Towards a Spatial Analysis of Shooting in Philippine Basketball: Applications in the University Athletics Association of the Philippines Men's Basketball Tournament (Season 81)

FIRST COMPLETE DRAFT (NOT FINAL COPY)

Ben Hur Pintor
MS Geomatics Engineering (GeoInformatics)
National Graduate School of Engineering
University of the Philippines

1. Introduction

1.1 Background

Basketball is spatial. Any event that occurs during a basketball game—a made shot, a missed shot, a rebound—has a corresponding spatial or spatio-temporal information embedded in it and, one can argue, that location oftentimes plays an important role in its occurrence or success.

If you think of the basketball court as a map, a parcel of the earth, or simply a cartesian coordinate plane then every location on the court can be specified by a coordinate pair. If we consider one type of basketball event—a shot or field goal—every occurrence of this event on the court will have its own corresponding coordinates. Aside from coordinates, these field goals can also have attributes or marks—the name of the player, the name of the team, the opponent, the time left on the clock, whether the shot was made or not, whether it was defended—that provide other information about the field goal. If we take this collection of field goals, what we actually have is a collection of points in space that is, similar to any spatial point dataset, susceptible to spatial analysis. This is why it makes sense to analyze basketball from a spatial perspective.

The advent of player tracking systems in basketball such as the SportVU Player Tracking System used in the National Basketball Association (NBA) has enabled research and studies that use location data to create a deeper understanding of the spatial nature of the game and even challenge conventional wisdom. Optical tracking data has been used to study shooting and introduce spatially-aware metrics for analysing shooting tendencies and potency—metrics like Spread and Range (Goldsberry, 2012) that measure how much of the court a player shoots and scores from, Spatial Shooting Effectiveness (SSE) and Points Over League Average (POLA) (Shortridge et al., 2014) that compare a player's actual and expected scoring performance based on the spatial distribution of his shots, and Lineups Points Lost (LPL) (Sandholtz et al., 2019) that looks at field goals as an optimal allocation problem and computes the difference between what a five-man lineup is expected to score if they optimized their choice of field goals versus what the same five-man lineup actually scores. Spatial analysis has also been used to study and deconstruct rebounding with new metrics for positioning, hustle, and conversion of rebounds generated using a Voronoi-tessellation approach combined with a spatial probability distribution (Maheswarean et al., 2012; Maheswaran et al., 2014). Defense has also been studied (Goldsberry et al., 2013; Franks et al., 2015) as well as the effects of player motion on creating open shots (D'Amour et al., 2015). Optical tracking data has even been used to determine the value of different areas on the court (Cervone et al., 2016a) and predict the outcomes of basketball possessions (Cervone et al., 2016).

Truly, the application of spatial analysis in basketball has added a new dimension to how we view the game.

BN HR

1.2 Basketball analysis in the Philippines

Advanced and spatial analytics do not seem to be part of mainstream basketball analysis in the country as evidenced by the lack of available data and information released by popular basketball leagues such as the Philippine Basketball Association (PBA), Maharlika Pilipinas Basketball League (MPBL), University Athletics Association of the Philippines (UAAP), and National Collegiate Athletics Association (NCAA). Compared to the NBA which releases a wide variety of statistics in its Official NBA Stats website (https://www.nba.com/stats/) and in other sites such as Basketball Reference (https://www.basketball-reference.com/), it's hard to find any publicly available advanced statistics websites for Philippine basketball. One example is HumbleBola (https://twitter.com/humblebola_) although at the time of this writing, their website (https://humblebola.com) is no longer available. Meanwhile, websites of mainstream media outlets such as ABS-CBN that cover collegiate basketball leagues and the official PBA (https://www.pba.ph/stats) and MPBL (https://mpbl.web.geniussports.com/competitions/) websites only show basic counting and rate statistics such as the number of field goals, field goal percentage, etc.

That's not to say that analytics isn't used in Philippine basketball. An article by Socamos (2018) highlighted the use of data and analytics by the Alaska Aces in the PBA and the National University Bulldogs in the UAAP. Another article by Murillo (2019) mentioned that "while use of sports analytics in the Philippines is acknowledged to be still in its infancy, it has nonetheless made in-roads with more teams and organizations recognizing its value and potential" noting that in the PBA, the Alaska Aces and the Ginebra San Miguel Kings were using sports analytics with designated personnel. The same article stated that "despite the pickup that sports analytics has gained in the Philippines,... its appreciation and use of it are still a way off as compared to those in other countries, especially in those with readily available technology and equipment" (Murillo, 2019). In the UAAP, personal correspondence with individuals who work on managing the UP Fighting Maroons Men's Basketball Team informed me that the team uses a combination of videos and statistics in their game preparation and analysis.

1.3 Incorporating the spatial nature of shooting

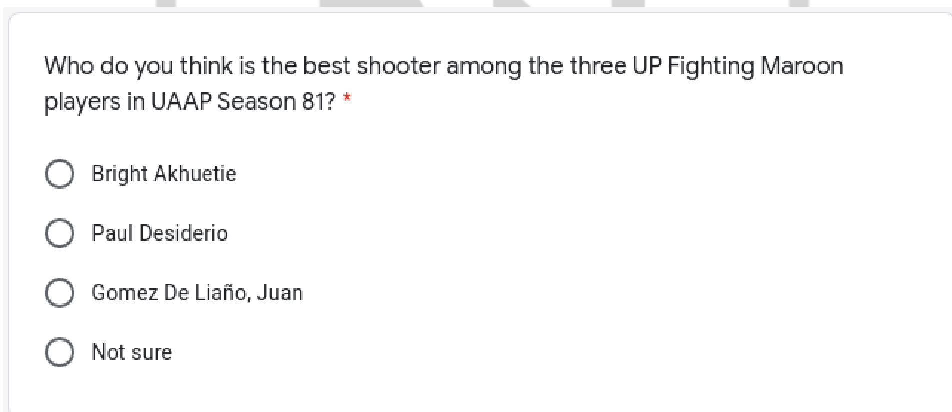
Players shoot differently at different locations on the court and this variation has effects on how players are rated, how they are trained, and how teams create strategies. Conventional shooting statistics such as the number of made field goals or the field goal percentage that do not account for the spatial nature of shooting provide incomplete and, at times, incorrect information about the phenomenon that can lead to incorrect conclusions.

Take for example two players who both attempted 100 field goals over the course of a season—Player A shot 50% from the court and Player B shot 40% from the court. When only this information is provided, it may be assumed that Player A is a better shooter than Player B but this conclusion fails to consider that the two players may not be shooting from the same locations on the court. This is problematic because there are several instances when Player B can be considered a better shooter than Player A in this scenario. One example is when Player A is only taking two-pointers and thus scoring 1 point per attempt while Player B is only taking three-pointers and thus scoring 1.2 points per attempt, outscoring Player A by 0.2 points with every shot. Another example is when Player A takes shots at areas where other players average shooting 60% from while Player B takes shots from areas where other players average shooting 33% from. This means that Player A is actually a below-average shooter while Player B is an above-average one.

So what makes a good shooter? Is it someone who simply makes a high percentage of his shots like a center or power forward who dominates the interior but rarely takes shots outside the paint? Is it someone who can make shots that are far from the basket, a three-point specialist perhaps? Or is it someone who is a threat to score everywhere on the court even though they may not shoot particularly well overall? When answering this question, conventional statistics such as FG%, 3P%, PPA, eFG, and TS% provide a generalized version of a player's shooting or scoring ability but they don't provide context as to the locations where players shoot and how well they shoot at these locations. This is disappointing because, when you think about it, being able to pinpoint where a player or team scores is a powerful tool to have when creating basketball strategies and game plans.

BN
HR

As part of the research, a survey shown in Figure 1.1 was conducted where respondents were asked to select who they thought was the best shooter among three players. They weren't explicitly told that they were selecting from the same three players in all the questions and in each of the questions, only some information was shown to them. In the first question they were given just names of players; in the second, the FG and FG% of the players; and in the third, the 3P and 3P% of the players. Table 1.1. summarizes the players and their shooting statistics.


Who do you think is the best shooter among the three UP Fighting Maroon players in UAAP Season 81? *

- Bright Akhuetie
- Paul Desiderio
- Gomez De Liaño, Juan
- Not sure

The image below shows the total number of field goals made [FGM] and the field goal percentage [FG%] (field goals made / field goals attempted) of three players. Which one do you think is the best shooter among them? *

PLAYER	Field Goals Attempted (FGA)	Field Goals Made (FGM)	Field Goal Percentage (FG%)
Player A	180	78	43.33%
Player B	177	106	59.89%
Player C	176	72	40.91%

- Player A
- Player B
- Player C

The image below shows the total number of three-pointers made [3PM] and the three-point percentage [3P%] (three-pointers made / three-pointers attempted) of three players. Which one do you think is the best shooter among them? *

A 3-point field goal is a field goal made beyond the three-point line, a designated arc (~6.75 m +) surrounding the basket.
A successful attempt is worth three points, in contrast to the two points awarded for field goals made within the three-point line.

PLAYER	3-point Field Goals Attempted (3PA)	3-point Field Goals Made (3PM)	3-point Field Goal Percentage (3P%)
Player A	73	23	31.51%
Player B	81	24	29.63%
Player C	3	0	0.00%

- Player A
- Player B
- Player C

Figure 1.1

Questions in the "Who is the best shooter?" survey

player	**team**	**2PA** # of 2pt FGs attempted	**2P** # of 2pt FGs made	**2P%** 2P / 2PA	**3PA** # of 3pt FGs attempted	**3P** # of 3pt FGs made	**3P%** 3P / 3PA
B. Akhuetie	UP	174	106	0.6092	3	0	0.0000
Ju. GDL	UP	107	55	0.5140	73	23	0.3151
P. Desiderio	UP	95	48	0.5003	81	24	0.2963

player	**team**	**FGA** # of FGs attempted	**FG** # of FGs made	**FG%** FG / FGA	**eFG%** (FG + 0.5 * 3P) / FGA	**PPA** ((2 * 2P) + (3 * 3P)) / FGA
B. Akhuetie	UP	177	106	0.5989	0.5989	1.20
Ju. GDL	UP	180	78	0.4333	0/4972	0.99
P. Desiderio	UP	176	72	0.4091	0.4773	0.95

Table 1.1

Shooting statistics of B. Akhuetie, Ju. Gomez de Liaño, and P. Desiderio for UAAP S81

The results of the survey were interesting in that the respondents rarely selected the same player for all three questions. Only 3/82 selected the same player in all three questions while 32/82 selected a different player for each of the questions. They were consistent in choosing the player with the best percentage as the best shooter. More than 90% of the time they selected the player with the best FG% in question 2 (73/82) and the best 3P% in question 3 (77/82) even though these were 2 different players.

Figure 1.2 shows the results of the survey while Figure 1.3 shows a shot chart of the three players.

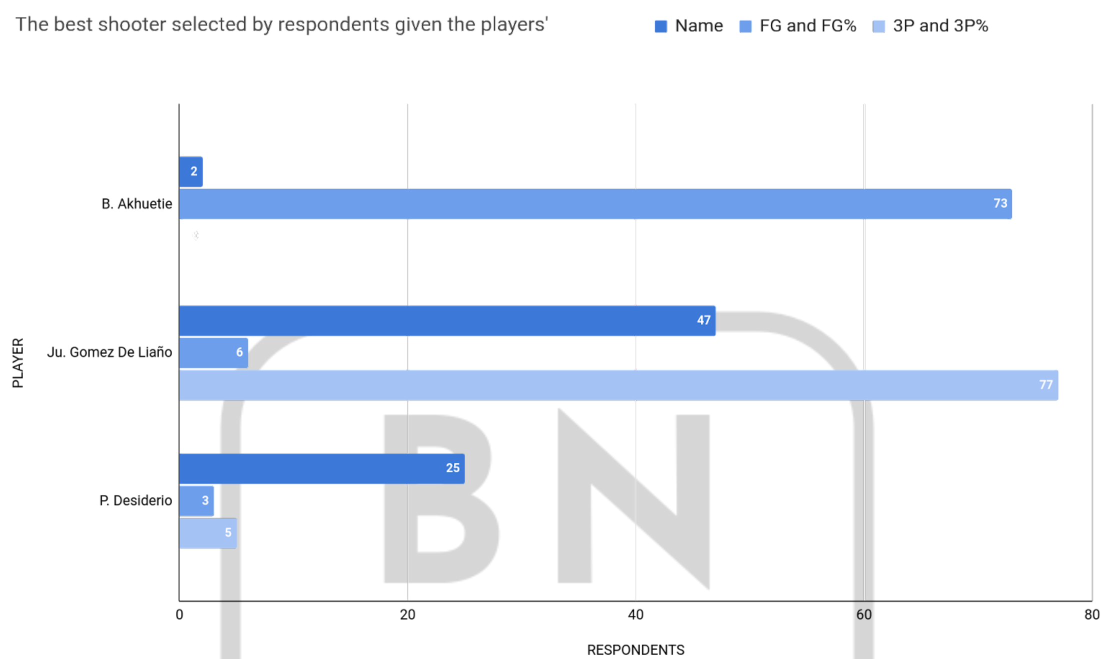


Figure 1.2

The best shooter as chosen by respondents

BN HR

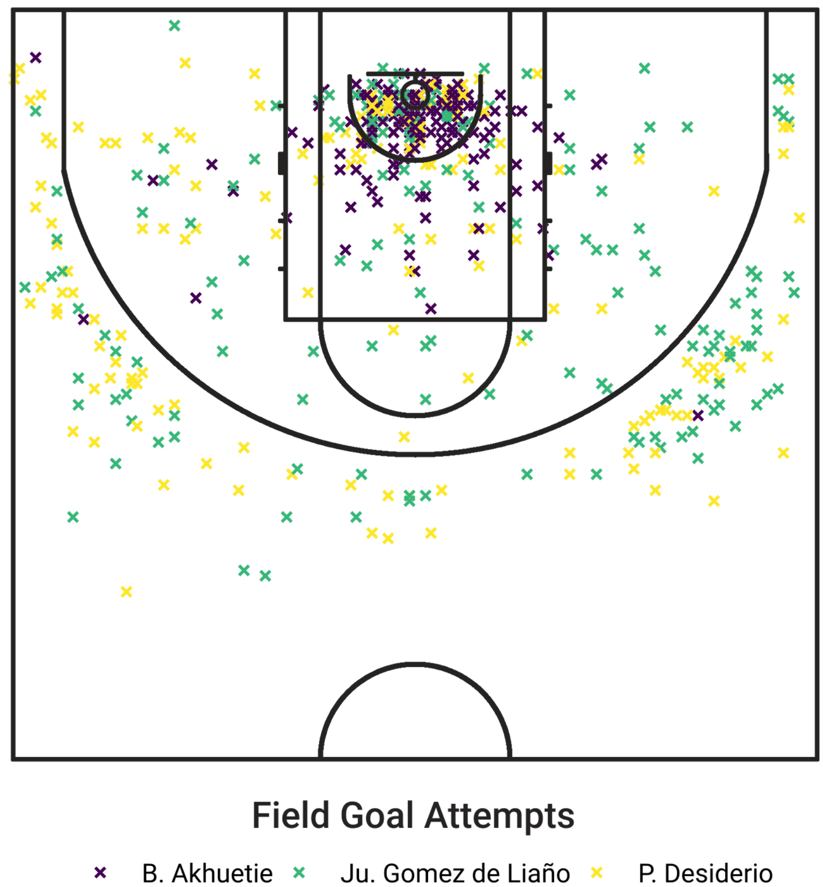


Figure 1.3

Shot chart of Bright Akhuetie, Juan GDL, and Paul Desiderio for UAAP S81

Bright was almost always selected as the best shooter when the respondents were only shown the Field Goal information as was the case with Juan in the question showing 3P information, but an interesting observation is that Desiderio was selected as the best shooter 25x just from name alone but was not selected more than 5x when the respondents were shown statistics -- probably a testament to the effect that reputation has when assessing players.

It's also worth noting that a lot of the respondents chose "Someone who can make a high percentage of his shots from different areas on the court." as their definition of a good shooter. This can be the reason why Bright was almost never chosen as the best shooter for questions 1 and 3. He had no 3P game to speak of as seen in the statistics and the distribution of his field goals. So even though he had the best metrics among the players in all but 3P shooting, respondents rarely chose him as their best shooter just from his name alone.

Conventional shooting statistics also have trouble differentiating players with similar shooting statistics. Table1.2 shows the shooting statistics of Paul Desiderio, Dave Ildefonso, and Sean Manganti—players of UP, NU, and Adamson respectively in Season 81. If you look at their stats, they are practically the same players with similar number of attempts, percentages, and metrics across the board. This might lead people to believe that they also shoot the same way but this isn't the case.

player	team	2PA # of 2pt FGs attempted	2P # of 2pt FGs made	2P% 2P / 2PA	3PA # of 3pt FGs attempted	3P # of 3pt FGs made	3P% 3P / 3PA
P. Desiderio	UP	95	48	0.5003	81	24	0.2963
D. Ildefonso	NU	110	57	0.5182	85	23	0.2706
S. Manganti	ADU	106	56	0.5283	66	15	0.2273

player	team	FGA # of FGs attempted	FG # of FGs made	FG% FG / FGA	eFG% (FG + 0.5 * 3P) / FGA	PPA ((2 * 2P) + (3 * 3P)) / FGA
P. Desiderio	UP	176	72	0.4091	0.4773	0.95
D. Ildefonso	NU	195	80	0.4103	0.4692	0.94
S. Manganti	ADU	172	71	0.4128	0.4564	0.91

Table 1.2

Shooting statistics of P. Desiderio, D. Ildefonso, and S. Manganti for UAAP S81

Figure 1.4 shows a Voronoi tessellation of the players' field goals using the mean coordinate of clusters computed using a K-means clustering with K=8. From these images, the nuances in their shooting styles that are not readily apparent from the conventional statistics can be seen. Paul's clusters are fairly symmetrical, Dave has a cluster of shots near the basket, while Sean has left and right clusters for shots in the paint. These maps show that even though the three players have similar shooting statistics, they have, in fact, different habits of shooting.

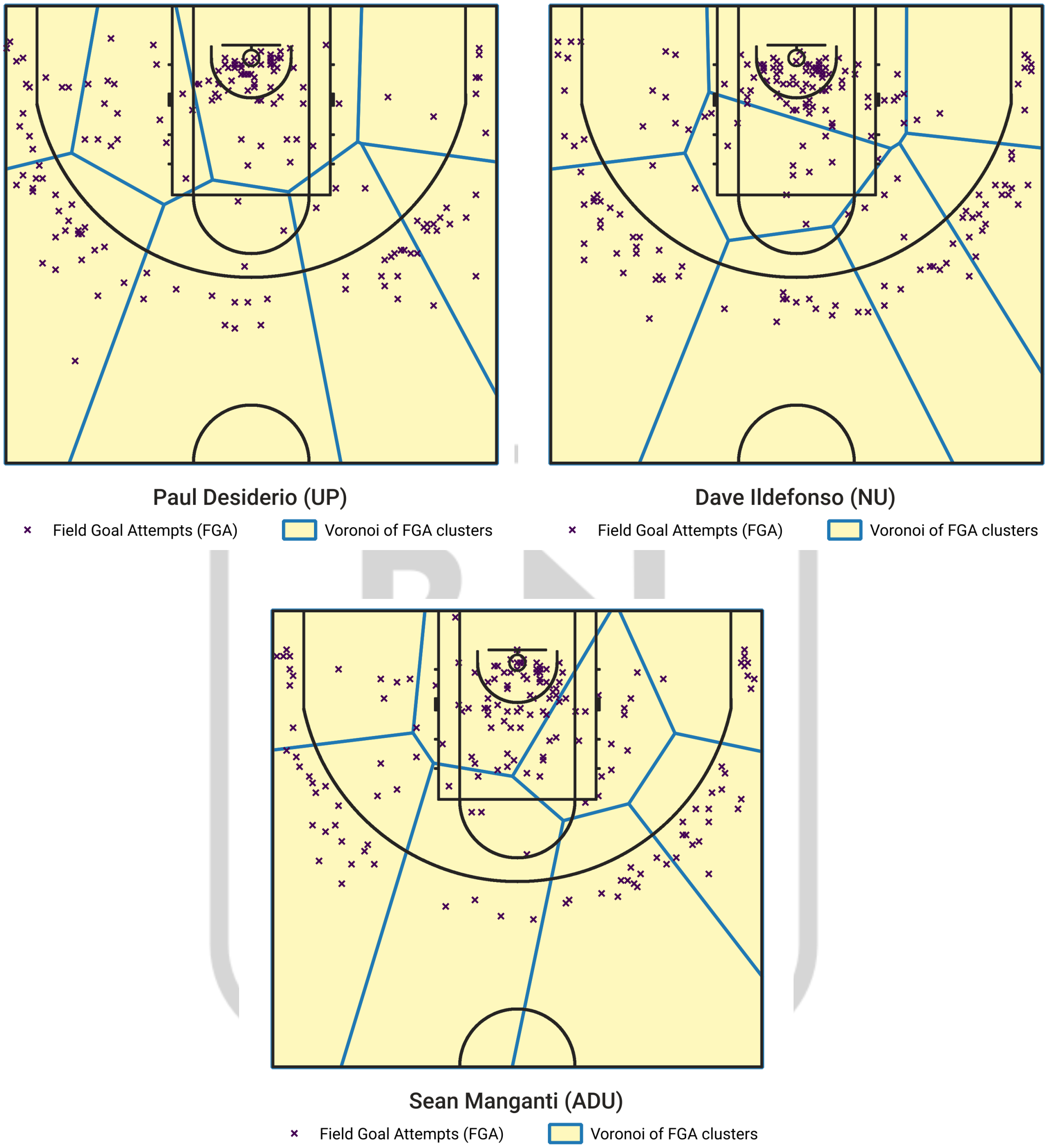


Figure 1.4

Field goal clusters and voronoi of P. Desiderio (top-left), D. Ildefonso (top-right), and S. Manganti (bottom) for UAAP S81

1.4 Overview of the research

The current state of basketball analysis in the Philippines provides an opportunity to utilize spatial analysis techniques in Philippine basketball and show not only that it can be done but also its applications in evaluating the abilities and performance of players and teams.

This research provides methods and metrics for analyzing, visualizing, and describing scoring in Philippine basketball—particularly the UAAP—in a manner that explicitly accounts for the spatial nature of shooting. It utilizes a matrix decomposition algorithm known as Non-negative Matrix Factorization (NMF) to divide the court into areas or shooting zones where field goals are commonly taken and provide information about how frequently players shoot from these shooting zones. Using this information, players are grouped and compared according to their shooting habits. It also presents spatial metrics to evaluate shooting and scoring that build on the SSE and POLA metrics introduced by Shortridge (2014). These metrics measure the scoring effectiveness at different locations on the court by comparing the expected number of points scored and the actual points scored at these locations.

Finally, the research applies these methods and metrics in a case study of the UAAP Men's Basketball Tournament (Season 81) and showcases the value of spatial analysis in evaluating players and teams in basketball.

1.5 Research objectives

The general goal of this research is to display and highlight the value of spatial analysis as applied in Philippine basketball. To meet this goal, the specific objectives of this research are:

1. To divide the court into shooting areas stochastically by using the spatial dataset of field goals to find commonly occurring patterns of where field goals are taken;
2. To group similar players based on their shooting tendencies;
3. To generate spatially-aware metrics of shooting and show how they can be applied to analyze players and teams; and
4. To create and share a spatial dataset of field goals that can be used for future research.

1.6 Thesis structure

This thesis is divided into five chapters. Chapter 1 provides a background of how spatial analysis has been used to study basketball abroad, the current use of analytics in Philippine basketball, and presents the overview and objectives of the study. Chapter 2 gives a review of relevant literature on the use of spatial analysis in Philippine basketball, the spatial characterization of field goals using Non-negative Matrix Factorization, and different spatial metrics of shooting. Chapter 3 describes the scope and delimitation of the study, the dataset used, the methods employed, and the spatial metrics generated. Chapter 4 presents a case study applying the methods and metrics generated by the research on the UAAP MBT (Season 81) and provides a detailed discussion of the results. Chapter 5 gives the summary and conclusions of the study and outlines recommendations for future research. The remaining sections contain the references, appendices, glossary of terms, and other supplementary materials related to the study.

BN
HR

2. Review of Literature

A review of methods for quantifying and characterizing basketball gameplay for both the player and team level is given by Tenner and Franks (2020) but most of the literature pertains to studies using data from the National Basketball Association (NBA). Although it was noted that most of the methods discussed there were relevant across all basketball leagues, looking at the current body of work around the spatial analysis of basketball, it becomes apparent that such studies are rarely done in the Philippine context.

2.1 Spatial analysis of Philippine basketball

The use of spatial analytics in mainstream Philippine basketball is not as common as it is abroad. This can be attributed to the lack of readily available spatial data that is needed for spatial analysis. Basketball leagues in the Philippines either do not have player tracking data or they do not share it publicly.

One study that looked at the entire pipeline of spatial analysis for Philippine basketball—from data extraction, storage, analysis, and presentation—was CourtVisionPH by Pintor and Cataniag (2014). Due to the unavailability of field goal location data, CourtVisionPH relied on extracting shooting locations from broadcast basketball videos by utilizing the concept of transformation between two coordinate reference systems—that is, the coordinates on the video frame where a shot is taken can be mapped to its corresponding location on a model of the basketball court whose dimensions and coordinates are known. The system took advantage of the fact that a regulation basketball court has a standard size and that there are easily-distinguishable markings on the court that could serve as control points for solving the coordinate transformation problem. These points included the corners of the baseline, intersections of court lines, the corners of the paint, and other markings whose positions on the court are known. After extracting the field goal locations, the data was stored in a database and spatial metrics were computed and visualized. Figure 2.1 details the CourtVisionPH system.

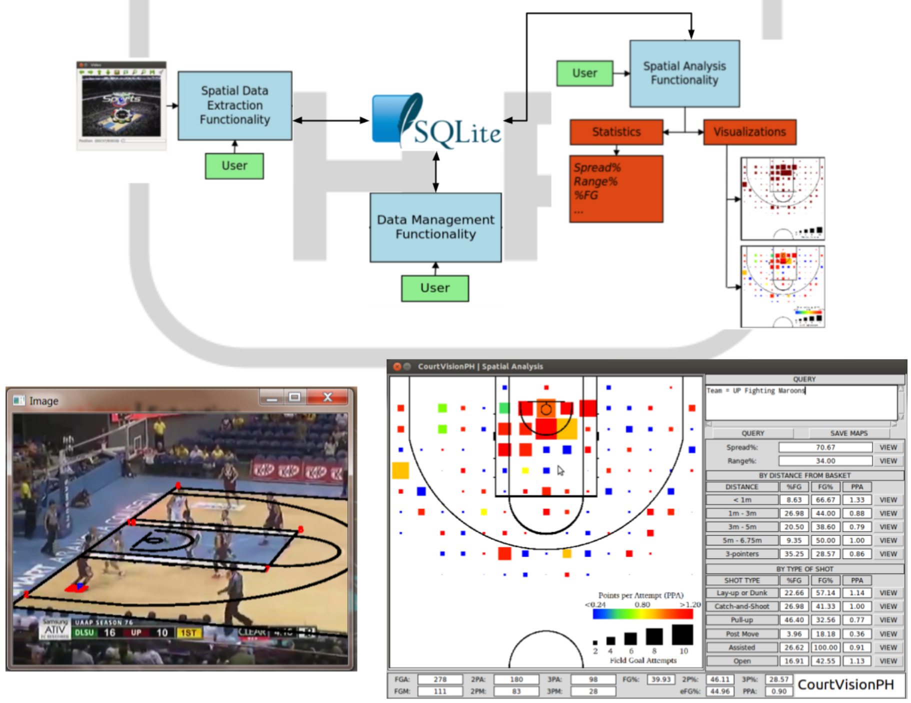


Figure 2.1

The CourtVisionPH System (Pintor and Cataniag, 2014)

Note: The system was divided into three primary modules: a data extraction, data storage, and data analysis/visualization module. Field goal data is extracted semi-automatically from a video source and stored in a database. The data analysis and visualization module uses the stored data to perform queries, compute metrics, and generate maps.

