

Research Article

Mirko Luigi Carlesso*, Andrea Cappozzo, Andrea Gilardi and Paola Zuccolotto*

A mixed-effects spatial point process framework for modeling player shooting behavior in basketball

<https://doi.org/10.1515/jqas-2025-0163>

Received October 28, 2025; accepted March 10, 2026;

published online April 27, 2026

Abstract: Understanding offensive player profiles in basketball requires analyzing both the spatial distribution of shot attempts and the corresponding scoring effectiveness across court locations. Motivated by this, we propose a semi-parametric spatial point process model that combines flexible spline-based intensity estimation with player-specific mixed effects to characterize shooting intensity as a function of distance and angle from the basket. The model balances flexibility and interpretability while accommodating the hierarchical structure of player-level data through random effects. We apply this framework to shooting data from the NBA 2024/2025 season, focusing on High-Volume Shooters who exhibit diverse spatial patterns. The model captures both population-level shooting tendencies and individual deviations, revealing substantial heterogeneity in offensive behavior across players. Extending the framework to marked point processes allows us to construct spatial scoring probability maps that quantify efficiency beyond shot frequency. Collectively, the proposed methodology offers a unified statistical approach for modeling spatial shooting behavior and provides actionable insights for player evaluation, tactical decision-making, and roster construction.

Keywords: basketball analytics; spatial point patterns; shooting intensity maps; random effects; relative risk

1 Introduction

In recent years, the statistical analysis of sports data has attracted increasing interest from both researchers and practitioners, with a wide range of diverse objectives. Statistical methods can be applied to assess individual and team performance, identify strengths and weaknesses, support strategic and tactical decisions, optimize training processes, and manage player scouting. In addition, predictive models are increasingly used to forecast match outcomes, estimate player contributions, evaluate injury risks, and inform betting and market-related decisions. For some recent overviews of the broad landscape of sports statistics, see, for example, Dominicy and Ley (2023), Davis et al. (2024), Mansurali and Mahmoud (2024). Among the various research directions in sports analytics, particular attention has been devoted to the quantitative evaluation of the most decisive technical actions, such as goal scoring in soccer, serving performance in tennis, shot conversion in handball, and penalty shooting in ice hockey.

In basketball, the act of shooting is arguably the most crucial and fundamental action a player performs. The success of a team is ultimately measured by its ability to convert shooting opportunities into points, so a player's shooting performance is a key indicator of their value on the court. To gain a deeper understanding of a player's offensive profile, it is essential to look beyond simple shooting percentages. A comprehensive analysis must consider several critical aspects related to shooting, namely the spatial intensity of attempts and the scoring probability of shots from different locations on the court. The probability of scoring on each shot attempt is a fundamental concept, commonly quantified using the effective field goal percentage (EFG %), since the establishment of a formal analytical framework for basketball by Kubatko et al. (2007). Analyzing the probability of scoring from a spatial perspective, *i.e.*, considering the specific location on the court where a shot is taken, provides a much richer and more granular evaluation, as pointed out in the spatial analysis of professional basketball by Miller et al. (2014).

*Corresponding authors: **Mirko Luigi Carlesso** and **Paola Zuccolotto**, Big & Open Data Innovation Laboratory (BODaI-Lab), University of Brescia, Brescia, Italy, E-mail: mirko.carlesso@gmail.com (M. L. Carlesso), paola.zuccolotto@unibs.it (P. Zuccolotto)

Andrea Cappozzo, Department of Statistical Sciences, Università Cattolica del Sacro Cuore, Milan, Italy, E-mail: andrea.cappozzo@unicatt.it

Andrea Gilardi, Department of Economics, Management and Statistics, University of Milano – Bicocca, Milan, Italy, E-mail: andrea.gilardi@unimib.it

However, a deep analysis of the scoring probability can only be performed after having accurately addressed the issue of the shooting intensity. In fact, in a spatial context, simply measuring the probability that a shot is going in can be misleading if the underlying distribution of shots is not taken into account. A player might appear to be highly efficient from a specific spot, but this could be of limited interest (and even scarcely reliable from a statistical point of view) if few shots have been taken from that location. By first mapping the shooting intensity, we establish a foundational “footprint” that reveals where the player tends to operate on the court. This gives us a more realistic and complete picture of their offensive behavior. Without this crucial step, an analysis of the scoring probability might give undue weight to statistically insignificant areas, leading to biased conclusions about true efficiency. Therefore, assessing intensity first provides the necessary context to accurately and reliably interpret scoring probability maps. Subsequently, the shooting intensity maps can be combined to obtain an estimate of the EFG %.

This two-step approach finds its natural methodological framework in spatial point process analysis (Baddeley et al. 2015; Diggle 2013). Specifically, shot attempts are treated as a spatial point pattern. The intensity function of this process describes the rate or density of points (shots) at any given location on the court. The outcome of each shot (made or missed) is considered a mark on that point. According to this framework, the analysis of marks (scoring probability) is often conditional on the underlying point pattern (shooting intensity).

In this context, Jiao et al. (2021) propose a Bayesian joint model for the mark and intensity analysis of marked point processes, where the intensity is incorporated in the mark model as a covariate. Yin et al. (2022) and Yin et al. (2023) employ Bayesian nonparametric learning for point processes, with a flexible modeling of the underlying spatial intensity of shot attempts built upon a combination of Dirichlet process and Markov random field, which allows a local spatial homogeneity when estimating a globally heterogeneous intensity surface. Narayanan et al. (2023) develop a new family of marked point processes, which, although applied to football data, can be of interest also in basketball, as they have been proven to be capable of quantifying characteristics of game dynamics, predict event occurrences, and extract event-specific team abilities. Furthermore, Wang and Zheng (2022a) resort to Bayesian hierarchical models to examine positional differences in shooting accuracy, acknowledging that players in different roles might exhibit varying spatial shooting profiles, and

Cao et al. (2025) develop a Bayesian log-Gaussian Cox process model to jointly analyze the spatial distribution of shot locations and outcomes across multiple games, where the log-intensity function is modeled hierarchically with Gaussian processes, incorporating spatially varying effects of game-specific covariates.

Other approaches in the context of spatial performance measurement include hierarchical spatial models for shot chart data, which allow for spatially varying effects of covariates (Reich et al. 2006), a model-based approach of Scrucca and Karlis (2025), who propose a statistical framework for shot charts that explicitly considers the physical boundaries of the basketball court by means of Gaussian mixtures for bounded data, and Shortridge et al. (2014), who propose methods to characterize and visualize relative spatial shooting effectiveness in basketball by developing metrics to assess spatial variability in shooting. Methods that directly address scoring probability without explicitly considering intensity include some proposals based on machine learning techniques such as CART, random forests, and extremely randomized trees, which have been applied for spatial performance analysis by Zuccolotto et al. (2023), building upon previous work on spatial performance indicators and graphs (Zuccolotto et al. 2021). In these works, machine learning methods have proved able to effectively capture non-linear relationships between shot location and scoring probability. More recently, Carlesso et al. (2024) explored the use of Indicator Kriging to generate scoring probability maps, providing an alternative that can account for spatial correlation. A more structured approach has been proposed by Sandholtz et al. (2020), where allocation efficiency is considered in a spatial context at a lineup level, with players' shooting performance estimated at every location in the offensive half court using a Bayesian hierarchical model.

In this paper, we make use of a spatial point process modeling framework to estimate the intensity of basketball shot attempts and, subsequently, to derive scoring probabilities through the analysis of marked point processes. In particular, we introduce a semiparametric specification for shooting intensity that coherently integrates spline-based smoothing and random effects, thereby striking a balance between model flexibility and interpretability. The use of splines provides a highly flexible yet computationally efficient way to model nonlinear relationships between the response and two covariates, namely distance from the basket and angle of the shot. By representing these relationships with smooth piecewise polynomials, splines capture complex spatial variations without enforcing a rigid functional form.

Moreover, we employ random effects to enhance the model through partial pooling of information between players, taking advantage of the inherent hierarchy in shots attempted by the same player. This hierarchical structure stabilizes individual parameter estimates, especially when data for certain players are sparse, while still preserving unit-specific variability. Random effects also account for unobserved heterogeneity and correlation within players, leading to more interpretable inference and improved generalizability. By combining splines and random effects within a spatial point process framework, we establish a methodology capable of capturing complex spatial patterns and reliably characterizing player-specific shooting behaviors across varying levels of data granularity. Only a limited number of studies have applied spatial point process models that incorporate mixed effects, particularly within a frequentist framework. This makes our approach methodologically innovative, as it extends a modeling strategy that has received relatively little attention in the existing literature. The few existing contributions, such as Bell and Grunwald (2004) and Illian and Hendrichsen (2010), represent early attempts in this direction, but comprehensive developments remain scarce and, to the best of our knowledge, no application has been made in the context of basketball.

The remainder of the article is structured as follows. In Section 2, we briefly describe the statistical approach to our problem, outlining how splines are combined with the mixed-effects component in the proposed spatial point process framework. Section 3 presents and discusses in detail the results of the case study based on data from the 2024–2025 NBA season, describing the impact of angle and distance on shooting intensity, and the patterns that characterize different players. The section continues validating the proposed approach analysing the model's residuals. Finally, we discuss strategic implications based on shooting intensity and scoring probability maps derived from the fitted model. Section 4 concludes the paper and outlines directions for future research.

