A Bayesian Joint Model for Spatial Point Processes with Application to Basketball Shot Chart

Jieying Jiao

University of Connecticut, Storrs, USA.

E-mail: jieying.jiao@uconn.edu

Guanyu Hu

University of Connecticut, Storrs, USA.

E-mail: guanyu.hu@uconn.edu

Jun Yan

University of Connecticut, Storrs, USA.

E-mail: jun.yan@uconn.edu

Summary. The success rate of a basketball shot may be higher at locations where a player makes more shots. In a marked spatial point process model, this means that the marks are dependent on the intensity of the process. We develop a Bayesian joint model of the mark and the intensity of marked spatial point process, where the intensity is incorporated in the model for the mark as a covariate. Further, we allow variable selection through the spike-slab prior. Inferences are developed with a Markov chain Monte Carlo algorithm to sample from the posterior distribution. Two Bayesian model comparison criteria, the modified Deviance Information Criterion and the modified Logarithm of the Pseudo-Marginal Likelihood, are developed to assess the fitness of different models focusing on the mark. The empirical performances of the proposed methods are examined in extensive simulation studies. We apply the proposed methods to the shot charts of four players in the NBA's 2017–2018 regular season to analyze the shot intensity in the field and the field goal percentage. The results suggest that the field goal percentages of three players are significantly positively dependent on their shot intensities, and that different players have different predictors for their field goal percentages.

Keywords: MCMC; Model Selection; Sports Analytic; Variable Selection

1. Introduction

Shot charts are important summaries for basketball coaches. A shot chart is a spatial representation of the location and the result of each shot attempt by one player. Good defense strategies depend on good understandings of the offense players' tendencies to shoot and abilities to score. Reich et al. (2006) proposed hierarchical spatial models with spatially-varying covariates for shot attempt frequencies over a grid on the court and for shot success with shot locations fixed. Spatial point processes are natural to model random locations (e.g., Cressie, 2015; Diggle, 2013). Miller et al. (2014) used a low dimensional representation of related point processes to analyze shot attempt locations.

Franks et al. (2015) combined spatial and spatio-temporal processes, matrix factorization techniques, and hierarchical regression models to analyze defensive skill. A shot chart can be viewed as a spatial point process with a binary mark, where the location and the mark are both random (e.g., Banerjee et al., 2014, Ch. 8). Many parametric models for spatial point process have been proposed in the literature, including the Poisson process (Geyer, 1999; Ord, 2004), the Gibbs process (Møller and Waagepetersen, 2003), and the log Gaussian Cox process (Møller et al., 1998). A marked point process model is a natural way to model a point process and its associated marks (Møller and Waagepetersen, 2003; Vere-Jones and Schoenberg, 2004).

The success rate of a basketball shot may be higher at locations in the court where a player makes more shots. In a marked spatial point process model, this means that the mark and the point pattern are dependent. There are two approaches to model how the mark depends on the point process. Location dependent models (Mrkvička et al., 2011) are observation driven, where the observed point pattern is incorporated into characterizing the spatially varying distribution of the mark. Intensity dependent models (Ho and Stoyan, 2008) are parameter driven, where the intensity instead of the observed point pattern of the point process characterizes the distribution of the mark at each point in the spatial domain. To the best of our knowledge, no work has been done to jointly model the intensity of the shot attempts and the results of the attempts.

The contribution of this paper is two-fold. First, we propose a Bayesian joint model of marked spatial point process to analyze the spatial intensity and the outcome of shot attempts simultaneously. In particular, we use a non-homogeneous Poisson point process to model the spatial pattern of the shot attempts and incorporate the intensity of the process as a covariate in the model for the success rate of each shot. Second, we propose the variable selection with the spike-slab prior to identify important covariates for the model of success rate. Inferences are made with Markov chain Monte Carlo (MCMC). The modified Deviance Information Criterion (mDIC) and the modified Logarithm of the Pseudo-Marginal Likelihood (mLPML) are introduced to assess the fitness of our proposed model.

The rest of the paper is organized as follows. In Section 2, the shot chart data of top players from the 2017–2018 NBA regular season are introduced. In Section 3, we develop the Bayesian joint model of marked spatial point process and variable selection with the spike-slab prior. Details of the Bayesian computation are presented in Section 4, including the MCMC algorithm and the two model selection criteria. Extensive simulation studies are summarized in Section 5 to investigate empirical performance of the proposed methods. Applications of the proposed methods to 4 NBA players are reported in Section 6. Section 7 concludes with a discussion.

2. Shot Chart Data

The website `stats.nba.com` provides shot data for NBA players. We focus on the 2017–2018 regular NBA season here. For each player, the dataset contains information about each of his shots in this season including game date, opponent team, game period when the shot was made (four quarters and a fifth period representing extra time), minutes and seconds left to the end of that period, success indicator or mark (0 for missed and 1 for
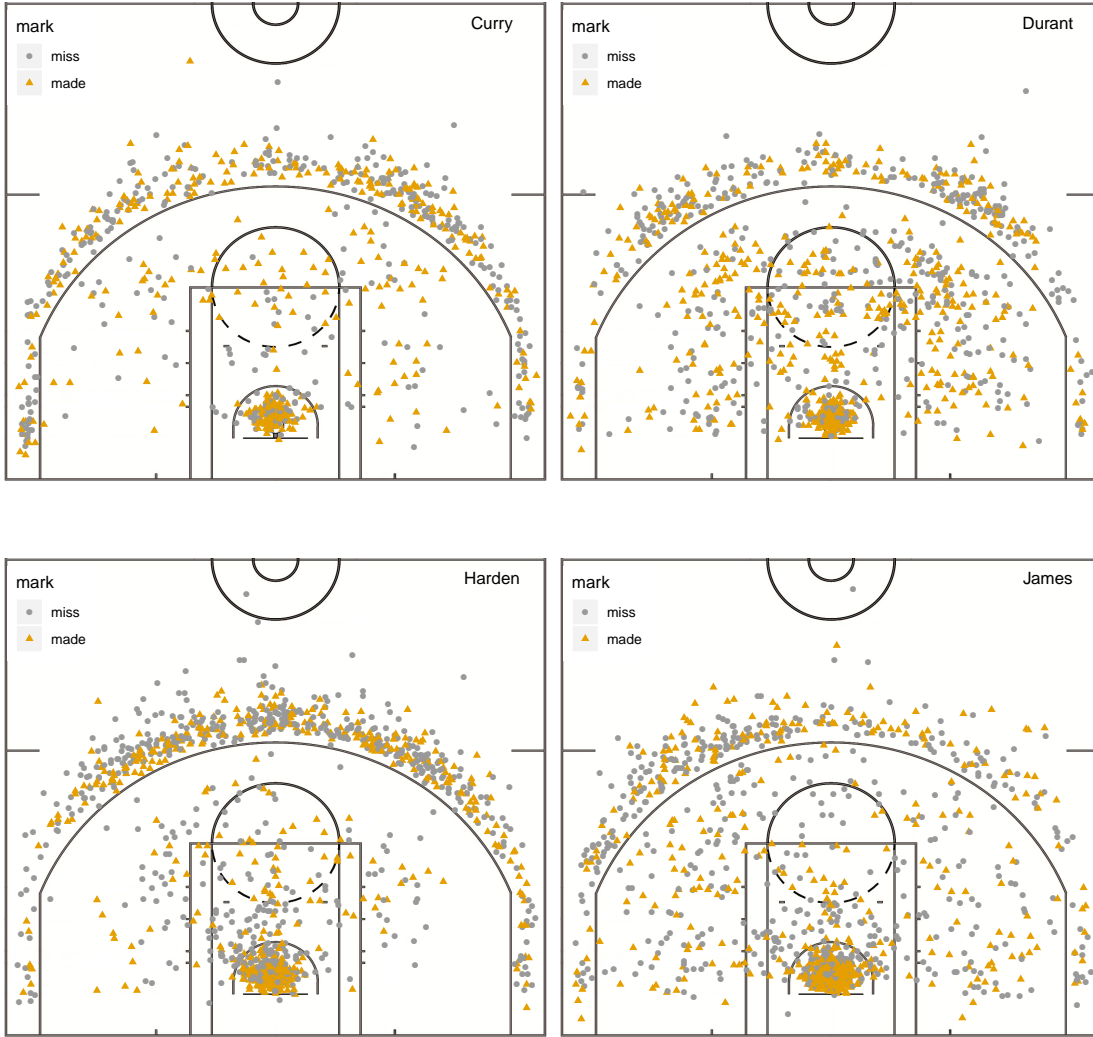
Fig. 1. Shot charts of Curry, Durant, Harden and James in the 2017–2018 regular NBA season.