Although crude, CourtVisionPH demonstrated that the spatial analysis of shooting is possible in the Philippine context. The system has received some updates since its release notably the use of a spatial database and existing geospatial applications such as QGIS for the analysis and presentation instead of the original standalone analysis/visualization application. At the time when CourtVisionPH was created, the Spread and Range metrics first introduced by Goldsberry (2012) and used by the system were still considerably novel but since then, the number of researches and studies that applied spatial analysis to shooting have grown and newer methods and metrics have been introduced. Some of these methods and metrics are discussed in the next section.

2.2 Spatial characterization of field goals

Dividing the court into shooting zones or hot zones is a common technique used to analyze shooting. Using this approach, analysts can compute and compare how players and teams shoot at predefined partitions of the court. Most of the time however, the way the court divisions are defined is arbitrary and dependent on the person doing the analysis. Even though common court divisions include general areas such as the paint, the mid-range area, the three-point area, the restricted area, the non-restricted area in the paint, the left and right baseline, the left and right block to elbow, the key, the top of the key three-pointer, the left and right wing three-pointers, the corner three-pointers, the way these areas are delineated on the court still vary. Figure 2.2 shows how these general areas can be divided differently on the court.

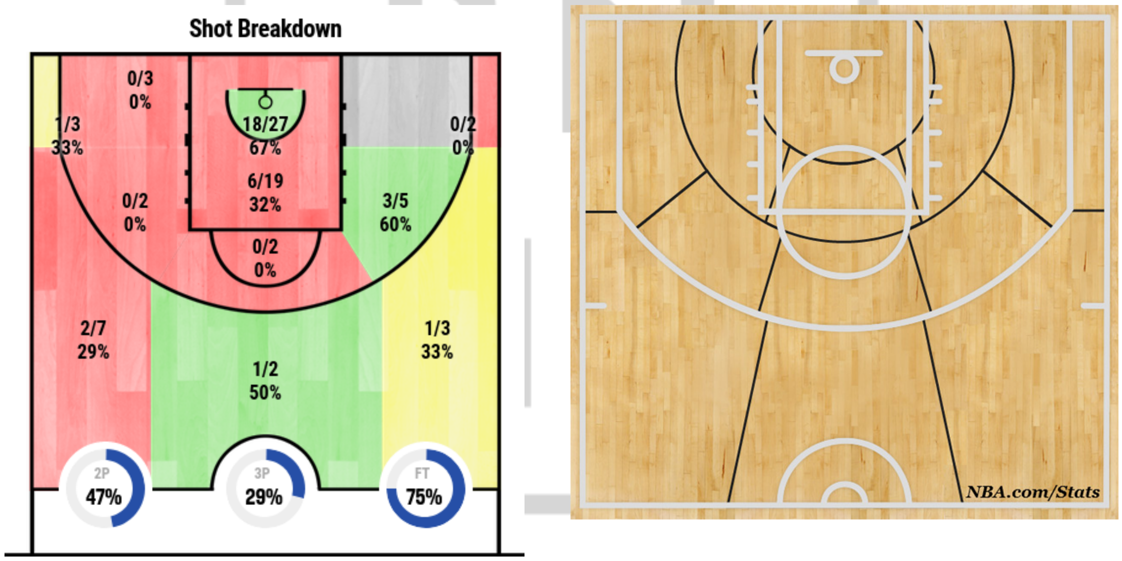


Figure 2.2

Shooting zone divisions (Top: FIBA LiveStats; Bottom: NBA)

Note: The FIBA LiveStats and NBA divisions partition the court into the same general areas but do so differently.

Although there is value in having predefined court divisions, we can't definitely say that these divisions characterize the shooting patterns present in a field goal dataset. What these divisions represent is an idealized version of where and how we think field goals will be grouped together before the fact but once players start taking shots, the actual patterns of these field goals and areas where they are commonly taken might partition the court differently. In order to characterize where players take shots and how frequently they take them, a data-driven approach that divides the court based on patterns in the field goal dataset is needed.

Finding patterns in spatial datasets can be done using clustering algorithms like K-means clustering or matrix decomposition algorithms like Principal Component Analysis (PCA) or Singular Value Decomposition (SVD). In recent years, the process of Non-negative Matrix Factorization (NMF) has gained traction in basketball analytics for its ability to divide the court into parts and provide an approximation of players' shooting habits in those parts.

2.2.1 Finding spatial basis vectors using Non-negative Matrix Factorization

Non-negative Matrix Factorization or NMF is a matrix decomposition algorithm that assumes some non-negative matrix V can be approximated by the product of two lower-rank non-negative matrices W and B where as seen in (2.1).

$$V = WB \tag{2.1}$$

The matrix $V \in R_{+}^{N \times X}$ is composed of N data points each of length X, the basis loadings or weight matrix $W \in R_{+}^{N \times K}$ consists of N non-negative weight vectors, and the basis matrix $B \in R_{+}^{K \times X}$ contains K basis vectors. Each data point in V can be reconstructed using a combination of W and B. (Miller, 2014)

A distinct characteristic of NMF is that it constrains all of its component matrices to be non-negative. Because of this, the resulting basis vectors B tend to be disjoint and exhibit a "parts-based" decomposition that corresponds to frequently occurring patterns in the sample. This has the advantage of avoiding the cancellation phenomenon exhibited by other non-constrained matrix factorization methods like PCA. This restrictive property of NMF also results in sparser and more interpretable basis vectors. (Lee & Seung, 1999)

In the case of basketball, using NMF to decompose a field goal dataset makes sense due to the following reasons: first, the field goal matrix—or the collection of field goals at different locations on the court—is always non-negative because it is impossible to attempt a negative number of shots; second, the outputs of NMF correspond intuitively to basketball concepts. The spatial basis vectors in B can be interpreted as disjoint sub-intensities or parts that represent shot-types or shooting zones on the court. Meanwhile, the player weights in W can be used to summarize the spatial shooting habits of individual players inside the spatial basis vectors in B (Miller et al., 2014; Franks et al., 2015).

The general steps in applying NMF for deconstructing field goals as presented in Miller et al. (2014), Franks et al. (2015) and Jiao et al. (2020) are:

1. Discretize the court using a regular tessellation into X shooting cells.
2. Construct a count matrix C where C_{nx} = the number of field goals by player k in cell x.
3. Fit an intensity surface $\lambda_n = (\lambda_{n1}, \dots, \lambda_{nX})^T$ for each player k over the discretized court.
4. Construct the data matrix (field goal matrix) $V = (\bar{\lambda}_i, \dots, \bar{\lambda}_N)^T$ where $\bar{\lambda}_n$ has been normalized such that it has a unit volume.
5. Solve the optimization problem $V = WB$ where W and B are lower-rank matrices and all matrices are non-negative.

To fit the intensity surface of a player's field goals, Miller et al. (2014) and Franks et al. (2015) modeled them as a Log Gaussian Cox Process (LGCP) after discretizing the court into 1 square foot tiles to gain computational tractability in fitting the LGCP surfaces. Meanwhile, Jiao et al. (2020) used a kernel estimation "which is easier to compute and more accurate in the sense of intensity fitting accuracy".

Figure 2.3 (Miller et al., 2014) shows a comparison of the resulting basis vectors generated by NMF with LGCP-fitted intensity surfaces using the (a) Kullback-Leibler (KL) and (b) Frobenius loss functions, (c) NMF with a discrete dataset, and (d) PCA with the LGCP-fitted intensity surfaces.

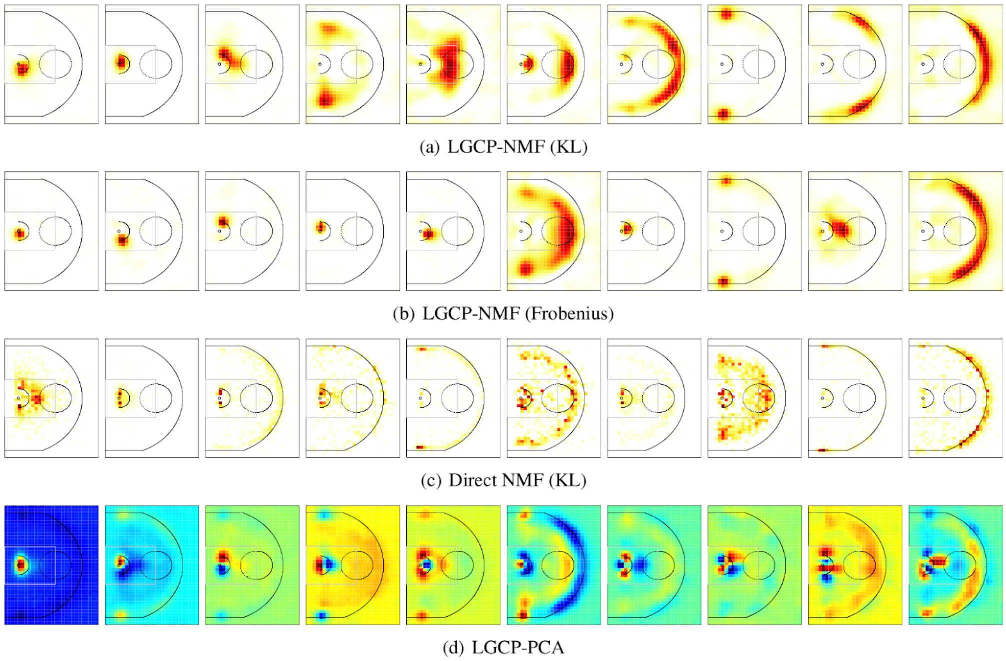


Figure 2.3

Visual comparison of basis resulting from various dimensionality reduction approaches.

Note: From "Factorized Point Process Intensities: A Spatial Analysis of Professional Basketball," by A. Miller et al., 2014, *Proceedings of The 31st International Conference on Machine Learning (ICML14)*, Beijing, China, June 22-24, 2014. Journal of Machine Learning Research: W&CP 32: 235-243.

According to Miller et al. (2014), the KL-based NMF resulted in a "more spatially diverse basis" compared to the Frobenius-based one which focused on "high-intensity areas near the basket" at the expense of other areas on the court. This can be attributed to the difference between the KL loss function—which includes a log ratio term that disallows large ratios between the original and reconstructed matrices—and the Frobenius loss function—which does not include a log ratio term and thus only disallows large differences. Meanwhile, the PCA basis vectors were uninterpretable as parts of the court due to the bases being unconstrained real numbers. The corner three-point feature that was salient in the LGCP-NMF decompositions appeared in several PCA vectors with positive and negative values that exhibited the cancelation phenomenon with PCA that NMF avoids. Subsequent studies using NMF by Franks et al. (2015), Sandholtz et al. (2019), and Jiao et al. (2020) also used the KL loss function.

Miller et al. (2014) also found that the LGCP-NMF method discovered a "shots-based decomposition" of NBA players where the resulting basis vectors B corresponded to "visually interpretable shot types"—one basis corresponded to corner three-point shots, another to wing three-point shots, and yet another to top of the key three point shots, etc.—while the player specific basis weights in W provided a concise characterization of player's offensive habits. The weight w_{nk} can be interpreted as the "amount player k takes shot type k".

Miller et al. (2014) also showed that, after a certain K, the low-rank NMF reconstructions had better predictive performance than independent LGCPs for player data with 10% of the shots held out. Figure 2.4 (Miller et al., 2014) shows the predictive likelihood for independent LGCP and LGCP-NMF at varying K.

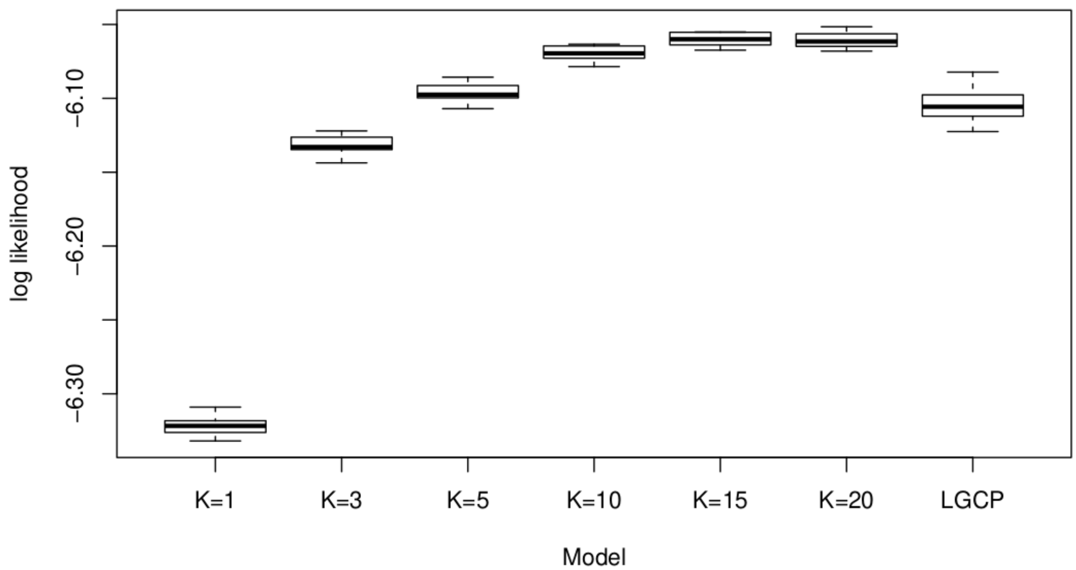


Figure 2.4

Predictive Likelihood (10-fold cv) of LGCP and LGCP-NMF at varying K

Note: From "Factorized Point Process Intensities: A Spatial Analysis of Professional Basketball," by A. Miller et al., 2014, *Proceedings of The 31st International Conference on Machine Learning (ICML14),* Beijing, China, June 22-24, 2014. Journal of Machine Learning Research: W&CP 32: 235-243.

The *K* values with better predictive performance can be used as the *K* values for the NMF decomposition. Miller et al. (2014) and Jiao et al. (2014) used *K=10* while Franks et al. (2015) used *K=6*.

Aside from using a lower number of bases, Franks et al. (2015) also discarded a residual basis from the six computed by NMF since, unlike PCA, NMF is not mean-centered and a residual basis appears regardless of the value of *K*. This residual basis captures the positive intensities outside the support of the relevant bases and is therefore not used in the analysis. Figure 2.5 (Franks et al., 2015) shows the spatial bases identified using LGCP-NMF with *K*=6.

Similar to Miller et al. (2014), Franks et al. (2015) also arrived at a "shot-based decomposition of NBA players" where the bases corresponded to shots in the restricted area (Basis 1), shots from the rest of the paint (Basis 2), mid-range shots (Basis 3), corner three-point shots (Basis 4), and center three-point shots (Basis 5).

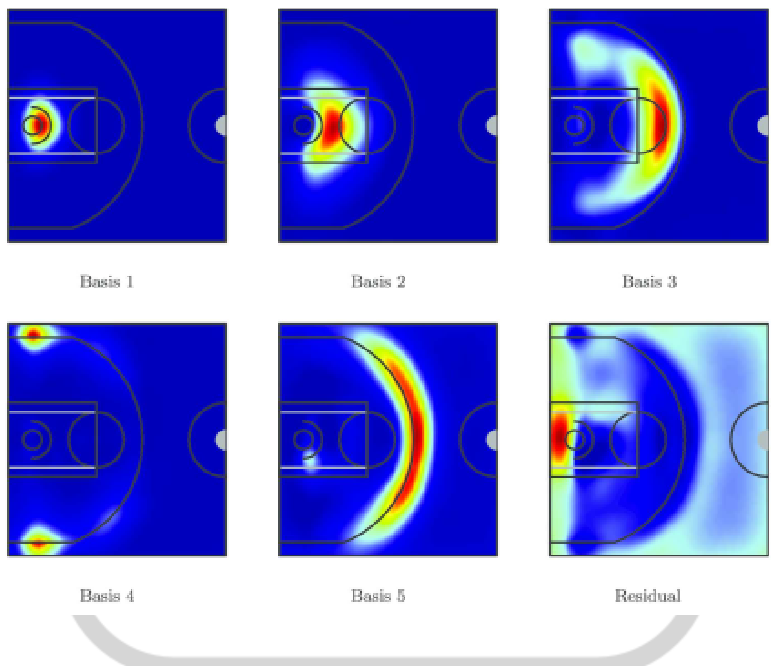


Figure 2.4

The basis vectors and the residual basis using LGCP-NMF with KL loss function and K=6

Note: From "Characterizing the Spatial Structure of Defensive Skill in Professional Basketball," by A. Franks et al., 2015, *The Annals of Applied Statistics*, 2015, Vol. 9, No. 1, 94–121.

The spatial basis and weights matrices generated by NMF have applications beyond simply providing court divisions and an approximation of player shooting tendencies. They have been used in conjunction with other models to "estimate the probability of a made shot for each point in the offensive half court for each individual player" (Miller et al., 2014), "to characterize how players affect both shooting frequency and efficiency of the player they are guarding" (Franks et al., 2015), to study the "optimal way to allocate shots within a lineup" and "measure how efficiently a lineup adheres to optimal allocative efficiency" (Sandholtz et al., 2019), and the "association between shooting frequency and accuracy" (Jiao et al., 2020).

2.2.1.1 Finding the optimal number of basis or factorization rank

The number of basis vectors or factorization rank (K) is a crucial parameter in NMF. The optimal number of basis vectors will allow the dataset to be decomposed into latent features without overfitting the model. Miller (2014) and Franks (2015) both used a 10-fold cross validation between the predictive performance of LGCP and LGCP-NMF over different values of k to select the optimal number of basis vectors.

Several other methods for determining the optimal factorization rank of NMF have been proposed especially when working with omic data. Brunet (2004) looked at the cophenetic coefficients and proposed to take the first value of K for which the coefficient starts decreasing. Hutchins (2008) showed that the first value where the residual sum of squares (rss) curve showed an inflection point provided a robust estimate of the proper number of vectors. Meanwhile, Frigyesi (2008) estimated the appropriate factorization rank by comparing the residual error of NMF reconstruction of data to that of NMF reconstruction of permuted or random data and suggested that K be considered as the smallest value where the decrease in RSS computed from the data is lower than the decrease in RSS computed from random data.

2.2.2 Grouping similar players based on shooting characteristics

Although individual players shoot differently, players with similar roles tend to have similar shooting characteristics. This is important when modelling the shooting characteristics of players in areas where they took a low volume of shots. Using the W and B outputs of NMF, it is possible to normalize the shooting characteristics of players at different areas on the court based on the characteristics of other similar players.

Franks et al. (2015) initially used SVD and graphed the first two principal components of the player weights matrix W to determine groupings but found that "the players do not cluster; specifically, there appears to be far more variability in offender types" and that "while players tend to be more similar to players with the same listed position, on the whole, position is not a good predictor of an offender's shooting characteristics." Franks et al. (2015) then applied a conditionally autoregressive (CAR) model on the player weights or basis loadings and identified "the 10 nearest neighbors in the space of shot selection weights" and connected "two players if, for either player in the pair, their partner is one of their ten closest neighbors."

Sandholtz et al. (2019) also applied a CAR prior on the player weights or basis loadings in order to "shrink the FG% estimates of players with similar shooting characteristics toward each other". This helped regularize the values by "borrowing strength from the player's neighbors in the estimation" especially for cells where a player attempted a low number of field goals. To find similar players, the Euclidean distance between the player weights were computed and the 5 players with the nearest distance from player k were determined as his neighbors. Symmetry was enforced in the nearest-neighbors relationship by assuming that "if player j is a neighbor of player l then player l is also a neighbor of player j".

2.3 Spatial metrics of shooting

Several metrics have been proposed that measure shooting effectiveness and also account for the effects of location. These "spatially-aware" metrics range from simple assessments of how players score at different areas on the court, comparisons between expected and actual points scored at different court locations, finding the optimal allocation of field goals for a team based on the shooting characteristics of a lineup, and even hierarchical and high-resolution data models of shooting. Some examples of these metrics are discussed in the next sections.

2.3.1 Spread and Range

Spread and Range, introduced by Goldsberry (2012) in the paper *CourtVision: CourtVision: New Visual and Spatial Analytics for the NBA*, are some of the earliest and most influential of these spatial shooting metrics. *CourtVision* tried to answer the question of who the best shooter was in the NBA at that time and argued that conventional metrics and evaluative approaches fail to provide a simple answer to this question—in essence, these conventional metrics failed to account for the spatial aspect of shooting—so it proposed a new way to quantify the shooting range of NBA players and measure shooting abilities using "spatially-aware" metrics. It found that from 2006-2011, more than 98% of the field goal attempts in the NBA occurred within a 1,284 ft^2 area in between the baseline and a relatively thin buffer around the 3-point arc which was designated as the "scoring area." This scoring area was divided into unique 1ft^2 cells and the Spread and Range metrics were computed using the cells.

Spread is simply a count of the unique shooting cells in which a player has attempted at least one field goal. It summarizes the diversity of a player's shooting attempts or the overall size of a player's shooting territory given by (2.2).

$$Spread = \sum_{i \in SA} FGA_i \tag{2.2}$$

where:

FGA_i *= 1 if at least one field goal was taken in cell i, 0 otherwise*

SA = the scoring area

Dividing Spread by 1284 resulted in Spread% which is the percentage of the scoring area where a player attempted at least 1 shot.

Range accounts for spatial influences on shooting effectiveness by counting the number of shooting cells in which a player averages more than 1 PPA as given by (2.3). PPA was chosen over FG% because it inherently accounts for the differences between two-point and three-point field goal attempts.

$$Range = \sum_{i \in SA} PPA_i \tag{2.3}$$

where:

PPA_i *= 1 if PPA >1 in cell i, 0 otherwise*

SA = the scoring area

Dividing Range by 1284 resulted in Range% which is the percentage of the scoring area where a player scored at least 1 PPA.

CourtVision's ideas were novel for its time and its approach served as the blueprint for succeeding studies that measured shooting spatially. However, putting Spread and Range in the current era of basketball shows some limitations in these metrics. First, the scoring area defined in the original study has, without a doubt, increased from the original 1284 ft^2. Three-pointers are more common now and have even become the staple of some offenses. In the NBA, there are some games where teams attempt more three-point field goals than two-point field goals. Players are shooting more three-pointers and shooting them from farther away than they were 10 years ago and this will have an effect on the definition of the "scoring area". Second, although the choice of PPA over FG% was a great move, estimating a player's shooting performance at different areas on the court by comparing it to a single value for all locations—in the case of Range, 1 PPA—runs contrary to the fact that the scoring geography on the basketball court is not flat. If we take all the field goals for a season or league and map them on the court, we will find that there are areas where, on average, players shoot and score more efficiently and there are other areas where, on average, they score less. The scoring geography of the basketball court is more similar to hills and valleys than it is to a plain.

BN HR

2.3.2 Spatial Shooting Effectiveness and Points Over League Average

Instead of comparing a player's scoring ability at different locations on the court to a single value, Shortridge et al. (2014) compared it to the expected points that a player should score at a location based on how other players in the league were scoring, on average, from that location. Similar to Goldsberry (2012), Shortridge et al. (2014) also divided the court into 1 ft^2 cells that served as the basis for computing the Spatial Shooting Effectiveness (SSE) and Points Over League Average (POLA) metrics albeit the division was for the entire half court and not just for the 1284 ft^2 "scoring area" described in the earlier study. SSE and POLA evaluate a player's scoring effectiveness by considering spatially the relative difficulty of their field goal attempts based on the locations they take them and the scoring effectiveness of other players in these same locations.

Another innovation introduced by Shortridge et al. (2014) was the use of an Empirical Bayes (EB) estimate for the FG rate (or FG%) at each cell instead of the raw rate. This was done to account for the uncertainty in the positional accuracy of the field goal dataset as well as the uncertainty in the estimated FG% for cells with a small number of attempts and resulted in an EB-estimated FG% map that is smoother and less noisy.

The Empirical Bayes approach computes an estimate of the true rate at each location as a weighted combination of a prior probability distribution function and the local raw rate. In Shortridge et al. (2014), a reasonable assumption was made that the prior distribution would include the locations that are about the same distance from the basket and those relatively close to the location being estimated. Mathematically, the prior distribution for cell i positioned d feet from the basket includes those cells satisfying both of the following conditions:

- equidistant from the basket: within 1.2 feet of d
- close: < 5 feet from i

Compared to other kernel smoothing methods, the Empirical Bayes approach "explicitly accounts for distance from the basket" and excludes locations that are substantially nearer or farther from the basket in the estimation of the rate. It also allows for the modification of the prior or local neighborhood to "ensure robust rate estimation" and "provides more adjustment to raw rates in locations where fewer shorts are attempted while maintaining local yet meaningful departures from the neighborhood rate in locations supported by a high local number of shots" (Shortridge et al., 2014).

Figure 2.5 shows a comparison between the raw FG% and the EB-estimated FG% in Shortridge et al. (2014).

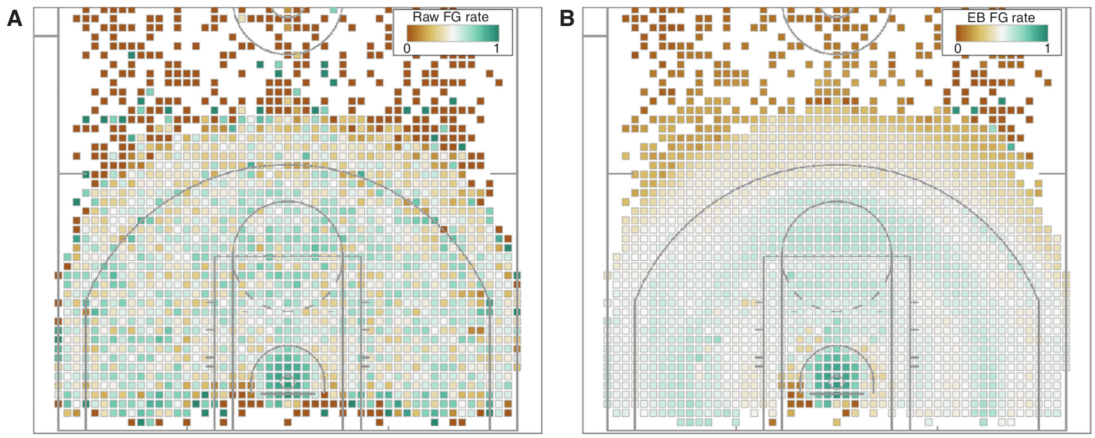


Figure 2.5

Mapped shooting patterns from the 2011 to 2012 NBA regular season: (A) raw field goal rate surface; (B) empirical Bayesian smoothed FG rate estimate, based on all shots taken in that season. Color/shade scales represent the same rates for both surfaces.

Note: From "Creating space to shoot: quantifying spatial relative field goal efficiency in basketball," by A. Shortridge et al., 2014, Journal of Quantitative Analysis in Sports, 10(3), 303-313.

The EB-estimated FG% ($\hat{\theta}$) was used to compute the number of points a player is expected to score at each cell (ELPTS) based on the number of his field goal attempts at each cell given by Shortridge et al. (2014) in (2.4)

$$ELPTS_{ik} = \hat{\theta}_i \times FGA_{ik} \times PBB_i \qquad (2.4)$$

where:

FGA_{ik} *= number of shots taken by player k at cell i*

PBB_i *= number of points a FG at cell i is worth*

These ELPTS values were then used to compute for SSE and POLA.

SSE is essentially a measure of how much a player is scoring per shot versus how much he is expected to score per shot given the spatial distribution of his field goals. It indicates the difference between a player's expected and actual points per attempt. Positive values indicate that the player is scoring more effectively than expected while negative values indicate that the player is scoring less effectively. The units of SSE are in points per shot.