2 Methodology

In the following, we will assume that the basketball shots taken by the j -th player, with $j \in \{1, \dots, J\}$ and J identifying the total number of players in a complete season, constitute a finite realization of a spatial point process X within a bounded window $W \subset \mathbb{R}^2$ (Diggle 2013). We denote the collection of shots for the player j as $\{\mathbf{s}_{1j}, \dots, \mathbf{s}_{n_jj}\}$, with $\mathbf{s}_{ij} = (x_{ij}, y_{ij})$, $i = 1, \dots, n_j$ representing the Cartesian coordinates of the shot and n_j the total number of shots taken by the j -th player within the basketball court. Assume that

there exists a function, say $\lambda(\mathbf{s})$, $\mathbf{s} \in W$, satisfying the following condition:

$$\mathbb{E}[N(A)] = \int_A \lambda(\mathbf{s}) \, d\mathbf{s}, \quad A \subset W,$$

where $N(A)$ denotes the number of shots falling in the set A . This function is usually termed the *intensity function* of the process and, broadly speaking, describes the rate at which events occur in the given region. Following a classical and pragmatic hypothesis in the literature of spatial point processes, we assume that X is an inhomogeneous Poisson point process on W . The log-likelihood function for this model is (up to a constant) equal to:

$$\ell = \sum_{j=1}^J \sum_{i=1}^{n_j} \log \lambda(\mathbf{s}_{ij}) - \int_W \lambda(\mathbf{s}) \, d\mathbf{s}. \quad (1)$$

We specify a semiparametric log-linear mixed-effects model for $\lambda(\mathbf{s})$, $\mathbf{s} = (x, y) \in W$ as a function of covariates related to the characteristics of the shot with respect to the basketball court. Specifically, the functional formulation considered in this paper is as follows:

$$\begin{aligned} \log \lambda(\mathbf{s}) = & \beta_0 + \beta_1(x) + \beta_2(y) + \beta_3(\text{distance}(\mathbf{s})) \\ & + \beta_4(\text{angle}(\mathbf{s})) + b_{0j} + b_{3j}(\text{distance}(\mathbf{s})) \\ & + b_{4j}(\text{angle}(\mathbf{s})). \end{aligned} \quad (2)$$

In Equation (2), β_0 represents the intercept, while the terms $\beta_c(\cdot)$, $c = 1, 2, 3, 4$, denote spline transformations that may capture potential non-linear relationships between the covariates and the shooting intensity (Friedman et al. 2009; Wood 2017). The terms $\text{distance}(\mathbf{s})$ and $\text{angle}(\mathbf{s})$, respectively, denote the Euclidean distance from the locations $\mathbf{s}$ to the basket and the shooting angle, measured in radians from 0 to π . Natural cubic splines are used to smooth the first three covariates, while a cyclic cubic spline is applied to the angular component to capture its periodic structure and enforce boundary continuity. These semiparametric transformations improve the flexibility of the regression model by capturing smooth spatial variations and incorporating the geometric properties of the basketball court. Lastly, the remaining three terms in Equation (2) represent the random-effects components specific to each player j in the log-linear intensity model. In detail, b_{0j} denotes the random intercept, while $b_{3j}(\cdot)$ and $b_{4j}(\cdot)$ represent the random slopes at the player-level associated with distance and angle, respectively. In line with the fixed component specification, the distance covariate is modeled through natural cubic splines, allowing flexible non-linear estimation while preserving smoothness at the boundaries. Similarly, the

angular random covariate is modeled using cyclic cubic splines, which appropriately capture the inherently periodic nature of shooting angles on the basketball court. Inclusion of random terms allows the model to account for player-specific heterogeneity beyond population-level effects. That is, each player is assigned individual coefficients that describe how their shooting intensity varies with respect to the common fixed structure that reflects the shared shooting tendencies. Specifically, we include random effects for distance and angle as these features characterize a player's shooting style, which varies markedly across players (see Sections 3.1 and 3.2). For a comprehensive discussion and detailed presentation of the mixed-effects methodology, interested readers are referred to, e.g., Pinheiro and Bates (2006) and Demidenko (2013).

Following the approach described in Bell and Grunwald (2004), the parameters in Equation (2) are estimated using a two-step procedure. First, the likelihood of the model in Equation (1) is approximated by employing a technique known as Berman-Turner Device (Berman and Turner 1992) via a discretization of the window W that uses approximately 1,600 dummy points for each player. Then, considering that our model contains both fixed and random effects, we adopt the strategy proposed by Wolfinger and O'Connell (1993) to iteratively estimate the two sets of spline parameters until convergence. The computations were performed using the R software (R Core Team 2024) and, in particular, the R package *spatstat* (Baddeley et al. 2015). It took approximately 1 h to perform the aforementioned procedure on the dataset described in Section 3.

As a last worthy note, although our primary focus here is on unmarked point processes for estimating the shooting intensity surface, the same framework can naturally be extended to marked processes, for example to investigate relative risks associated with shot outcomes. More formally, let $\lambda_0(\mathbf{s})$ and $\lambda_1(\mathbf{s})$ denote the intensity function of missed and successful shots, respectively. The spatial EFG %, usually named (normalized) relative risk function in the point process literature, can be defined as

$$\rho(\mathbf{s}) = \frac{\lambda_1(\mathbf{s})}{\lambda_0(\mathbf{s}) + \lambda_1(\mathbf{s})}, \quad \mathbf{s} \in W. \quad (3)$$

This function represents the probability that a shot made at the location $\mathbf{s}$ is successfully converted. Clearly, $\rho(\mathbf{s}) \in [0, 1]$. Values of $\rho(\mathbf{s})$ near 0.5 indicate that shots made and missed are equally likely; values approaching 1 indicate that almost all shots taken at $\mathbf{s}$ go in; and values near 0 imply that almost every shot is missed. A plug-in estimator of $\rho(\mathbf{s})$ can be obtained by replacing $\lambda_0(\mathbf{s})$ and $\lambda_1(\mathbf{s})$ with their parametric estimates derived using the procedure described in the previous paragraph. Scoring probability maps derived

from the relative risk definition, together with ad hoc visual representations of player-specific distance and angle shooting patterns, provide substantial analytical value in characterizing individual shooting styles across the basketball league. A comprehensive analysis is presented in the following section.

3 Case study

Building on the theoretical framework introduced above, we now turn to its practical application by analyzing real-world basketball data. Our empirical study is based on a comprehensive dataset from the NBA 2024/2025 season, which includes 219,527 shot events.¹ Each record in the dataset corresponds to a shot attempt and contains all the information required to estimate the spatial model detailed in the previous section: the two-dimensional coordinates of the attempt on the half-court and other contextual attributes related to the shots, such as name of the player who attempted it or the team in which he plays for. The other two covariates were calculated from the coordinates of the location $\mathbf{s}$. In addition, the last part of this section incorporates the outcomes of the shot (i.e., made or missed) to construct relative risk maps that characterize the shooting performances of the players.

As discussed in the previous section, each shot is represented as a point located in the two-dimensional space of the basketball court, thereby producing a spatial point pattern. A crucial element of this framework is the definition of the observation window W which specifies the region where points might be observed. Rather than adopting the entire half-court as W , we construct a tailored observation window that reflects the areas of the court where shots typically occur. Specifically, regions corresponding to highly improbable attempts are removed from the analysis, such as those near the half-court line or behind the basket, which is an area where shots typically occur in desperate or late-clock situations. These zones contribute little to understanding the offensive behavior of players and, if included, could introduce noise and distortions in the statistical analysis. Restricting W in this way ensures that our analysis remains focused on meaningful shooting behavior and improves both the interpretability and robustness of the results.

¹ Data were downloaded in June 2025 and are publicly available at the following link: https://github.com/shufinskiy/nba_data. We also thank BigDataBall (www.bigdataball.com), a data provider which leverages computer-vision technologies to enrich and extend sports datasets with lots of unique metrics, for supplying play-by-play data used to test the model on additional datasets.

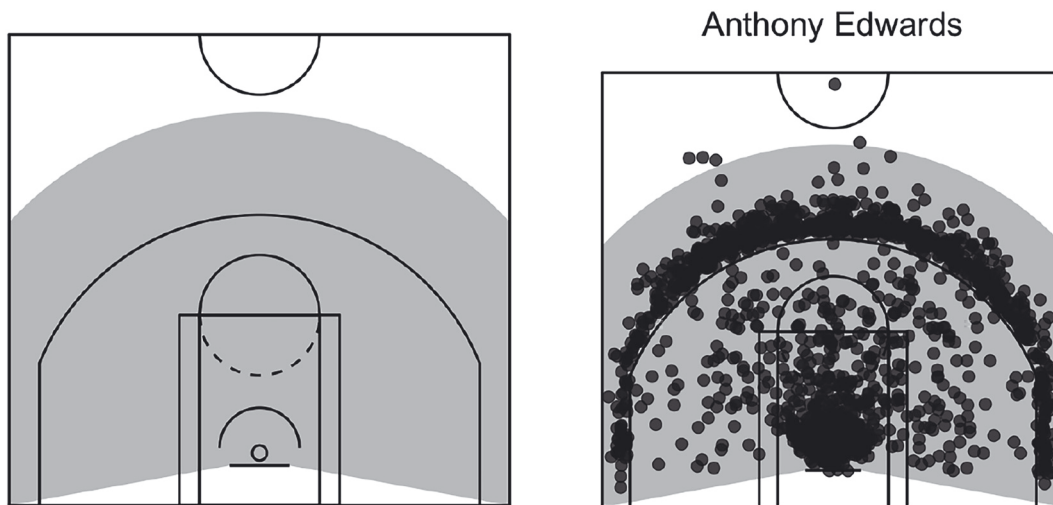


Figure 1: Left: The outermost rectangle denotes the half-court, whereas the gray polygon denotes the chosen observation window W . Right: Shotchart of 1,612 shots – data of Anthony Edwards (Minnesota Timberwolves), season 2024/2025.

Figure 1 illustrates this setup. The left panel shows the observation window W (the gray polygon) within the half-court, while the right panel presents a scatterplot of Anthony Edwards' shot attempts, a leading player of the Minnesota Timberwolves. Analogous maps can be created for each player in the NBA league. It is immediately evident that shots are not randomly distributed across the court: some areas exhibit a high density of attempts, while others remain scarcely used. This non-uniformity underscores the importance of studying the spatial distribution of shots, as it provides insight into players' tendencies, preferred scoring zones, and strategic decision-making during the game. Furthermore, we can clearly see that the vast majority of shots are taken within the observation window W , highlighting the adequacy of our proposal.

As already mentioned, we decided to focus on the analysis of individual players. To ensure that the estimated shooting patterns reflect stable and representative behaviors, we only included players who appeared in at least 50 % of the regular-season (41 games out of 82). This criterion reduces the dataset from 650 to 291 players, effectively excluding those whose limited appearances could produce shooting distributions overly influenced by short-term variability or small-sample noise. Globally, the 291 considered players (45 % of the total) represent 182,025 shot attempts (83 % of the total), indicating that these players are the main contributors driving the league.