made), shot type (2-point or 3-point shot), shot distance, and shot location coordinates, among others. From the data, the half court is positioned using a Cartesian coordinate system (x, y) with the origin placed at the center of basket rim, x ranging from -25 to 25 feet and y ranging from -5 to 42 feet. Euclidean distance of a location to the origin is rounded to foot. Shots made beyond the half court, which are very rare and not of our interest, are excluded.

From the top 20 players ranked by the website, we chose four players with quite different styles: Stephen Curry, Kevin Durant, James Harden and LeBron James. Figure 1 shows their shot locations with the shot success indicators. Table 1 summarizes the count, success rate, percentage of 2-point shots, and percentage of the shots in each

Table 1. Shot data summary for four players in the 2017–2018 regular NBA season. The period includes 4 quarters and over time.

Player	# Shots	Success rate (%)	2-point shot (%)	Period(%)
Curry	740	50.0	43.4	(35.0, 20.7, 34.1, 9.9, 0.4)
Durant	1032	52.5	66.7	(30.9, 23.6, 30.4, 14.7, 0.3)
Harden	1286	45.6	50.6	(28.7, 22.3, 27.7, 21.0, 0.3)
James	1409	54.3	74.6	(29.1, 21.5, 25.3, 23.6, 0.4)

period. As shown in Figure 1, most of the shots were made close to the rim or out of but close to the 3-point line. This is expected since shorter distance should give higher shot accuracy for either 2-point or 3-point shots. The overall shot success rates for the four players were all close to 50%. James had the highest percentage (74.6%) of 2-point shot while Curry had the highest percentage (56.6%) of 3-point shot. About half and 2/3 of the shots made by Harden and Durant, respecitively were 2-point shots. Curry and Durant's shot percentages were very similar for the first three periods and smaller for the fourth period. Harden and James had relatively similar shot percentages across four quarters.

3. Bayesian Joint Model of Marked Spatial Point Process

The observed shot chart of a player is represented by $(\mathbf{S}, \mathbf{M})$, where $\mathbf{S}$ is the collection of the locations of shot attempts (x and y coordinates) and $\mathbf{M}$ is the vector of the corresponding marks (1 means success and 0 means failure). Assuming that N shots were observed, we have $\mathbf{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_N)$ and $\mathbf{M} = (m(\boldsymbol{s}_1), m(\boldsymbol{s}_2), \ldots, m(\boldsymbol{s}_N))$.

3.1. Marked Spatial Point Process

We propose to model $(\mathbf{S}, \mathbf{M})$ by a marked spatial point process. The shot locations $\mathbf{S}$ are modeled by a non-homogeneous Poisson point process (e.g., Diggle, 2013). Let $\mathcal{B} \subset \mathbb{R}^2$ be a subset of the half basketball court on which we are interested in modeling the shot intensity. A Poisson point process is defined such that $N(A) = \sum_{i=1}^{N} 1(\boldsymbol{s}_i \in A)$ for any $A \subset \mathcal{B}$ follows a Poisson distribution with mean $\lambda(A) = \int_A \lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s}$, where $\lambda(\cdot)$ defines an intensity function of the process. The likelihood of the observed locations $\mathbf{S}$ is

$$\prod_{i=1}^{N} \lambda(\boldsymbol{s}_i) \exp\left(-\int_{\mathcal{B}} \lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right).$$

Covariates can be incorporated into the intensity by setting

$$\lambda(\boldsymbol{s}_i) = \lambda_0 \exp\left(\mathbf{X}^{\top}(\boldsymbol{s}_i)\boldsymbol{\beta}\right), \tag{1}$$

where λ_0 is the baseline intensity, $\mathbf{X}(\boldsymbol{s}_i)$ is a $p \times 1$ spatially varying covariate vector, and $\boldsymbol{\beta}$ is the corresponding coefficient vector.

Next we consider modeling the success indicator (mark). It is natural to suspect that the success rate of shot attempts is higher at locations with higher shot intensity,

suggesting an intensity dependent mark model. In particular, the success indicator is modeled by a logistic regression

$$
\begin{aligned}
m(\boldsymbol{s}_i) \mid \mathbf{Z}(\boldsymbol{s_i}) &\sim \text{Bernoulli}\big(\theta(\boldsymbol{s}_i)\big), \\
\text{logit}\big(\theta(\boldsymbol{s}_i)\big) &= \xi\lambda(\boldsymbol{s}_i) + \mathbf{Z}^{\top}(\boldsymbol{s}_i)\boldsymbol{\alpha},
\end{aligned}
\tag{2}
$$

where $\lambda(\boldsymbol{s}_i)$ is the intensity defined in (1) with a scalar coefficient ξ, $\mathbf{Z}(\boldsymbol{s}_i)$ is a $q \times 1$ covariate vector evaluated at i-th data point ($\mathbf{Z}$ does not need to be spatial, like period covariates), and $\boldsymbol{\alpha}$ is a $q \times 1$ vector of coefficient.