To compute the SSE of player k given by (2.6), the Expected Points Per Shot (*EPPS*) given by (2.5) is first computed. *EPPS* is a summary measure characterizing the average difficulty of the spatial distribution of player k's field goal and is simply the sum of the Expected Local Points (ELPTS) across all cells N where player k attempted at least 1 FGA divided by the total number of field goals attempted by player k anywhere on the court. A high EPPS suggests that the player takes shots that are, based on league average, easy to convert while a low EPPS suggests that the player takes shots at areas that are, based on league average, difficult to score from. (Shortridge et al., 2014).

$$EPPS_k = \frac{\sum_{i=1}^{N} ELPTS_{ik}}{FGA_k} \tag{2.5}$$

where:

N = all the cells where player k has at least 1 FGA

FGA_k = the number of field goal attempts by player K

$$SSE_k = PPS_k - EPPS_k \tag{2.6}$$

where:

PPS_k = the actual points per shot of player k (PPA)

$EPPS_k$ = the estimated points per shot of player k based on the spatial distribution of his shots.

Meanwhile, POLA is a measure of the total number of points that a player scored compared to the number of points he is expected to score given the spatial distribution of his field goals. Similar to SSE, positive values indicate that the player is scoring more than expected while negative values indicate that the player is scoring less than expected. Its units are in points. In Shortridge et al. (2014), the POLA of player *k* is given by (2.7).

$$POLA_k = PTS_k - \sum_{i=1}^{N} ELPTS_k \tag{2.7}$$

where:

N = the cells where a player takes at least 1 FGA

PTS_k = the actual points per scored by player k

Both SSE and POLA can be computed globally resulting in a single number that summarizes a player's shooting or scoring effectiveness based on the spatial distribution of his field goal attempts or locally on a per-cell basis which enables it to be mapped and show the spatial distribution of the metric.

The assessment and evaluation of SSE and POLA were done two ways by Shortridge et al. (2014): comparing the actual and observed points, and direct comparison of SSE or POLA values of two players. In the first case, they utilized a weighted, paired t-test to evaluate whether the observed points differ significantly from the expected points. In the second case, they used the same weighted t-test to determine whether a field goal distribution's POLA or SSE values are significantly larger than another's. In the latter case, the t-test cannot be paired since the two field goal distributions cannot be assumed to be identical. The weighted mean (SSE_{ik}) is given by (2.8); the weighted variance (s^2) is provided by (2.9); and the weighted sum of squares (WSS_k) is computed using (2.10).

$$SSE_k = \frac{\sum_{i=1}^{N} FGA_{ik} \times LSSE_{ik}}{\sum_{i=1}^{N} FGA_{ik}} \tag{2.8}$$

where:

N = the cells where a player takes at least 1 FGA

$LSSE_{ik}$ = local SSE value for the cell

FGA_{ik} = number of field goal attempts at the cell

$$s^2 = \frac{WSS_k - (N \times SSE_k^{\ 2})}{(N - 1)} \tag{2.9}$$

where:

N = the cells where a player takes at least 1 FGA

WSS_k = weighted sum of squares

$$SSE_k = \frac{\sum_{i=1}^{N} FGA_{ik} \times LSSE_{ik}^{\ 2}}{\sum_{i=1}^{N} FGA_{ik} / N} \tag{2.10}$$

where:

N = the cells where a player takes at least 1 FGA

$LSSE_{ik}$ = local SSE value for the cell

FGA_{ik} = number of field goal attempts at the cell

2.3.3 Lineups Points Lost

Sandholtz (2019) studied shooting efficiency as an optimal allocation problem by comparing the shooter's FG% to his field goal attempt (FGA) rate in the context of his other four teammates on the court and the spatial distribution of his shots. First, the court was divided into areas by applying NMF. The FG% of players in each of the spatial bases were then modelled. A conditionally autoregressive (CAR) prior on the players loadings or basis weights in W was used to "shrink the FG% estimates of players with similar shooting characteristics toward each other". The players in a five-man lineup were then ranked based on how well they scored from each of the spatial basis vectors computed by NMF. Using this ranking, a value was calculated corresponding to the expected number of points scored by a five-man lineup if they allocated their field goals in an area according to the ranks of the players in that area (i.e. the best shooter takes the most shots and the worst shooter takes the least). This expected value was then subtracted to the actual number of points scored by the lineup to generate the metric Lineup Points Lost (LPL). LPL is defined as the difference in expected points between a lineup's actual distribution of FG attempts, A, and a proposed redistribution, A^*, that has perfect rank correspondence.

Figure 2.6 shows the rank correspondence of players from the starting lineup of the 2016 Cleveland Cavaliers at different areas on the court. A positive rank correspondence means areas of under-utilization while negative values indicate potential over-utilization. A +4 means the best shooter took the fewest shots in the area while a -4 means the worst shooter took the most shots in the area.

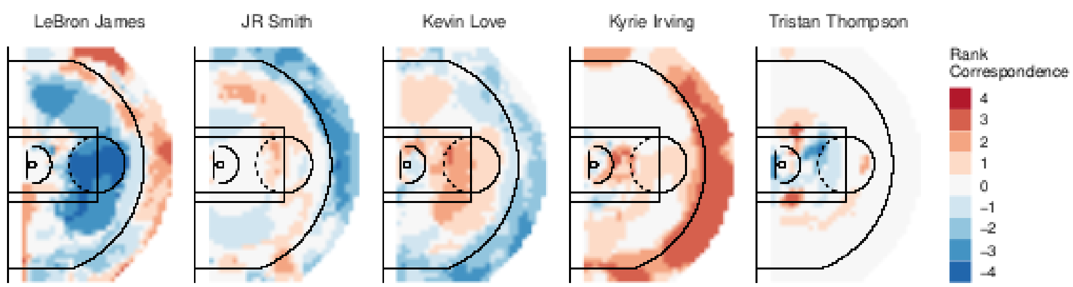


Figure 2.6

Rank correspondence surfaces for the Cleveland Cavaliers' starting lineup.

Note: From "Measuring Spatial Allocative Efficiency in Basketball," by N. Sandholtz et al., 2019, https://arxiv.org/abs/1912.05129v1.

Sandholtz et al. (2019) found that lower LPL is associated with increased offensive production but cautioned that there are game scenarios where minimizing LPL is sub-optimal especially in cases where there are confounding variables such as defensive pressure, expiring shot clock, or clutch situations. Strict adherence to LPL minimizing could also lead to a more predictable offense and thus make it easier to defend.

2.4 Hierarchical and high-resolution data models

Although not used in this study, hierarchical models have also been developed that target increasingly context-specific scenarios in basketball.

Both Franks et al. (2015) and Cervone et al. (2016) utilized hierarchical logistic regression models to estimate the probability of making a shot given attributes such as shooter identity, defender distance, and shot location. Franks et al. (2015) also used a hierarchical multinomial logistic regression to "predict who will take and where a shot will be taken given defensive matchup information" while Cervone et al. (2016) used the Expected Possession Value (EPV) framework and introduced the concept of shot satisfaction. Shot satisfaction is computed using several contextual information such as shooter identity and all player locations and abilities to indicate, per shot, how satisfied a player was with his decision to shoot. Meanwhile, Jiao et al. (2020) modeled shooting using a "Bayesian joint model for the mark and the intensity of marked point processes where the intensity is incorporated in the mark model as a covariate."

High-resolution spatio-temporal data which can include the full three-dimensional trajectories of the ball while being shot have also been used to study shooting. Using ball tracking data, Marty (2018) and Daly-Grafstein & Bornn (2019) were able to show that the optimal entry location for a shot is about 2 inches from the center of the basket at an entry angle of about 45°. Daly-Grafstein & Bornn (2019) utilized a technique known as Rao-Blackwellization (RB) to generate lower error estimates of shooting skill and demonstrated that RB estimates were better at predicting the three-point percentages of players from limited data compared to empirical make percentages. By integrating the RB approach into a hierarchical model, they were able to achieve further variance reduction. Bornn & Daly-Grafstein (2019) extended the research and studied the effects that defenders had on shot trajectories.

2.5 Summary

The research identified the following gaps in the literature:

1. There are no recent studies that utilize spatial concepts to analyse basketball in the Philippines.
2. There is no readily available basketball spatial data in the Philippines that can be used for analysis.
3. Several spatial analysis techniques and methodologies have been developed and introduced but they commonly use optical tracking data not available in the Philippines.
4. Some of these spatial analysis techniques and methodologies can be modified to utilize simpler inputs that are applicable to countries like the Philippines that do not have the aforementioned optical tracking data.

The research aims to address these gaps by:

1. Generating an open dataset of field goal locations from available shot chart data online.
2. Developing and releasing open source code for:
    a. the generation of a field goal dataset based on online shot chart data, and
    b. the spatial analysis of the generated field goal dataset using modified versions of previously introduced techniques and methodologies.
3. Applying the metrics developed in the research to a case study using Philippine basketball data.

3. Methodology

3.1 Scope and delimitation

The research dataset included the field goals from the elimination round games in the UAAP MBT Season 81 (AY 2018-2019). The data was sourced from FIBA LiveStats shot charts available online at https://www.fibalivestats.com. The information included in the dataset are:

- Location (x, y coordinates on the basketball court)
- Points (2 or 3)
- Made (1 or 0)
- Player Information (Name, Number, Team)
- Team Information (Name)
- Opposing Team Information (Name)
- Date & Venue
- Shot type

Free throws and missed shots due to fouls were not included in the analysis.

Field goal location was the only variable considered for shooting tendency and ability. Other contextual information (covariates) that may affect shooting—player height, time remaining in game, type of shot, whether the player is defended or not—were not considered in this research but may be the subject of further research into hierarchical spatial models of shooting in the UAAP.

3.2 The data

Shot charts for UAAP Season 81 are available online from the FIBA LiveStats website at https://www.fibalivestats.com/u/UAAP/<gameid>/sc.html where <*gameid*> is the identification code for a specific game during the season.

Although the data exists, there were some issues that needed to be addressed before it became usable for the research:

1. The <*gameids*> for the different UAAP games were not known.
2. The default formatting of the data wasn't suitable for the research.
    a. The data was in HTML.
    b. The field goal locations cover the whole court but for the study we needed to map the shots to just a single half court.
3. The positional accuracy of the field goal locations was not provided.

Figure 3.1 shows an example of the shot chart and underlying data.

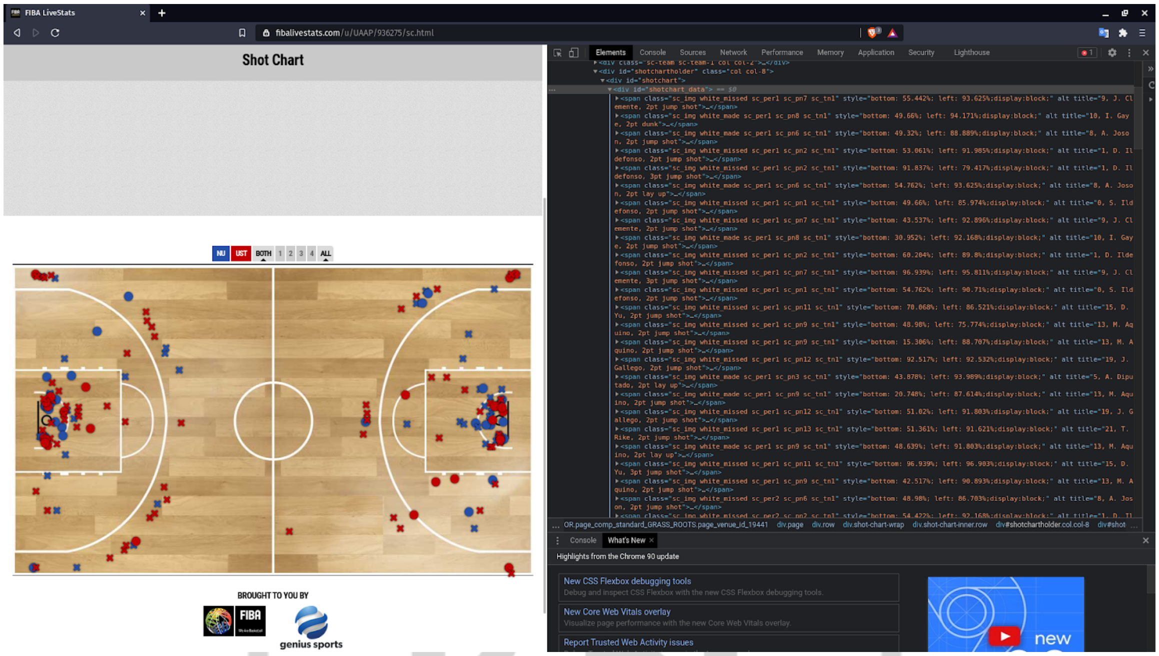


Figure 3.1

Shot chart of NU vs UST UAAP MBT Game (Season 81)

Note: The <gameid> is 936275. The field goal locations for both teams are mapped to the whole court (left side) and the underlying data (right side) is available as HTML text. Source: https://www.fibalivestats.com/u/UAAP/936275/sc.html

To get the data from FIBA LiveStats shot charts and make it useable for the purpose of the study, a web scraper was developed in Python that:

1. Looked for the <gameid> of each UAAP MBT Season 81 game.
2. For each game:
    a. extracted the information from the shot chart HTML
    b. mapped the location of the field goals into a single half court
3. Saved the scraped data into a CSV file.

The resulting file was a field goal dataset that contained the information mentioned in the previous section.

The resulting raw data included:

- 7619 FGA
- 55 games (1 missing game; 1st game of the season between UP and UE)
- 120+ players

The scraper and dataset were both released freely and openly online at:

- FIBA LiveStats Shot Chart Scraper - https://github.com/benhur07b/ms-thesis-spatial-analysis-shooting-philippine-basketball/tree/main/code
- UAAP MBT Season 81 Shot Chart Data - https://github.com/benhur07b/ms-thesis-spatial-analysis-shooting-philippine-basketball/tree/main/data

As mentioned above, UP and UE are both missing one game—the game they played against each other during round 1 of the elimination round could not be found in FIBA LiveStats. There might be some minute differences in the computed statistics if that game was included but they are not expected to be significant enough to change the results of the analysis.

BN HR

Players with less than 28 field goal attempts (or average less than 2 FGA/game if they played all games) were removed in the NMF computations. This left a total of 7105 FGA distributed among 82 unique players. The study also divided the regulation FIBA half-court, a 14m x 15m rectangle, into 50cm x 50cm cells. It's worth noting that the dimensions of the UAAP court which follow FIBA standards are slightly different than those used in the NBA (37ft x 50ft or ~14.23m x 15.4m). It is assumed that the 50cm^2 tile size captures all interesting spatial variations in the data.

Figure 3.2 shows a sample of (a) the raw data mapped onto the court and (b) the corresponding raw field goal grids.

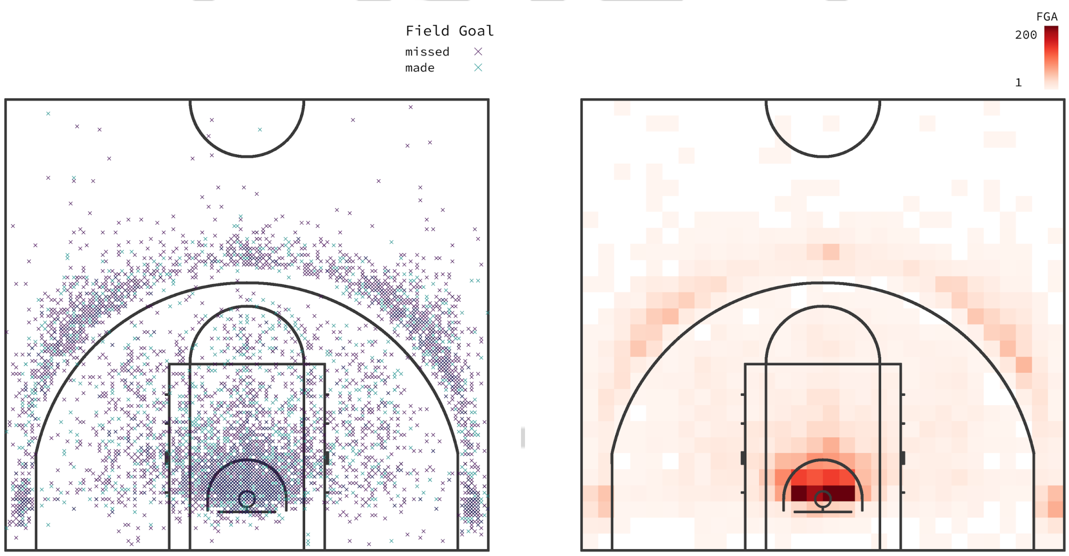


Figure 3.2

Map of the field goal (point) dataset and the field goal (discretized) grid

3.3 Spatial characterization of field goals using NMF

Non-negative Matrix Factorization (NMF) was used to determine the different spatial basis and individual player basis loadings based on the collected field goal dataset.

The formula for NMF is given in (2.1) and the general steps using NMF in basketball is given in Section 2.2.1.

3.3.1 Finding the spatial basis and basis loadings

In the study, the court was discretized into 50cm x 50cm cells resulting in a 28 x 30 court matrix. A subset of the total FG dataset was used for the computation of the spatial basis vectors. This subset included players with at least 28 total field goal attempts—or an average of two attempts if they played all 14 games—in order to minimize the effects of players with low number of field goal attempts skewing the computation.

In order to solve $V = WB$ using NMF, the number of output bases, K, must first be defined. Franks et al. (2015) and Sandholtz et al. (2019) looked at the predictive performance of NMF-LGCP at varying K against independent LGCP using a 10-fold cross validation. This study used a simpler approach which applied the elbow method on the residual sum of squares of different K values to find the optimal number of bases or latent features to describe the dataset as suggested by Frigyesi (2008).

Aside from the optimal value for K, results of using different models and initialization modes for NMF were also compared to determine the best ones to use. This was done by looking at which combination of K, model, and initialization parameters resulted in spatial basis vectors that made the most sense or the one that divided the court into visually-interpretable shot types.

Once the number of bases, the model, and the initialization method were identified, NMF was done using the field goal dataset, where, for each player k:

1. A field goal matrix X_{ki} was generated where x_{ki} = the number of field goal attempts by player k at cell i.
2. An intensity surface λ_k was generated by fitting X_{ki} to the discretized court using Kernel Density Estimation.
3. The data matrix$V = (\bar{\lambda}_1 \dots , \bar{\lambda}_K)^T$ was constructed where $\bar{\lambda}_k$ was normalized such that it had unit volume.
4. The optimization problem of $V = WB$ was solved to find the low-rank matrices W and B.

B was mapped to show the spatial bases or the frequent shooting areas on the court based on the field goal dataset. W was then used to rank and compare how frequent players take shots from these areas.

3.3.2 Grouping similar players based on their shooting habits

Aside from ranking and comparing the shooting frequency of players, *W* was also used to group similar players based on their shooting habits. The individual w_{kb} values in *W* represent how frequent player *k* shoots at spatial basis *b*. For each player *k*, a K-nearest-neighbor algorithm was used to find player *k*'s 5 nearest neighbors. To enforce symmetry of the nearest-neighbor relationship, it was assumed that if player *k* was a neighbor of player *l* then player *l* was also a neighbor of player *k*.

The number of neighbors and the average distance of the neighbors were also plotted to see how common or unique a player's shooting habit is.

3.4 Spatial metrics of shooting

Finding the spatial basis vectors that define a field goal dataset is good especially when we want to identify the common shooting areas where players take field goals but there is also value in dividing the court into a regular tessellation or grid for analysis.

Let's say the court has been divided into cells and we want to represent shooting ability and performance inside a cell. There are two ways we can go about this—the first one is by looking at the success rate or the percentage of field goals that a player makes inside the cell and the second is by looking at the average number of points that a player scores for each field goal attempt inside the cell. In a sense, we're choosing between Field Goal Percentage (FG%) or Effective Field Goal Percentage (eFG%) for the former and Points Per Attempt (PPA) for the latter. In this research, PPA was used for its simplicity and the elegance of its computation. PPA inherently accounts for the difference between two-point field goals and three-point field goals, an advantage it shares with eFG% over FG%; but it also holds an advantage over eFG% in that it uses a more understandable unit of measure—points per attempt instead of percentage.

BN HR

3.4.1 Modeling scoring ability

In the study, the offensive half-court was divided into 50cm x 50cm shooting cells resulting in a 28 x 30 grid and a player's scoring ability at each shooting cell was modeled by considering two main variables.

1. The expected points per attempt at the cell based on the league average.
2. The points per attempt of the player at the cell.

Consider that there exists a spatial surface overlying the offensive half-court that represents the background local values of the number of points scored per attempt at that location on the court. Although these local values cannot be directly observed, they can be inferred from a large sample of field goals. This sample, K, can include all the field goals attempted by all players during a season but it can also be constrained to include just a specific set of field goals depending on the analysis and evaluation being done. For example, K can include just the field goals against specific teams, field goals during a specific time or scenario in the game, field goals by specific players, etc. In this study, K is the set of field goals by all players during UAAP Season 81.

A simple and intuitive approach to estimate these background values is to divide the court into grids and compute the raw points per attempt scored at that cell (3.1). Although simple, this approach is problematic because at cells where there are a small number of attempts or no attempts made then the actual background values are unlikely to be well-estimated.

An Empirical Bayes estimator similar to that used by Shortridge et al. (2014) was used in this study but instead of computing for the expected field goal rate, this study proposes to compute for the expected Points Per Attempt (PPA) directly using the EB estimator. This approach should provide better values for pixels divided by the three point line and offer a more elegant solution that allows for different kinds of cell sizes regardless of whether they divide the two-point and three-point field goals areas on the court perfectly.

The EB estimator accounted for the uncertainty in the raw PPA for cells with a small number of attempts and incorporated the PPA at nearby and equidistant cells as the prior distribution in order to compute the expected PPA at a cell. It worked under the assumption that players score in a similar manner in an area around a location as well as in other areas that are the same distance from the basket as the location. In a sense, there was also an assumption that shooting ability is symmetric—i.e. a player's shooting ability at locations *d* meters to the right of the basket is similar to locations *d* meters to the left of the basket.

BN HR

Mathematically, the computation of the EB estimate PPA was:

1. For each cell i, the raw PPA was computed using (3.1).

$$PPA_i = \frac{PTS_i}{FGA_i} \tag{3.1}$$

where:

PTS_i = the number of points scored at cell i

FGA_i = the number of field goal attempts at cell i.

2. For each cell i, the Empirical Bayes estimate of the PPA ($\hat{\theta}_i$) given by (3.4) was computed using a prior distribution (j) which included:
   a. a 7x7 grid around cell i (< 1.5m from i)
   b. cells that are equidistant from the basket

The prior mean $\hat{\gamma}_i$ for the neighborhood cells j of cell i was computed using (3.2) similar to Shortridge et al. (2014).

$$\hat{\gamma}_i = \frac{\sum PTS_j}{\sum FGA_j} \tag{3.2}$$

where:

PTS_j = points scored at cell j

FGA_j = field goal attempts at cell j

The prior variance $\hat{\varphi}_i$ for the neighborhood cells j of cell i was computed using (3.3) similar to Shortridge et al. (2014).

$$\hat{\varphi}_i = \frac{\sum n_j (p_j - \hat{\gamma}_j)^2}{\sum n_j} - \frac{\hat{\gamma}_i}{\bar{n}_j} \tag{3.3}$$

where:

n_j = *number of shots taken at cell j*

p_j = *the raw PPA observed at cell j*

$\bar{n}_j$ = *the sample mean of the shots taken within all neighborhood cells j*

The weighting factor or shrinking factor $\hat{W}_i$ which is used to shrink the effects of the prior mean in the EB estimation was computed using (3.4) similar to Shortridge et al. (2014).

$$\hat{W}_i = \frac{\hat{\varphi}_i}{\hat{\varphi}_i + \hat{\gamma}_i / n_i} \tag{3.4}$$

where:

$\hat{\gamma}_i$ = *the prior mean*

$\hat{\varphi}_i$ = *the prior variance*

n_i = *the number of field goal attempts at cell i*

BN HR

BN
HR

Finally, the Empirical Bayes estimate of the PPA at cell i ($\hat{\theta}_{pi}$) was computed using (3.5) similar to Shortridge et al. (2014).

$$\hat{\theta}_{pi} = \hat{W}_i p_i + (1 - \hat{W}_i)\hat{\gamma}_i \tag{3.5}$$

where:

W_i = *the weighting factor*

p_j = *the raw PPA observed at cell i*

$\hat{\gamma}_i$ = *the prior mean (PPA) at cell i*

For a player k, the $ELPTS_{in}$ or the Expected Local Points of player k at cell i was computed using (3.6). Because PPA was used instead of FG%, there was no need to multiply the EB-estimated rate with the number of points a FG at cell i is worth.

$$ELPTS_{ki} = \hat{\theta}_i \times FGA_{ki} \tag{3.6}$$

where:

$\hat{\theta}_i$ = *the EB estimate PPA for cell i*

FGA_{ki} = *the number of field goal attempts of player k at cell i.*

The total expected points ($EPTS_k$) that a player k will score based on his field goal distribution is just the sum of all $ELPTS_{ki}$ across all cells i where player k has at least 1 FGA.

Player k's Local Points Per Attempt ($LPPA_{ki}$) at cell i was calculated using (3.7) similar to Shortridge et al. (2014). *LPPA* is a disaggregated version of (3.2) computed on a per-cell basis.

$$LPPA_{ki} = \frac{PTS_{ki}}{FGA_{ki}} \tag{3.7}$$

where:

PTS_{ki} = *the number of points scored by player k at cell i*

FGA_{ki} = *the number of field goal attempts off player k at cell i.*

3.4.2 Spatial Scoring Effectiveness (SScE)

The metric Spatial Scoring Effectiveness (*SScE*) which is a measure of scoring effectiveness based on the spatial distribution of field goals was introduced and defined as the difference between the player's points per attempt (PPA_n) given by (3.1) and the expected points per attempt ($EPPA_n$), given by (3.8) which is the same as (2.6) by Shortridge et al. (2014).

$$EPPA_k = \frac{\sum_{i=1}^{N} ELPTS_{ki}}{FGA_k} \tag{3.8}$$

where:

N = *all the cells where player k had at least 1 FGA*

FGA_k = *the total number of field goal attempts by k anywhere on the court*

Similar to the *EPPS* metric of Shortridge et al. (2014), *EPPA* is a summary measure characterizing the average difficulty of the spatial distribution of player *k*'s field goals. A high EPPA indicates a player takes shots at easy spots on the court while a low EPPA might indicate that a player takes shots at areas that are, on average, more difficult to score from.

The global SScE for player k was computed using (3.9).

$$SScE_k = PPA_k - EPPA_k \tag{3.9}$$

This global SScE value is a metric for showing how much more or less a player is scoring per field goal attempt based on the spatial distribution of his field goals. Positive values indicate players are scoring more than expected while negative values indicate the opposite.

The local SScE ($LSScE_{ki}$) for player k at cell i was computed using (3.10). This is simply the difference between the *LPPA* of player k and the EB-estimate PPA at cell i.

$$LSScE_{ki} = LPPA_{ki} - \hat{\theta}_{pi} \tag{3.10}$$

Because the *LSScE* is computed on a per-cell basis, it was mapped and used to show the spatial distribution of a player's *SScE*—i.e. at what areas on the court is a player scoring more or less than expected.

3.4.3 Points Relative to League Average (PRLA)

The metric Points Relative to League Average (PRLA) was introduced and defined as the difference between the points scored (*PTS*) by a player k and his expected points scored based on the spatial distribution of field goals.

The global *PRLA* for player k was computed using (3.11).

$$PRLA_{ki} = PTS_{k} - \sum_{i=1}^{N} ELPTS_{ki} \tag{3.11}$$

where:

PTS_k = *the number of points scored by player k*

N = *all the cells where player k has at least 1 FGA*

$ELPTS_{ki}$ = *the expected local points of player k at cell i*

If the player *SScE* and *FGA* of player k are known, then the global *PRLA* can be computed using (3.12).

$$PRLA_{k} = SScE_{k} \times FGA_{k} \tag{3.12}$$

This global *PRLA* value indicates the number of points a player is scoring above or below what's expected of him based on the spatial distribution of his field goals.

The local *PRLA* (*LPRLA*) for player k at cell i was computed using (3.12). This was simply the difference between the points scored by player k at cell i and the expected points scored by player k at cell i.

