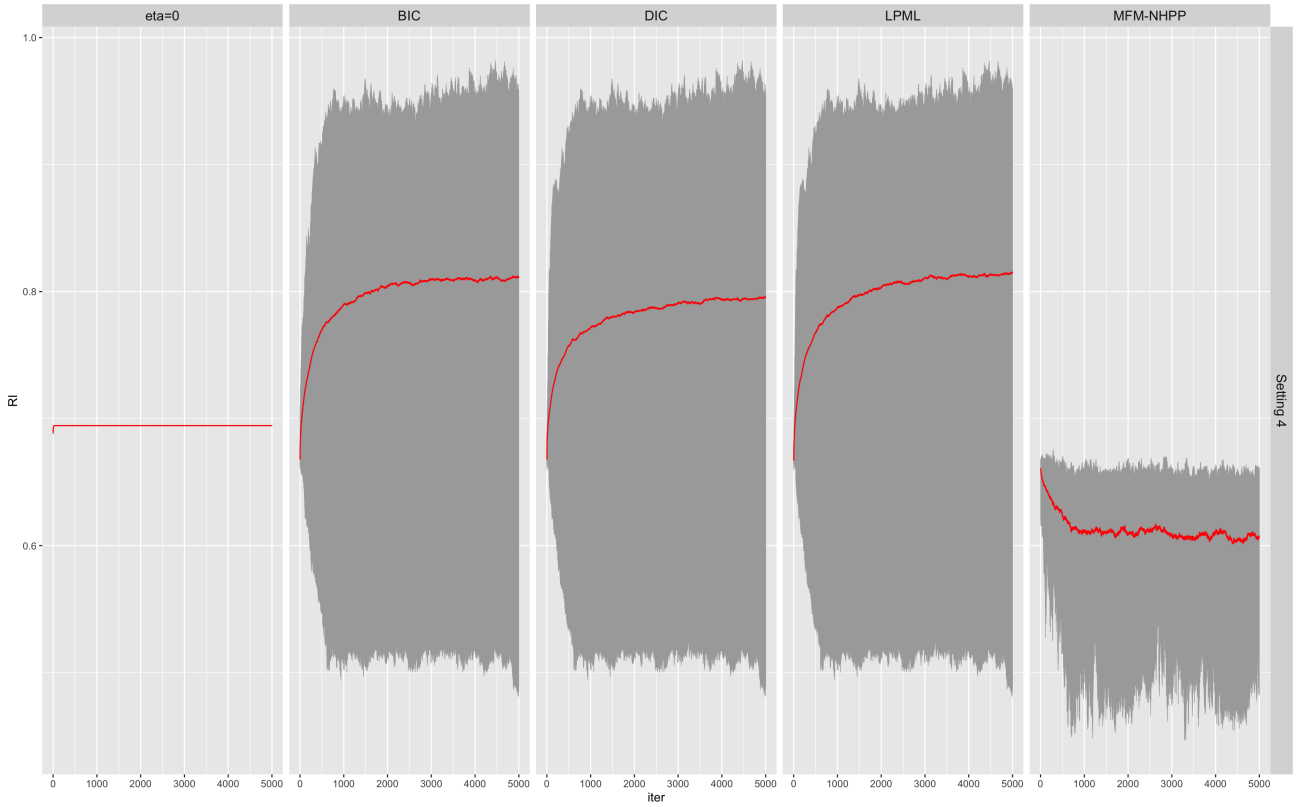

Figure 10: Overlaid trace plots of RI between the truth and the partition obtained at each iteration from 100 replicates for the simulation setting under grid size 50×35. The thick lines are the average traces over the 100 replicates.

G.3. Traceplots of estimated intensity at pixel (25,6) in NBA data analysis - full data