With $\boldsymbol{\Theta} = (\lambda_0, \boldsymbol{\beta}, \xi, \boldsymbol{\alpha})$, the joint likelihood for the observed marked spatial point process $(\mathbf{S}, \mathbf{M})$ is

$$
\begin{aligned}
L(\boldsymbol{\Theta} \mid \mathbf{S}, \mathbf{M}) \propto &\prod_{i=1}^{N} \theta(\boldsymbol{s}_i)^{m(\boldsymbol{s}_i)} \left(1 - \theta(\boldsymbol{s}_i)\right)^{1-m(\boldsymbol{s}_i)} \\
&\times \left(\prod_{i=1}^{N} \lambda(\boldsymbol{s}_i)\right) \exp\left(-\int_{\mathcal{B}} \lambda(\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right).
\end{aligned}
\tag{3}
$$

3.2. Prior Specification

Vague priors are specified for the model parameters. For λ_0, the gamma distribution is a conjugate prior (e.g., Leininger et al., 2017). For $\boldsymbol{\beta}$, ξ, or $\boldsymbol{\alpha}$, there is no simple conjugate prior and we specify a vague, independent normal prior. In summary, we have

$$
\begin{aligned}
\lambda_0 &\sim \text{G}(a, b), \\
\boldsymbol{\beta} &\sim \text{MVN}(\mathbf{0}, \sigma^2\mathbf{I}_p), \\
\xi &\sim \text{N}(0, \delta^2), \\
\boldsymbol{\alpha} &\sim \text{MVN}(\mathbf{0}, \delta^2\mathbf{I}_q),
\end{aligned}
\tag{4}
$$

where $\text{G}(a, b)$ represents a Gamma distribution with shape a and scale b, respectively, $\text{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$ is a multivariate normal distribution with mean vector $\mathbf{0}$ and variance matrix $\boldsymbol{\Sigma}$, a, b, σ^2 and δ^2 are hyper-parameters to be specified, and $\mathbf{I}_k$ is the k-dimensional identity matrix.

3.3. Bayesian Variable Selection

The variable selection procedure aims to find those coefficients that are significantly different from zero. For our problem, we focus on variable selection for the mark model (2). The same method can be applied to the intensity model (1). In the intensity dependent mark model, the intensity λ is always present and, hence, is not selected. Also kept is the intercept in $\boldsymbol{\alpha}$.

A widely used Bayesian variable selection method is to use the spike-slab prior independently on covariate coefficients (?). This prior is a mixture of a nearly degenerated distribution at zero (the spike) and a flat distribution (the slab). One simple choice for the spike and slab is mean zero normal distributions with certain small and large variances, respectively. The ratio of the small variance to the large variance should

not be too small to avoid MCMC getting stuck in the spike component (Malsiner-Walli and Wagner, 2018). George and McCulloch (1993) recommended a ratio of $1/10,000$. A common choice is 0.01 and 100 for the small and large variances, respectively. To summarize, the spike-slab prior for each α_i we want to select is specified by

$$
\begin{aligned}
\alpha_i &\sim \mathrm{N}(0, \delta_i^2), \\
\delta_i^2 &= 0.01(1 - \gamma_i) + 100\gamma_i, \\
\gamma_i &\sim \mathrm{Bernoulli}(\phi_i), \\
\phi_i &\sim \mathrm{Beta}(0.5, 0.5).
\end{aligned}
\tag{5}
$$

MCMC sampling is used to draw posterior samples for each of these parameters. Since γ_i is a binary random variable, its posterior mode is used to decide on the importance of Z_i. A posterior mode of zero means that α_i is not significant and Z_i should not be included in the model.

4. Bayesian Computation

4.1. The MCMC Sampling Schemes
The posterior distribution of $\boldsymbol{\Theta}$ is

$$
\pi(\boldsymbol{\Theta}|\mathbf{S}, \mathbf{M}) \propto L(\boldsymbol{\Theta} \mid \mathbf{S}, \mathbf{M})\pi(\boldsymbol{\Theta}),
\tag{6}
$$

where $\pi(\boldsymbol{\Theta}) = \pi(\lambda_0)\pi(\boldsymbol{\beta})\pi(\xi)\pi(\boldsymbol{\alpha})$ is the joint prior density as specified in (4) or (5). In practice, we used vague priors with hyper-parameters $\sigma^2 = \delta^2 = 100$ and $a = b = 0.01$ in (4).

To sample from the posterior distribution of $\boldsymbol{\Theta}$ in (6), an Metropolis–Hasting within Gibbs algorithm is facilitated by R package nimble (de Valpine et al., 2017). The loglikelihood function of the joint model used in the MCMC iteration is directly defined using the RW_llFunction() sampler. To use the spike-slab prior in (5) for variable selection, we need to specify the sampling methods for hyper-parameters ϕ_i's and γ_i's. The RW() sampler can be used to sample ϕ_i's and the binary() sampler can be used for γ_i's.

The integration in the likelihood function (3) does not have a closed-form. It needs to be computed with a Riemann approximation by partitioning $\mathcal{B}$ into a grid with a sufficiently fine resolution. Within each grid box, the integrand $\lambda(\boldsymbol{s})$ is approximated by a constant. Then the integration of $\lambda(\boldsymbol{s})$ becomes an integration of a piece-wise constant function, which is easily computed by a summation over all of the grid boxes.

4.2. Bayesian Model Selection for the Mark Model
Within the Bayesian framework, deviance information criterion (DIC; Spiegelhalter et al., 2002) and logarithm of pseudo-marginal likelihood (LPML; Geisser and Eddy, 1979; Gelfand and Dey, 1994) are two well-known Bayesian criteria for model comparison. A smaller DIC and larger LPML indicates a better model. When assessing only a specific component of a full model, as is the case with our application where we are interested in only the mark model, these global criteria need to be modified to reflect the focus. Besides, in contrast to DIC, the LPML for the joint model is hard to calculate

since the number of points N is random and, hence, the standard computing approaches for known sample sizes do not apply. If we focus only on the mark model, however, then N can be treated as a fixed sample size and the existing approaches can then be applied.

To assess the intensity dependent mark model, namely, whether the intensity helps improve the fitting of the mark, one needs to compare two mark models with and without intensity as a covariate. We consider modifying DIC and LPML to only focus on the mark model. Using the idea from Ma et al. (2018), we first define the following deviance function:

$$\text{Dev}(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \xi \mid \mathbf{M}) = -2 \sum_{i=1}^{N} \log f(m(\boldsymbol{s}_i) \mid \lambda(\boldsymbol{s}_i), \boldsymbol{\alpha}, \xi, \mathbf{Z}(\boldsymbol{s}_i)),$$

where $\boldsymbol{\lambda} = (\lambda(\boldsymbol{s}_1), \lambda(\boldsymbol{s}_2), \ldots, \lambda(\boldsymbol{s}_N))$, and $f(m(\boldsymbol{s}_i) \mid \lambda(\boldsymbol{s}_i), \boldsymbol{\alpha}, \xi, \mathbf{Z}(\boldsymbol{s}_i))$ is the conditional probability mass function of $m(\boldsymbol{s}_i)$ given $(\lambda(\boldsymbol{s}_i), \boldsymbol{\alpha}, \xi, \mathbf{Z}(\boldsymbol{s}_i))$. The effective number of parameters for this conditional model is defined as

$$p_D = \overline{\text{Dev}}(\boldsymbol{\lambda}, \boldsymbol{\alpha}, \xi \mid \mathbf{M}) - \text{Dev}(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\alpha}}, \bar{\xi} \mid \mathbf{M}),$$

where $\overline{\text{Dev}}$ is the mean of the deviance evaluated at each posterior draw of the parameters, and $\bar{\boldsymbol{\lambda}}$, $\bar{\boldsymbol{\alpha}}$, and $\bar{\xi}$ are, respectively, the posterior mean of $\boldsymbol{\lambda}$, $\boldsymbol{\alpha}$, and ξ. A modified DIC (mDIC; Ma et al., 2018) for the mark model is

$$\text{mDIC} = \text{Dev}(\bar{\boldsymbol{\lambda}}, \bar{\boldsymbol{\alpha}}, \bar{\xi} \mid \mathbf{M}) + 2p_D. \tag{7}$$