BN HR

$$LPRLA_{ki} = PTS_{ki} - ELPTS_{ki} \tag{3.13}$$

where:

PTS_{ki} = points scored by player k at cell i

$ELPTS_{ki}$ = expected points scored by player k at cell i

If the $LSSE_{ki}$ and FGA_{ki} values are known for player k at each cell i, the local *PRLA* can also be computed using (3.14).

$$LPRLA_{ki} = SScE_{ki} \times FGA_{ki} \tag{3.12}$$

These *LPRLA* values were then mapped to show the spatial distribution of a player's PRLA.

3.4.4 Player analysis

Aside from the global and local values of SScE and PRLA, player Spread and Spread% were also calculated as provided in (2.2). The scoring area is defined as the shooting cells with at least 1 FGA. Instead of Range and Range%, new metrics were introduced called Effective Range (ERNG), Net Effective Range (NERNG), Player Effective Range% (PERNG), and Total Effective Range% (TERNG).

Effective Range is the number of shooting cells where a player has a positive SScE and is computed using (3.13). This provides a summary value of how large an area on the court a player scores effectively or more than expected—the larger the Effective Range, the more areas on the court a player scores effectively. Net Effective Range is the Effective Range subtracted by the number of shooting cells where a player has negative SScE—i.e. shooting cells where a player scored less than expected—and is computed using (3.14). A positive Net Effective Range means that there are more shooting cells where a player is scoring better than expected than shooting cells where he is scoring worse than expected. A negative Net Effective Range means the opposite: that a player is scoring worse than expected in more shooting cells than he is scoring better than expected.

It's good to look at Effective Range in conjunction with Net Effective Range. A player with high Effective Range but negative Net Effective Range means that even though he is effective at a lot of areas on the court, he also takes shots at a lot of areas where he isn't effective. This is indicative of a player who isn't shy at taking the shot anywhere on the court. Meanwhile, a player with a low Effective Range but a high Net Effective Range means that he is only taking shots from a few areas on the court but he is effective at a lot of them. This is indicative of a specialist such as a player who specializes in corner 3-pointers or points at the rim—e.g. dominant big men, a player who only scores on putbacks.

$$Effective\ Range_k = \sum_{i \in N} A_i \tag{3.13}$$

where:

$A_i = 1$ *if* $SScE_i > 0$ *in cell i, 0 otherwise*

N = all shooting cells where player k has at least 1 FGA

$$Net\ Effective\ Range = Effective\ Range - \sum_{i \in N} B_i \tag{3.14}$$

where:

$B_i = 1$ *if* $SScE <= 0$ *in cell i, 0 otherwise*

N = all shooting cells where player k has at least 1 FGA

Player Effective Range % and Total Effective Range % are essentially a proportion of the Effective Range relative to a defined set of shooting cells or area on the court. For Player Effective Range %, this set of shooting cells are those where he attempted at least 1 FGA. Thus, Player Effective Range % is equal to a player's Effective Range divided by his Spread as shown in (3.15). It indicates how much of his individual scoring area he is effective from. Player Effective Range % is conceptually similar to Net Effective Range. High values mean that a player is effective at a lot of the areas where he takes a field goal attempt but can also be indicative of a specialist—a player who only takes shots at specific locations on the court. Meanwhile, Total Effective Range % is equal to the Effective Range divided by

BN HR

the total number of cells in the scoring area and is computed using (3.16). It indicates how much of the court a player scores effectively from. Similar to Effective Range and Net Effective Range, Player Effective Range % and Total Effective Range % should also be taken in conjunction with each other. For example, between two players with the same or similar Player Effective Range %, the player with the larger Total Effective Range % is effective at more locations on the court than the other player. Meanwhile, between two players with the same or similar Total Effective Range %, the player with a lower Player Effective Range % is taking more field goals at locations where he isn't effective.

$$Player\ Effective\ Range\%_{\ k} = \frac{Effective\ Range_{\ \ k}}{Spread_k} \tag{3.15}$$

where:

Effective Range$_k$ = *Effective Range of player k*

Spread$_k$ = *Spread of player k*

$$Total\ Effective\ Range\%_{\ k} = \frac{Effective\ Range_{\ \ k}}{N} \tag{3.16}$$

where:

Effective Range$_k$ = *Effective Range of player k*

N = # of shooting cells in the league shooting area

3.4.5 Team analysis, Team opponents' analysis, NetSScE, and NetPRLA

For a team *m*, team-wide SScE and PRLA were introduced and computed by applying the SScE and PRLA formulas to the team's entire field goal dataset. These values indicated the scoring effectiveness of a team at different areas on the court.

The metrics oppSSE and oppPRLA were also introduced and computed by applying the team-wide SScE and PRLA formulas to the field goal dataset of team *m*'s opponents. These values indicated the defensive ability of a team or how well they limited the scoring effectiveness of their opponents at different areas on the court.

Local values of team-wide SScE, PRLA, oppSScE, and oppPRLA were also computed and mapped.

3.4.6 Assessment and evaluation of SScE and PRLA

The study used the same weighted t-test provided by equations (2.8), (2.9), and (2.10) as utilized by Shortridge et al. (2014) to assess and evaluate SScE. Paired weighted t-test was done to determine if there was significant difference between the expected and observed points scored by individual players and teams as well as to compute for the confidence interval of the SScE. Unpaired weighted t-test was used to compare the SScE of two players and teams.

The top and bottom players in terms of metrics introduced in this study values were analysed and compared. The same was done for the Final Four and non-Final Four teams.

Lastly, the correlation between the metrics introduced in this study with other conventional statistics were also computed.

3.5 Case study: the UAAP MBT 2018-2019 (Season 81)

The new metrics introduced in this study were used to analyze the field goals during the UAAP MBT 2018-2019 Season 81.

Data was obtained from scraping the FIBA LiveStats website using Python 3.8 (Python Software Foundation, 2021), JupyterLab 2.1.2 (Project Jupyter, 2021), and the BeautifulSoup (Richardson, 2021) and requests (Reitz, 2021) libraries.

Analysis and visualization was conducted using R 4.1 (R Core Team, 2021) and RStudio (RStudio Team, 2021) with nmf (Gaujoux R. and Seoighe C., 2010) library. Python was also used with numpy (Harris et al., 2020), scipy (Virtaten et al., 2020), scikit-learn (Pedragosa, 2011), pandas (McKinney, 2010), and matplotlib (Hunter, 2007). Other analysis and visualization were done with QGIS 3.20-3.22 (QGIS.org, 2021).

The computer used for processing runs on a 64-bit Pop! OS operating system with 32GB of RAM, 6GB RTX 2060 graphics card, and Intel® Core™ i7-9750H CPU @ 2.60GHz × 12 cores.

The data and code are all available at:

https://github.com/benhur07b/ms-thesis-spatial-analysis-shooting-philippine-basketball

BN HR

4. Results and Discussion

This part of the research will present the results obtained from computing spatial metrics for UAAP Season 81 and discuss their applications in analyzing player and team performance.

4.1 Conventional shooting statistics

Table 4.1 shows the conventional shooting statistics (FG%, 2P%, 3P%, eFG, PPA) of the teams and their opponents during UAAP Season 81. These team-wide conventional shooting statistics provide an overview of the shooting performance of teams in Season 81. For example, let's compare the two teams that met in the Finals that year: the ADMU Blue Eagles and the UP Fighting Maroons. By looking at the PPA and opp_PPA of the two teams, we can say that both have efficient offenses with UP ranking 1st in points scored per attempt (0.994) and ADMU ranking 3rd (0.920) but there was a significant difference in their defense with ADMU allowing the least number of points per attempt (0.814) and UP lagging behind ranking 6th (0.942).

Table 4.1

Conventional shooting statistics of teams for UAAP Season 81.
2P = 2-point field goals made; 2PA = 2-point field goals attempted; 2P% = 2P/2PA
3P = 3-point field goals made; 3PA = 3-point field goals attempted; 3P% = 3P/3PA
FG = total field goals made; FGA = total field goals attempted; FG% = FG/FGA
*eFG = effective field goal percentage = (FG + 0.5*3P)/FGA*
PPA = points per attempt = total points scored / FGA
opp_* = statistics by team opponents

team	2P	2PA	2P%	3P	3PA	3P%	FG	FGA	FG%
ADMU	288	584	49.3	115	417	27.6	403	1001	40.3
ADU	297	679	43.7	88	302	29.1	385	981	39.2
DLSU	300	665	45.1	85	294	28.9	385	959	40.1
FEU	272	590	46.1	115	342	33.6	387	932	41.5
NU	295	671	44.0	79	297	26.6	374	968	38.6
UE	288	638	45.1	81	333	24.3	369	971	38.0
UP	363	697	52.1	88	299	29.4	451	996	45.3
UST	210	527	39.8	121	421	28.7	331	948	34.9

team	opp_ 2P	opp_ 2PA	opp_ 2P%	opp_ 3P	opp_ 3PA	opp_ 3P%	opp_ FG	opp_ FGA	opp_ FG%
ADMU	258	592	43.6	70	300	23.3	328	892	36.8
ADU	271	646	42.0	81	291	27.8	352	937	37.6
DLSU	264	616	42.9	101	383	26.4	365	999	36.5
FEU	279	614	45.4	104	370	28.1	383	984	38.9
NU	310	655	47.3	97	361	26.9	407	1016	40.1
UE	290	610	47.5	125	363	34.4	415	973	42.7
UP	287	620	46.3	108	333	32.4	395	953	41.4
UST	354	698	50.7	86	304	28.3	440	1002	43.9

team	eFG	PPA	opp_ eFG	opp_ PPA	net_ PPA
ADMU	46.0	0.920	40.7	0.814	0.106
ADU	43.7	0.875	41.9	0.838	0.037
DLSU	44.6	0.892	41.6	0.832	0.060
FEU	47.7	0.954	44.2	0.884	0.070
NU	42.7	0.854	44.8	0.897	-0.042
UE	42.2	0.843	49.1	0.982	-0.138
UP	49.7	0.994	47.1	0.942	0.052
UST	41.3	0.826	48.2	0.964	-0.138

Note: Data includes the UP-UE game without a shot chart/spatial data. Check the glossary (Appendix 1) for the definition of the statistics.

Table 4.2 shows the conventional shooting statistics (FG, 2P, 3P, EFG, PPA) of the top three players per team with the most number of field goal attempts during UAAP Season 81. An interesting thing to note in Table 4.2 is the distribution of field goals by the top three players with the most number of attempts per team. ADMU's three players with the most attempts—Thirdy Ravena, Ange Kouame, and Raffy Verano—only accounted for 36.7% of the team's total number of field goals. This is the least among the eight teams. In comparison, UP's top three players with the most field goal attempts—Juan Gomez de Liaño, Bright Akhuetie, and Paul Desiderio—accounted for 57.3% of the team's total field goal attempts which is the 2nd most among all the teams. This indicates that ADMU did not rely that heavily on Ravena, Kouame, and Verano for taking field goals and that scoring opportunities were more distributed among the other players of the team. Meanwhile, UP favored having Juan GDL, Akhuetie, and Desidero take a majority of the teams field goals.

Table 4.2

Conventional shooting statistics of top 3 players with the most field goal attempts per team for UAAP Season 81
2P = 2-point field goals made; 2PA = 2-point field goals attempted; 2P% = 2P/2PA
3P = 3-point field goals made; 3PA = 3-point field goals attempted; 3P% = 3P/3PA
FG = total field goals made; FGA = total field goals attempted; FG% = FG/FGA
%FG = percentage of team's FGA that the player takes
*eFG = effective field goal percentage = (FG + 0.5*3P)/FGA*
PPA = points per attempt = total points scored / FGA

player	team	FG	FGA	FG%	%FG	eFG	PPA
T. Ravena	ADMU	54	139	38.8	13.9	43.5	0.871
A. Kouame	ADMU	80	132	60.6	13.2	61.0	1.220
R. Verano	ADMU	40	96	41.7	9.6	45.8	0.917
J. Ahanmisi	ADU	86	194	44.3	19.8	54.6	1.093
S. Manganti	ADU	71	172	41.3	17.5	45.6	0.913
P. Sarr	ADU	54	146	37.0	14.9	37.0	0.740
A. Melecio	DLSU	78	203	38.4	21.2	46.3	0.926
J. Baltazar	DLSU	71	149	47.7	15.5	48.3	0.966
L. Santillian	DLSU	61	141	43.3	14.7	45.4	0.908
H. Cani	FEU	49	120	40.8	12.9	45.4	0.908
W. Comboy	FEU	42	119	35.3	12.8	42.9	0.857
A. Tolentino	FEU	51	117	43.6	12.6	54.3	1.085

D. Ildefonso	NU	80	195	41.0	20.1	46.9	0.938
J. Clemente	NU	65	165	39.4	17.0	48.8	0.976
S. Ildefonso	NU	32	90	35.6	9.3	36.1	0.722
A. Pasaol	UE	130	312	41.7	32.1	45.7	0.913
P. Manalang	UE	47	147	32.0	15.1	37.1	0.741
J. Varilla	UE	41	120	34.2	12.4	39.6	0.792
Ju. Gomez de Liaño	UP	84	189	44.4	19.0	51.3	1.026
B. Akhuetie	UP	112	188	59.6	18.9	59.6	1.191
P. Desiderio	UP	77	194	39.7	19.5	46.1	0.923
R. Subido	UST	62	197	31.5	20.8	42.1	0.843
M. Lee	UST	59	176	33.5	18.6	46.0	0.920
C. Cansino	UST	51	115	44.3	12.1	50.4	1.009

player	**team**	**2P**	**2PA**	**2P%**	**3P**	**3PA**	**3P%**
T. Ravena	ADMU	41	92	44.6	13	47	27.7
A. Kouame	ADMU	79	121	65.3	1	11	9.1
R. Verano	ADMU	32	72	44.4	8	24	33.3
J. Ahanmisi	ADU	46	99	46.5	40	95	42.1
S. Manganti	ADU	56	106	52.8	15	66	22.7
P. Sarr	ADU	54	144	37.5	0	2	0.0
A. Melecio	DLSU	46	111	41.4	32	92	34.8
J. Baltazar	DLSU	69	129	53.5	2	20	10.0
L. Santillian	DLSU	55	116	47.4	6	25	24.0
H. Cani	FEU	38	79	48.1	11	41	26.8

W. Comboy	FEU	24	61	39.3	18	58	31.0
A. Tolentino	FEU	26	58	44.8	25	59	42.4
D. Ildefonso	NU	57	110	51.8	23	85	27.1
J. Clemente	NU	34	78	43.6	31	87	35.6
S. Ildefonso	NU	31	82	37.8	1	8	12.5
A. Pasaol	UE	105	208	50.5	25	104	24.0
P. Manalang	UE	32	85	37.6	15	62	24.2
J. Varilla	UE	28	62	45.2	13	58	22.4
Ju. Gomez de Liaño	UP	58	111	52.3	26	78	33.3
B. Akhuetie	UP	112	185	60.5	0	3	0.0
P. Desiderio	UP	52	105	49.5	25	89	28.1
R. Subido	UST	20	69	29.0	42	128	32.8
M. Lee	UST	15	38	39.5	44	138	31.9
C. Cansino	UST	37	77	48.1	14	38	36.8

Note: Data includes the UP-UE game without a shot chart/spatial data. Check the glossary (Appendix 1) for the definition of the statistics.

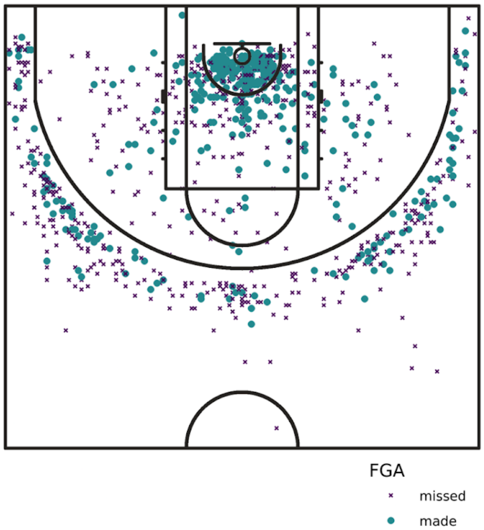
As useful as the information provided by the conventional statistics shown in Tables 4.1 and 4.2 are, they still do nothing to give insight as to how these teams and players performed at specific areas on the court. For that, we go to the spatial shooting statistics.

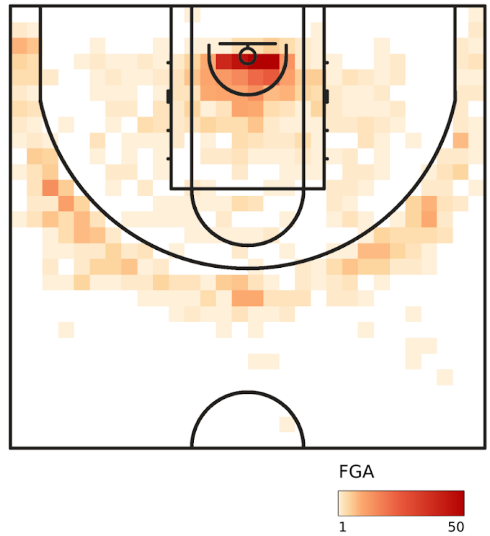
BN HR

4.2 The spatial field goal dataset

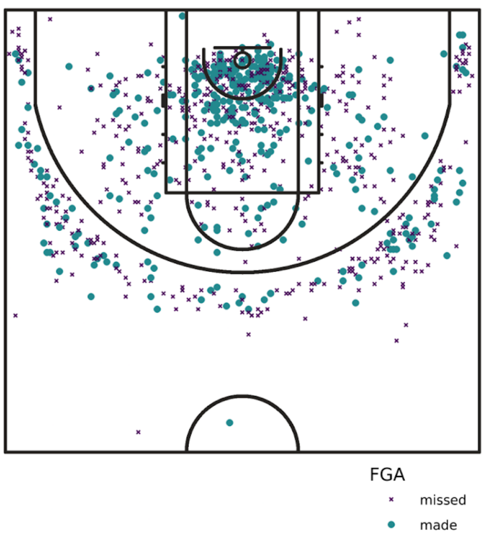
To recap, the raw data extracted from FIBA LiveStats included 7619 FGA distributed among 55 games (1 missing game; 1st game of the season between UP and UE) and 123 unique players. The court was divided into a 50cm x 50cm grid consisting of 840 shooting cells. The 7619 FGA of UAAP Season 81 were distributed among 587 unique shooting cells of the said grid. These 587 shooting cells were defined as the scoring area.

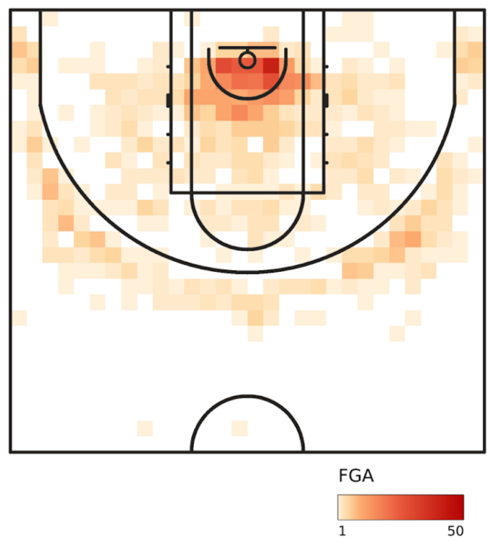
Figure 4.1 shows the field goal distribution map and the corresponding field goal grids for each of the UAAP teams. These maps already provide more nuance and information about the shooting tendencies of UAAP teams compared to the conventional shooting statistics shown in Table 4.1. For example, the maps show us that UP attempted shots close to the basket more than any other team. Another observation is that the corner three-pointer—touted as one of the most efficient shots in basketball—wasn't widely used that season with FEU and UST as the only teams that showed prominent use of the shot.