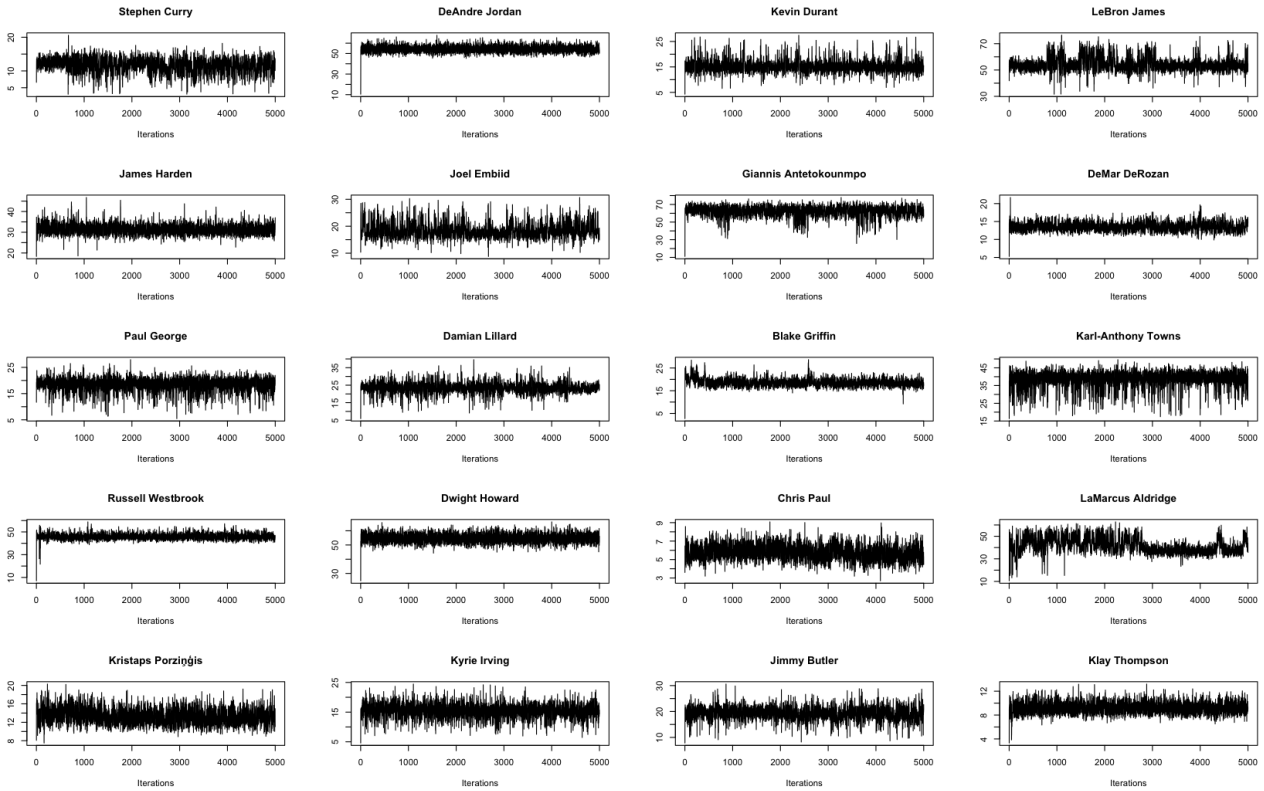

Figure 11: Traceplots of estimated intensity at pixel (25,6) for NBA data analysis after burnin and thinning for two independent chains.

G.4. Traceplots of Adjusted RI between successive partitions in NBA data analysis - full data