Similarly, a modified LPML (mLPML) is defined using a modified conditional predictive ordinate (mCPO), which can be calculated using a Monte Carlo estimation (Chen et al., 2012, Ch. 10). For the i-th data point, define

$$\widehat{\text{mCPO}}_i^{-1} = \frac{1}{B} \sum_{b=1}^{B} \frac{1}{f\left(m(\boldsymbol{s}_i) \mid \lambda^{(b)}(\boldsymbol{s}_i), \boldsymbol{\alpha}^{(b)}, \xi^{(b)}, \mathbf{Z}(\boldsymbol{s}_i)\right)}$$

where $\{\lambda^{(b)}(\boldsymbol{s}_i), \boldsymbol{\alpha}^{(b)}, \xi^{(b)} : b = 1, 2, \ldots, B\}$ is a posterior sample of size B of the unknown parameters. Then the mLPML is

$$\widehat{\text{mLPML}} = \sum_{i=1}^{N} \log(\widehat{\text{mCPO}}_i). \tag{8}$$

5. Simulation Studies

5.1. *Estimation*

To investigate the performance of the estimation, we generated data from a non-homogeneous Poisson point process defined on a square $\mathcal{B} = [-1, 1] \times [-1, 1]$ with intensity $\lambda(\boldsymbol{s}_i) = 100\lambda_0 \exp(\beta_1 x_i + \beta_2 y_i)$, where $\boldsymbol{s}_i = (x_i, y_i) \in \mathcal{B}$ is the location for every data point. For each $\boldsymbol{s}_i$, $i = 1, \ldots, N$, the mark $m(\boldsymbol{s}_i)$ follows a logistic model with two covariates in addition to λ and intercept:

$$\begin{aligned} m(\boldsymbol{s}_i) &\sim \text{Bern}(p_i), \\ \text{logit}(p_i) &= \xi\lambda(\boldsymbol{s}_i) + \alpha_0 + \alpha_1 Z_{1i} + \alpha_2 Z_{2i}. \end{aligned} \tag{9}$$

The parameters of the model were designed to give point counts that are comparable to the basketball shot chart data. We fixed $(\beta_1, \beta_2) = (2, 1)$, $\xi = 0.5$, $\alpha_0 = 0.5$ and $\alpha_2 = 1$. Three levels of α_1 were considered, $\alpha_1 \in \{0.8, 1, 2\}$, in order to compare the performance of the estimation procedure under different magnitudes of the coefficients in the mark model. Two levels of λ_0 were considered, $\lambda_0 \in \{0.5, 1\}$, which controls the mean of the number of points on $\mathcal{B}$. It is easy to integrate in this case the intensity function over $\mathcal{B}$ to get the average number of points being 850 and 1700, respectively, for $\lambda_0 = 0.5$ and 1. The numbers are approximately in the range of the NBA basketball shot chart in Section 2. In the mark model, covariate Z_1 was generated from the standard normal distribution; two types of Z_2 were considered, standard normal distribution or Bernoulli with rate 0.5. The resulting range of the Bernoulli rate of the marks was within $(0.55, 0.78)$ for all the scenarios.

For each setting, 200 data sets were generated. R package **spatstat** (Baddeley et al., 2005) was used to generate the Poisson point process data with the given intensity function. The priors for the model parameters were set to be (4) with the hyper-parameters $\sigma^2 = \delta^2 = 100$ and $a = b = 0.01$. The grid used to calculate the integration in likelihood function had resolution 100×100. For each data set, a MCMC was run for 20,000 iterations with the first 10,000 treated as burn-in period. For each parameter, the posterior mean was used as the point estimate and the 95% credible interval was constructed with the 2.5% lower and upper quantiles of the posterior sample. Empirical standard deviation (SD) of the 200 point estimates and the mean of the posterior standard deviations ($\widehat{\text{SD}}$) were also reported.

Table 2–3 summarize the simulation results for the scenarios of standard normal Z_2 and Bernoulli Z_2, respectively. The empirical bias for all the settings are close to zero. The average posterior standard deviation from the 200 replicates is very close to the empirical standard deviation of the 200 point estimates for all the parameters, suggesting that the uncertainty of the estimator are estimated well. Consequently, the empirical coverage rates of the credible intervals are close to the nominal level 0.95. As α_1 increases, the variation increases in the mark parameter estimates but does not change in the intensity parameter estimates. As λ_0 increases, the variation of the estimates for both intensity and mark parameters gets lower. Between the continuous and binary cases of Z_2, the variation in the estimates is higher in the latter case, especially for the coefficient of Z_2.

5.2. *Variable Selection*

To assess the performance of the variable selection method, we generated data with additional covariates and different scales in one of the covariate coefficients in the mark model. The intensity model and the coefficients value remained the same as in the Section 5.1. The mark model had six covariates $(Z_1, Z_2, \ldots, Z_6)$. Except for Z_2, these covariates were generated independently from the standard normal distribution. Covariate Z_2 again had two forms: standard normal or Bernoulli with rate 0.5. The values of $\xi = \alpha_0 = 0.5$ and $\alpha_2 = 1$ were fixed. Three levels of α_1 were considered: $\alpha_1 \in \{0.8, 1, 2\}$. Three scenarios of $(\alpha_3, \ldots, \alpha_6)$ were considered, $(0, 0, 0, 0)$, $(1, 1, 0, 0)$, and $(1, 1, 1, 1)$, representing more and more covariates that are present in the data generation model. For each setting, 200 datasets were generated. For each dataset, the spike-slab priors

Table 2. Summaries of the bias, standard deviation (SD), average of the Bayesian SD estimate ($\widehat{\text{SD}}$), and coverage rate (CR) of 95% credible intervals when Z_2 is continuous: $\xi = \alpha_0 = 0.5$, $\alpha_2 = 1$, $(\beta_1, \beta_2) = (2, 1)$ and $Z_2 \sim N(0, 1)$.