Based on this sample, we initially constructed two non-parametric shooting intensity maps using nonparametric kernel-based methods (Silverman 2018). The results are shown in Figure 2. The first map includes the 30 players with the highest shot volume (with an average of about 1,200 shots per player in the whole season), henceforth denoted as *High-Volume Shooters*, while the second group, termed as *Secondary Shooters*, aggregates all remaining players (with an average of about 560 shots per player). High-Volume Shooters have been selected as the players with the highest number of field goal attempts per game. Field goal attempts, rather than points scored, are used as the selection criterion because shot intensity maps are intended to model shooting behavior and spatial shot frequency, independently of the outcome. Using per-game averages instead of total attempts allows for comparability across players by controlling for differences in playing time and game participation, and ensures that the selected sample reflects consistently high offensive involvement rather than cumulative volume driven by availability. Alternative selection criteria could have been considered, including total field goal attempts, minutes played, usage rate, or points per game. Each of these measures captures different aspects of offensive contribution, but places emphasis on cumulative volume, playing time, or scoring efficiency rather than on the spatial distribution and frequency of shot attempts, which is the primary focus of the present analysis. Overall, the 30

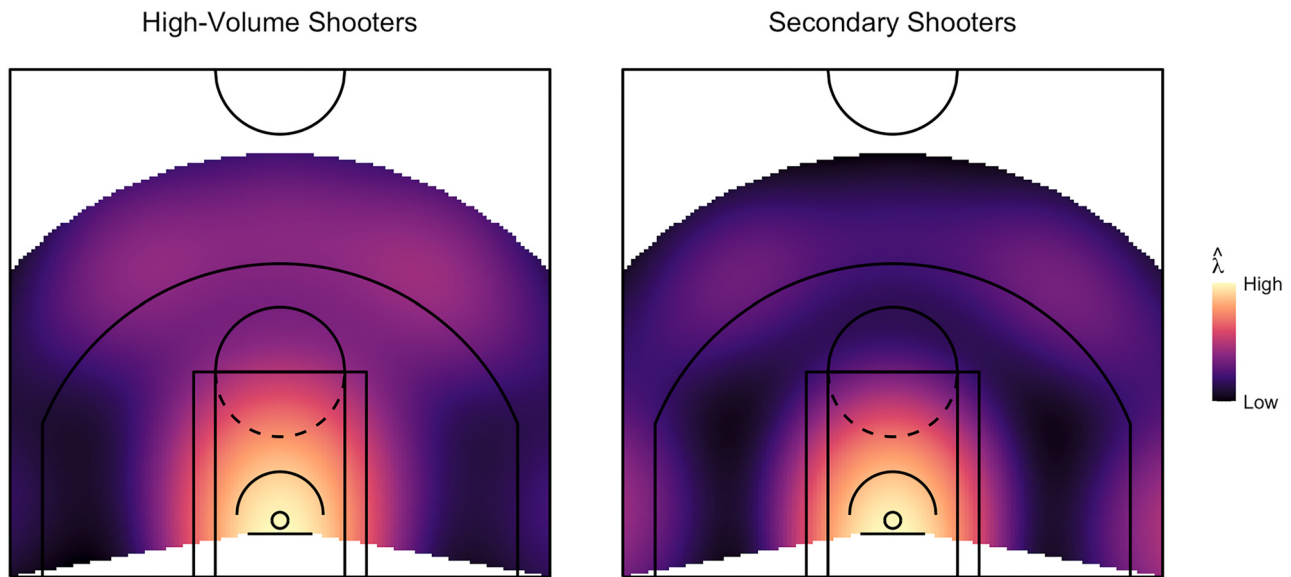


Figure 2: Nonparametric estimation of shooting intensity $\hat{\lambda}$ for High-Volume Shooters (left) and Secondary Shooters (right). Values of $\hat{\lambda}$ are displayed on a log-scale.

High-Volume Shooters² belong to 23 teams.³ The contrast between these two maps is striking. High-Volume Shooters make extensive use of the central court area: intensity decreases gradually with distance from the basket, yet it remains relatively high even in the mid-range. In contrast, Secondary Shooters show a much more polarized distribution, concentrating their attempts either near the rim or from the corners and wings beyond the three-point line.

This distinction highlights the structural organization of modern NBA offenses around two principal roles. High-Volume Shooters function as primary creators, generating scoring opportunities through distinctive shot patterns, whereas other players typically exploit the spacing and playmaking facilitated by the former (Wang and Zheng 2022b; Zajac et al. 2023). Given the extent of this spatial divergence, the remainder of the analysis focuses on High-Volume Shooters (so $J = 30$). This group captures the most meaningful spatial patterns and individual variations in shooting behavior, while its analysis is especially

important for decision-making processes that guide team composition (see Section 3.4 for further discussion).

Building on the methodological framework outlined in Section 2, our objective is to model the overall shooting intensity by accounting for the individual contributions of the 30 High-Volume Shooters. Specifically, we aim to determine whether these players exhibit shooting density patterns that closely resemble the aggregate heatmap, or whether they display distinct spatial tendencies that can be grouped into meaningful subcategories. The unified specification introduced in Equation (2) allows us to examine how the shooting behavior of each player deviates from the general trend in the two primary spatial dimensions: the distance and angle of the shot relative to the basket.

Following the approach adopted by Illian and Hendrichsen (2010) and to clarify how the model fitting is implemented in practice, we report below a simplified pseudo-code summarizing the main step of the estimation procedure, that was run using the *spatstat* R package.

² The High-Volume Shooters include: Anthony Davis, Anthony Edwards, Cade Cunningham, CJ McCollum, Damian Lillard, De'Aaron Fox, DeMar DeRozan, Devin Booker, Donovan Mitchell, Franz Wagner, Giannis Antetokounmpo, Ja Morant, Jalen Brunson, Jalen Green, Jaylen Brown, Jayson Tatum, Kevin Durant, Kyrie Irving, LaMelo Ball, LeBron James, Miles Bridges, Nikola Jokić, Paolo Banchero, RJ Barrett, Shai Gilgeous-Alexander, Stephen Curry, Trae Young, Tyler Herro, Tyrese Maxey, Victor Wembanyama.

³ The teams in which the high-volume shooters play are: Atlanta Hawks (ATL), Boston Celtics (BOS), Charlotte Hornets (CHA), Cleveland Cavaliers (CLE), Dallas Mavericks (DAL), Denver Nuggets (DEN), Detroit Pistons (DET), Golden State Warriors (GSW), Houston Rockets (HOU), Los Angeles Lakers (LAL), Memphis Grizzlies (MEM), Miami Heat (MIA), Milwaukee Bucks (MIL), Minnesota Timberwolves (MIN), New Orleans Pelicans (NOP), New York Knicks (NYK), Oklahoma City Thunder (OKC), Orlando Magic (ORL), Philadelphia 76ers (PHI), Phoenix Suns (PHO), Sacramento Kings (SAC), San Antonio Spurs (SAS), Toronto Raptors (TOR).

```

mppm(
  Y ~ ns(x, df = 7) +
    ns(y, df = 7) +
    ns(dist.m, df = 3) +
    bsp(angle.m, degree = 3,
        df = 3, periodic = TRUE),
  # fixed effects
  data = hy.data,
  # hyperframe with all shots
  random = ~ ns(dist.m, df = 3) +
    bsp(angle.m, degree = 3,
        df = 3, periodic = TRUE) |
    PLAYER_NAME,
  # player-specific deviations
  nd = c(40, 40)
  # spatial grid resolution
)

```

Here, `mppm()` denotes the fitting function. The object `hy.data` is a hyperframe containing the full set of shot locations and associated covariates, which is required to jointly estimate fixed and random effects. The spatial coordinates (x, y) are included only in the fixed-effects component. The covariates `dist.m` (distance to the basket) and `angle.m` (shooting angle) are included both as fixed effects, representing the overall trend, and as random effects grouped by player, allowing each shooter to deviate from the population-level pattern. The argument `nd` controls the spatial discretization used for numerical integration.

When spline bases are employed, the choice of the spline degree and the number of degrees of freedom is guided by two main considerations. The spline degree is selected based on domain knowledge. For example, shooting intensity as a function of distance from the basket is expected to be high close to the rim, decrease with distance, and increase again around the three-point area, a pattern that can be reasonably approximated by a cubic functional form. The number of degrees of freedom is instead chosen with computational complexity in mind. Given the high dimensionality of the mixed-effects Poisson point process model, limiting the flexibility of the spline bases helps ensure numerical stability and feasible computation.

In the following subsections, we explore the two estimated random components separately. First, we investigate the random effect on distance by fixing the shooting angle, assessing how the shooting intensity of each player varies with proximity to the basket. Second, we analyze the random effect on angle by fixing the shooting distance to identify potential asymmetries or directional preferences in individual shooting patterns. Section 3.3 discusses the problem of model validation and reports a series of analyses that were run to test the adequacy of our proposal

focusing on a few relevant examples. Finally, in Section 3.4 we present two examples of player-specific shooting intensity and scoring probability maps. These maps highlight individual spatial tendencies as deviations from the population-level trend and spatial scoring effectiveness, demonstrating the practical applicability of the model to generate actionable insights.

3.1 Player-specific effects along the distance dimension

To gain a deeper understanding of how players differ in their shooting behavior with respect to proximity to the basket, we first focus on the random effect associated with the shooting distance. By examining this component, while maintaining the shooting angle fixed as perpendicular to the baseline, we can explore how the intensity-distance relationship varies between players. This highlights individual shooting preferences and allows us to distinguish between players who rely heavily on close-range or long-range attempts and those who make more balanced use of intermediate areas. To this aim, we generate player-level prediction curves. Specifically, we select a grid of values that span the full range of observed shooting distances from the basket at an angle of $\pi/2$ and predicted the shooting intensity for each player at each point on the grid. This procedure yields, as shown in Figure 3, intensity-distance profiles for the 30 High-Volume Shooters.