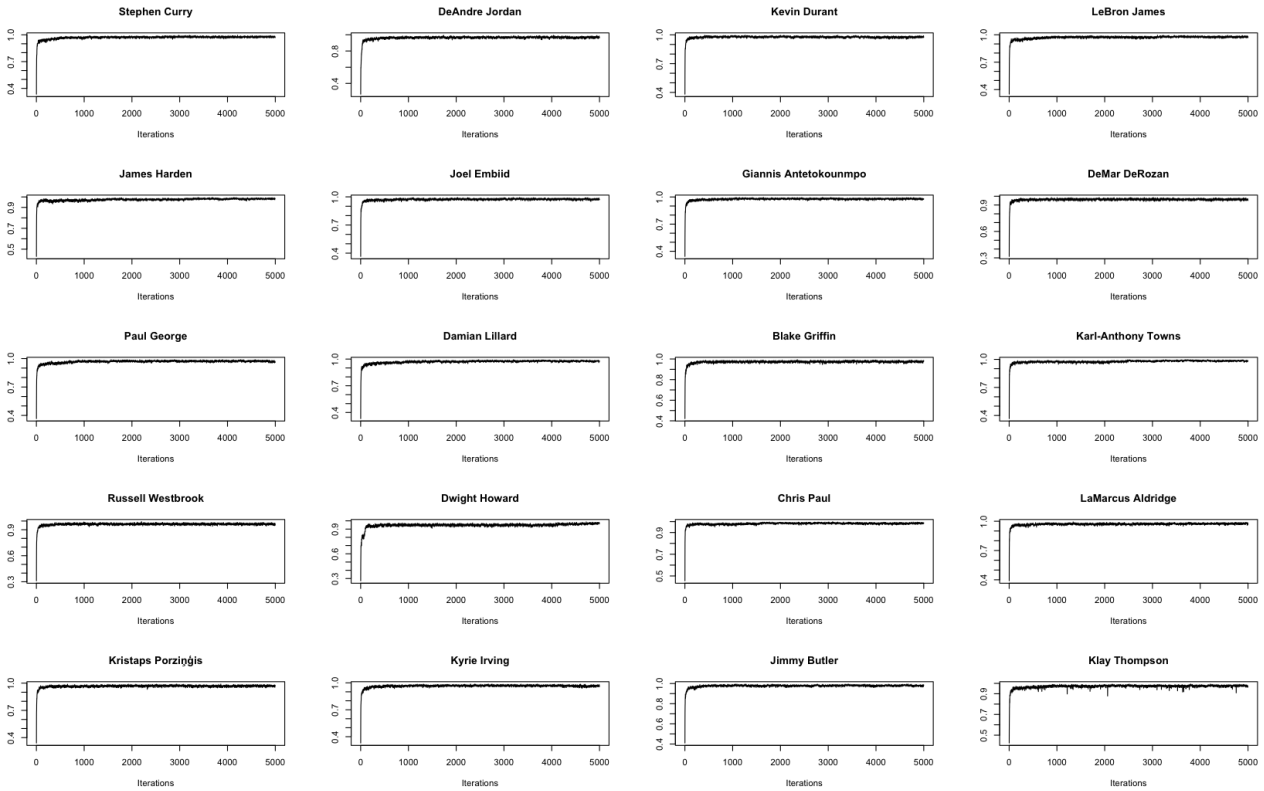

Figure 12: Traceplots of adjusted RI between successive partitions for NBA data analysis after burnin and thinning for two independent chains.

G.5. Traceplots of estimated intensity at pixel (25,6) in NBA data analysis - training data