α_1	Model	Para	$\lambda_0 = 0.5$				$\lambda_0 = 1$			
			Bias	SD	$\widehat{\text{SD}}$	CR	Bias	SD	$\widehat{\text{SD}}$	CR
0.8	Intensity	λ_0	0.01	0.04	0.04	0.96	0.01	0.06	0.06	0.93
		β_1	−0.06	0.11	0.11	0.90	−0.06	0.09	0.08	0.88
		β_2	−0.05	0.09	0.09	0.94	−0.03	0.06	0.06	0.92
	Mark	ξ	0.11	0.57	0.60	0.97	0.04	0.22	0.22	0.96
		α_0	0.01	0.20	0.20	0.95	0.00	0.14	0.14	0.97
		α_1	0.03	0.13	0.13	0.94	0.01	0.09	0.09	0.94
		α_2	0.03	0.14	0.14	0.95	0.01	0.10	0.10	0.93
1	Intensity	λ_0	0.00	0.04	0.04	0.94	0.00	0.06	0.06	0.94
		β_1	−0.05	0.11	0.11	0.94	−0.05	0.08	0.08	0.91
		β_2	−0.03	0.10	0.09	0.92	−0.04	0.07	0.06	0.92
	Mark	ξ	0.03	0.60	0.61	0.95	0.05	0.21	0.22	0.96
		α_0	0.00	0.20	0.20	0.95	−0.01	0.14	0.14	0.95
		α_1	0.01	0.13	0.14	0.97	0.01	0.10	0.10	0.96
		α_2	0.03	0.14	0.14	0.95	0.01	0.10	0.10	0.94
2	Intensity	λ_0	0.00	0.04	0.04	0.94	0.00	0.06	0.06	0.95
		β_1	−0.06	0.12	0.11	0.91	−0.04	0.08	0.08	0.93
		β_2	−0.03	0.09	0.09	0.94	−0.03	0.07	0.06	0.91
	Mark	ξ	0.04	0.71	0.69	0.94	0.05	0.23	0.24	0.96
		α_0	0.02	0.22	0.23	0.95	−0.01	0.15	0.16	0.97
		α_1	0.08	0.23	0.21	0.93	0.03	0.15	0.15	0.94
		α_2	0.03	0.17	0.16	0.93	0.02	0.11	0.11	0.95

in (5) were specified for parameters α_i, $i \in \{1, 2, 3, 4, 5, 6\}$. The priors for the other parameters were kept the same as in (4). Other settings such as integration grid and MCMC were the same as in the last study.

Table 4–5 summarize the percentages of the 200 replicates in which the variable selection decision was correct for each covariate and for all covariates as a whole. Interestingly, the unimportant covariates are correctly excluded in all cases. The accuracy rates of correctly selecting the important variables individually decrease as more candidate variables are included. Changing Z_2 from continuous to binary gives less accurate selection decisions for this variable. More points with higher λ_0 improves the accuracy rate. The selection performance of all variables as a whole depends on the worst individual variable selection results, with much lower accuracy for binary Z_2 than for continuous Z_2, especially when λ_0 is lower. Lower magnitude of α_1 leads to lower accuracy in selecting Z_1 as well as the whole model correctly. We also experimented with an even lower $\alpha_1 = 0.5$ (results not shown), in which case, the correct selection percentage for the whole model reduces to as low as 10% in some settings.

Table 3. Summaries of the bias, standard deviation (SD), average of the Bayesian SD estimate ($\widehat{\text{SD}}$), and coverage rate (CR) of 95% credible intervals when Z_2 is binary: $\xi = \alpha_0 = 0.5$, $\alpha_2 = 1$, $(\beta_1, \beta_2) = (2, 1)$ and $Z_2 \sim Bernoulli(0.5)$.

α_1	Model	Para	$\lambda_0 = 0.5$				$\lambda_0 = 1$			
			Bias	SD	$\widehat{\text{SD}}$	CR	Bias	SD	$\widehat{\text{SD}}$	CR
0.8	Intensity	λ_0	0.00	0.04	0.04	0.94	0.01	0.06	0.06	0.95
		β_1	−0.05	0.12	0.11	0.88	−0.05	0.08	0.08	0.90
		β_2	−0.02	0.10	0.09	0.94	−0.03	0.06	0.06	0.95
	Mark	ξ	0.07	0.62	0.61	0.94	0.04	0.20	0.23	0.96
		α_0	0.00	0.23	0.22	0.93	0.01	0.16	0.16	0.95
		α_1	0.03	0.13	0.13	0.96	0.01	0.10	0.10	0.94
		α_2	0.03	0.25	0.24	0.96	0.01	0.19	0.18	0.95
1	Intensity	λ_0	0.00	0.04	0.04	0.94	0.00	0.06	0.06	0.94
		β_1	−0.06	0.11	0.11	0.93	−0.04	0.09	0.08	0.89
		β_2	−0.04	0.08	0.09	0.94	−0.02	0.06	0.06	0.93
	Mark	ξ	0.10	0.64	0.63	0.94	0.09	0.22	0.23	0.94
		α_0	0.01	0.23	0.23	0.97	−0.03	0.16	0.16	0.94
		α_1	0.03	0.15	0.14	0.92	0.01	0.11	0.10	0.92
		α_2	0.02	0.27	0.25	0.93	0.02	0.17	0.18	0.96
2	Intensity	λ_0	0.00	0.04	0.04	0.95	0.01	0.06	0.06	0.94
		β_1	−0.05	0.11	0.11	0.92	−0.05	0.08	0.08	0.90
		β_2	−0.04	0.09	0.09	0.91	−0.04	0.06	0.06	0.92
	Mark	ξ	0.06	0.73	0.70	0.94	0.07	0.28	0.25	0.93
		α_0	0.03	0.29	0.26	0.93	−0.01	0.20	0.19	0.94
		α_1	0.06	0.21	0.21	0.94	0.05	0.15	0.15	0.94
		α_2	0.03	0.31	0.28	0.93	0.03	0.19	0.20	0.94

6. NBA Players Shot Chart Analysis

The proposed methods were applied to analyze the shot chart of each of the four players in Figure 1. Because there were few shots placed behind the backboard line or too far beyond the 3-point line, we focused on a region of the half court with vertical coordinate $y \in [-0.75, 30]$ feet, where $y = -0.75$ is the line of the backboard. In evaluating the joint log-likelihood, this region was evenly partitioned into a 123×200 grid for the numerical integration in Equation (3).

Table 6 summarizes the covariates used in the intensity model and the mark model. We defined two distance covariates: distance to basket and distance to 3-point line. The former one is set to 0 for locations outside of the 3-point line and the latter one is set to 0 for locations inside the 3-point line. Both of them were rounded to the nearest foot. In order to improve the convergence of the MCMC, these variables were standardized (centered by the mean and scaled by the standard deviation). Intensity may also be related to angle (using the center of the basket rim as origin). We divided the court to six areas according to the shot angle relative to the backboard line: $[-\pi/2, \pi/6]$, $[\pi/6, \pi/3]$, $[\pi/3, \pi/2]$, $[\pi/2, 2\pi/3]$, $[2\pi/3, 5\pi/6]$, $[5\pi/6, 3\pi/2]$. Dummy variables were

Table 4. Percentages of correct decision of the variables in 200 replicates with continuous Z_2. The significant parameters except α_1 all equal to 1. All covariates were generated with standard Normal distribution.