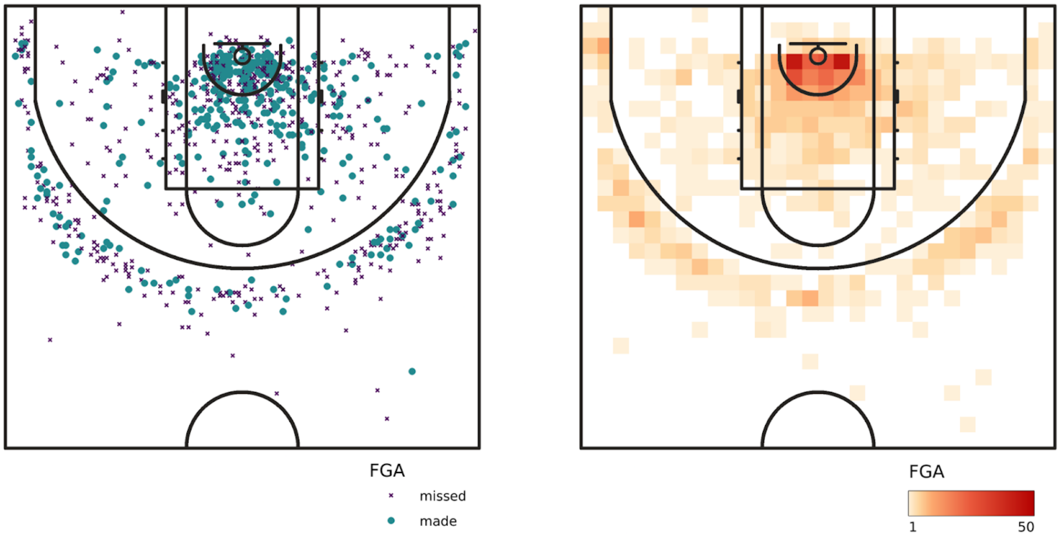


ADMU

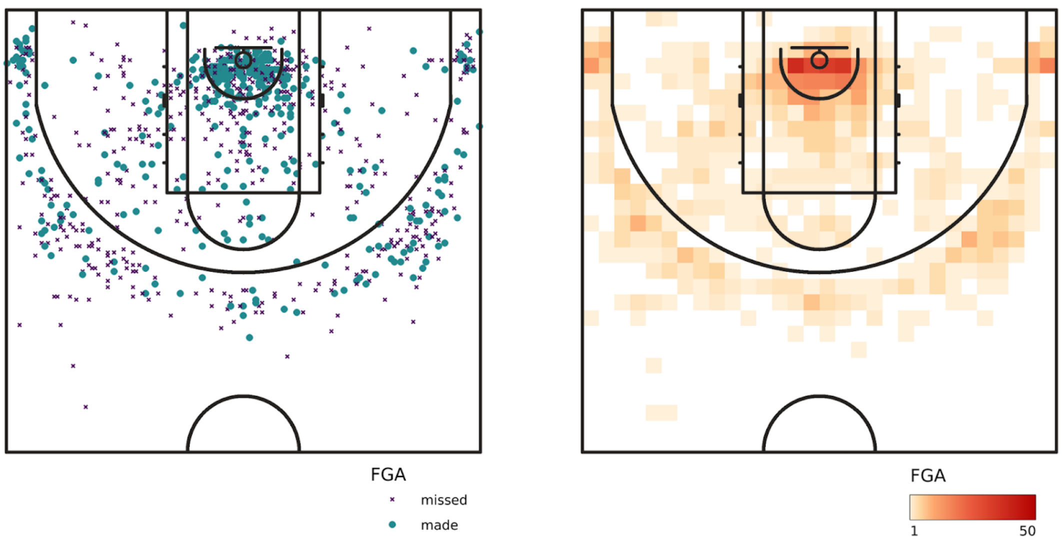




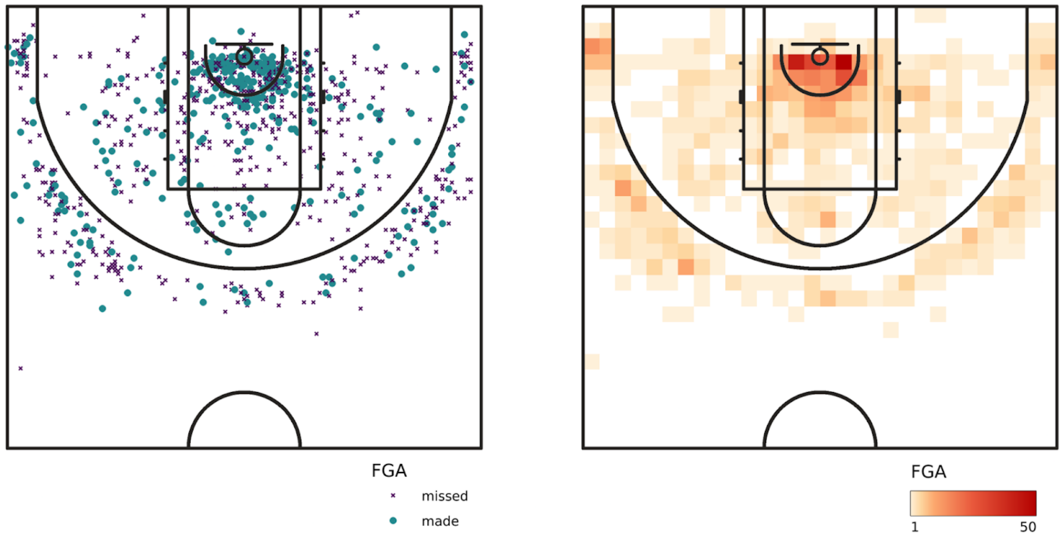
ADU



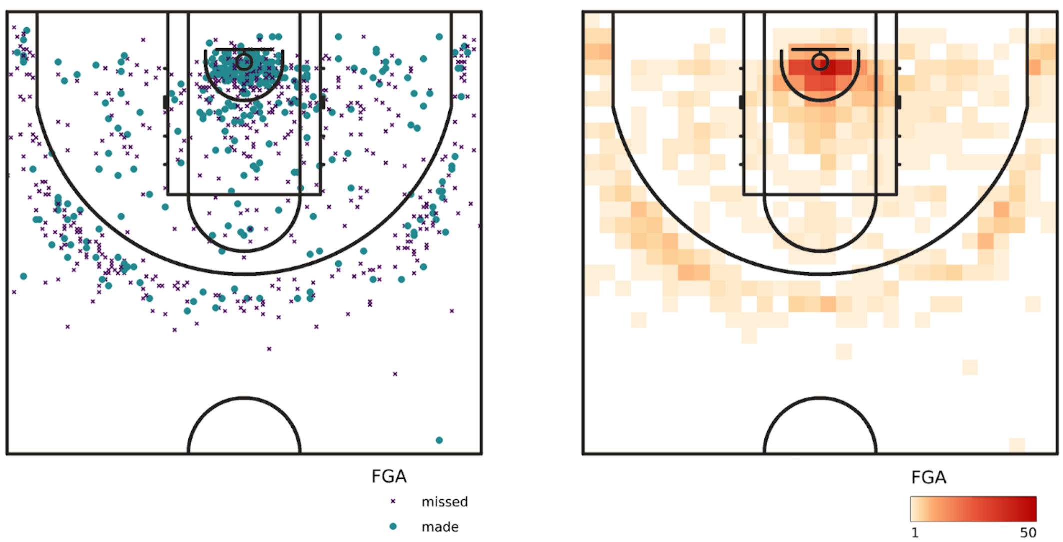
DLSU



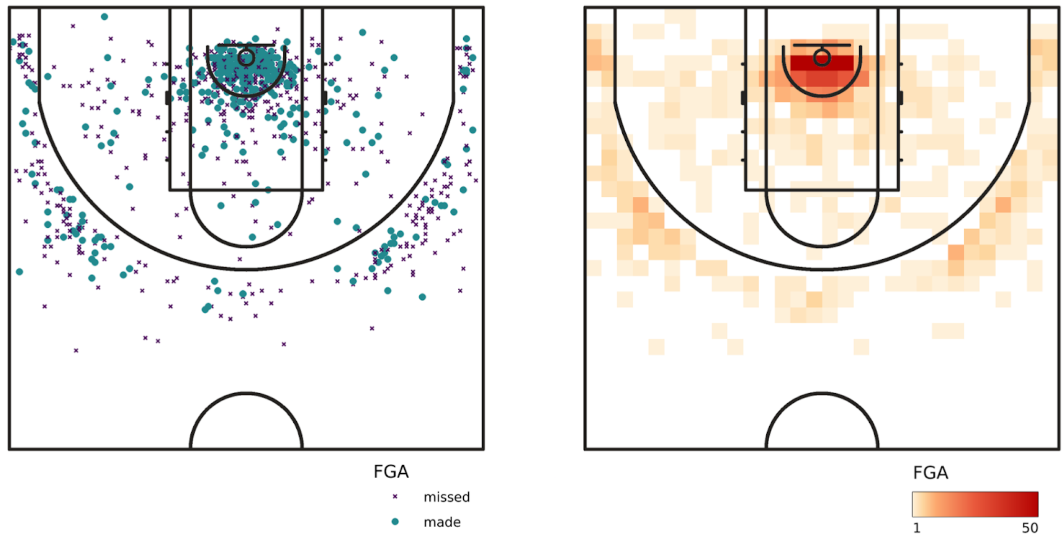
FEU



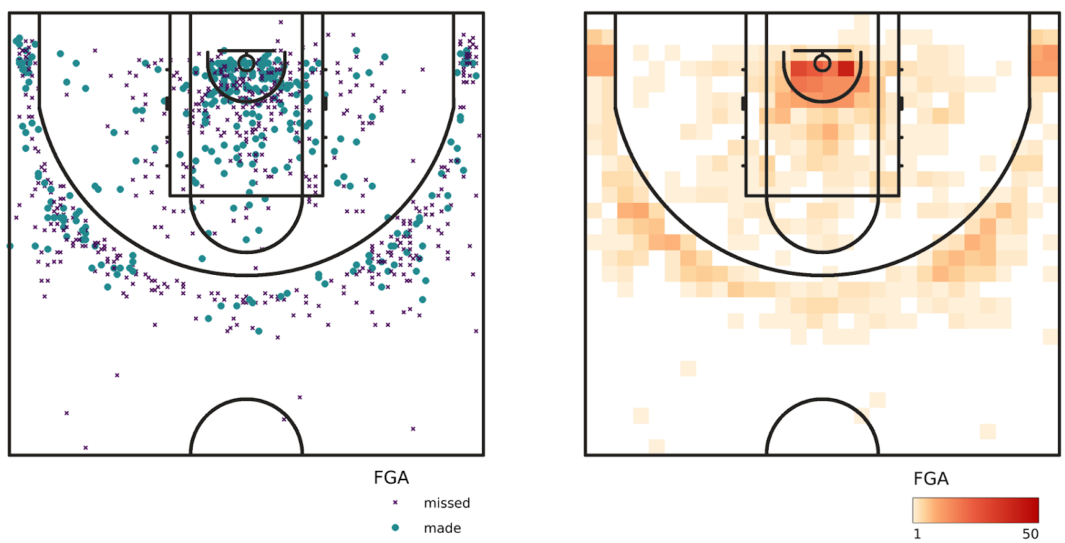
NU



UE



UP

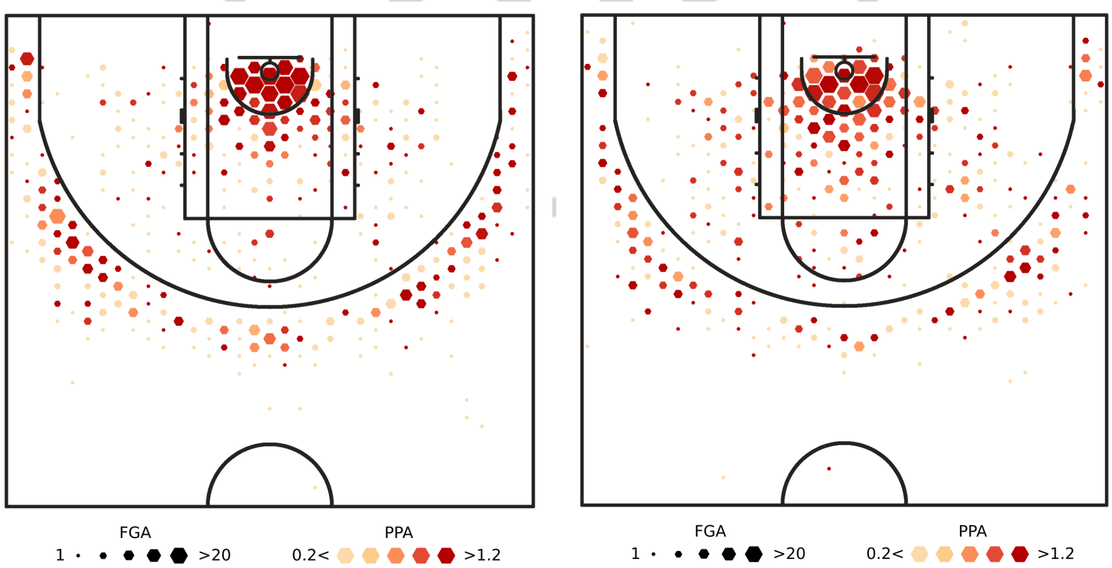


UST

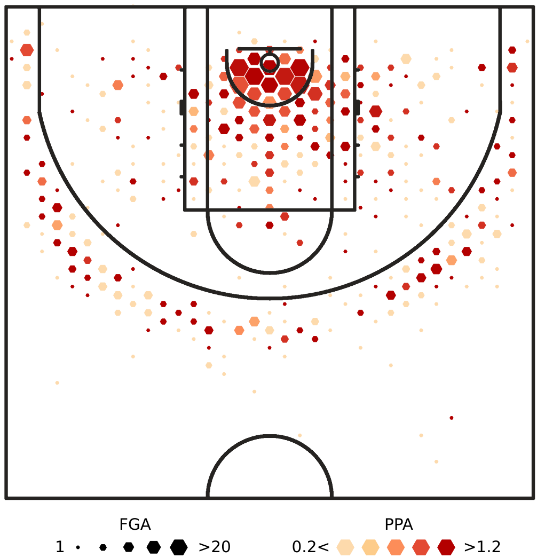
Figure 4.1

Maps of the team field goal (point) datasets and the team field goal (discretized) grids

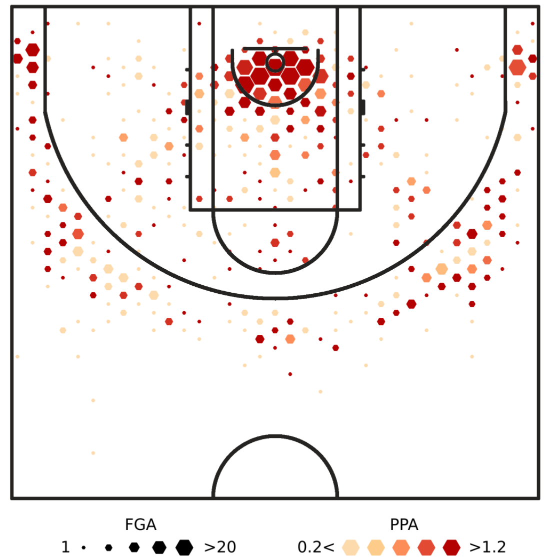
Aside from showing the spatial distribution of shots, the spatial field goal dataset can also show the spatial distribution of points scored by each team. Figure 4.2 shows the Range maps of the UAAP teams. The team Range maps provide more context about the shooting performance of teams compared to the conventional statistics in Table 4.1. We can see that both ADMU and UP attempted and converted a high number of field goals close to the basket evidenced by the large red hexagons near the basket in their Range maps. Both teams also avoided taking mid-range field goals but ADMU appeared to be more willing to take three pointers than UP. UST is also an interesting case. The team attempted and scored the least amount of field goals near the basket but attempted and scored more than the other teams from corner, wing, and top of the key three-pointers—an indication that the team was highly dependent on these three areas rather than the paint for their scoring.



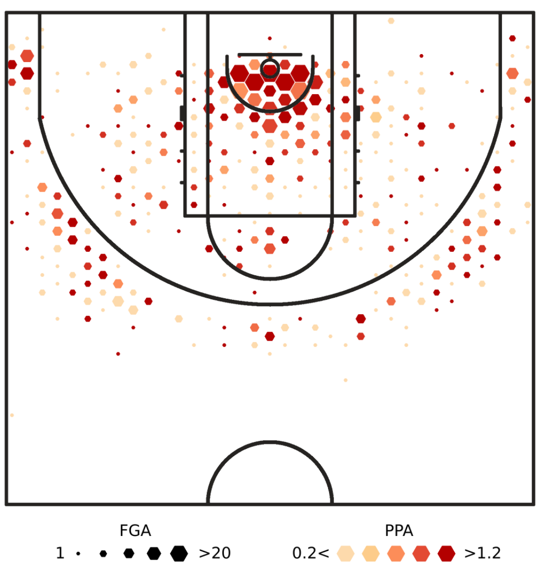
ADMU **ADU**



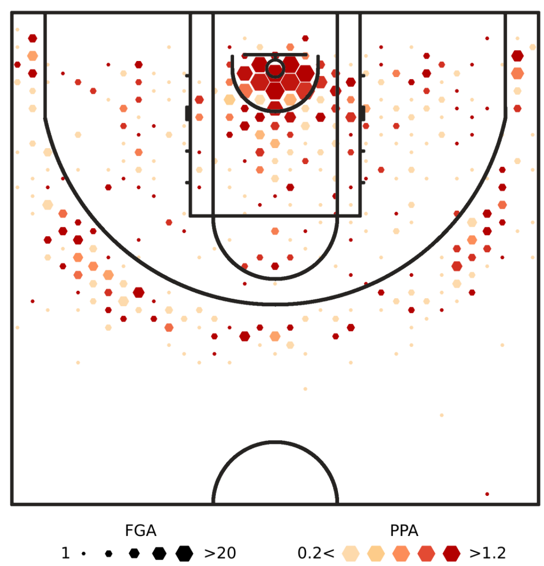
DLSU



FEU



NU



UE

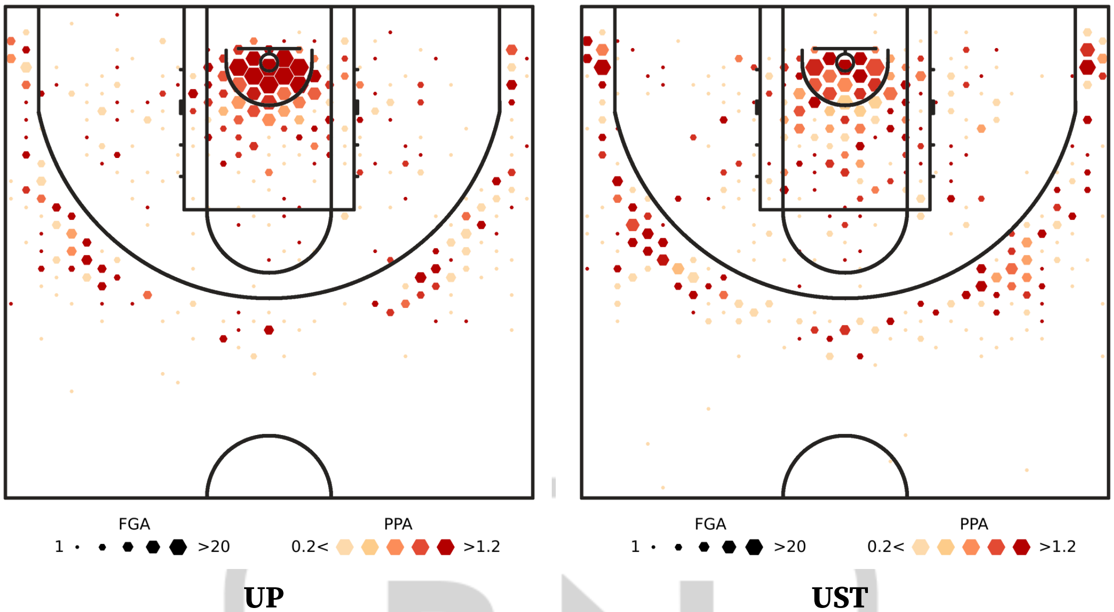


Figure 4.2

Range of the teams in UAAP Season 81

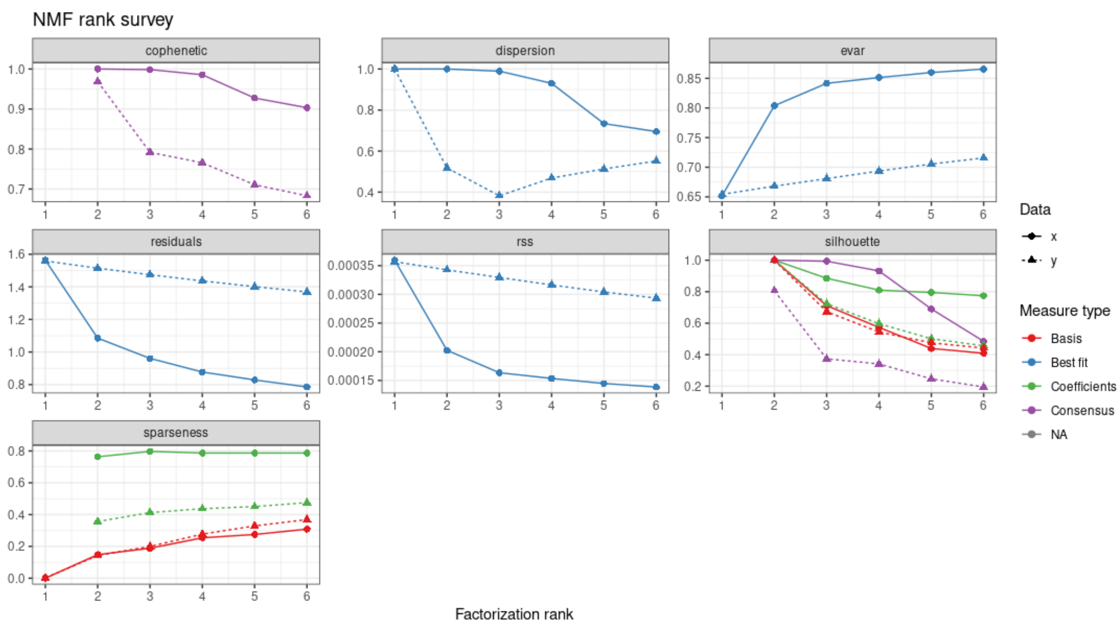
Note: The size of the hexagon pertains to the number of field goals attempted in the area while the color pertains to the points scored per attempt.

The next sections will show the spatial basis vectors and the SScE and PRLA metrics computed from the spatial field goal dataset.

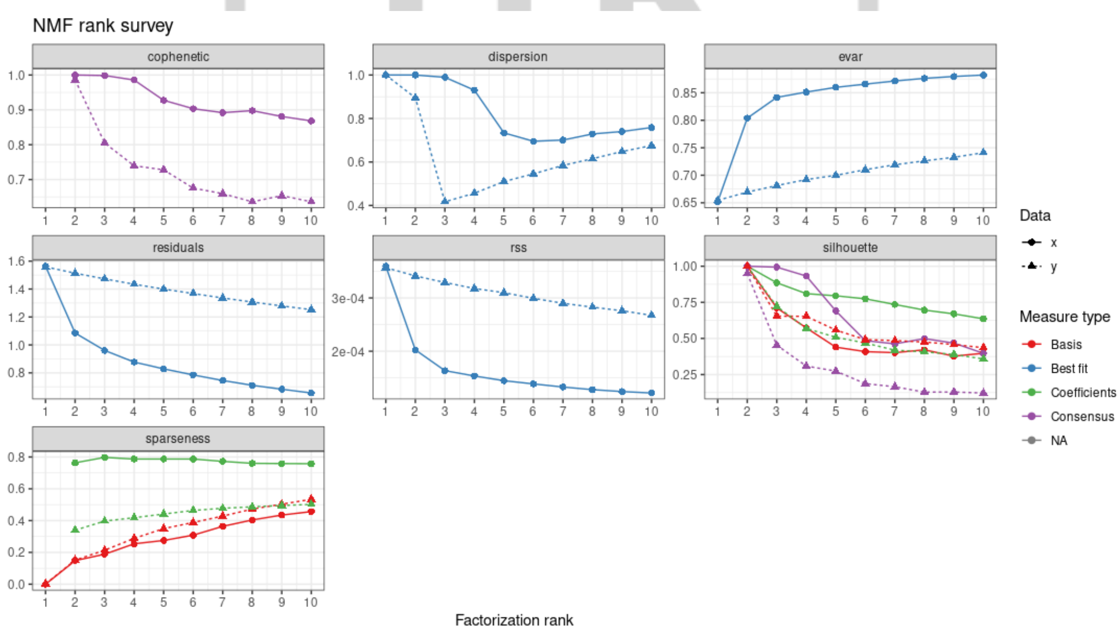
4.3 The spatial basis vectors of field goals

Using the spatial field goal dataset, the spatial basis vectors and their corresponding basis weights were computed using the NMF formula given in (2.1) and the general steps outlined in section 3.3.1.

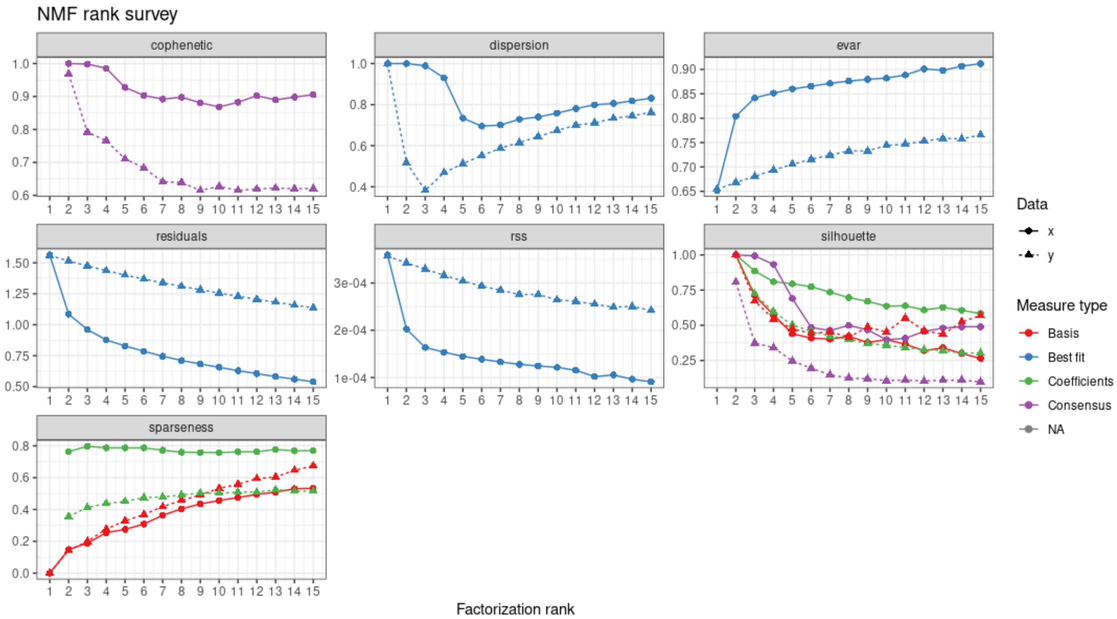
First the optimal number of ranks was estimated. Figure 4.3 shows the graphs of the different measures at varying values of K (1-6, 1-10, 1-15) using the Kullback-Leibler model at 30 runs and using a random seed of 42.



(a) K=1-6, model=KL, run=30, seed=42, y=random data



(b) K=1-10, model=KL, run=30, seed=42, y=random data



(c) K=1-15, model=KL, run=30, seed=42, y=random data

Figure 4.3

Measures of NMF at varying values of K

Looking at the different measures computed from the field goal dataset and referring to the methods introduced by Brunet (2004), Hutchins (2008), and Frigyesi (2008), it can be inferred that the optimal number of bases for the field goal dataset is from 3-5.

Next, the different models and initialization modes were compared. Figure 4.4 shows the comparison of the measures of the Kullback-Leibler and Frobenius models using the same parameters as above.

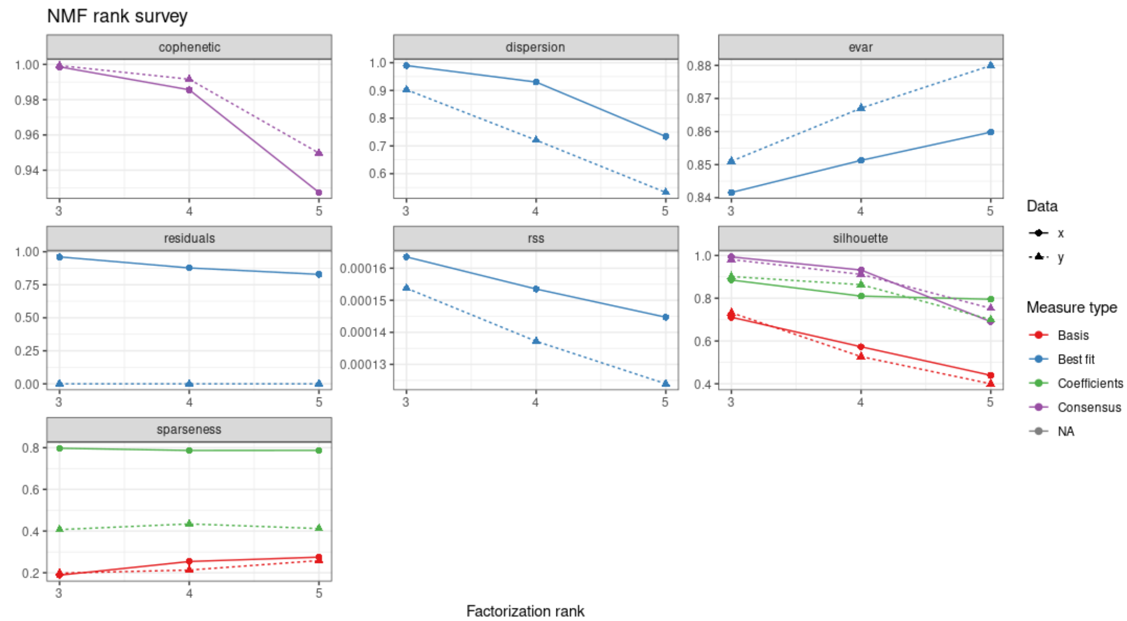
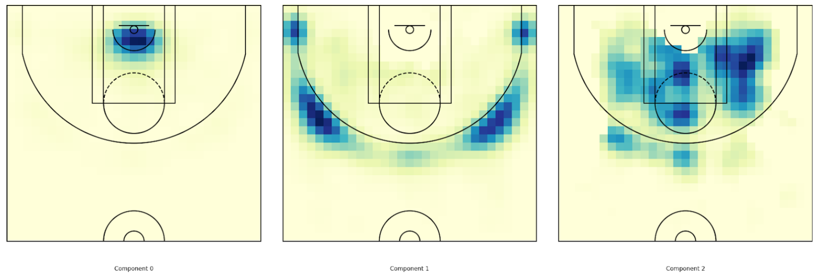


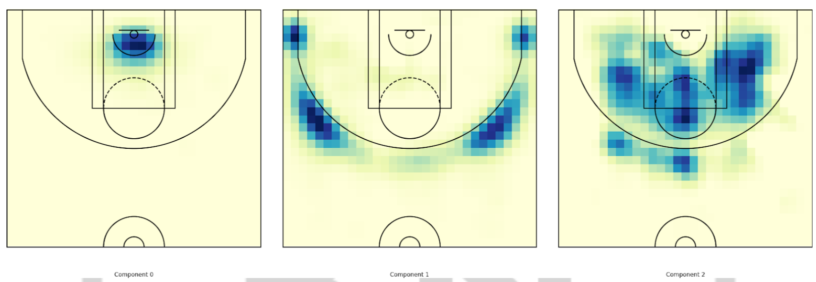
Figure 4.4

Measures of NMF for the Kullback-Leibler and Frobenius models at K = 3, 4, 5

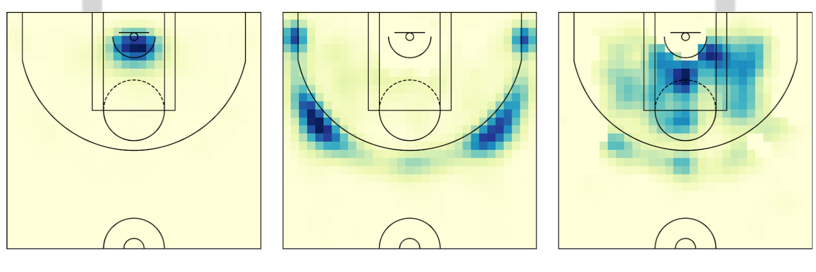
Aside from just numerical measures, it is also important to ensure that the spatial basis vectors resulting from the NMF decomposition correspond to visually interpretable shot types. Figures 4.5-4.7 show the resulting spatial basis vectors for different combinations of K, model, solver, and initialization methods.