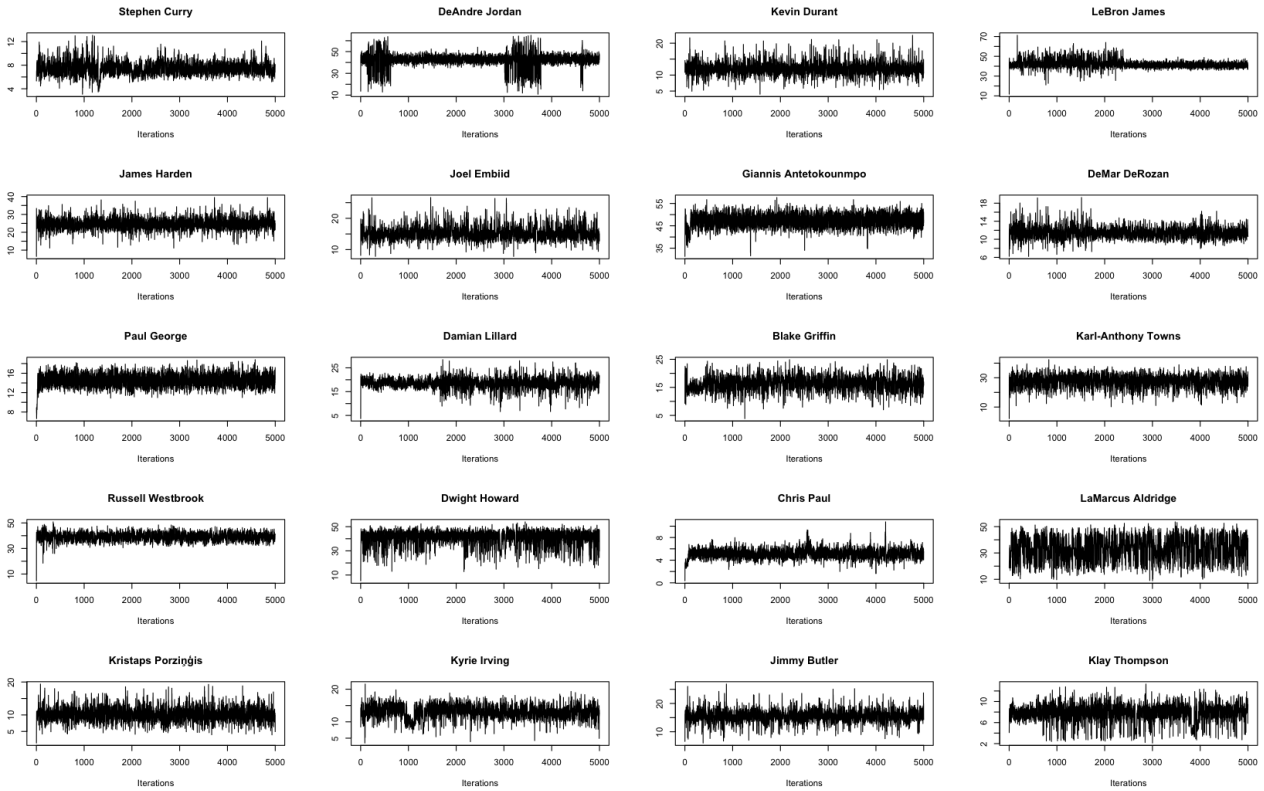

Figure 13: Traceplots of estimated intensity at pixel (25,6) for NBA data analysis, training data, after burnin and thinning for two independent chains.

G.6. Traceplots of Adjusted RI between successive