Parameter	Non-Zero	$\alpha_1 = 0.8$		$\alpha_1 = 1$		$\alpha_1 = 2$	
		$\lambda_0 = 0.5$	$\lambda_0 = 1$	$\lambda_0 = 0.5$	$\lambda_0 = 1$	$\lambda_0 = 0.5$	$\lambda_0 = 1$
		$(\alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0, 0, 0, 0)$					
α_1	Yes	84	89	95	96	99	100
α_2	Yes	96	97	94	98	94	92
α_3	No	100	100	100	100	100	100
α_4	No	100	100	100	100	100	100
α_5	No	100	100	100	100	100	100
α_6	No	100	100	100	100	100	100
$\boldsymbol{\alpha}$	–	80	86	89	94	93	92
		$(\alpha_3, \alpha_4, \alpha_5, \alpha_6) = (1, 1, 0, 0)$					
α_1	Yes	77	87	94	94	100	100
α_2	Yes	94	96	95	96	92	94
α_3	Yes	94	96	93	96	94	98
α_4	Yes	91	96	93	95	91	98
α_5	No	100	100	100	100	100	100
α_6	No	100	100	100	100	100	100
$\boldsymbol{\alpha}$	–	60	78	78	81	78	90
		$(\alpha_3, \alpha_4, \alpha_5, \alpha_6) = (1, 1, 1, 1)$					
α_1	Yes	74	85	90	96	100	100
α_2	Yes	95	93	94	95	94	94
α_3	Yes	97	98	94	97	95	92
α_4	Yes	90	93	96	97	91	96
α_5	Yes	92	95	89	96	89	95
α_6	Yes	93	94	93	96	87	97
$\boldsymbol{\alpha}$	–	61	71	76	87	80	83

created using $[-\pi/2, \pi/6]$ as reference.

The joint model in (1)–(2) was fitted with the covariates in Table 6. For the mark model, the spike-slab prior (5) was imposed on each element in $\boldsymbol{\alpha}_i$ except the intercept. Other parameters' priors were set to be (4) with the hyper-parameters $\sigma^2 = \delta^2 = 100$ and $a = b = 0.01$. To check the importance of intensity as covariate in the mark component, we fitted the model with and without restricting $\xi = 0$. The two mark models were compared with mDIC and mLPML. Besides, the DIC for the full joint model, which, unlike the LPML, can be easily computed, was also obtained to compare the joint fit of both the intensity and the mark models with and without $\xi = 0$. For each model fitting, the trace plot of the MCMC was checked and the convergence of all the parameters were confirmed.

Table 7 summarizes the mDIC and mLPML for the mark model and the DIC for the full joint model. The smallest difference is 3.1 in mDIC and 0.4 in mLPML for Curry; the largest difference is 89.1 in mDIC and 57.8 in mLPML for Harden. The DIC has a rule of thumb similar to AIC to make decision (Spiegelhalter et al., 2002, Page. 613):

Table 5. Percentages of correct decision of the variables in 200 replicates with binary Z_2. The significant parameters except α_1 all equal to 1. Z_2 was generated using Bernoulli(0.5), while other covariates were from standard normal.

Parameter	Non-Zero	$\alpha_1 = 0.8$		$\alpha_1 = 1$		$\alpha_1 = 2$	
		$\lambda_0 = 0.5$	$\lambda_0 = 1$	$\lambda_0 = 0.5$	$\lambda_0 = 1$	$\lambda_0 = 0.5$	$\lambda_0 = 1$
		$(\alpha_3, \alpha_4, \alpha_5, \alpha_6) = (0, 0, 0, 0)$					
α_1	Yes	80	87	95	96	100	100
α_2	Yes	64	90	71	87	57	83
α_3	No	100	100	100	100	100	100
α_4	No	100	100	100	100	100	100
α_5	No	100	100	100	100	100	100
α_6	No	100	100	100	100	100	100
$\boldsymbol{\alpha}$	—	52	79	67	83	57	83
		$(\alpha_3, \alpha_4, \alpha_5, \alpha_6) = (1, 1, 0, 0)$					
α_1	Yes	77	89	94	98	100	100
α_2	Yes	65	85	62	82	58	97
α_3	Yes	92	95	96	96	95	95
α_4	Yes	94	94	94	96	93	94
α_5	No	100	100	100	100	100	100
α_6	No	100	100	100	100	100	100
$\boldsymbol{\alpha}$	—	44	68	52	74	53	76
		$(\alpha_3, \alpha_4, \alpha_5, \alpha_6) = (1, 1, 1, 1)$					
α_1	Yes	69	85	93	97	99	100
α_2	Yes	62	78	55	76	47	75
α_3	Yes	93	95	92	97	91	97
α_4	Yes	92	96	91	97	85	96
α_5	Yes	92	96	93	93	89	96
α_6	Yes	94	95	93	92	92	97
$\boldsymbol{\alpha}$	—	36	63	42	70	34	70

a difference larger than 10 is substantial and a difference about 2–3 does not give an evidence to support one model over the other. For LPML, difference less than 0.5 is "not worth more than to mention" and larger than 4.5 can be considered "very strong" (Kass and Raftery, 1995). With these guidelines applied to mDIC and mLPML, the mark model with shot intensity included as a covariate has a clear advantage relative to the model without it for all players except Curry. Although Curry's result favors the intensity independent mark model, the evidence is "not worth more than to mention". We also obtained the DIC for the whole joint model for all four players. The differences in the DIC of the joint models are almost the same as those in the mDIC of the mark models, suggesting that the DIC for the intensity model are almost the same with and without $\xi = 0$. This is expected because the marks may contain little information about the intensities.

Figure 2 presents the fitted shot intensities of the four players on the same scale. The fitted intensities appear to capture the spatial patterns of the shot charts in Figure 1. The intensities are highest at the origin and gradually decreases as the distance

Table 6. Covariates used in the intensity model and the mark model.

Model	Covariates	Explanation
intensity	beyond 3-point line	indicator (1 = beyond 3-point line)
	distance to basket	standardized; 0 for 3-point shots
	distance to 3-point line	standardized; 0 for 2-point shots
	shot angle	relative to backboard line; 6 levels with $[-\frac{\pi}{2}, \frac{\pi}{6})$ as reference
mark	intensity	shot intensity
	game period	5 levels with the first period as reference
	seconds left	time in seconds left towards the end of the period, divided by 100
	opponent	indicator (1 = opponent made playoff last season)
	beyond 3-point line	same as above
	distance to basket	same as above
	distance to 3-point line	same as above
	shot angle	same as above

Table 7. Summaries of mDIC and mLPML for the mark model and DIC for the full joint model.

	Mark model				Joint model	
	mDIC		mLPML		DIC	
Player	$\xi \neq 0$	$\xi = 0$	$\xi \neq 0$	$\xi = 0$	$\xi \neq 0$	$\xi = 0$
Curry	1041.7	1038.8	−520.9	−521.3	151.1	148.2
Durant	1386.4	1454.1	−693.2	−729.1	335.0	402.8
Harden	1720.1	1809.2	−860.2	−918.0	1516.7	1605.9
James	1758.6	1809.2	−879.3	−909.1	1750.6	1802.0

increasing. An obvious increase of intensities is observed at the 3-point line, followed by a faster decreasing speed as the distance increases further. Regions with angles closer to vertical relative to the backboard have higher intensities than regions with angles closer to horizontal. The fitted intensities are the lowest in just inside the 3-point line where the angles are tough. Between the players, it is obvious that Curry has higher intensity just outside of the 3-point line than James, although James made almost twice as many total shots as Curry.