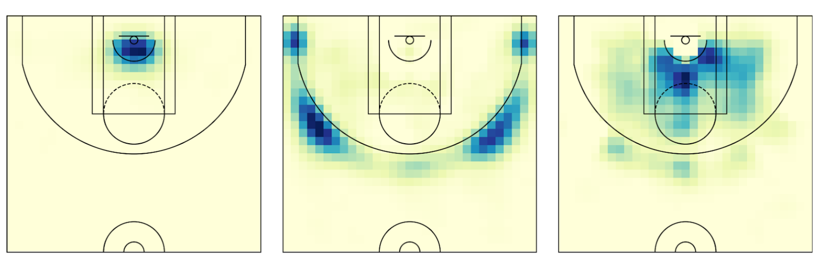
(a) Kullback-Leibler, K=3, solver=multiplicative update, init=None



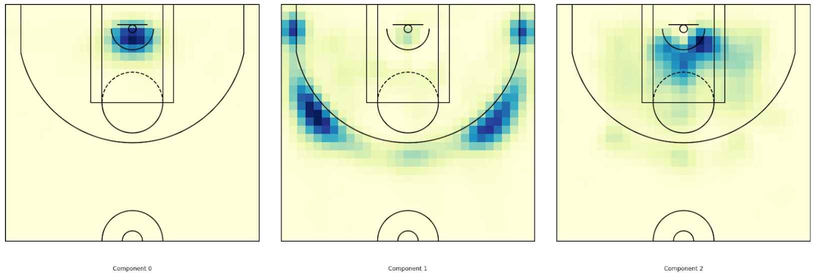
(b) Kullback-Leibler, K=3, solver=multiplicative update, init=nndsvda



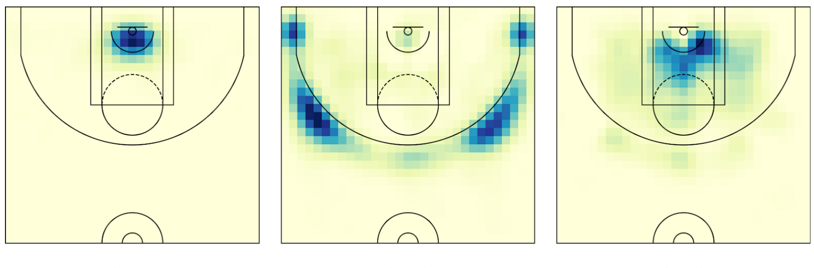
(c) Frobenius, K=3, solver=multiplicative update, init=None



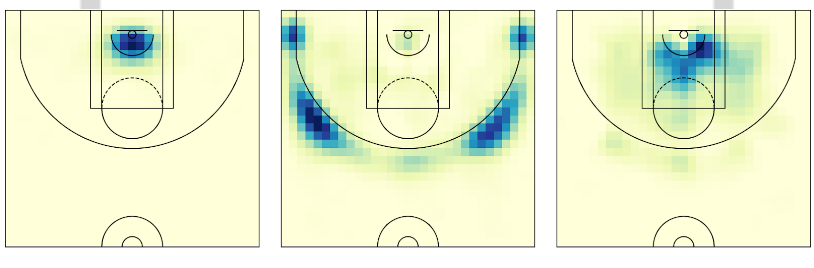
(d) Frobenius, K=3, solver=multiplicative update, init=nndsvda



(e) Frobenius, K=3, solver=coordinate descent, init=None



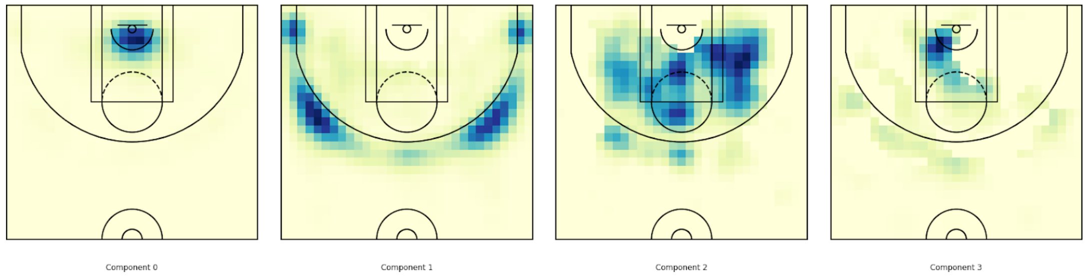
(f) Frobenius, K=3, solver=coordinate descent, init=nndsvd



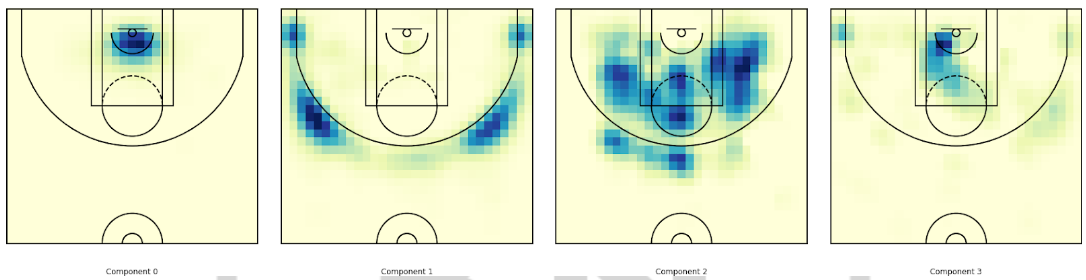
(g) Frobenius, K=3, solver=coordinate descent, init=nndsvda

Figure 4.4

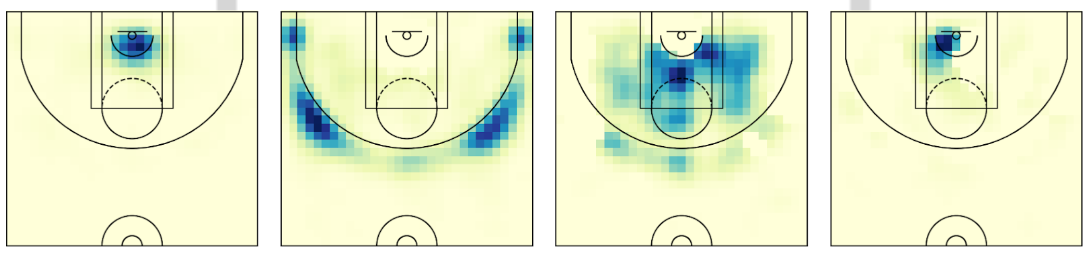
Resulting basis vectors from NMF using K = 3; model = Kullback-Leibler or Frobenius; solver = coordinate descent or multiplicative update; and init = None, nndsvd, or nndsvda



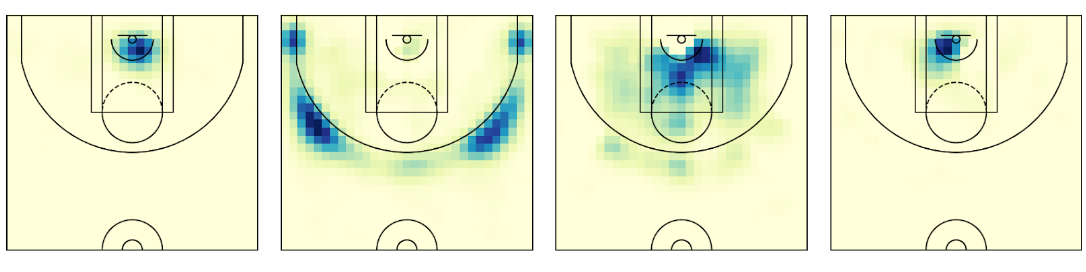
(a) Kullback-Leibler, K=4, solver=multiplicative update, init=None



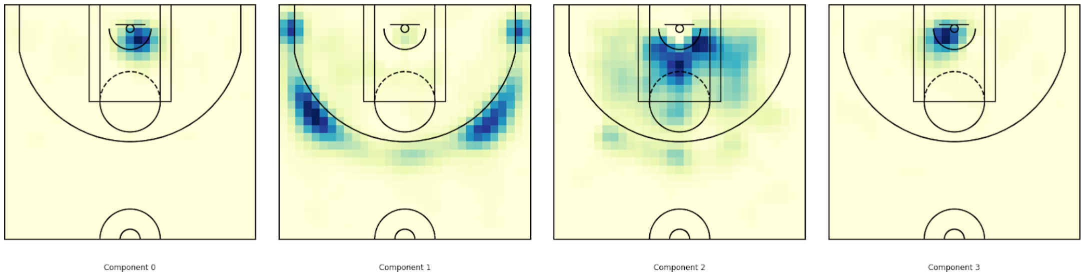
(b) Kullback-Leibler, K=4, solver=multiplicative update, init=nndsvda



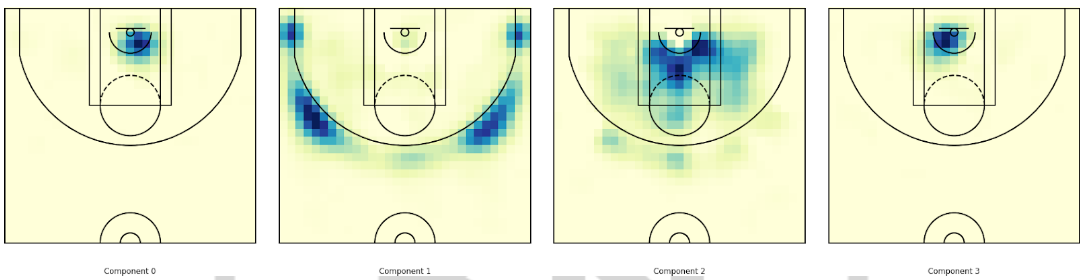
(c) Frobenius, K=4, solver=multiplicative update, init=None



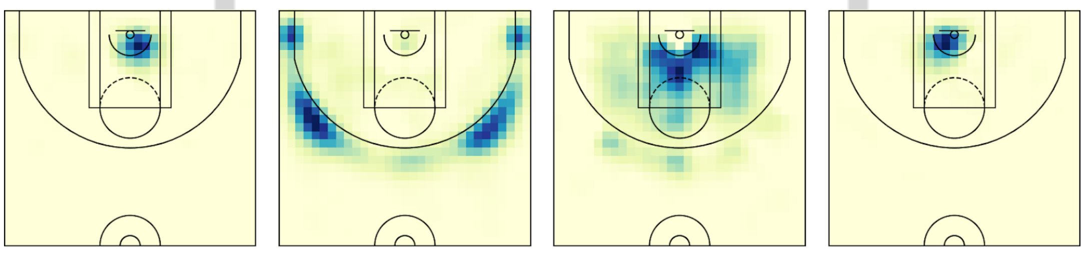
(d) Frobenius, K=4, solver=multiplicative update, init=nndsvda



(e) Frobenius, K=4, solver=coordinate descent, init=None



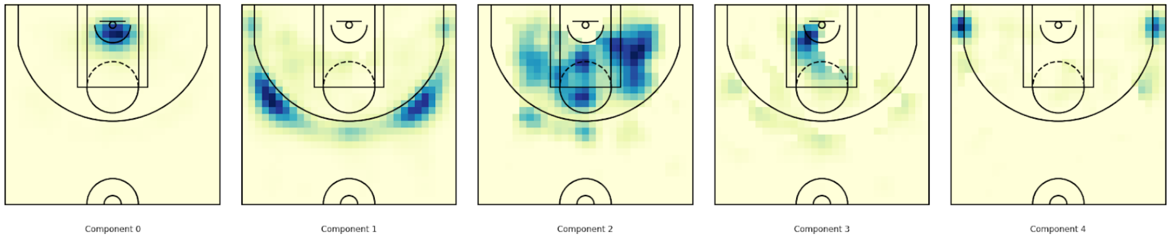
(f) Frobenius, K=4, solver=coordinate descent, init=nndsvd



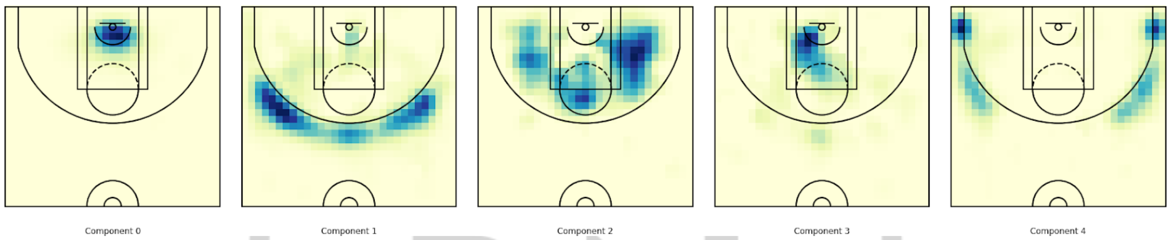
(g) Frobenius, K=4, solver=coordinate descent, init=nndsvda

Figure 4.6

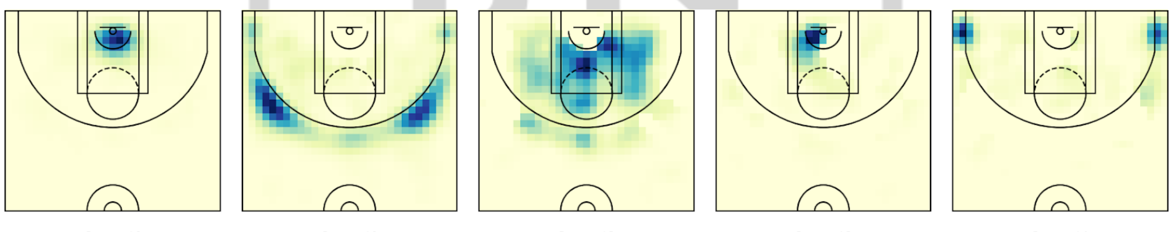
Resulting basis vectors from NMF using K = 4; model = Kullback-Leibler or Frobenius; solver = coordinate descent or multiplicative update; and init = None, nndsvd, or nndsvda



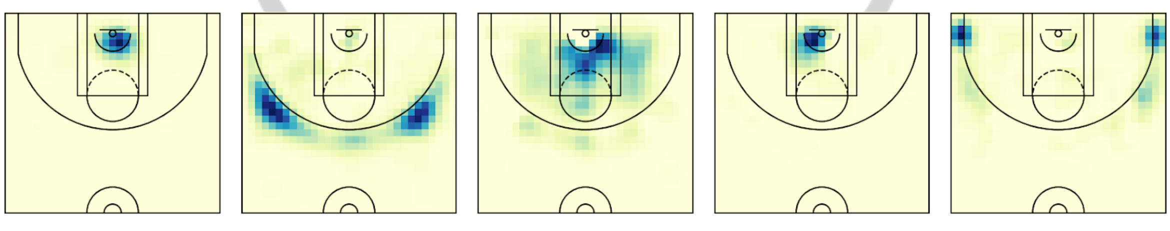
(a) Kullback-Leibler, K=5, solver=multiplicative update, init=None



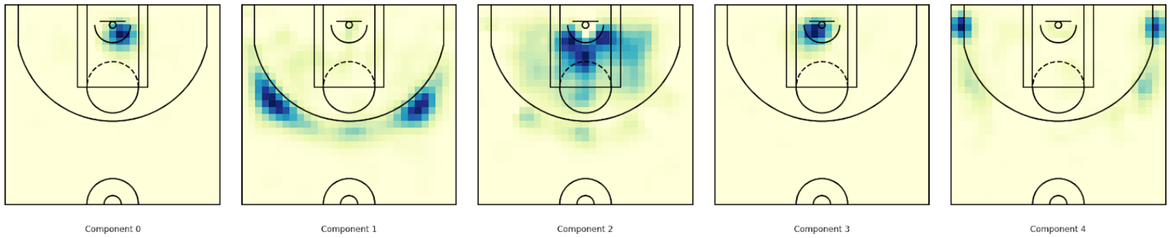
(b) Kullback-Leibler, K=5, solver=multiplicative update, init=nndsvda



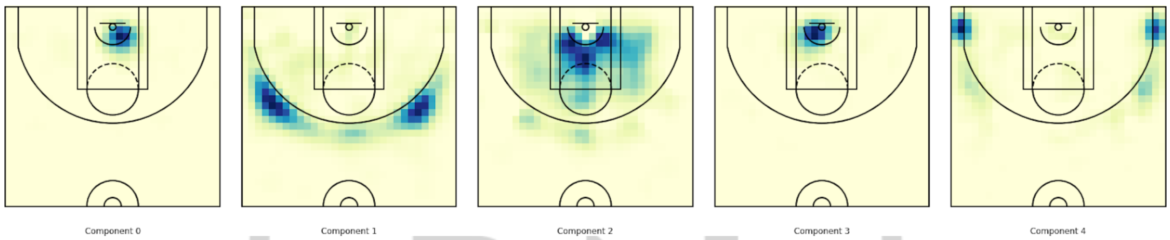
(c) Frobenius, K=5, solver=multiplicative update, init=None



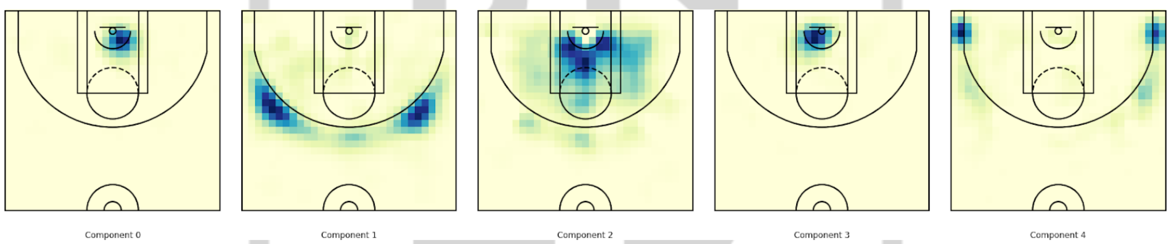
(d) Frobenius, K=5, solver=multiplicative update, init=nndsvda



(e) Frobenius, K=5, solver=coordinate descent, init=None



(f) Frobenius, K=5, solver=coordinate descent, init=nndsvd



(g) Frobenius, K=5, solver=coordinate descent, init=nndsvda

Figure 4.7

Resulting basis vectors from NMF using K = 5; model = Kullback-Leibler or Frobenius; solver = coordinate descent or multiplicative update; and init = None, nndsvd, or nndsvda

The following observations were made based on the outputs of different combinations of K, model, solver, and initialization methods:

1. The resulting basis vectors are more or less similar for the same value of *K* regardless of the *model*, *solver*, and *initialization* parameter used. There is always a basis for field goals near the basket (*Component 0 for K=3,4,5*), three-pointers (*Component 1 for K=3,4; Component 1 and 4 for K=5*), and mid-range shots (*Component 2 for K=3,4,5*).
2. The results of *init=nndsvd* or *init=nndsvda* are nearly identical when the same model and solver are used as shown in Figure 4.5 (f) and (g), Figure 4.6 (f) and (g), and Figure 4.7 (f) and (g).
3. Using *solver=multiplicative update* results in denser bases compared to using *solver=coordinate descent*.
4. At *K=3 (Figure 4.5)* and *K=4 (Figure 4.6)*, the main difference between the computed spatial basis vectors is that the "mid-range basis vectors" (*Component 2*) computed by the Kullback-Leibler (KL) model are denser than those computed using the Frobenius model.
5. Similar to the findings of Miller (2014), at *K=5 (Figure 4.7)*, the KL model resulted in more "spatially diverse" basis vectors compared to those computed using the Frobenius model. The mid-range shots outside the paint (*Component 2*) are more pronounced in the outputs of KL (*a, b*) and are understated in those of Frobenius (*c, d, e, f, g*).

6. The resulting basis vectors from the Frobenius model appear to be more focused on high-intensity areas of field goal attempts as observed when comparing *Component 3* obtained using KL and Frobenius. *Component 3* in KL includes locations outside of the restricted area while the same component in Frobenius computations only include areas in the restricted area of the paint.

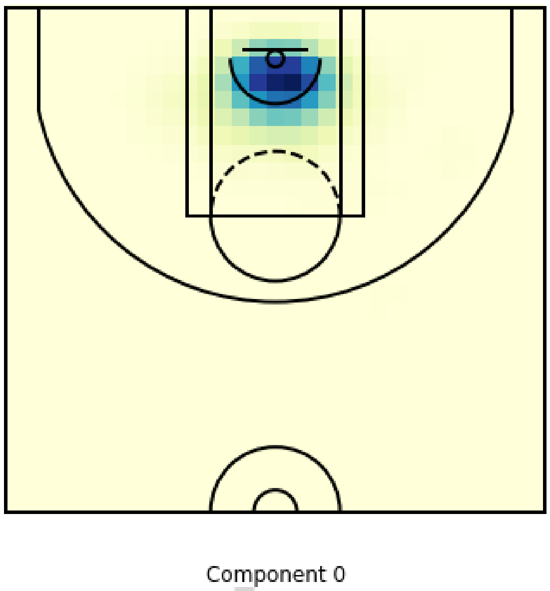
Based on the considerations that NMF should provide a good parts-based decomposition of the field goal dataset and avoid overfitting, Figure 4.7 (b) was chosen as the combination to use in the research because:

1. it separates the corner three-pointers (*Component 4*) from the other three-pointers (*Component 1*);
2. it does not understate the mid range field goals (*Component 2*).

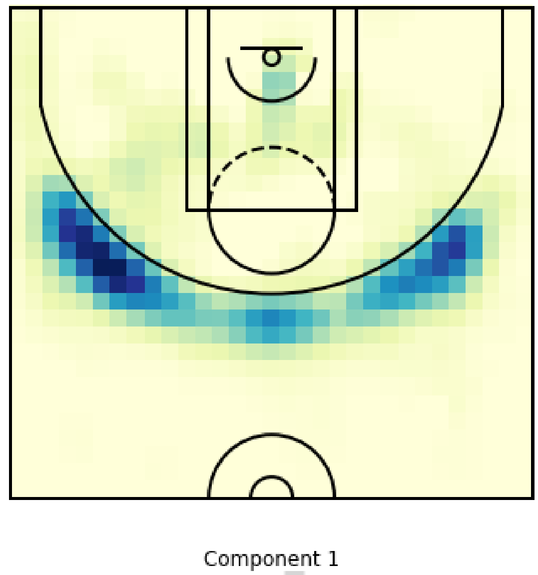
   The parameters for Figure 4.7 (b) are:

- model/loss function: *Kullback-Leibler*
- K / factorization rank: *5*
- solver: *multiplicative update*
- init: *nndsvda* or Non-negative Double Singular Value Decomposition (NNDSVD) initialization with zeros filled with the average of the non-negative matrix V

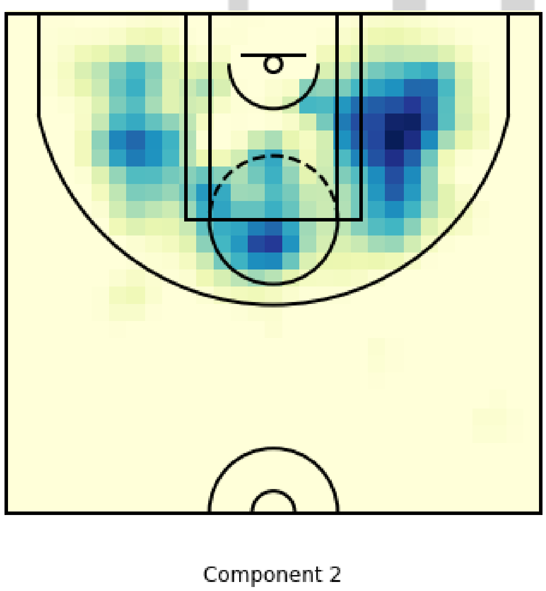
Figure 4.8 shows the spatial basis vectors of the field goal dataset and their corresponding "shot types".



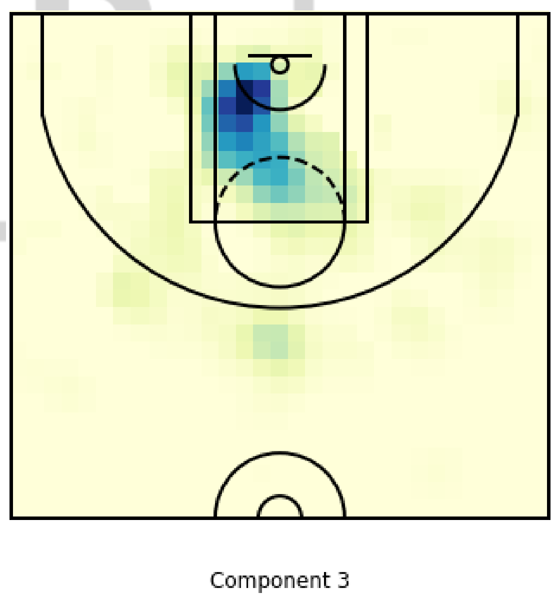
(a) restricted area / at-rim



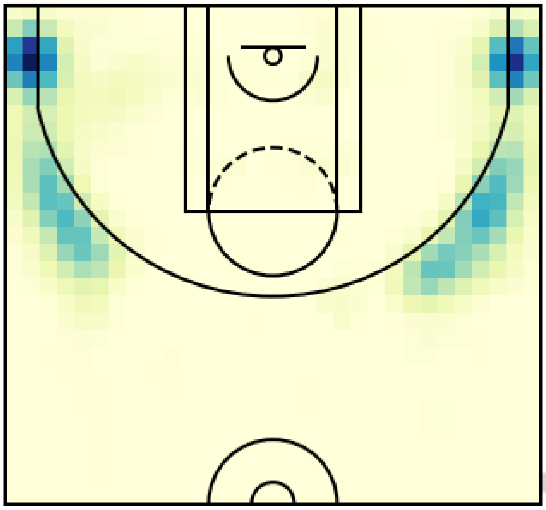
(b) wing + key three-pointers



(c) mid-range + some paint



(d) left-block + some paint



Component 4

(e) corner three-pointers
+ some wing three-pointers

Figure 4.8

The spatial basis vectors of the field goal dataset and their corresponding shot types

Player-specific basis weights computed by NMF provide a concise characterization of a player's shooting habits as these weights correspond to the frequency by which the player takes certain types of field goals. Table 4.3 compares the normalized basis weights of the top 20 players with the most number of field goal attempts from UAAP Season 81.

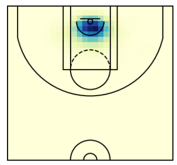
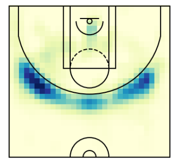
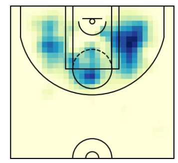
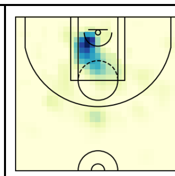
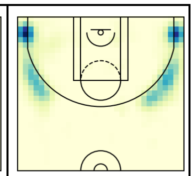
		 Component 0	 Component 1	 Component 2	 Component 3	 Component 4
1	A. Pasaol	**0.47**	**0.27**	0.11	0.00	0.14
2	A. Melecio	0.18	**0.27**	**0.20**	**0.12**	**0.22**
3	R. Subido	0.08	**0.41**	0.12	0.10	**0.29**
4	D. Ildefonso	0.31	**0.26**	0.11	0.09	**0.22**
5	J. Ahanmisi	0.04	**0.32**	**0.16**	**0.36**	0.12
6	Ju. Gomez de Liaño	0.26	**0.21**	**0.21**	0.06	**0.27**
7	B. Akhuetie	**0.94**	0.00	0.03	0.00	0.03
8	P. Desiderio	0.29	**0.26**	0.07	0.11	**0.28**
9	M. Lee	0.05	**0.25**	0.07	0.06	**0.57**
10	S. Manganti	0.27	0.15	0.14	**0.19**	**0.25**
11	J. Clemente	0.24	**0.28**	0.12	0.00	**0.35**
12	J. Baltazar	**0.43**	0.13	**0.19**	**0.25**	0.00
13	P. Sarr	**0.49**	0.00	**0.26**	**0.24**	0.01
14	L. Santillian	**0.44**	0.17	**0.20**	**0.19**	0.00
15	T. Ravena	**0.49**	**0.42**	0.09	0.01	0.00
16	P. Manalang	**0.38**	**0.39**	0.07	0.05	0.12
17	A. Kouame	**0.78**	0.08	0.03	0.11	0.00
18	A. Caracut	0.20	**0.23**	**0.20**	**0.25**	0.12
19	H. Cani	0.22	0.15	**0.29**	0.09	**0.26**
20	W. Comboy	0.12	**0.22**	0.12	**0.29**	**0.26**
	Mean	***0.35***	***0.20***	***0.15***	***0.12***	***0.18***

Table 4.3

Normalized player weights for each basis for the top 20 players with the most field goal attempts from UAAP Season 81

BN HR

Note: The dataset of normalized basis weights for all players can be found at:

https://github.com/benhur07b/ms-thesis-spatial-analysis-shooting-philippine-basketball/blob/main/outputs/spatial-basis/nmf_weights-players-retained.csv

Highlighted values in Table 4.3 indicate weights that are greater than the mean for the corresponding spatial basis and reveal the general type of shooter a player is. In broad strokes, we can identify several types of shooters:

- Three-point specialists are those with high weights for Components 1 and 4. Players such as R. Subido (UST), M. Lee (UST), J. Clemente (NU) are examples.
- Players who only take shots near the basket are those with unusually high weights for Component 0. Players like B. Akhuetie (UP) and A. Kouame (ADMU) fit the bill.
- Scorers who attack the basket or take wing/key three-pointers are those with high weights for Components 0 and 1. This includes A. Pasaol (UE), T. Ravena (ADMU), and P. Manalang (UE).
- Two-point scorers who have high weights for Components 0, 2, and 3 such as J. Baltazar (DLSU), P. Sarr (ADU), and L. Santillan (DLSU).

These player types correspond to general intuitions about player shooting habits while also providing more context.

BN HR

4.4 Groups of players with similar shooting habits

Players with similar shooting habits were determined by computing the Euclidean distance of the player basis weights. Using this distance, a player's five nearest neighbors were identified. If player k is the neighbor of player l, it is assumed that player l is also a neighbor of player k to enforce symmetry. Table 4.4 shows the nearest neighbors of the top 20 players with the most field goal attempts in UAAP Season 81.

Table 4.4

Top 20 players with the most field goal attempts and the players with similar shooting habits

#	Player	Similar Players
1	A. Pasaol	A. Diputado, W. Navarro, P. Manalang, R. Escoto, C. Conner, T. Ravena, J. Manzo, Ja. Gomez de Liaño
2	A. Melecio	J. Manuel, Ju. Gomez de Liaño, J. Varilla, D. Ildefonso, H. Cani, R. Subido, A. Caracut, W. Comboy, M. Aquino
3	R. Subido	J. Mendoza, A. Wong, A. Melecio, J. Varilla, D. Dario, K. Zamora, T. Tio, A. Inigo
4	D. Ildefonso	P. Desiderio, Ju. Gomez de Liaño, W. Navarro, J. Cullar, S. Manganti, A. Melecio, F. Serrano, C. Conner, J. Lastimosa, M. Nieto, C. Vito
5	J. Ahanmisi	W. Comboy, T. Tio, A. Caracut, I. Batalier, A. Inigo, J. Gallego
6	Ju. Gomez de Liaño	H. Cani, J. Parker, A. Melecio, D. Ildefonso, J. Manuel, S. Manganti, J. Clemente, J. Varilla, C. Conner, J. Espeleta, M.