The first key finding is that the dependence of the shooting intensity on the distance does not exhibit the same pattern between players. Instead, the 30 profiles reveal heterogeneous tendencies that reflect distinct offensive identities. To identify groups of players described by similar curves, we perform a cluster analysis of the 30 trajectories using a partitioning clustering procedure equivalent to a functional k -means algorithm. In this framework, each trajectory is represented as a vector of log-intensities, computed at the same distances from the basket. The similarity between two trajectories is quantified using the Euclidean distance between the corresponding vectors. This distance captures pointwise differences in amplitude between trajectories and assumes that all trajectories are aligned with respect to the distance from the basket and of equal length. To determine the optimal number of clusters, the clustering procedure is repeated for values of k ranging from 2 to 10. For each solution, three internal validation indices are computed, namely the Silhouette index (Rousseeuw 1987), the Dunn index (Dunn 1974) and the Modified Davies–Bouldin index (Kim and Ramakrishna 2005). These indices provide complementary assessments of clustering quality, taking into

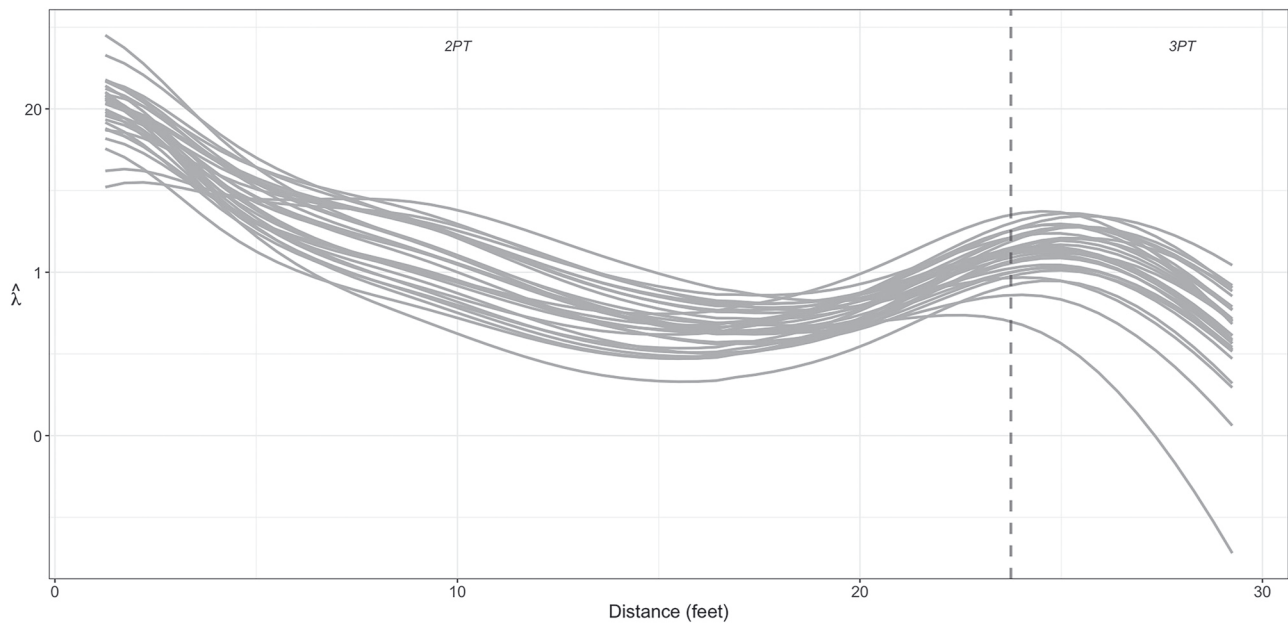


Figure 3: Relationship between fitted intensity $\hat{\lambda}$ and distance from the basket for each player of the High-Volume Shooters set. Values of $\hat{\lambda}$ are displayed on a log-scale. The dashed vertical line denotes the 3 points area.

account both intra-cluster compactness and inter-cluster separation. The results are reported in Table 1.

While more parsimonious solutions are supported by the Silhouette and Modified Davies–Bouldin indices, which suggest three clusters as the optimal solution, the six-cluster solution provided by the Dunn index is selected to improve interpretability and to capture subtle but meaningful variations in shooting patterns. Given the intrinsic similarity of shooting behavior among High-Volume Shooters, these differences are necessarily nuanced. Visual inspection of the resulting cluster plots confirms that the three-cluster solution produces a cluster that contains highly heterogeneous trajectories, representing players with markedly different

shooting patterns, while the six-cluster solution provides a more coherent grouping, with clusters corresponding to players exhibiting more comparable trajectory profiles and shooting styles (Figure 4). This observation is also confirmed by the resulting cluster-level intensity maps (Figure 5) which show distinct spatial structures across clusters and provide empirical support for the chosen level of granularity.

Cluster 1, composed of Miles Bridges (CHA) and RJ Barrett (TOR), includes players who concentrate a very high share of their attempts close to the rim, with shooting intensity that drops abruptly as distance increases. Only at three-point range their intensity rises again, producing a distinctly two-zone shot profile focused on athletic finishes at the basket and perimeter attempts, with virtually no usage of the mid-range area.

Cluster 2 groups DeMar DeRozan (SAC) and Kevin Durant (PHO), two of the few NBA players who continue to rely heavily and effectively on the mid-range area. Their trajectories highlight a deliberate preference for this zone, making them different in an era where mid-range shots have become increasingly rare.

Cluster 3, which includes Anthony Davis (LAL-DAL) and Giannis Antetokounmpo (MIL), represents physically dominant players whose offensive game naturally develops around the basket. Their shooting intensity peaks in the immediate vicinity of the rim, followed by a sharp decline with increasing distance, and no notable rise from three-point range. This pattern stems from their exceptional size,

Table 1: Internal validation indices for different values of k when clustering the intensity-distance curves. For each index, optimal values are reported in *italic*.

k	Silhouette	Modified Davies–Bouldin	Dunn
2	0.3252	1.1114	0.1422
3	<i>0.3417</i>	<i>0.9479</i>	0.1762
4	0.2941	1.1290	0.1360
5	0.1862	1.1558	0.2691
6	0.2437	1.0910	<i>0.3107</i>
7	0.2882	1.2296	0.2639
8	0.2502	1.4109	0.2133
9	0.3027	1.3822	0.2214
10	0.2901	1.4920	0.2214
11	0.2745	1.4819	0.2214
12	0.2512	1.4898	0.2214

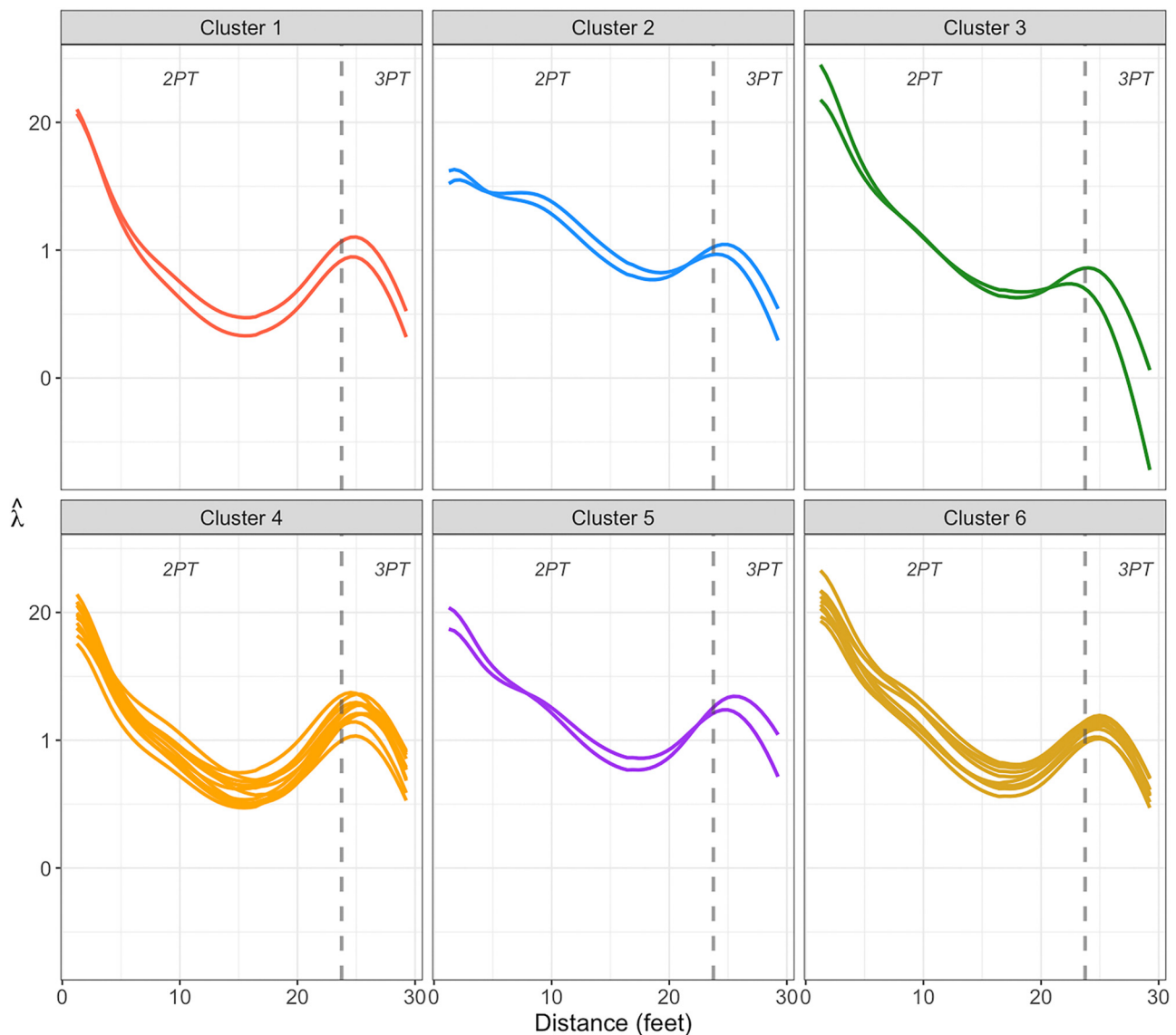


Figure 4: Relationship between $\hat{\lambda}$ and distance from the basket for each player of the High-Volume Shooters set. Players are grouped in 6 categories following the indications of the Dunn index (see Table 1).

strength, and finishing ability, which make close-range scoring their primary and most efficient offensive option.