Estimates of model parameters for the four players are summarized in Table 8–9, including posterior mean, posterior standard deviation, and 95% credible interval (constructed with the lower and upper 2.5% percentiles of the posterior sample). The results from the intensity models provide more detailed insights on how the covariates affect the shot intensity as visualized in Figure 2. Regions with shot angles in $[\pi/6, 5\pi/6]$ have higher shot intensities than those with poorer angles for all four players. The estimated coefficient of the 3-point indicator suggests that Curry and Harden intend to shoot for 3-point more often than James and Durant. The estimated coefficient of the distance to
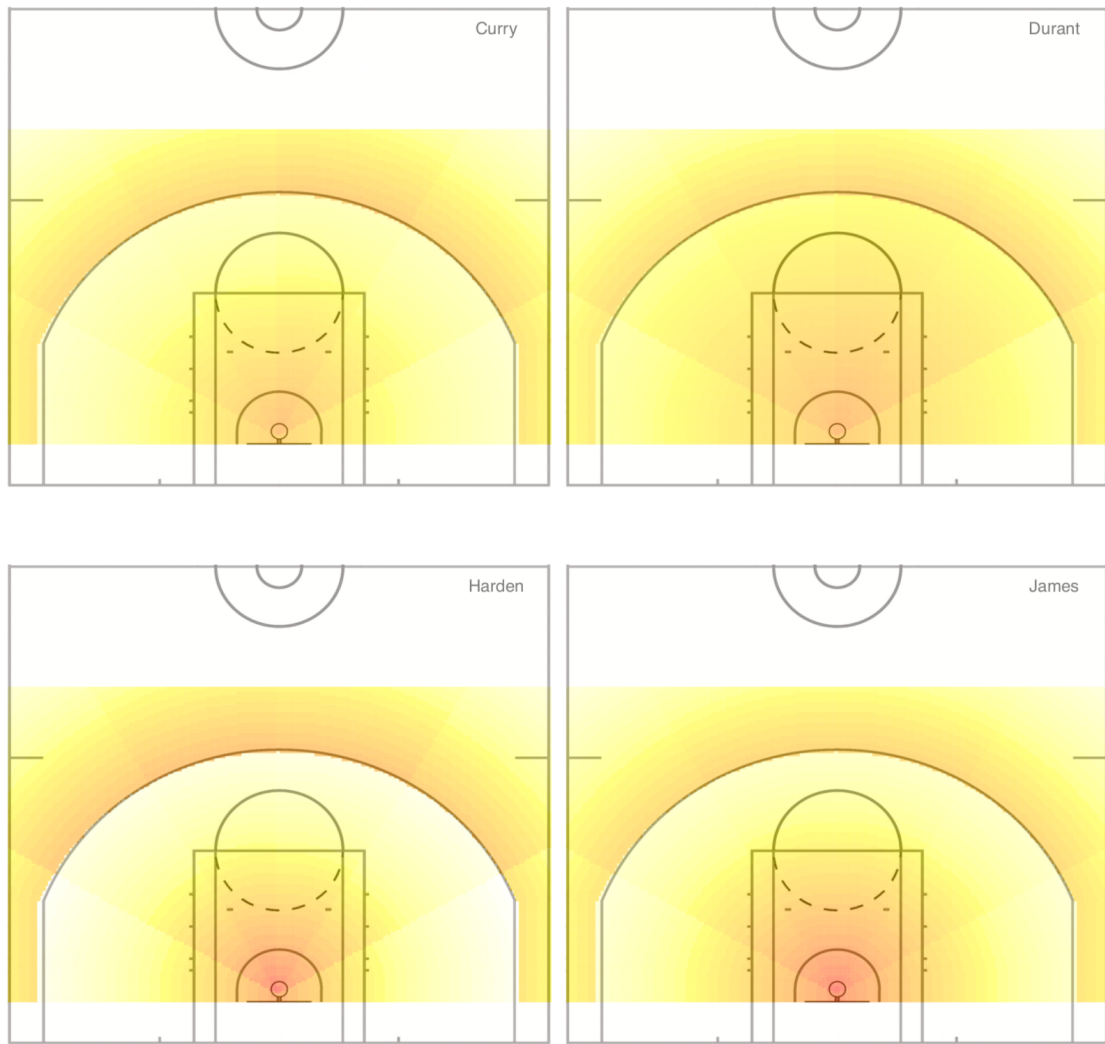
Fig. 2. Intensity fit results of Curry, Durant, Harden and James on the same scale. Red means higher intensity.

bakset are closer to zero for Curry and Durant than for Harden and James, suggesting that the 2-point shot intensities for the former decreases at a slower rate than the latter. For the mark model, only results for the variables selected by the variable selection procedure are reported in addition to the intercept and the shot intensity. Shot intensity does not influence Curry's shot accuracy directly. The coefficient of the intensity is significantly positive for the other three players. That is, higher field goal rate are expected at locations where the players make more shots. For Durant and James, no other covariates are selected after the intensity is included in the mark model. Curry has marginally higher field goal percentage during the extra times than during the first

Table 8. Data analysis result using intensity dependent model for Curry and Durant.

Player	Model	Covariates	Posterior Mean	Posterior SD	95% Credible Interval
Curry	Intensity	baseline (λ_0)	0.25	0.03	(0.19, 0.32)
		beyond 3-point line	-2.12	0.22	(-2.56, -1.70)
		distance to basket	-1.39	0.07	(-1.54, -1.25)
		distance to 3-point line	-2.28	0.17	(-2.61, -1.96)
		angle $[\pi/6, \pi/3]$	0.69	0.15	(0.40, 0.98)
		angle $[\pi/3, \pi/2]$	1.03	0.14	(0.75, 1.30)
		angle $[\pi/2, 2\pi/3]$	0.78	0.15	(0.49, 1.06)
		angle $[2\pi/3, 5\pi/6]$	0.58	0.15	(0.29, 0.88)
		angle $[5\pi/6, 3\pi/2]$	0.06	0.18	(-0.28, 0.40)
	Mark	intercept	0.32	0.18	(-0.02, 0.67)
		beyond 3-point line	-0.88	0.37	(-1.91, -0.21)
		fifth period	7.10	6.40	(-0.10, 22.09)
Durant	Intensity	baseline (λ_0)	0.65	0.07	(0.53, 0.79)
		beyond 3-point line	-3.21	0.22	(-3.64, -2.79)
		distance to basket	-0.98	0.05	(-1.07, -0.89)
		distance to 3-point line	-2.32	0.17	(-2.65, -1.99)
		angle $[\pi/6, \pi/3]$	0.71	0.12	(0.48, 0.95)
		angle $[\pi/3, \pi/2]$	0.84	0.12	(0.61, 1.07)
		angle $[\pi/2, 2\pi/3]$	0.65	0.12	(0.42, 0.90)
		angle $[2\pi/3, 5\pi/6]$	0.42	0.13	(0.17, 0.67)
		angle $[5\pi/6, 3\pi/2]$	-0.21	0.15	(-0.49, 0.08)
	Mark	intercept	-0.39	0.18	(-0.73, -0.01)
		intensity (λ)	0.33	0.07	(0.19, 0.46)

quarter, and his success rate in 3-point shots is significantly lower than 2-point shots. Harden has a marginally higher field goal percentage with shot angles in $[5\pi/6, 3\pi/2]$ than with angles in $[-\pi/2, \pi/6]$.