		Aquino
7	B. Akhuetie	S. Akomo, B. Ebona, V. Magbuhos, P. Orizu, D. Murrell, A. Kouame, S. Camacho
8	P. Desiderio	D. Ildefonso, J. Clemente, S. Manganti, M. Nieto, C. Vito, F. Serrano, J. Lastimosa, A. Inigo, J. Manuel, D. Dario
9	M. Lee	A. Asistio, B. Bienes, E. Mojica, J. Go, D. Yu, K. Tuffin
10	S. Manganti	J. Manuel, P. Desiderio, F. Serrano, D. Ildefonso, Ju. Gomez de Liaño, A. Caracut, W. Comboy, A. Inigo
11	J. Clemente	D. Dario, Ju. Gomez de Liaño, P. Desiderio, J. Varilla, M. Maloles, M. Nieto, C. Vito
12	J. Baltazar	L. Santillian, M. Dyke, P. Sarr, R. Escoto, L. Gonzales, C. Catapusan, J. Sinclair, I. Batalier
13	P. Sarr	M. Dyke, J. Sinclair, J. Baltazar, L. Santillian, G. Mahinay, I. Gaye
14	L. Santillian	J. Baltazar, R. Escoto, M. Dyke, J. Cullar, R. Verano, P. Sarr, L. Gonzales
15	T. Ravena	P. Manalang, A. Pasaol, I. Go, W. Navarro, J. Manzo
16	P. Manalang	I. Go, W. Navarro, T. Ravena, A. Pasaol, J. Lastimosa, A. Tolentino
17	A. Kouame	V. Magbuhos, B. Akhuetie, D. Murrell, S. Akomo, B. Ebona, J. Manzo, S. Ildefonso
18	A. Caracut	L. Gonzales, R. Verano, J. Manuel, A. Melecio, S. Manganti, J. Ahanmisi, W. Comboy, T. Tio, J. Gallego, I. Batalier
19	H. Cani	Ju. Gomez de Liaño, J. Parker, J. Manuel, A. Melecio, J. Espeleta, J. Varilla
20	W. Comboy	A. Inigo, J. Manuel, A. Caracut, S. Manganti, A. Melecio, J. Ahanmisi

BN HR

Figure 4.9 shows the plot of the number of neighbors (similar shooting habits) a player has versus the average distance between him and his neighbors. The size of the circle indicates the number of field goal attempts by the player. This plot can identify players with unique shooting habits (low number of neighbors and high average distance) and those with common shooting habits (high number of neighbors with low average distance). There are two groups of players that are interesting in Figure 4.9. The first one is the group of players with average distance near or less than 0.1. This indicates that they have very similar shooting habits to their neighbors. Players from this group include P. Orizu, B. Ebona. S. Akomo, and B. Akhuetie. What's characteristic of these four players is that they almost exclusively take shots from near the basket as evidenced by their unusually high weights in Component 1 of the spatial basis vectors. The second group are those whose average distance to their neighbors is greater than 0.3. All the players that belong to this group only have five to six neighbors each. This combination of low number of neighbors and long distance between them and their neighbors could indicate that the players belonging to this group have uncommon shooting habits relative to the rest of the league. The players in this group are G. Mamuyac, J. Pingoy, F. Jaboneta, and T. Tio.

BN HR

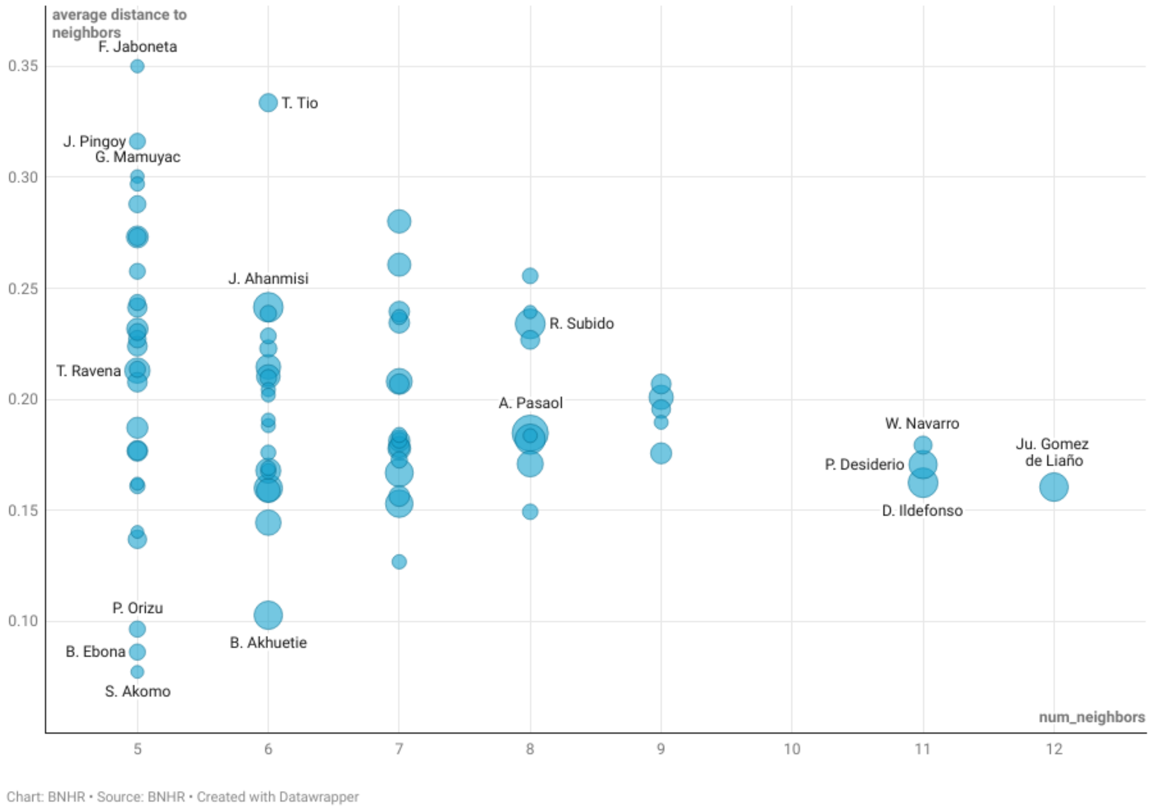


Figure 4.9

Plot of the number of player's neighbors vs average distance between the player and his neighbors. The size of the circle corresponds to the number of field goal attempts by the player.

Note: The number of neighbors indicate the number of players around the league with similar shooting habits as the player. The average distance indicates how similar the player is to his neighbors (shorter distance = more similar).

An interactive online-version of the chart can be found at:

https://datawrapper.dwcdn.net/HhqG5/1/

Figure 4.10 shows the same graph as Figure 4.9 but adds information about the player's team.

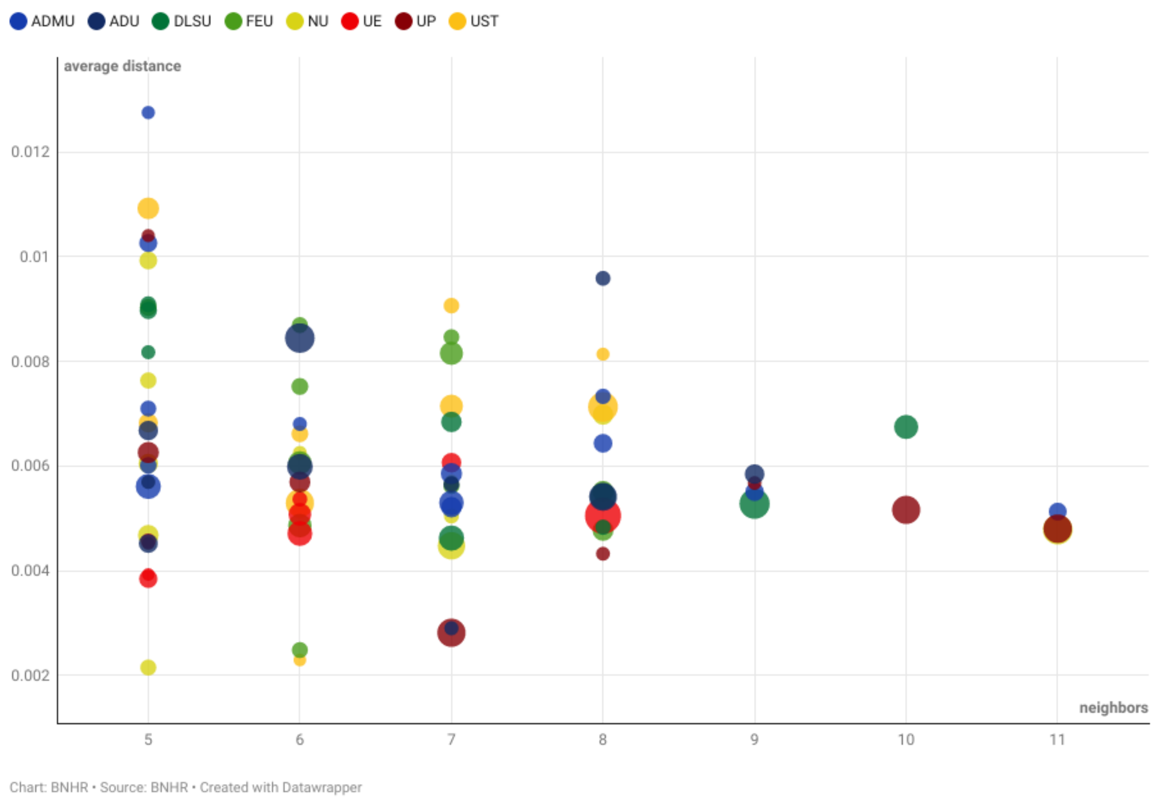


Figure 4.10

Plot of the number of player's neighbors vs average distance between the player and his neighbors. The size of the circle corresponds to the number of field goal attempts by the player. The color corresponds to the team.

Note: An interactive online version of the chart can be found at:

https://datawrapper.dwcdn.net/ReKGQ/1/

Another way to look at the data is to look at the similarities of players with their teammates. For each team, a nearest neighbor analysis was performed to compute the average distance between a player and all his teammates. Figure 4.11 shows the variation of player shooting habits per team. An interesting observation here are the plots of the UE Red Warriors and DLSU Green Archers. In the case of both teams, the distances between the shooting habits of their players with the rest of their teammates are small which means that their players have similar shooting habits and can indicate that both teams run a predictable offense regardless of the combination of players on the court. For both teams, the players with the most field goal attempts are also those with the most similar shooting habits to the rest of their team indicated by the large circles at the lower end of their graphs.

In contrast, the other teams have a significantly larger range of values in terms of the similarities in the shooting habits of their players. All of them have one or two players whose shooting habits are very different compared to the rest of the team—G. Mamuyac and A. Kouame for ADMU, J. Pingoy and V. Magbuhos for ADU, P. Orizu and B. Ebona for FEU, D. Yu for NU, B. Akhuetie for UP, and S. Akomo for UST. Unlike UE and DLSU, the players with the most field goal attempts for the rest of the teams do not always have similar shooting habits with the rest of their teammates which can indicate that the teams can run a more varied offense.

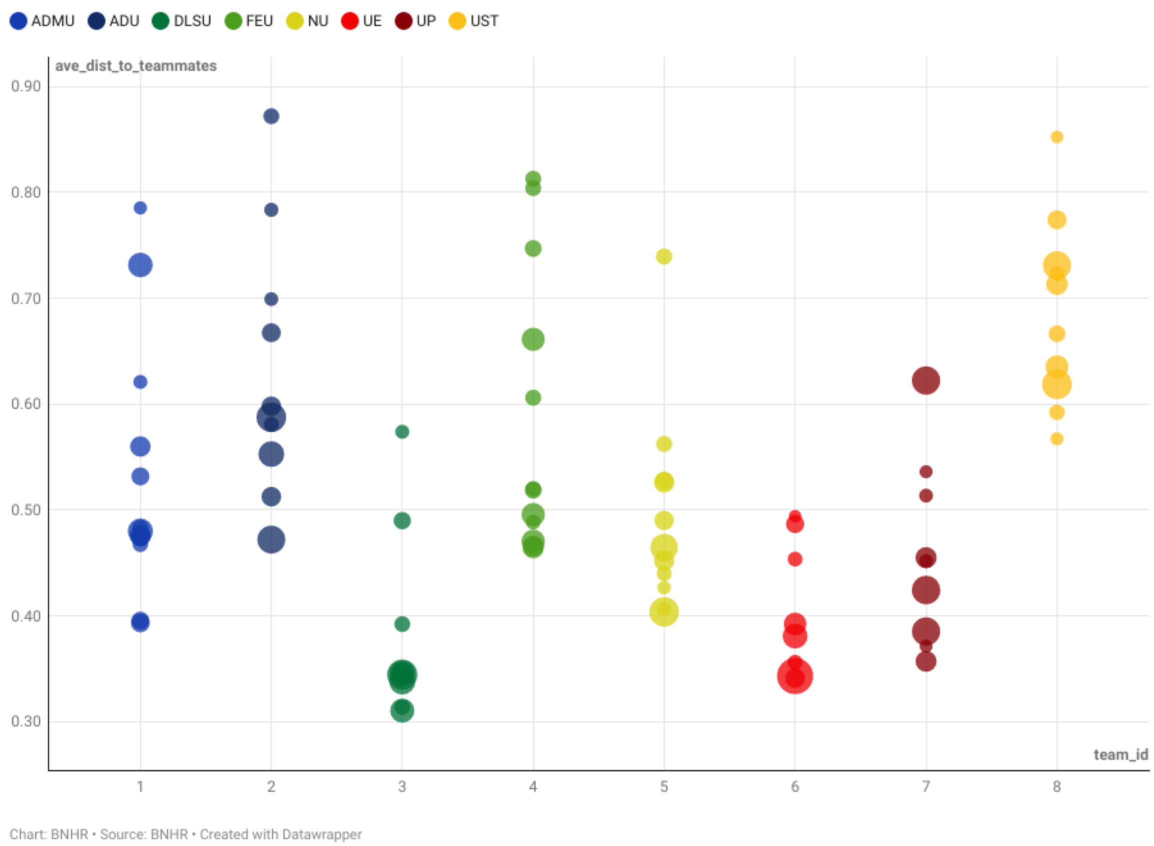


Figure 4.11

Plot of the average distance between a player's shooting habits with the rest of his teammates. Shorter distance = greater similarity.

Note: An interactive online version of the chart can be found at: https://datawrapper.dwcdn.net/2UphE/1/

The complete results of the nearest neighbors analysis are found below: https://github.com/benhur07b/ms-thesis-spatial-analysis-shooting-philippine-basketball/tree/main/outputs/nearest-neighbor

4.5 Empirical Bayes estimate of Points Per Attempt

The raw Points Per Attempt (PPA) map of UAAP Season 81 is shown in Figure 4.12. This raw PPA map is spatially noisy and contains cells that have very high or very low values compared to their neighbors that can be attributed to the fact that the PPA is estimated poorly at cells with low number of field goal attempts.

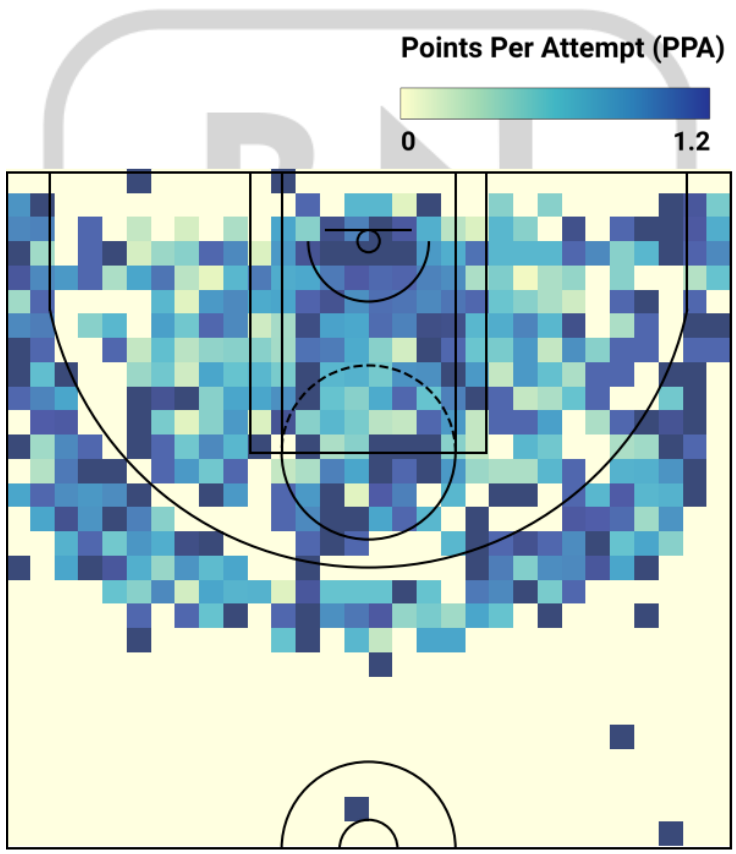


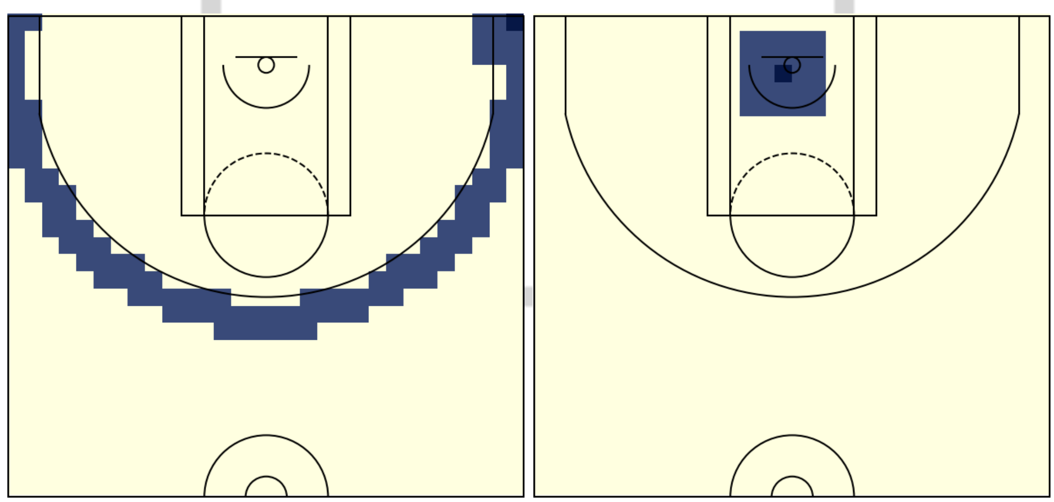
Figure 4.12

Raw Points Per Attempt (PPA) map for all field goals of UAAP Season 81

To create a better estimate of the PPA at each cell, an EB rate estimate of the PPA was created using the method described in Section 3.4.1.

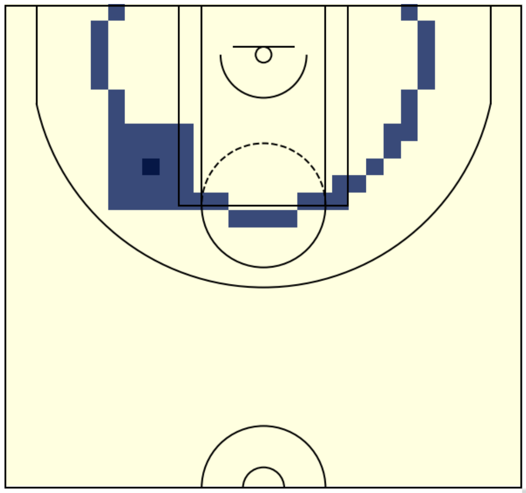
For each cell, a prior distribution which includes nearby and equidistant cells was defined. Figure 4.13 shows some examples of priors for different cells on the court. The dark pixel indicates the location/shooting cell and the lighter pixels indicate the nearby and equidistant cells used as the prior for the EB-estimation.

More than 90% of the shooting cells had at least 40 cells included in their prior distribution while around 80% of the shooting cells had at least 40 field goal attempts as part of their prior distribution. All of the shooting cells with a prior consisting of at least 40 field goal attempts are located within 10 meters from the basket.

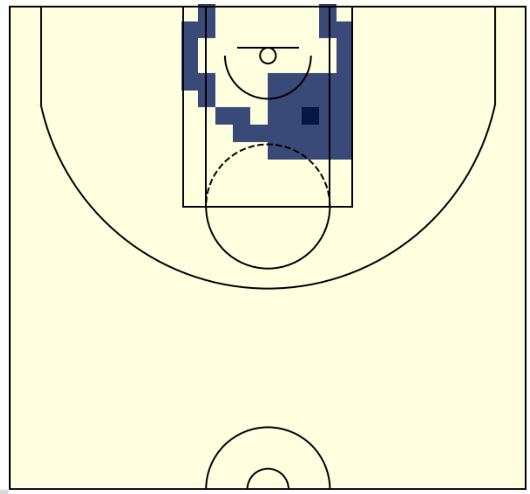


NEARBY AND EXIDISTANT CELLS (row=0, col=0 from top-right corner)

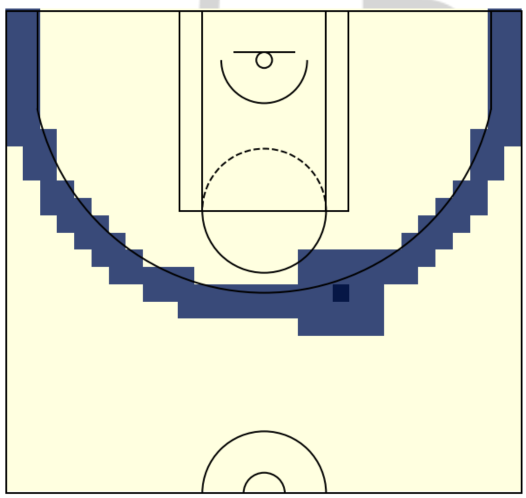
NEARBY AND EXIDISTANT CELLS (row=3, col=15 from top-right corner)



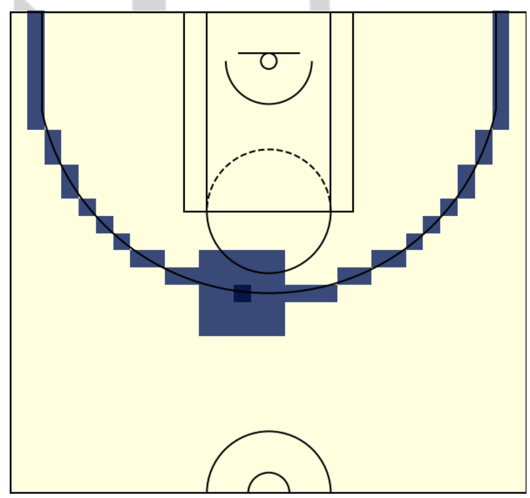
NEARBY AND EXIDISTANT CELLS (row=9, col=21 from top-right corner)



NEARBY AND EXIDISTANT CELLS (row=6, col=12 from top-right corner)



NEARBY AND EXIDISTANT CELLS (row=16, col=10 from top-right corner)



NEARBY AND EXIDISTANT CELLS (row=16, col=16 from top-right corner)

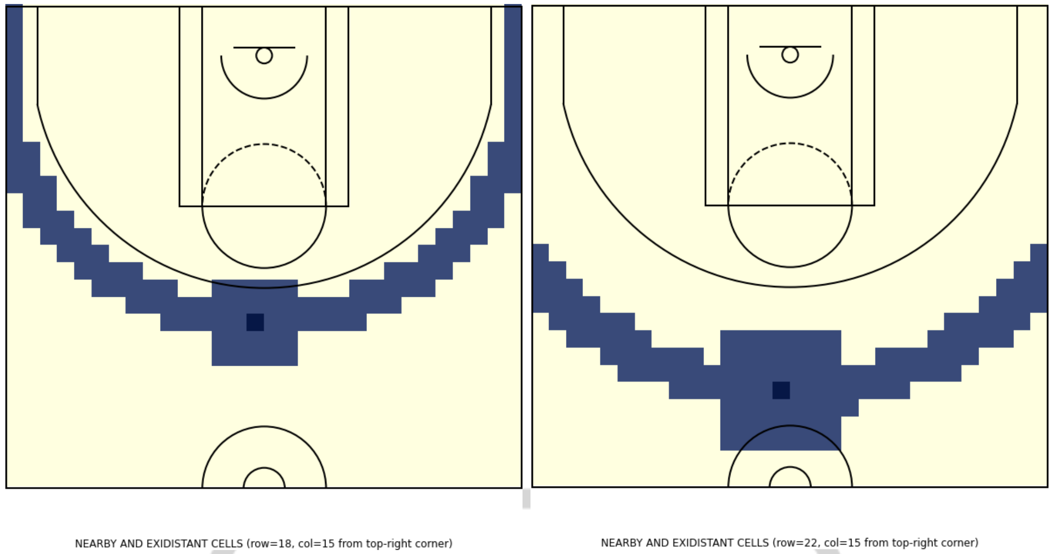


Figure 4.13

Priors of different shooting cells

BN HR

The prior means, prior variances, and shrinking factors were then computed. These are shown in Figures 4.14, 4.15, and 4.16 respectively.

The prior means in Figure 4.14 show considerable symmetry and a pattern where the points per attempt values are high near the basket then slowly decreases from outside the restricted area until near the three-point line where it increases again until about 2 meters from the line where it starts decreasing once more. The prior means beyond 10 meters from the basket are close to zero because almost no field goal attempts are taken from this distance.

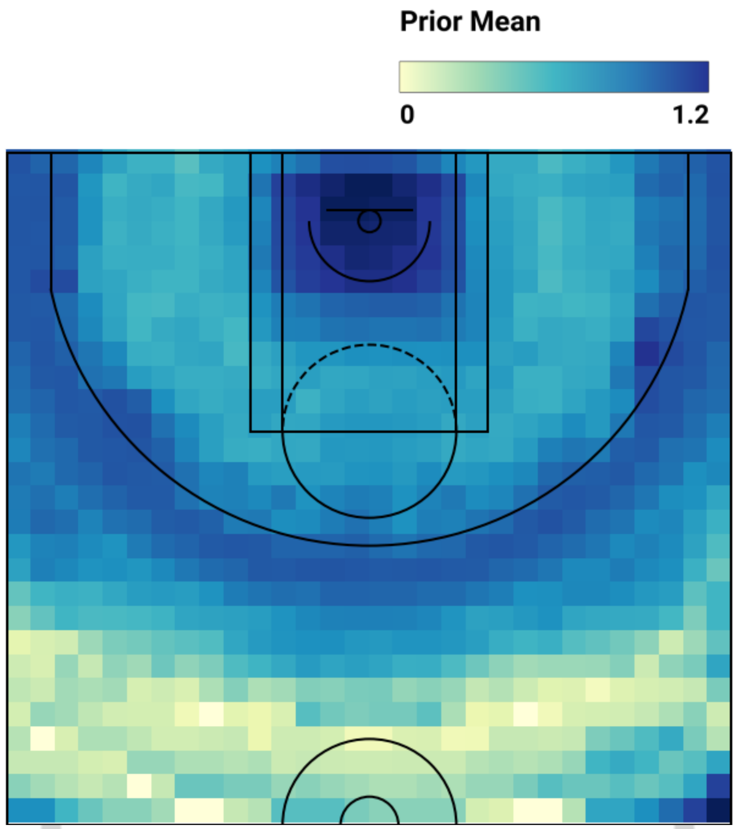


Figure 4.14

Map of prior means per shooting cell

Figure 4.15 shows a distinct demarcation at around 10 meters from the basket where the prior variances suddenly increase in value. This can be attributed to the fact that more than 99% of field goal attempts are taken within 10 meters from the basket. Beyond 10 meters, the shooting cells have a high number of cells but a low number of field goals that are part of their prior distribution.