Cluster 4, the largest group with 12 players,⁴ presents a more balanced version of the modern shots distribution seen with cluster 1. Their curves still show a clear emphasis on rim attempts and a prominent peak at the three-point line, but the decline in intensity between these two

areas is less abrupt than cluster 1, revealing a moderate amount of mid-range activity. This group includes players such as Anthony Edwards (MIN) and Tyrese Maxey (PHI), whose versatility allows them to combine efficient three-point shooting with strong finishing ability at the rim and occasional use of the mid-range area.

Cluster 5, consisting of Devin Booker (PHO) and Trae Young (ATL), features players with a more balanced distribution. They attempt fewer shots at the rim but frequently rely on pull-up jumpers and mid-range actions (roughly 8–15 feet from the basket). Compared with DeRozan and Durant, they exhibit greater three-point and rim volume,

⁴ Players belonging to cluster 4: Anthony Edwards (MIN), CJ McCollum (NOP), Damian Lillard (MIL), Donovan Mitchell (CLE), Jalen Green (HOU), Jayson Tatum (BOS), Kyrie Irving (MIL), LaMelo Ball (CHA), Stephen Curry (GSW), Tyler Herro (MIA), Tyrese Maxey (PHI), Victor Wembanyama (SAS).

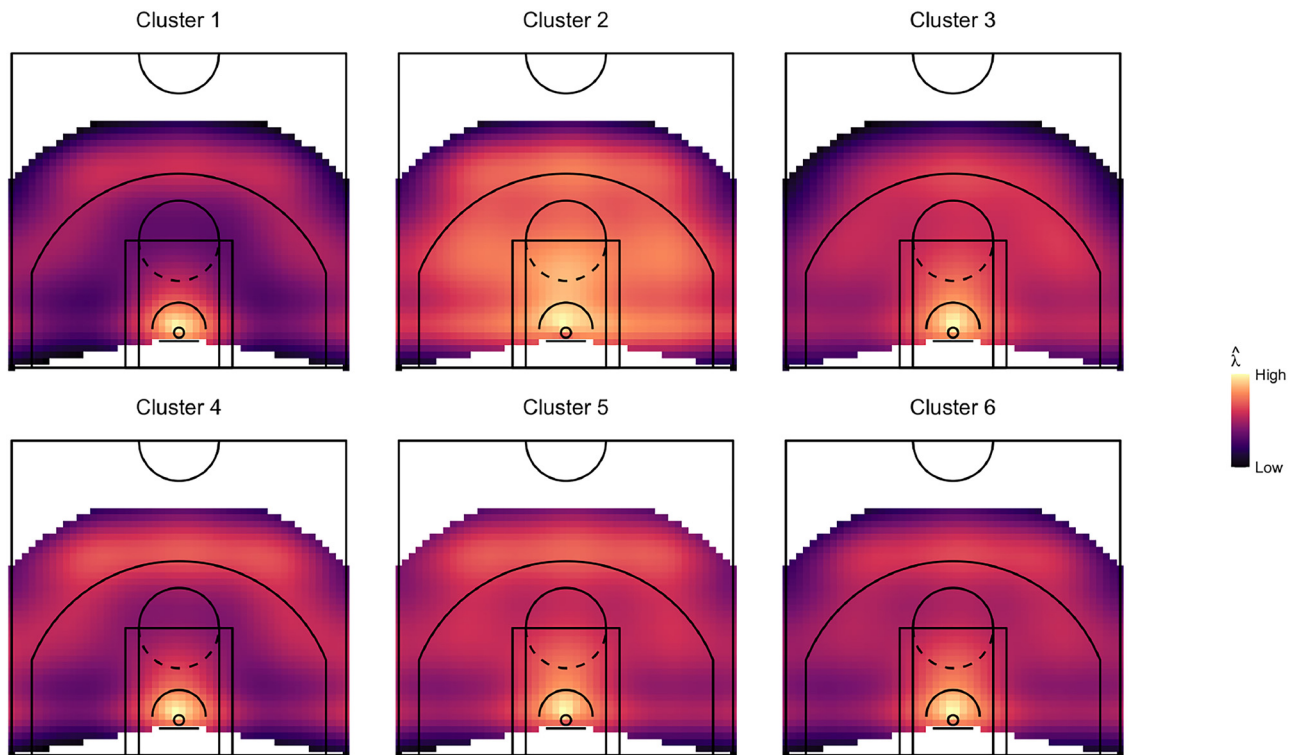


Figure 5: Cluster-level intensity $\hat{\lambda}$ maps. Values of $\hat{\lambda}$ are displayed on a log-scale.

placing them in an intermediate position between traditional mid-range scorers and the modern perimeter or rim oriented style.

Finally, Cluster 6, which includes 10 players,⁵ displays a trajectory similar to Cluster 4 but with a slightly lower three-point peak and a noticeable bump in intensity around 8–10 feet. This feature corresponds to players skilled at creating and converting tough shots such as fadeaways, turnarounds, or contested pull-ups. Among them are Jalen Brunson (NYN), Nikola Jokić (DEN), LeBron James (LAL), and the 2024/25 regular-season MVP Shai Gilgeous-Alexander (OKC).

All the aforementioned patterns, observed in the trajectories partitioned into the six clusters shown in Figure 4, are clearly reflected in the intensity maps presented in Figure 5.

Overall, this clustering reveals how diverse offensive profiles coexist even among High-Volume Shooters, from interior-dominant scorers to perimeter specialists and versatile mid-range creators, illustrating the rich spectrum of spatial shooting behaviors captured by the model. From a

practical perspective, this analysis offers valuable insight for player evaluation and team construction, allowing coaches and analysts to understand not only individual shooting profiles but also how players spatially complement each other on the court. For instance, our analysis shows that the Orlando Magic have two High-Volume Shooters assigned to Cluster 6 (Franz Wagner and Paolo Banchero). Although this may suggest potential predictability in their offensive patterns, it also highlights how their strengths can be mutually reinforced. Such insights derived from mixed-effects spatial modeling provide actionable information for coaching decisions, roster design, and strategic planning. Albeit to a lesser extent, additional game-related implications can be drawn by examining the player-specific effects also along the angle dimension, as discussed in the following subsection.

3.2 Player-specific effects along the angle dimension

We now turn to the random effects associated with the shooting angle. While the previous section examined how the shooting intensity varies with distance from the basket, this analysis focuses on its variation between different angular positions on the court. In doing so, our aim is to

⁵ Players belonging to cluster 6: Cade Cunningham (DET), De'Aaron Fox (SAC-SAS), Franz Wagner (ORL), Ja Morant (MEM), Jalen Brunson (NYK), Jaylen Brown (BOS), LeBron James (LAL), Nikola Jokić (DEN), Paolo Banchero (ORL), Shai Gilgeous-Alexander (OKC).

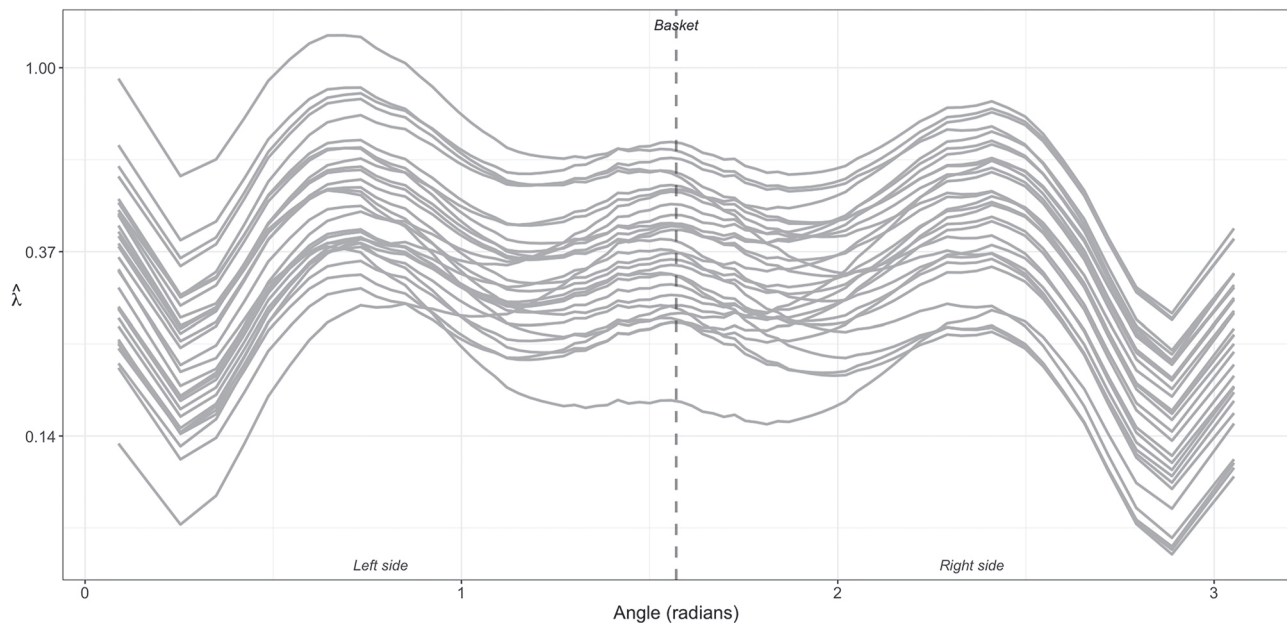


Figure 6: Relationship between intensity $\hat{\lambda}$ and the angle with respect to the basket for each player of the High-Volume Shooters set. Values of $\hat{\lambda}$ are displayed on a log-scale.

understand how players distribute their shooting attempts across the court, whether they maintain a balanced pattern between the two sides or display a clear preference for one. Such asymmetries often reveal key aspects of the offensive profile, including dominant-hand tendencies, preferred driving directions, and typical movement patterns.

To isolate the effect of the shooting angle, we visualize as in the previous section player-level prediction curves, derived from the model introduced in Section 2. The results are displayed in Figure 6. For each angular value, the estimated shooting intensity is reported for all players in the High-Volume Shooters group, fixing the distance at 18 feet from the basket. The overall pattern reveals two pronounced peaks at angles of approximately 0.8 and 2.4 radians. These peaks correspond to shots taken from positions that enable optimal use of the backboard in close-range finishes, a common feature of attempts near the rim.