7. Discussion

We proposed to use a marked spatial point process to jointly model the locations and success rate of basketball players' shot charts in a Bayesian framework. Covariates are incorporated in the intensity of the Poisson point process model for the shot locations as well as in the logistic model for the success of the shots. One attraction of the model is that the intensity of the shot locations is incorporated as a covariate in the success rate model. Variable selection in the mark model is done with the spike-slab prior facilitated by MCMC. The proposed methods performed well in parameter estimation and variable selection in simulation studies under various settings. In applications to the shot charts of four NBA players, the intensity was found to be positively significant in the success rate model of three players in terms of Bayesian model assessment criteria. The fitted results of the intensity dependent mark model suggest that the three players had higher success rate where they made more shots. The fourth player's results are not different

Table 9. Real data analysis results using intensity dependent model for Harden and James.

Player	Model	Covariate	Posterior Mean	Posterior SD	95% Credible Interval
Harden	Intensity	baseline (λ_0)	0.18	0.02	(0.14, 0.23)
		beyond 3-point line	-2.12	0.18	$(-2.46, -1.77)$
		distance to basket	-2.44	0.07	$(-2.58, -2.31)$
		distance to 3-point line	-2.49	0.13	$(-2.75, -2.24)$
		angle $[\pi/6, \pi/3]$	1.02	0.14	(0.76, 1.29)
		angle $[\pi/3, \pi/2]$	1.41	0.13	(1.16, 1.67)
		angle $[\pi/2, 2\pi/3]$	1.57	0.13	(1.32, 1.82)
		angle $[2\pi/3, 5\pi/6]$	1.27	0.14	(1.01, 1.54)
		angle $[5\pi/6, 3\pi/2]$	0.29	0.16	$(-0.02, 0.60)$
	Mark	intercept	-0.58	0.17	$(-0.94, -0.26)$
		intensity (λ)	0.07	0.01	(0.04, 0.09)
		angle $[5\pi/6, 3\pi/2]$	0.57	0.35	(0.00, 1.18)
James	Intensity	baseline (λ_0)	0.74	0.06	(0.62, 0.87)
		beyond 3-point line	-3.15	0.22	$(-3.59, -2.72)$
		distance to basket	-2.10	0.05	$(-2.20, -2.01)$
		distance to 3-point line	-2.32	0.17	$(-2.66, -1.98)$
		angle $[\pi/6, \pi/3]$	0.42	0.09	(0.24, 0.61)
		angle $[\pi/3, \pi/2]$	0.58	0.09	(0.40, 0.76)
		angle $[\pi/2, 2\pi/3]$	0.58	0.09	(0.40, 0.77)
		angle $[2\pi/3, 5\pi/6]$	0.41	0.10	(0.22, 0.60)
		angle $[5\pi/6, 3\pi/2]$	-0.21	0.11	$(-0.43, 0.00)$
	Mark	intercept	-0.72	0.15	$(-1.01, -0.43)$
		intensity (λ)	0.08	0.01	(0.07, 0.11)

by much between the mark models with and without the intensity as a covariate.

A few directions of further work are worth investigating. Our proposed model is univariate. Each player is modeled separately. The summary statistics or estimators from the separate players may be used as inputs in a second stage analysis to cluster the players and analyze the similarities among groups of players. A multivariate model for multiple players jointly may be useful in capturing more game dynamics, provided that such data are available. Our spatial Poisson process model formulates a linear relationship between the spatial covariates and the log intensity, which cannot capture more complicated spatial trend of the intensity of spatial point pattern. Including some Bayesian non-parametric methods like finite mixture model (Miller and Harrison, 2018) may help increase the accuracy of the estimation of spatial point pattern. Model assessment criteria (Zhang et al., 2017; Hu et al., 2019) for the joint model is also an interesting future direction.

Acknowledgments

References

Baddeley, A., Turner, R. et al. (2005) spatstat: An R package for analyzing spatial point patterns. *Journal of Statistical Software*, **12**, 1–42.

Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2014) *Hierarchical Modeling and Analysis for Spatial Data.* Chapman and Hall/CRC.

Chen, M.-H., Shao, Q.-M. and Ibrahim, J. G. (2012) *Monte Carlo Methods in Bayesian Computation.* Springer Science & Business Media.

Cressie, N. (2015) *Statistics for Spatial Data.* John Wiley & Sons.

Diggle, P. J. (2013) *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns.* Chapman and Hall/CRC.

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

Geisser, S. and Eddy, W. F. (1979) A predictive approach to model selection. *Journal of the American Statistical Association*, **74**, 153–160.

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

George, E. I. and McCulloch, R. E. (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, **88**, 881–889.

Geyer, C. J. (1999) Likelihood inference for spatial point processes. In *Stochastic Geometry: Likelihood and Computation* (eds. O. Barndorff-Nielsen, W. Kendall and M. van Lieshout), vol. 80, 79–140. CRC Press.

Ho, L. P. and Stoyan, D. (2008) Modelling marked point patterns by intensity-marked Cox processes. *Statistics & Probability Letters*, **78**, 1194–1199.

Hu, G., Huffer, F. and Chen, M.-H. (2019) New development of Bayesian variable selection criteria for spatial point process with applications. *Tech. Rep. 18-05*, University of Connecticut, Department of Statistics.

Kass, R. E. and Raftery, A. E. (1995) Bayes factors. *Journal of the american statistical association*, **90**, 773–795.

Leininger, T. J., Gelfand, A. E. et al. (2017) Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. *Bayesian Analysis*, **12**, 1–30.

Ma, Z., Chen, M.-H. and Hu, G. (2018) Bayesian hierarchical spatial regression models for spatial data in the presence of missing covariates with applications. *Tech. Rep. 18-22*, University of Connecticut, Department of Statistics.

Malsiner-Walli, G. and Wagner, H. (2018) Comparing Spike and Slab priors for Bayesian variable selection. *arXiv preprint arXiv:1812.07259*.

Miller, A., Bornn, L., Adams, R. and Goldsberry, K. (2014) Factorized point process intensities: A spatial analysis of professional basketball. In *Proceedings of the 31st International Conference on Machine Learning — Volume 32*, ICML'14, 235–243.

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

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

Møller, J. and Waagepetersen, R. P. (2003) *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC.

Mrkvička, T., Goreaud, F. and Chadœuf, J. (2011) Spatial prediction of the mark of a location-dependent marked point process: How the use of a parametric model may improve prediction. *Kybernetika*, **47**, 696–714.

Ord, J. K. (2004) Spatial processes. *Encyclopedia of Statistical Sciences*, **12**.

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

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

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T. and Bodik, R. (2017) Programming with models: Writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, **26**, 403–413.

Vere-Jones, D. and Schoenberg, F. P. (2004) Rescaling marked point processes. *Australian & New Zealand Journal of Statistics*, **46**, 133–143.

Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E. and Shen, W. (2017) Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. *Journal of Computational and Graphical Statistics*, **26**, 121–133.