partitions in NBA data analysis - training data

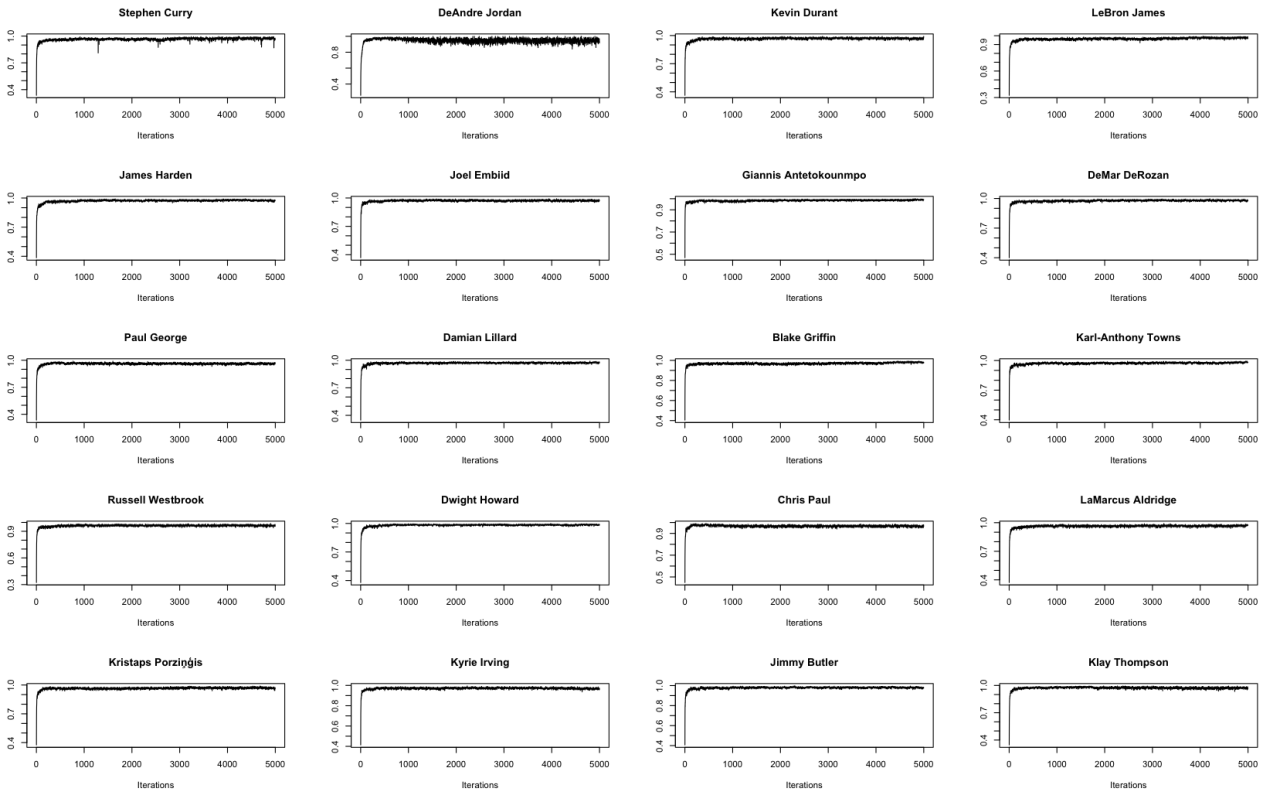

Figure 14: Traceplots of adjusted RI between successive partitions for NBA data analysis, training data, after burnin and thinning for two independent chains.