This analysis becomes particularly insightful when examining individual players, as deviations from this general pattern may indicate asymmetrical or distinctive shooting behaviors. Given that variations in shooting angle have a more limited impact on overall intensity compared to distance, focusing on selected player profiles offers a clearer and more meaningful way of interpreting these directional effects.

Taking Shai Gilgeous-Alexander (OKC) and Giannis Antetokounmpo (MIL) as illustrative examples, we complement their intensity-angle profiles with corresponding shooting charts to further interpret the player-specific

random effects on the shooting angle. By jointly examining the predicted intensity-angle relationship of each player and their actual shot distribution on the court, we can directly visualize how the model captures their characteristic offensive behaviors.

The first case concerns Shai Gilgeous-Alexander, the 2024/2025 NBA regular-season MVP. His intensity angle profile (Figure 7, left) reveals a clear asymmetry, with the intensity of the shots concentrated on the left side of the court. This model-derived evidence, combined with inspection of his shot chart (Figure 7, right), allows us to characterize Shai Gilgeous-Alexander's offensive behavior more precisely than either sources alone. Gilgeous-Alexander frequently operates in isolation or pick-and-roll actions, where defenses are structured to try to deny him drives toward his dominant right hand. As a result, many of his scoring opportunities arise on the left side, where he effectively creates separation for step-back or turnaround jumpers. This behavior is typical among right-handed scorers, who often find a more natural body alignment and shooting rhythm when taking jump shots while dribbling with the left hand.

In contrast, Giannis Antetokounmpo exhibits a markedly different profile (Figure 8, left). His intensity-angle curve reveals a concentration of shooting intensity around the central and slightly left portions of the court. This information, together with his shot chart (Figure 8, right), offers a detailed picture of his offensive behavior that closely aligns with his real on-court tendencies. The scoring production of Antetokounmpo is

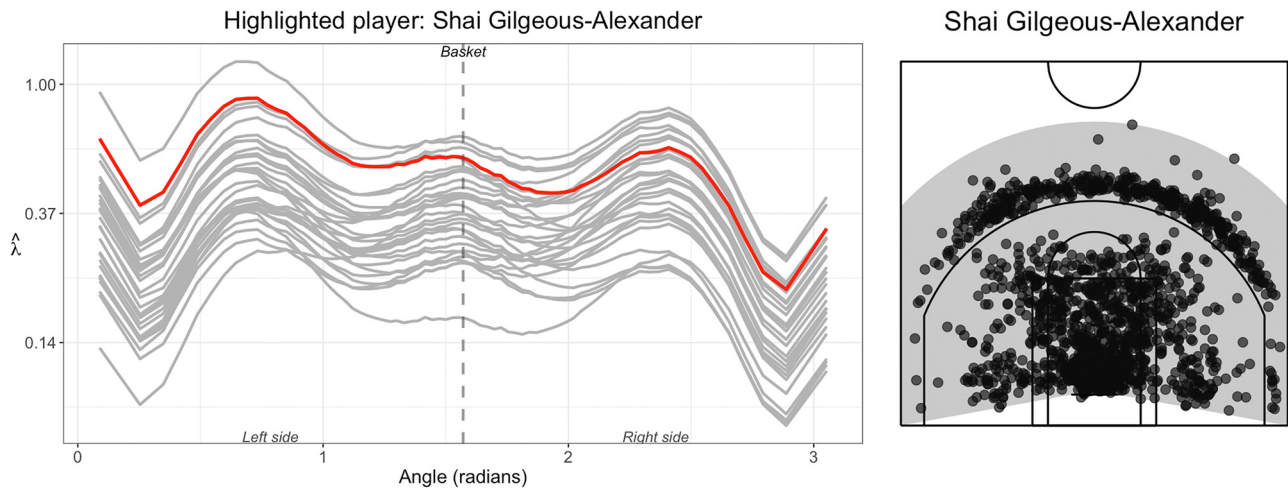


Figure 7: Left: Relationship between intensity $\hat{\lambda}$ and the angle with respect to the basket. Profile of Shai Gilgeous-Alexander (OKC) displayed in red versus all the other High-Volume Shooters in gray. Values of $\hat{\lambda}$ are displayed on a log-scale. Right: Shotchart of 1,656 shots.

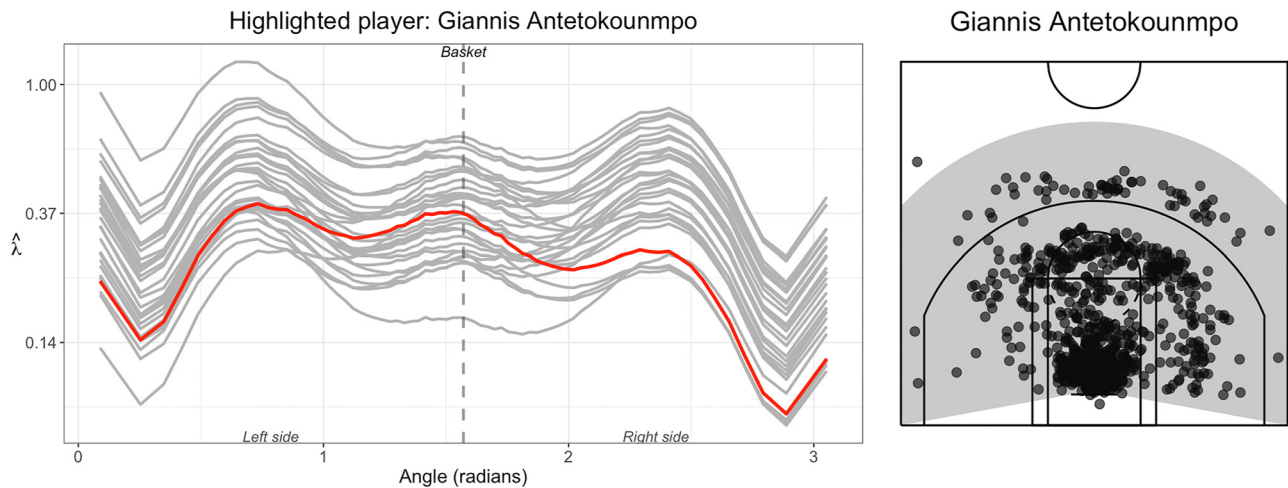


Figure 8: Left: Relationship between intensity $\hat{\lambda}$ and the angle with respect to the basket. Profile of Giannis Antetokounmpo (MIL) displayed in red versus all the other High-Volume Shooters in gray. Values of $\hat{\lambda}$ are displayed on a log-scale. Right: Shotchart of 1,319 shots.

heavily concentrated near the rim, where he dominates through direct drives and powerful finishes. His attacks are typically front-facing, leading to dunks or layups from the center of the paint rather than angled finishes off the backboard. The slight prevalence of left-side attempts corresponds to his frequent low-post isolations on that side, a common preference among right-handed players. Indeed, from that position they can face the middle of the floor with their strong hand and execute more natural fadeaway shots. These aspects explain the shape of his intensity-angle profile and confirm how the model effectively captures his distinctive, physically driven offensive style.

3.3 Model validation and analysis of residuals

After exploring the player-specific effects along the distance and angle dimensions derived from the proposed random-effects point process framework, we now focus on the problem of model validation and residuals' analysis. In fact, model validation represents a fundamental step for developing a proper data analysis pipeline as it ensures reliability and accuracy of the model's prediction, detects misspecifications, and strengthens the robustness of our analytical findings.

Following the approach and the ideas introduced in Baddeley et al. (2005), we define the *Pearson residual measure* for the j -th NBA player as:

$$\mathcal{R}_j^p(B) = \sum_{s_{ij} \in B} \hat{\lambda}(s_{ij})^{-1/2} - \int_B \hat{\lambda}^{1/2}(s) \, ds, \quad j = 1, \dots, J \quad (4)$$

where B denotes any possible region of W , $\hat{\lambda}(\cdot)$ is the fitted intensity, and the notation $s_{ij} \in B$ implies that summation is restricted only to the shots taken by the j -th player from region $B \subseteq W$. Intuitively, the first term in Equation (4) computes the (weighted) observed intensity in B , while the second term represents the (weighted) fitted intensity under the estimated model. Ideally, these two quantities should be close if the model fits the observed data adequately, producing a residual measure of approximately zero for every $B \in W$. More precisely, in Baddeley et al. (2005) it is shown that the weighting schema included in Equation (4) implies that the residual measure $\mathcal{R}_j^p(B)$ has mean 0 and variance equal to the area of B if the model is correctly specified. For a more complete and thorough introduction to residual analysis for spatial point patterns we refer to Baddeley et al. (2005) and Baddeley et al. (2015).

As an illustrative example, a smoothed version of the residual measure is reported in Figure 9, focusing on three representative players: LeBron James, Stephen Curry, and Jalen Brunson. As we can see, the residuals for LeBron James and Stephen Curry are concentrated around zero with no evident spatial patterns, highlighting that the proposed model adequately fits their shooting intensity. Such behavior is highly consistent among the majority of NBA players in our dataset, supporting the model's reliability and broader applicability. However, a handful of players, such as Jalen Brunson, exhibit slightly higher residuals and reveal local areas where the model overestimates or underestimates the true intensity. A plausible explanation for the spatial pattern observed for Jalen Brunson may be related to his left-handed shooting style. Left-handed players often display asymmetric shot selection, for instance taking a larger share of long-range shots from the right side of the

court and attempting more shots close to the basket on the left side. This behavior is consistent with the residual field shown in Figure 9, where the model tends to underestimate shooting intensity far from the basket on the right side of the court and close to the basket on the left. These localized discrepancies suggest that factors such as whether a player shoots primarily with the left or right hand may influence individual shooting patterns and may warrant further exploration in future work.

All in all, the statistical approach considered and validated here not only provides a coherent and realistic tool for modeling real-world dynamics, but can also be readily employed by managers and team builders to effectively construct a winning team. A qualitative discussion of these aspects is presented in the following section.