H. Simulation results, grid size: 50×35

H.1. Estimated intensity surfaces

Figure 15 displays the displays the averages of the median, 2.5th percentile, and 97.5th percentile of the estimated intensity surface obtained with the optimal η selected by BIC from the 100 replicates, in comparison with the true surfaces, for the three settings. The median surface agrees with true surface well. The 2.5th and 97.5th percentiles of the estimated intensity surfaces over 100 replicates have higher uncertainties occasionally at the boundaries where the true intensities jump, but in general are not far from the true surfaces.

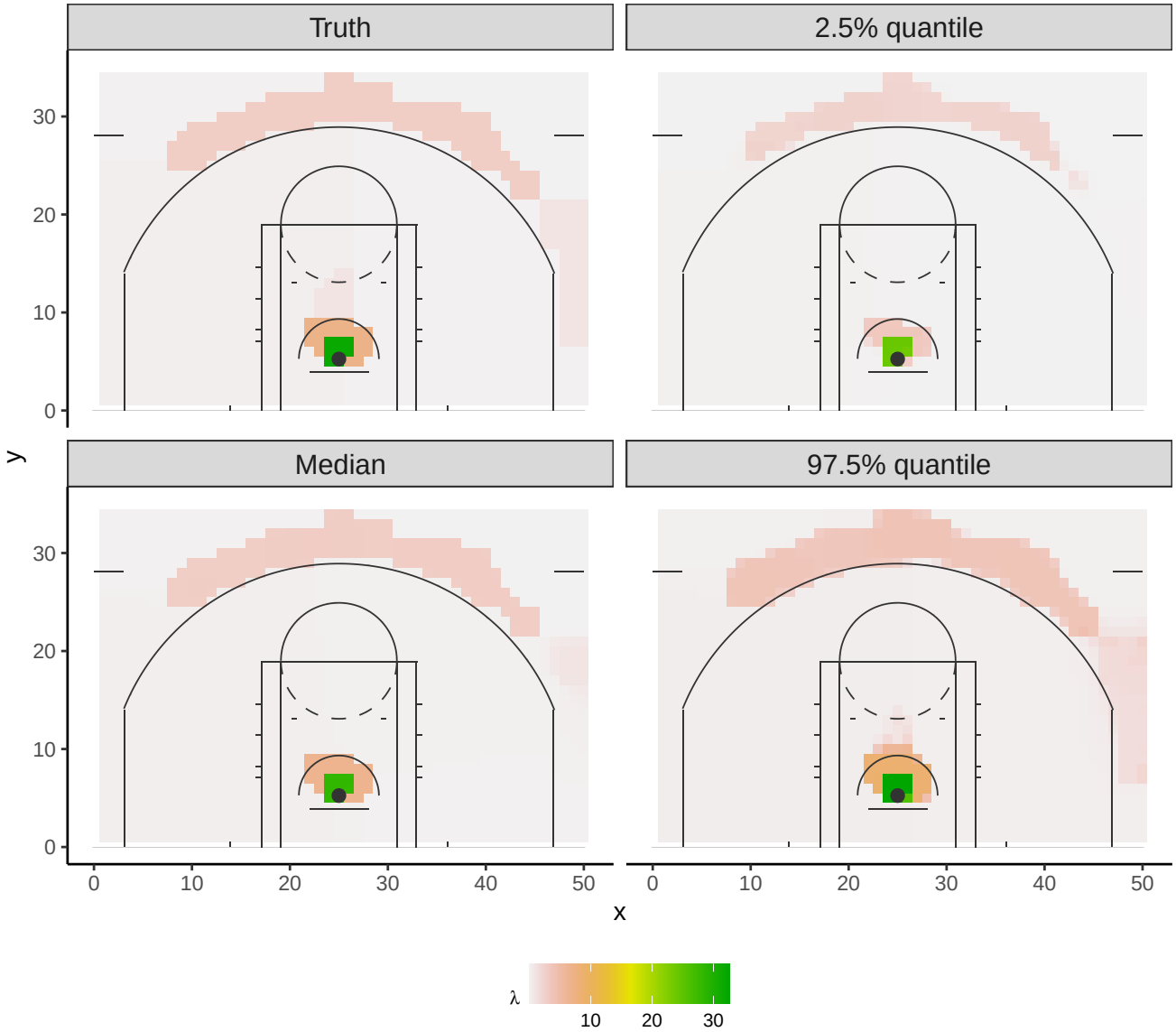

Figure 15: Simulation configurations for intensity surfaces under grid size of 50×35, with fitted intensity surfaces. Element-wise median and quantiles are calculated out of 100 replicates.

H.2. Absolute value of relative bias

Figure 16 shows the absolute value of relative bias of element-wise posterior mean estimates under the MFM-DPM-NHPP and other competing methods. We note that MFM-DPM-NHPP yields the smallest biases.

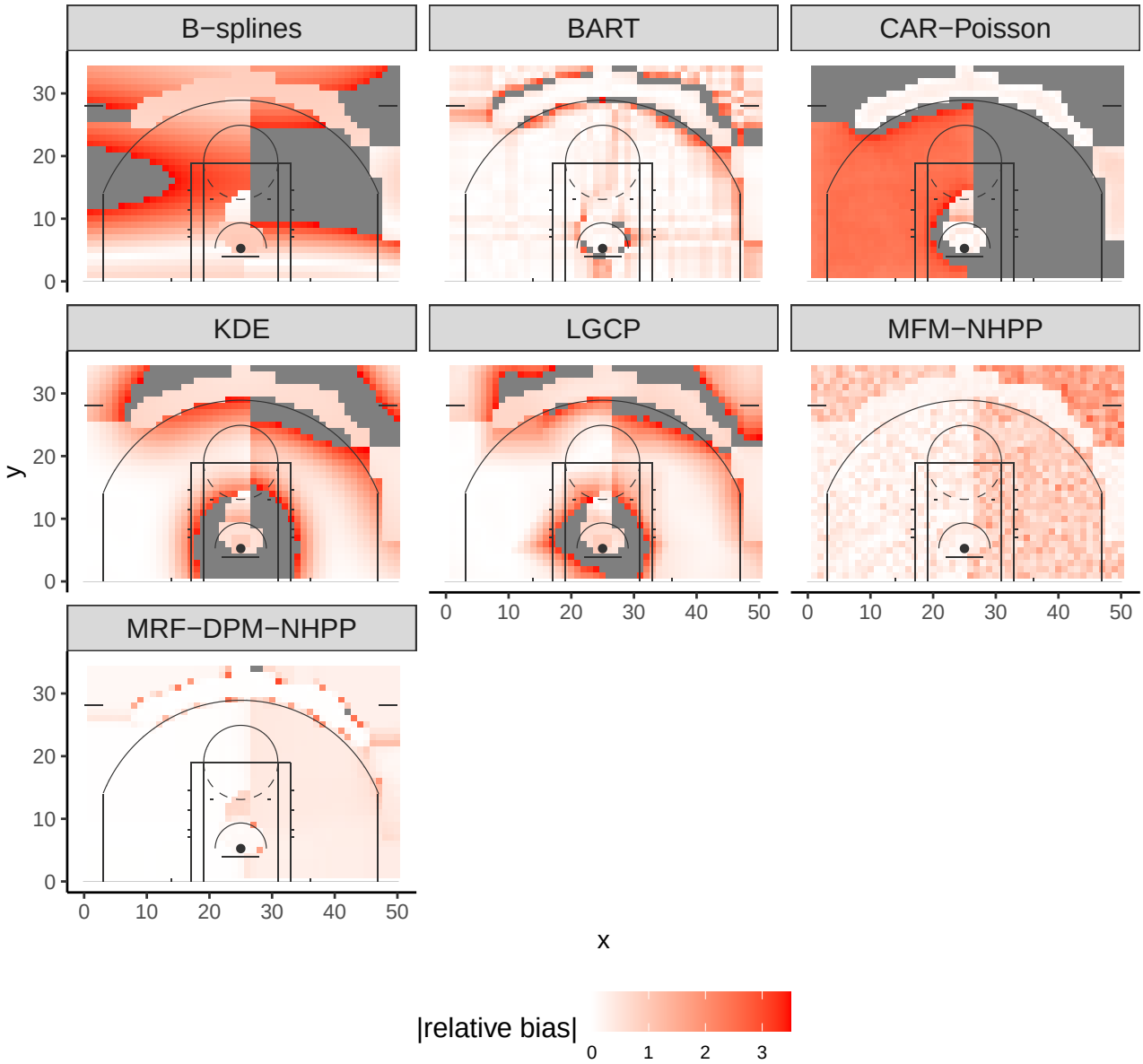

Figure 16: Absolute value of relative bias of element-wise posterior mean estimates for intensity surfaces (grid size: 50×35). Dark grey values correspond to regions with absolute bias beyond 3.5.

I. Model assessment for the case study - MAE results

To examine the intensity fitness of the proposed method compared to several classical alternative methods, we consider a predictive assessment criteria based on p-thinning approach (Illian et al., 2008). Letting p denotes the retention probability, the p-thinning procedure proceeds by first independently moving each point $\boldsymbol{s}_i, i = 1, \ldots, N$ with probability $1 - p$ from training data to test data. Given a partition of $\mathcal{B}$, that is, $A_1, A_2, \ldots, A_n$, $\cup_{i=1}^n A_i = \mathcal{B}$ and $A_i \cap A_j = \emptyset, \forall i \neq j$. We consider the mean absolute error (MAE) defined as MAE $= \frac{1}{n}\sum_{i=1}^n \left|\frac{1-p}{p}\widehat{\lambda}(A_i) - N(A_i)\right|$, where $\widehat{\lambda}(A_i)$ is the estimated intensity of region A_i based on the training data and $N(A_i)$ is the number of observed points falling into region A_i in the test data. Here we fix $p = 0.8$ as Geng et al. (2021) and note that the model class with smaller MAE value fits the data

better from a prediction perspective.

In addition to the methods listed in Table 1, we also consider the method introduced in Miller et al. (2014), where we fit a non-homogeneous Poisson process with their basis as the trend terms using function `ppm` from package `spatstat`. For MRF-DPM-NHPP and MFM-NHPP, we find the MCMC setting detailed at the beginning of this section can work well for the training data obtained after p-thinning (see Appendix G for traceplots). The MAE results are summarized in Table 6 and we find that the proposed method is comparable to BART and MFM-NHPP and clearly superior to all other methods. Moreover, we run a simulation study using the fitted intensity of James Harden (Figure 5) under the grid size of 50ft $\times$ 35ft, and the results also confirm the advantage of the proposed method over a wide range of alternative methods. We summarize the MAE results in the model assessment section in Table 6, and we note that the proposed method (MRF-DPM-NHPP) is comparable to BART and MFM-NHPP and clearly superior to all other methods in terms of predictive performance.

J. Additional simulation results when KDE is the true data generating model

We fit kernel density estimate to the Durant's shot location data and use that model as the proxy to a continuous intensity surface. We generate synthetic data from this model and evaluate the performance of different methods following the predictive assessment criteria (Table 7). We see that the proposed method can yield comparable performance to other methods under this scenario.

K. Sensitivity analysis

We performed analysis on the shot location data under different grid partitions 2ft $\times$ 2ft, 2ft $\times$ 1ft, 1ft $\times$ 1ft, and we converted the resulting intensity surfaces to be under the grid size 2ft $\times$ 2ft using function `as.im` in **R** package `spatstat`. As shown in Figure 17, the results are not sensitive to the grid partitions.

Table 6: MAE results. The MAE for B-splines fitted to the training data from DeAndre Jordan is very large due to its poor fit at the boundary.