3.4 Comparative analysis and strategic implications

Having examined the random components associated with distance and shooting angle and assessed the model fit, we now apply the full methodological framework to construct player-specific intensity maps. We illustrate the procedure with two representative examples: Kevin Durant (PHO) and Miles Bridges (CHA). For each player, the intensity surface is obtained by generating a fine grid of points within the observation window. At each grid location, the corresponding covariates, namely distance and shooting angle, are computed, and the estimated model is used to predict the expected shooting intensity based on both the spatial coordinates and the player-specific random effects. This procedure produces a smooth, player-level surface that displays how the fitted model translates into spatial shooting behavior.

The resulting maps, shown in Figure 10, clearly reflect the distinctive spatial tendencies of each player's offensive behavior. Kevin Durant, as discussed in Section 3.1, belongs to Cluster 2, characterized by sustained midrange usage. His intensity surface confirms this tendency, showing a broad area of elevated values across the midrange region. At the

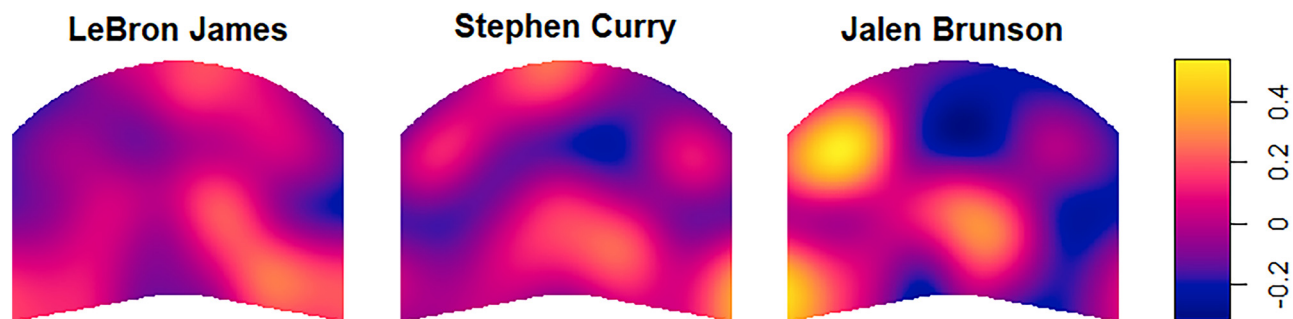


Figure 9: Smoothed Pearson residual field for the three representative players: LeBron James, Stephen Curry, and Jalen Brunson.

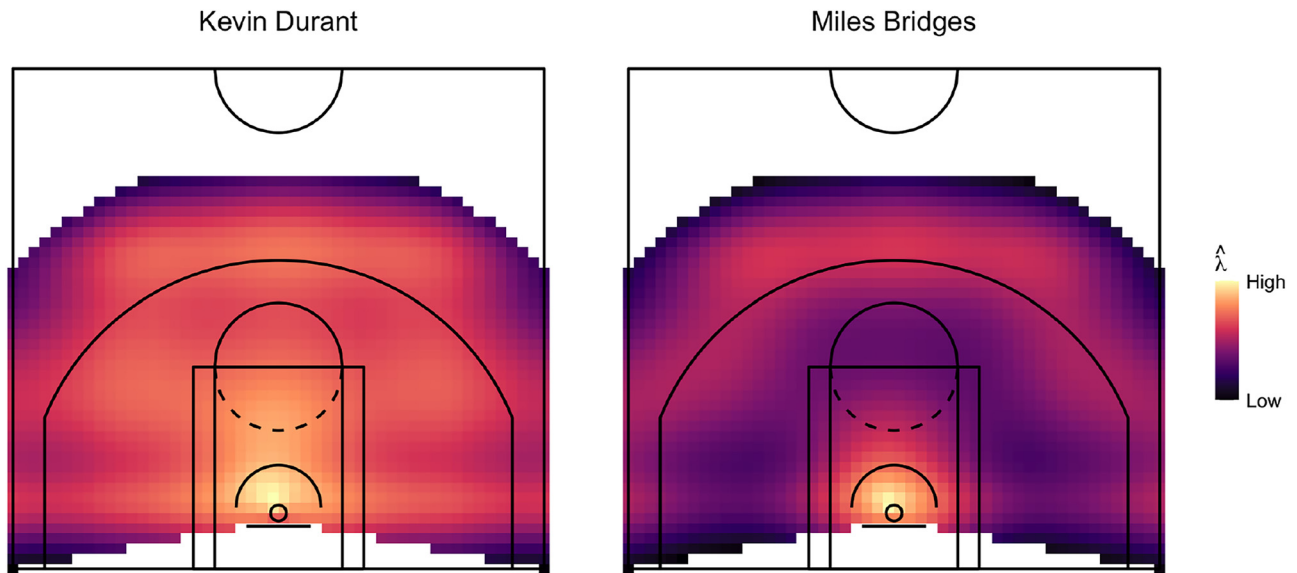


Figure 10: Predicted intensity $\hat{\lambda}(s)$ for Kevin Durant (left) and Miles Bridges (right).

same time, a very mild increase in intensity appears on the right side of the court, reflecting a subtle directional tendency captured by the random effect on shooting angle. Conversely, Miles Bridges (Cluster 1 with respect to distance) displays a sharply polarized distribution, with two dominant hotspots corresponding to finishes near the rim and three-point attempts. Although his overall pattern emphasizes these interior and perimeter areas, a slight increase in intensity can also be observed on the right side of the court, reflecting the subtle influence of the player-specific random effect on the shooting angle. Overall, as already anticipated, these results highlight that distance from the basket is the dominant covariate in explaining shooting intensity, while angle provides a secondary, finer-scale characterization of directional tendencies.

Beyond shooting intensity, the same framework can be extended to capture spatial variations in scoring efficiency through the construction of scoring probability maps. As already mentioned, these maps can be derived from the relative risk formulation introduced in Equation (3), which quantifies the probability of successfully converting a shot from any given location on the court.

To obtain these estimates, two separate mixed-effects models are estimated: one for made shots only and one for missed shots only, each sharing the same spatial structure adopted in the previous analysis. Following this, a fine grid of points is generated within the observation window. For every grid location, the corresponding covariates (distance and angle) are computed, and the two models are used to predict the expected intensities of shots made and missed.

Finally, the relative risk is calculated as the ratio between the predicted intensities of the successful shots and the sum of the intensities of made and missed shots, providing the estimated probability of making a shot from each position on the court.

This approach shifts the analytical perspective from where players shoot to how effectively they score in different areas. To illustrate these ideas, scoring probability maps are presented in Figure 11 for the same two previously considered players, namely Miles Bridges (CHA) and Kevin Durant (PHO). The contrast between the two players clearly emerges. Durant displays higher scoring probabilities across most shooting zones, reflecting his superior efficiency and shot-making versatility. Bridges, on the contrary, shows lower overall values, with effectiveness concentrated near the rim and selected three-point areas.

Despite these differences, both maps share a broadly consistent spatial structure, shaped by the design of the mixed-effects models and the influence of player-specific random effects on distance and angle. This common foundation ensures model stability while preserving meaningful inter-player variability.

The inclusion of a relative risk estimate completes the case study, extending the analytical framework from the shot frequency to the shot efficiency. In conclusion, the proposed methodology provides a comprehensive and fine-grained assessment of shooting performance, delivering both descriptive and comparative insights into the offensive characteristics of the considered High-Volume Shooters players.

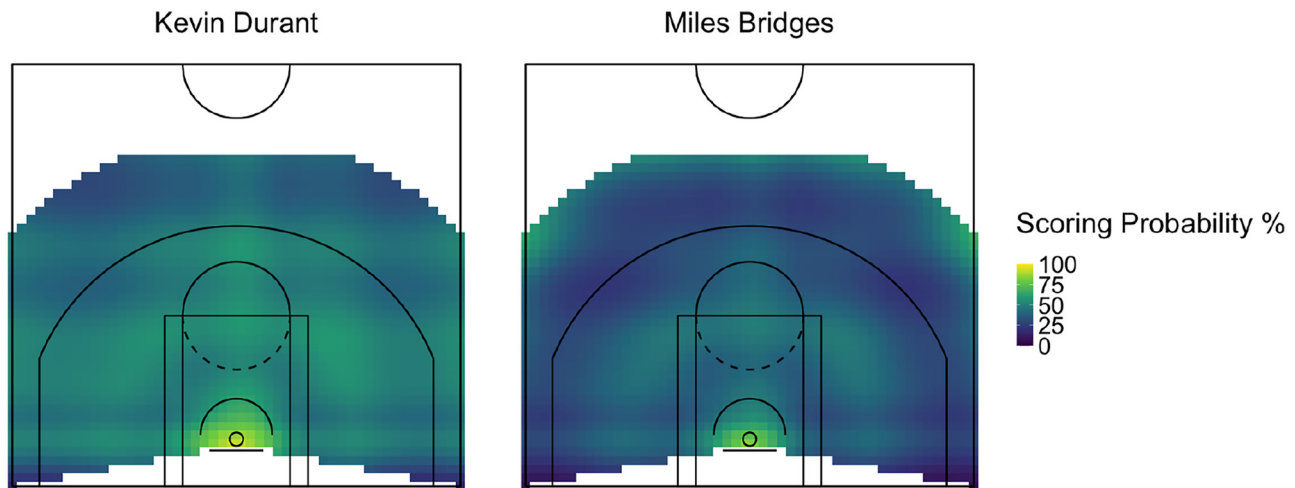


Figure 11: Relative risk maps $\hat{\rho}(s)$ for Kevin Durant (left) and Miles Bridges (right).

4 Conclusions and future works

The paper has been devoted to the practical and methodological value of applying semiparametric mixed-effects spatial models to basketball shooting analysis. Incorporating players as random effects to model individual deviations, while capturing nonlinear shooting trends with flexible splines, has enabled direct and interpretable comparisons among athletes within a unified statistical framework. This comparative perspective is crucial in sports analytics, where variations in spatial behavior provide actionable insights for applications ranging from game planning to scouting.

By incorporating two key spatial covariates, namely distance from the basket and shooting angle, the proposed model has captured fundamental aspects of shot creation and selection. Inspecting player-specific trajectories along these covariates has proved particularly informative, as it summarizes each athlete's shooting style in a concise and comparable manner, enabling the identification of characteristic patterns across the league.