	position	MRF-DPM-NHPP	MFM-NHPP	LGCP	CAR-Poisson	BART	KDE	B-splines	Miller
DeAndre Jordan	C	0.041	0.044	0.077	0.300	0.038	0.100	$>> 1$	0.109
Joel Embiid	C	0.147	0.143	0.150	0.304	0.137	0.167	0.181	0.181
Karl-Anthony Towns	C	0.162	0.156	0.193	0.317	0.149	0.215	0.241	0.245
Dwight Howard	C	0.077	0.083	0.117	0.300	0.075	0.144	0.171	0.172
Giannis Antetokounmpo	PF	0.158	0.153	0.201	0.321	0.153	0.228	0.252	0.255
Blake Griffin	PF	0.141	0.140	0.151	0.312	0.137	0.171	0.182	0.186
LaMarcus Aldridge	PF	0.181	0.199	0.207	0.317	0.176	0.226	0.232	0.236
Kristaps Porziņģis	PF	0.169	0.175	0.174	0.318	0.168	0.183	0.183	0.183
Stephen Curry	PG	0.137	0.134	0.146	0.302	0.131	0.158	0.160	0.152
Damian Lillard	PG	0.192	0.201	0.229	0.331	0.190	0.251	0.269	0.269
Chris Paul	PG	0.128	0.126	0.130	0.296	0.129	0.138	0.138	0.138
Kyrie Irving	PG	0.167	0.177	0.186	0.315	0.167	0.200	0.209	0.208
Kevin Durant	SF	0.190	0.192	0.204	0.327	0.185	0.218	0.215	0.213
LeBron James	SF	0.200	0.199	0.241	0.329	0.181	0.273	0.292	0.298
Paul George	SF	0.200	0.200	0.233	0.334	0.194	0.250	0.255	0.249
Jimmy Butler	SF	0.149	0.156	0.162	0.311	0.150	0.174	0.181	0.181
James Harden	SG	0.192	0.196	0.237	0.330	0.182	0.264	0.287	0.280
DeMar DeRozan	SG	0.215	0.218	0.221	0.330	0.206	0.237	0.247	0.249
Russell Westbrook	SG	0.200	0.209	0.261	0.340	0.207	0.294	0.300	0.305
Klay Thompson	SG	0.188	0.190	0.191	0.321	0.182	0.204	0.203	0.200

Table 7: MAE results when the underlying true data generating process is KDE.

MRF-DPM-NHPP	LGCP	CAR-Poisson	BART	KDE	B-splines	Miller
0.228	0.217	0.322	0.216	0.219	0.220	0.220

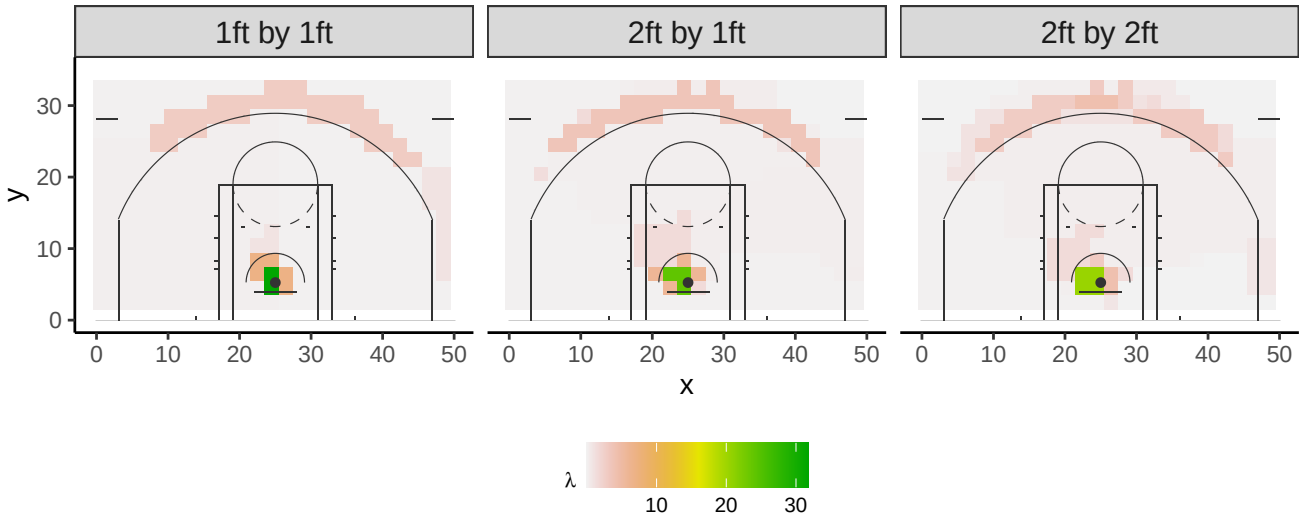

Figure 17: Sensitivity analysis based on different grid partitions, 2ft $\times$ 2ft, 2ft $\times$ 1ft, 1ft $\times$ 1ft for the shooting location data of James Harden.