At the same time, the model formulation allows the construction of player-specific shooting intensity maps, which provide a visually intuitive depiction of shooting tendencies across the court. Although these maps are less suitable for direct quantitative comparison than trajectory-based profiles, they complement the statistical analysis by illustrating how such tendencies manifest over the entire court surface. The analysis continues demonstrating the validity of model's prediction via the Pearson residual measure and concludes with the introduction of relative risk maps, which shift the focus from shot frequency to shot efficiency. These maps model the spatial probability of converting a field

goal, capturing how shooting success varies across the court. This perspective allows for evaluating players not only by where they shoot, but also by how effectively they score from different locations. Overall, the combination of shooting intensity modeling and relative risk estimation demonstrates how the spatial point pattern framework offers a comprehensive and versatile approach for the analysis of basketball shooting performance. It allows researchers and practitioners to progress from describing spatial tendencies to assessing their effectiveness, providing a complete spatial perspective on offensive behavior in the modern game.

The present study opens several avenues for future research. One line of work involves developing techniques to quantify the uncertainty associated with intensity, distance curves and their clustering, and relative risk maps. Methodologically, the statistical framework introduced herein could be extended in multiple directions. A particularly promising development is a spatio-temporal alternative of the model, which would capture temporal variations in players' shooting intensities throughout an NBA season. Such dynamics are likely influenced by factors such as physical condition, game scheduling, and injuries. Another line of investigation concerns improving the computational efficiency of the proposed methodology, scaling it to the complete set of NBA players (approximately 300 grouping levels). Alternative strategies for constructing spatial features may also be worth exploring. Data-driven approaches such as Nonnegative Matrix Factorization (NMF) have proven effective in basketball analytics for identifying archetypal shooting patterns or "shot types" from estimated intensity surfaces (Franks et al. 2015; Jiao et al. 2021; Miller et al. 2014). These methods could provide a complementary perspective to the distance-angle

parametrization adopted in this work. At the same time, NMF-based approaches entail their own methodological considerations: they require an initial intensity estimation step and involve design choices regarding the construction of the input matrix, both of which may influence the resulting components. A systematic comparison between spline-based approaches and data-driven decompositions would therefore be a valuable research direction, as it could help clarify the trade-offs between interpretability and predictive performance in the analysis of players shooting behavior. Finally, relaxing the Poisson process assumption would allow the model to better capture spatially clustered shooting behavior, for example, by employing Log Gaussian Cox processes (Cao et al. 2025; Møller et al. 1998), which could further improve the model's fit and address the sporadic cases with elevated residuals. However, estimating such models within a random-effects framework remains computationally demanding. Several of these extensions are currently under investigation and will be addressed in future work.

Acknowledgments: We thank BigDataBall, a data provider which leverages computer-vision technologies to enrich and extend sports datasets with lots of unique metrics, for supplying play-by-play data used to test the model on additional datasets. The authors gratefully acknowledge Marica Manisera for her valuable discussions and constructive input during the preparation of the initial version of this manuscript.

Research ethics: Not applicable.

Informed consent: Not applicable.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission. All authors have contributed equally. MLC, AC, AG, PZ: conceptualization, methodology, formal analysis, investigation and writing.

Use of Large Language Models, AI and Machine Learning Tools: None declared.

Conflict of interest: The authors state no conflict of interest.

Research funding: M.L. Carlesso, P. Zuccolotto: Research Project PRIN 2022, granted by European Union – Next Generation EU, “Statistical Models and AlgoRiThms in sports (SMARTsports). Applications in professional and amateur contexts, with able-bodied and disabled athletes”, project nr. 2022R74PLE, CUP: D53D23005950006. A. Cappozzo: partial support from UCSC D1 research grant.

Data availability: Data were downloaded in June 2025 and are publicly available at the following link: https://github.com/shufinskiy/nba_data.

References

- Baddeley, A., Turner, R., Møller, J., and Hazelton, M. (2005). Residual analysis for spatial point processes (with discussion). *J. R. Stat. Soc. Ser. B: Stat. Methodol.* 67: 617–666.
- Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial point patterns: methodology and applications with R*. CRC Press, New York, NY.
- Bell, M.L. and Grunwald, G.K. (2004). Mixed models for the analysis of replicated spatial point patterns. *Biostatistics* 5: 633–648.
- Berman, M. and Turner, T.R. (1992). Approximating point process likelihoods with glim. *J. R. Stat. Soc. Ser. C: Appl. Stat.* 41: 31–38.
- Cao, J., Cai, Q., Waller, L.A., Hickson, D.A., Hu, G., and Kang, J. (2025). What influences the field goal attempts of professional players? Analysis of basketball shot charts via log Gaussian cox processes with spatially varying coefficients. *arXiv preprint arXiv:2503.02137*.
- Carlesso, M.L., Cappozzo, A., Manisera, M., and Zuccolotto, P. (2024). Scoring probability maps in the basketball court with indicator kriging estimation. *Comput. Stat.*: 1–21, <https://doi.org/10.1007/s00180-024-01564-4>.
- Davis, J., Bransen, L., Devos, L., Jaspers, A., Meert, W., Robberechts, P., Van Haaren, J., and Van Roy, M. (2024). Methodology and evaluation in sports analytics: challenges, approaches, and lessons learned. *Mach. Learn.* 113: 6977–7010.
- Demidenko, E. (2013). *Mixed models: theory and applications with R*. John Wiley & Sons, Hoboken, New Jersey.
- Diggle, P.J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC Press, New York, NY.
- Dominicy, Y. and Ley, C. (2023). *Statistics meets sports: what we can learn from sports data*. Cambridge Scholars Publishing, London, UK.
- Dunn, J.C. (1974). Well-separated clusters and optimal fuzzy partitions. *J. Cybern.* 4: 95–104.
- Franks, A., Miller, A., Bornn, L., and Goldsberry, K. (2015). Characterizing the spatial structure of defensive skill in professional basketball. *Ann. Appl. Stat.* 9: 94–121.
- Friedman, J.H., Hastie, T., and Tibshirani, R. (2009). *The elements of statistical learning*. Springer Series in Statistics, New York, NY, USA.
- Illian, J.B. and Hendrichsen, D.K. (2010). Gibbs point process models with mixed effects. *Environmetrics* 21: 341–353.
- Jiao, J., Hu, G., and Yan, J. (2021). A Bayesian marked spatial point processes model for basketball shot chart. *J. Quant. Anal. Sports* 17: 77–90.
- Kim, M. and Ramakrishna, R. (2005). New indices for cluster validity assessment. *Pattern Recognit. Lett.* 26: 2353–2363.
- Kubatko, J., Oliver, D., Pelton, K., and Rosenbaum, D. (2007). A starting point for analyzing basketball statistics. *J. Quant. Anal. Sports* 3: 1.
- Mansurali, A. and Mahmoud, A.B. (2024). Sports analytics: data-driven sports and decision intelligence. In: *Sports analytics: data-driven sports and decision intelligence*. Springer, pp. 1–17.
- Miller, A., Bornn, L., Adams, R., and Goldsberry, K. (2014). Factorized point process intensities: a spatial analysis of professional basketball. In: *International conference on machine learning*. PMLR, pp. 235–243.
- Møller, J., Syversveen, A.R., and Waagepetersen, R.P. (1998). Log Gaussian cox processes. *Scand. J. Stat.* 25: 451–482.
- Narayanan, S., Kosmidis, I., and Dellaportas, P. (2023). Flexible marked spatio-temporal point processes with applications to event

- sequences from association football. *J. R. Stat. Soc. Ser. C: Appl. Stat.* 72: 1095–1126.
- Pinheiro, J. and Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media, New York, NY.
- R Core Team (2024). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reich, B.J., Hodges, J.S., Carlin, B.P., and Reich, A.M. (2006). A spatial analysis of basketball shot chart data. *Am. Stat.* 60: 3–12.
- Rousseeuw, P.J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20: 53–65.
- Sandholtz, N., Mortensen, J., and Bornn, L. (2020). Measuring spatial allocative efficiency in basketball. *J. Quant. Anal. Sports* 16: 271–289.
- Scrucca, L. and Karlis, D. (2025). A model-based approach to shot charts estimation in basketball. *Comput. Stat.*: 1–18, <https://doi.org/10.1007/s00180-025-01599-1>.
- Shortridge, A., Goldsberry, K., and Adams, M. (2014). Creating space to shoot: quantifying spatial relative field goal efficiency in basketball. *J. Quant. Anal. Sports* 10: 303–313.
- Silverman, B.W. (2018). *Density estimation for statistics and data analysis*. Routledge, New York, NY.
- Wang, F. and Zheng, G. (2022a). Examining positional difference in basketball players' field goal accuracy using Bayesian hierarchical model. *Int. J. Sport Sci. Coach.* 17: 848–859.
- Wang, F. and Zheng, G. (2022b). What are the changes in basketball shooting pattern and accuracy in national basketball association in the past decade? *Front. Psychol.* 13: 917980.
- Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models a pseudo-likelihood approach. *J. Stat. Comput. Simul.* 48: 233–243.
- Wood, S.N. (2017). *Generalized additive models: an introduction with R*. Chapman and Hall/CRC, New York, NY.
- Yin, F., Jiao, J., Yan, J., and Hu, G. (2022). Bayesian nonparametric learning for point processes with spatial homogeneity: a spatial analysis of nba shot locations. In: *International conference on machine learning*. PMLR, pp. 25523–25551.
- Yin, F., Hu, G., and Shen, W. (2023). Analysis of professional basketball field goal attempts via a Bayesian matrix clustering approach. *J. Comput. Graph. Stat.* 32: 49–60.
- Zajac, T., Mikołajec, K., Chmura, P., Konefał, M., Krzysztofik, M., and Makar, P. (2023). Long-term trends in shooting performance in the nba: an analysis of two-and three-point shooting across 40 consecutive seasons. *Int. J. Environ. Res. Public Health* 20: 1924.
- Zuccolotto, P., Sandri, M., and Manisera, M. (2021). Spatial performance indicators and graphs in basketball. *Soc. Indic. Res.* 156: 725–738.
- Zuccolotto, P., Sandri, M., and Manisera, M. (2023). Spatial performance analysis in basketball with cart, random forest and extremely randomized trees. *Ann. Oper. Res.* 325: 495–519.