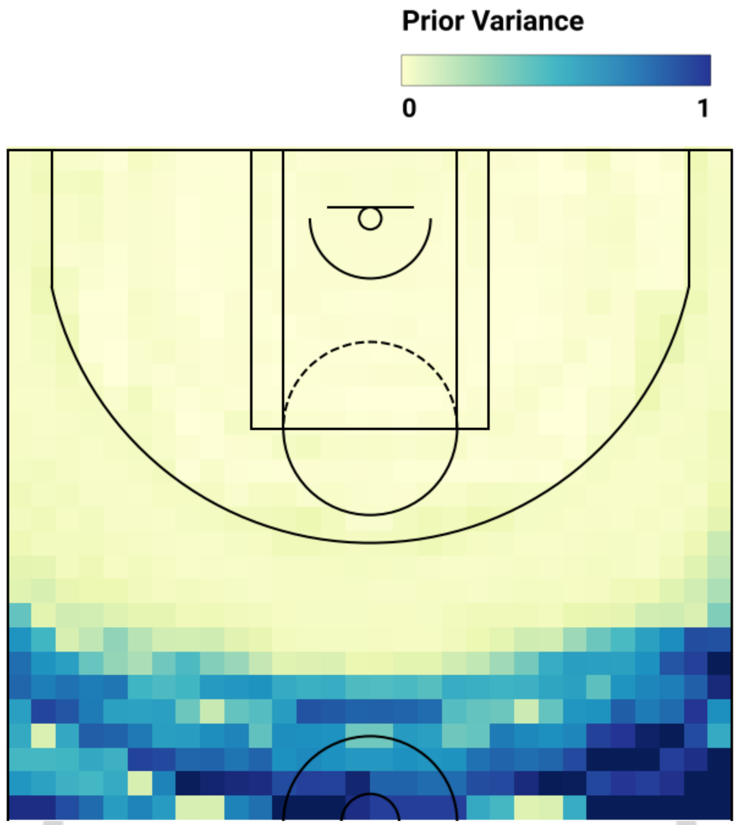


Figure 4.15

Map of prior variances per shooting cell

Figure 4.16 shows that shooting cells with low number of field goal attempts or high variance have low shrinking factors while those with high number of field goal attempts or low variance have high shrinking factors.

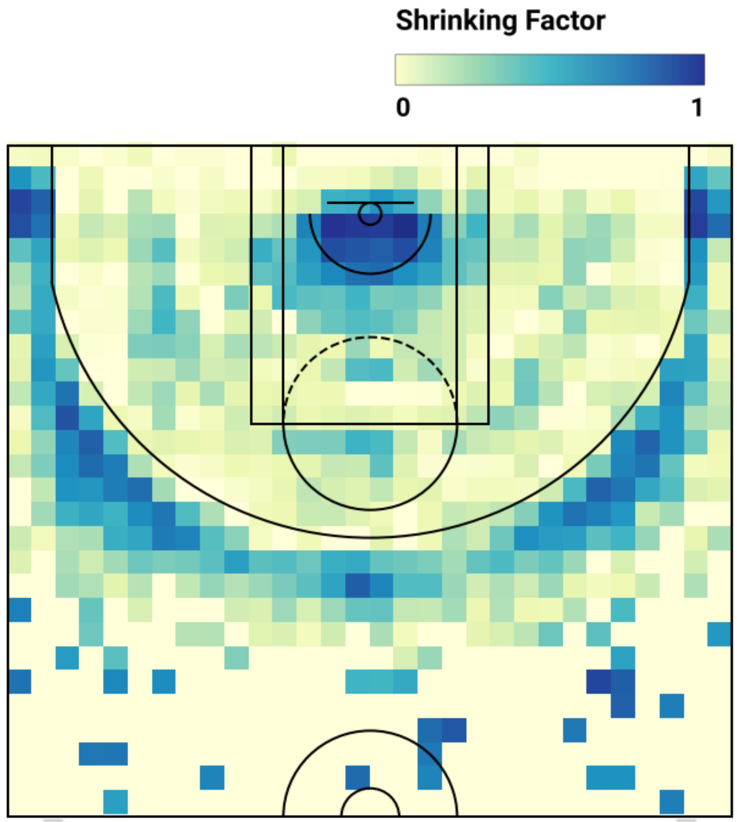


Figure 4.16

Map of shrinking factors per shooting cell

Finally, the Empirical Bayes estimate of the PPA was computed and mapped as shown in Figure 4.17 and Figure 4.18 shows the raw PPA map and the EB-estimated PPA map side-by-side. The EB-estimated PPA is smoother and less noisy than the raw PPA. There is also an apparent pattern in the PPA—high values near the basket and within the restricted area, an arc from 2m to 6m from the basket where the PPA values are low, another arc near the three-point line where the PPA values increase, and really low values beyond 10 meters from the basket.

Of the original 587 shooting cells with at least 1 field goal attempt, the difference between the raw PPA and the EB-estimated PPA were as much as 2.07 below and 0.70 above the raw PPA. These shooting cells with a high difference between the raw and estimated PPA are usually three-pointers with only 1 attempt and 1 made shot or areas near the basket with low number of attempts and made shots. All in all, the mean difference in the PPA (EB - raw) is -0.09 with more than 70% of the differences falling between -0.25 and 0.25.

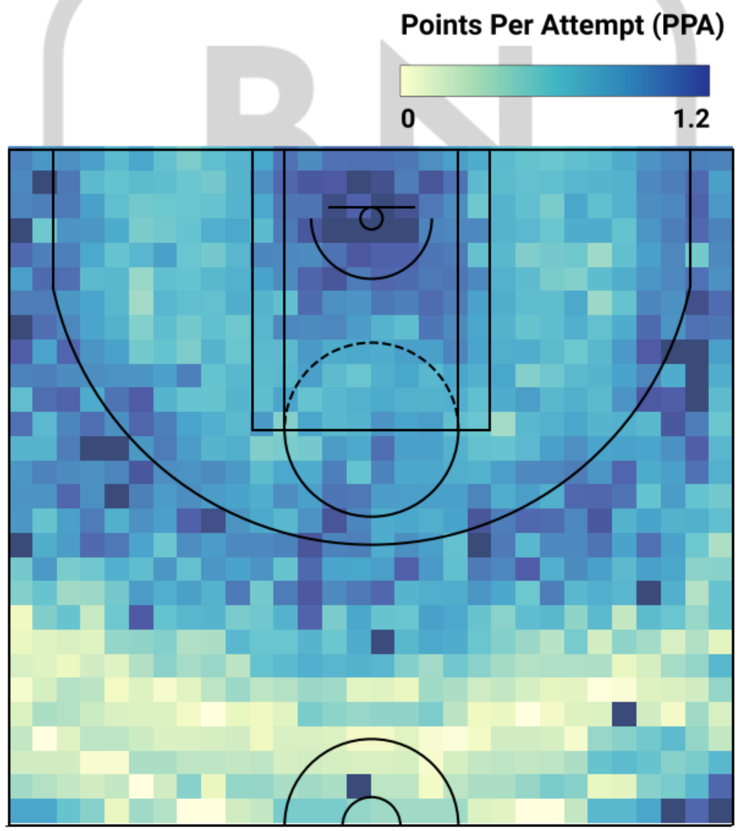
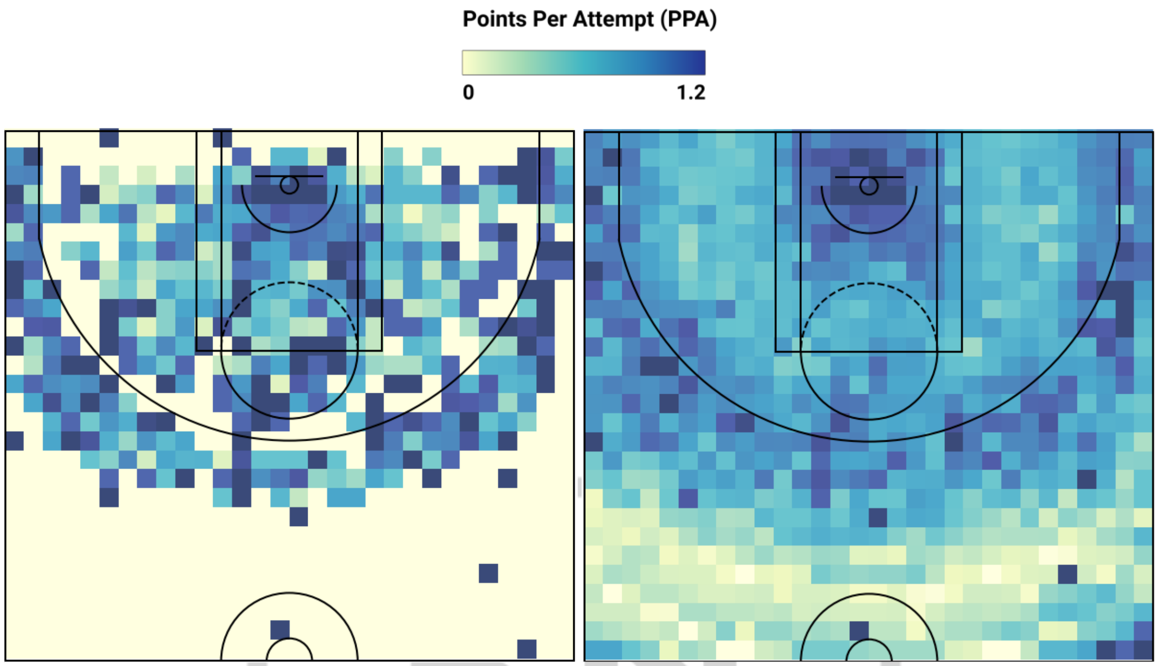


Figure 4.17

Empirical Bayes estimate of Points Per Attempt



A. Raw PPA **B. EB-estimated PPA**

Figure 4.18

Points Per Attempt for UAAP Season 81: (A) raw PPA; (B) Empirical Bayes-estimated PPA

The outputs of the Empirical Bayes estimation are found at:

https://github.com/benhur07b/ms-thesis-spatial-analysis-shooting-philippine-basketball/tree/main/outputs/ebppa

4.6 Player analysis

The EB-estimated PPA is used to compute for the Spatial Scoring Effectiveness (SScE) and Points Relative to League Average (PRLA).

Table 4.5 shows the shooting statistics of players with FGA >= 28 for UAAP Season 81. The range of values for PPA is considerably wider than that of EPPA but their means are similar. SScE ranges almost 1 point per shot showing considerable variability in shooting effectiveness of the players even when considering the locations of their shots. The standard deviation of the total points scored by the players is 54 while the standard deviation of the difference between their total and expected points (PRLA) is just 12 suggesting that the variability in points scored can be attributed to differences in the location of their field goal attempts.

Table 4.5

UAAP Season 81 shooting statistics for players with FGA >= 28

	PPA	EPPA	SSE	PTS	PRLA
Minimum	0.468	0.813	-0.400	18.000	-26.767
Mean	0.884	0.917	0.034	78.134	-1.046
Maximum	1.451	1.101	0.377	273.000	48.011
Standard deviation	0.180	0.059	0.159	54.187	11.890

Before looking at the SScE of players, a good metric to look at first is the Estimated Points Per Attempt (EPPA). EPPA is an indicator of how easy or difficult a shooting constellation is—high EPPA values indicate that a player takes shots from areas on the court that are, on average, easy to score from while low EPPA values indicate the opposite.

Table 4.6 shows the players with the ten highest and ten lowest EPPA values for UAAP Season 81. A noteworthy observation is that half of the players with high EPPA values are listed as Centers—a position that traditionally plays inside the paint/restricted area and takes high percentage shots near the basket. Only 2 of the players with the highest EPPA are listed as guards. This is a striking contrast with the players with the ten lowest EPPA—8 of whom are listed are guards. It is intriguing that two centers are among the ten players with the lowest EPPA but this shouldn't be surprising if we consider the spatial distribution of these two players' field goals. Both I. Gaye and E. Caunan took a considerable amount of their field goals outside of the paint which is different from the shooting tendencies of fellow Centers such as P. Orizu, B. Akhuetie, and A. Kouame

BN HR

Table 4.6

UAAP Season 81 players with highest and lowest EPPA values

Highest					Lowest				
	Player	**Team**	**Pos**	**EPPA**		**Player**	**Team**	**Pos**	**EPPA**
1	P. Orizu	FEU	C	1.101	82	J. Gallego	NU	SG	0.813
2	B. Akhuetie	UP	C	1.084	81	S. Belangel	ADMU	PG	0.821
3	A. Kouame	ADMU	C	1.066	80	T. Tio	ADMU	G	0.839
4	S. Akomo	UST	C	1.059	79	J. Espeleta	ADU	SG	0.844
5	B. Ebona	FEU	C	1.051	78	J. Ahanmisi	ADU	SG	0.845
6	F. Jaboneta	UP	SF	1.021	77	A. Melecio	DLSU	SG	0.847
7	D. Murrell	UP	PF	1.020	76	I. Gaye	NU	C	0.851
8	C. Cansino	UST	SG	1.001	75	J. Varilla	UE	G	0.858
9	A. Joson	NU	PG	0.991	74	E. Caunan	UST	C	0.858
10	V. Magbuhos	ADU	PF	0.987	73	K. Zamora	UST	SG	0.865

Table 4.7 shows the ten players with the highest and lowest PPA. We now see a lot more Guards, in fact five of them, in the list of players with the highest PPA in the league. This indicates that even though the field goal distribution of these Guards estimate that they should score less, they are actually performing better than expected. Meanwhile, we still have a good mix of Centers, Forwards, and Guards in the list of players with the lowest PPA.

Table 4.7

UAAP Season 81 players with highest and lowest PPA values

Highest					**Lowest**				
	Player	**Team**	**Pos**	**PPA**		**Player**	**Team**	**Pos**	**PPA**
1	P. Orizu	FEU	C	1.451	82	A. Wong	ADMU	G	0.468
2	J. Go	DLSU	SG	1.297	81	E. Mojica	ADU	SG	0.500
3	D. Dario	UP	G	1.257	80	K. Zamora	UST	SG	0.513
4	A. Kouame	ADMU	C	1.220	79	R. Acuno	UE	C	0.643
5	B. Ebona	FEU	C	1.216	78	I. Batalier	UST	GF	0.645
6	B. Akhuetie	UP	C	1.198	77	W. Navarro	ADMU	F	0.652
7	K. Tuffin	FEU	SF	1.158	76	M. Aquino	NU	C	0.667
8	A. Asistio	ADMU	G	1.101	75	K. Montalbo	DLSU	PG	0.672
9	G. Mamuyac	ADMU	SG	1.094	74	C. Catapusan	ADU	PF	0.674
10	J. Ahanmisi	ADU	SG	1.093	73	J. Gallego	NU	SG	0.677

Table 4.8 shows the ten players with the highest and lowest SScE. These are the players who over-performed or under-performed what was expected of their individual field goal distributions. The table also shows the 95% confidence interval computed using the weighted paired t-test of SScE. Most of the players with the highest and lowest SScE have low field goal attempts. Ten players—3 in the highest and 7 in the lowest—have less than 50 FGA. Only 2 of the 20 players listed have more than 100 FGA. This could explain the large confidence intervals, most of which include 0. In the top 10 and bottom 10 players, only two from the top list had 95%

confidence intervals above zero while only six from the bottom list had 95% confidence intervals below zero. All in all, out of the 82 players who were part of the analysis, 73 had confidence intervals that spanned zero, two had positive SScE with confidence intervals entirely above zero, and seven had negative SScE with confidence intervals entirely below zero. The intervals are relatively wide with a lot of overlap between the players. The results of weighted paired two one-sided T-tests (TOST) on the local SScE of the players also agree with this observation. At $p = 0.05$, the hypothesis that the SScE of a player is lower than 0 could be rejected for only six players. This includes the two players with entirely positive SScE confidence intervals. Meanwhile, at $p = 0.05$, the hypothesis that the SScE of a player is higher than 0 could be rejected for 12 players including all the seven players whose SScE had entirely negative confidence intervals.

Table 4.9 shows the top SScE of players with FGA >= 100. This list includes most of the top shot takers for each team shown in Table 4.2. For this list, the intervals are relatively narrower even though most of them still overlap. This suggests that adding more FG data might improve the accuracy of the SScE computation—e.g. using data from multiple seasons or multiple leagues.

In terms of the range of SScE values, these also become narrower as more field goals are attempted. For players with at least 50 FGA: 40% (23/58) had SScE values from -0.05 to 0.05, 55% (32/58) had SScE values from -0.1 and 0.1, and 84% (49/58) had SScE values between -0.2 and 0.2. For players with at least 70 FGA: 46%

(19/41) had SScE values from -0.05 to 0.05, 61% (25/41) had SScE values from -0.1 and 0.1, and 90% (37/41) had SScE values from -0.2 and 0.2. For players with at least 100 FGA: 50% (12/24) had SScE values from -0.05 to 0.05, 71% (17/24) had SScE values from -0.1 and 0.1, and 96% (23/24) had SScE values between -0.2 and 0.2. This indicates that as a player takes more shots, it becomes increasingly less likely for the player to score 0.2 points more or less than expected.

Table 4.8

UAAP Season 81 players with highest and lowest SScE (FGA >= 28)

Highest (FGA >= 28)							
	Player	**Team**	**Pos**	**FGA**	**EPPA**	**PPA**	**SScE (95% CI)**
1	J. Go	DLSU	SG	37	0.920	1.297	0.38 (0.821, -0.062)
2	D. Dario	UP	G	35	0.900	1.257	0.356 (0.833, -0.120)
3	P. Orizu	FEU	C	51	1.101	1.451	0.352 (0.623, 0.081)
4	J. Ahanmisi	ADU	SG	194	0.845	1.093	0.247 (0.427, 0.068)
5	K. Tuffin	FEU	SF	57	0.948	1.158	0.210 (0.526, -0.105)
6	A. Tolentino	FEU	PF	117	0.892	1.085	0.194 (0.422, -0.035)
7	A. Asistio	ADMU	G	89	0.907	1.101	0.193 (0.464, -0.078)
8	G. Mamuyac	ADMU	SG	32	0.914	1.094	0.177 (0.603, -0.249)
9	J. Lastimosa	ADU	PG	81	0.889	1.062	0.172 (0.431, -0.088)
10	B. Ebona	FEU	C	51	1.051	1.216	0.167 (0.466, -0.132)

Table 4.8 (cont)

Lowest (FGA >= 28)							
	Player	**Team**	**Pos**	**FGA**	**EPPA**	**PPA**	**SScE (95% CI)**
82	A. Wong	ADMU	G	47	0.868	0.468	-0.400 (-0.125, -0.676)
81	E. Mojica	ADU	SG	36	0.878	0.500	-0.378 (-0.029, -0.727)
80	K. Zamora	UST	SG	76	0.865	0.513	-0.352 (-0.112, -0.593)
79	C. Catapusan	ADU	PF	43	0.986	0.674	-0.310 (-0.003, -0.617)
78	F. Jaboneta	UP	SF	31	1.021	0.742	-0.279 (0.161, -0.718)
77	R. Acuno	UE	C	28	0.900	0.643	-0.255 (0.097, -0.607)
76	I. Batalier	UST	GF	31	0.879	0.645	-0.233 (0.121, -0.586)
75	W. Navarro	ADMU	F	66	0.879	0.652	-0.228 (-0.001, -0.455)
74	S. Ildefonso	NU	PF	90	0.944	0.722	-0.224 (-0.042, -0.405)
73	M. Maloles	UE	G	42	0.908	0.690	-0.218 (0.136, -0.573)

Table 4.9

SScE of UAAP Season 81 players with FGA >= 100

FGA >= 100							
	Player	**Team**	**Pos**	**FGA**	**EPPA**	**PPA**	**SScE (95% CI)**
1	J. Ahanmisi	ADU	SG	194	0.845	1.093	0.247 (0.427, 0.068)
2	A. Tolentino	FEU	PF	117	0.892	1.085	0.194 (0.422, -0.035)
3	A. Kouame	ADMU	C	132	1.066	1.22	0.155 (0.313, -0.003)
4	B. Akhuetie	UP	C	177	1.084	1.198	0.116 (0.276, -0.045)
5	Ju. Gomez de Liaño	UP	G	180	0.888	0.994	0.107 (0.276, -0.062)
6	A. Melecio	DLSU	SG	203	0.847	0.926	0.078 (0.244, -0.088)
7	J. Clemente	NU	SG	165	0.919	0.976	0.057 (0.237, -0.123)
8	P. Desiderio	UP	SG	175	0.921	0.960	0.04 (0.224, -0.145)
9	J. Baltazar	DLSU	C	149	0.929	0.966	0.038 (0.207, -0.131)
10	H. Cani	FEU	G	120	0.874	0.908	0.034 (0.235, -0.166)
11	M. Lee	UST	PG	175	0.878	0.909	0.03 (0.205, -0.144)
12	D. Ildefonso	NU	SF	195	0.913	0.938	0.026 (0.184, -0.132)
13	S. Manganti	ADU	SF	172	0.892	0.913	0.021 (0.185, -0.144)
14	L. Santillian	DLSU	PF	141	0.898	0.908	0.008 (0.184, -0.167)
15	C. Cansino	UST	SG	115	1.001	1.009	0.007 (0.243, -0.229)
16	W. Comboy	FEU	G	119	0.874	0.857	-0.017 (0.209, -0.244)
17	A. Pasaol	UE	F	294	0.952	0.929	-0.023 (0.098, -0.143)
18	R. Subido	UST	PG	197	0.867	0.843	-0.023 (0.148, -0.195)

19	J. Varilla	UE	G	111	0.858	0.82	-0.038 (0.171, -0.247)
20	A. Caracut	DLSU	PG	127	0.877	0.819	-0.059 (0.134, -0.252)
21	T. Ravena	ADMU	F	139	0.936	0.871	-0.066 (0.122, -0.253)
22	Z. Huang	UST	F	101	0.893	0.802	-0.09 (0.104, -0.285)
23	P. Sarr	ADU	C	146	0.897	0.74	-0.158 (-0.006, -0.311)
24	P. Manalang	UE	PG	135	0.935	0.763	-0.172 (0.018, -0.363)

Figure 4.19 shows the relationship between EPPA and SScE for all 82 players in UAAP Season 81 with at least 28 field goal attempts. The size of the circles indicate the number of field goals by the players. Most of the players with a high number of FGA (>=100) cluster near the center of the graph indicating that the more players shoot, the more they tend to score as expected. L. Santillan earns the distinction of having the PPA (0.908) that is most similar to his EPPA (0.898) with an SScE of just 0.01 in 144 FGA.

Players who scored more than expected are on the upper half of the graph while those who scored less are in the bottom half. Players in the upper right quadrant are those with already high expected points but still manage to score beyond that. These are players who have good shot selection and good shooting percentages. An example in the chart is B. Akhuetie with an EPPA of 1.084 and SScE of 0.114 in his 177 FGA. The top 5 players with the highest EPPA, all of whom are Centers, as shown in Table 4.6 are all in this quadrant.

Those in the upper left are players with low expected points based on their field goal distribution but manage to score more than expected. The players in this quadrant take difficult shots—maybe because they were forced by the defense, they had poor shot selection, or they just tend to take shots where other players struggle—but still manage to make them. An example is J. Ahanmisi with an EPPA of just 0.845 in 212 FGA but an SScE of 0.248 which is 30% of his expected points per attempt.

The players in the lower right quadrant are players with high expected points but failed to score as expected. This indicates that these players are already taking field goals at the right locations on the court—i.e. good shot selection—and they just need to work on making the shots they take. An example is F. Jaboneta with an EPPA of 1.021 but an SScE of -0.279 in his 31 FGA.

The lower left quadrant is the quadrant of woe. This includes players with low expected points—i.e. difficult field goal attempts distribution—and negative SScE. Players in this quadrant should work both on taking better field goals and making them. An example is K. Zamora with an EPPA of 0.865 and an SScE of -0.352 over 76 FGA.

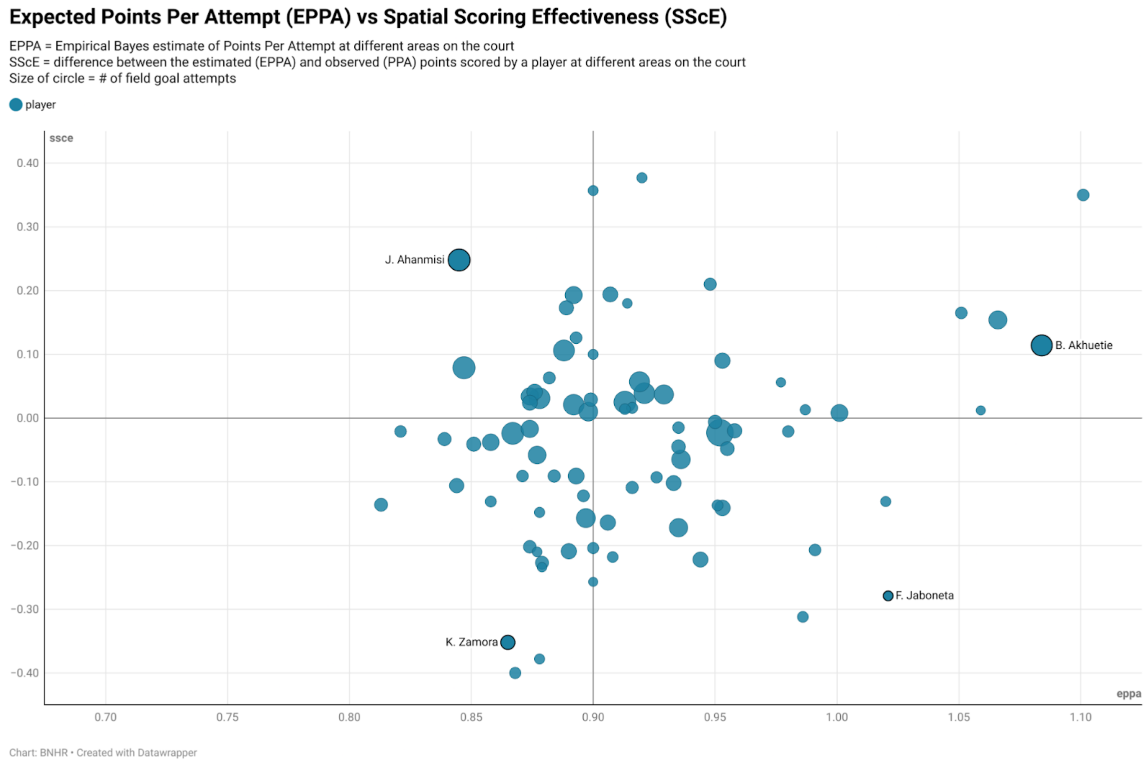


Figure 4.19

Chart of Expected Points Per Attempt vs Spatial Scoring Effectiveness, UAAP Season 81. Selected players are labeled.

An online version of the chart can be found here: https://datawrapper.dwcdn.net/u3gMr/1/

Similar to SScE, PRLA also compares expected and observed values but instead of points per attempt, PRLA compares the expected points scored by a player based on his field goal distribution against his actual points scored. SScE and PRLA are interrelated such that a player with positive SScE will have positive PRLA while a player with negative SScE will have negative PRLA. However, because PRLA isn't a rate statistic like SScE but a counting statistic, it is affected by the number of field goal attempts by a player. Between two players with the same SScE, the player with more FGA will have a higher or lower PRLA depending on whether the SScE is positive or negative. PRLA is essentially a raw count of how many points a player scored above or below expected over the course of the season.

Table 4.10 shows the top 10 players (FGA >= 28) with the highest and lowest PRLA. J. Ahanmisi and A. Tolentino had the two highest PRLA owing to their high SScE and FGA—the same players are ranked 1st and 2nd in SScE for players with FGA>=100. All five players with the highest PRLA and two of the five players with the lowest PRLA have FGA>=100 showing how PRLA is greatly affected by the number of FGA by a player. Of note are J. Go (9th highest PRLA) and E. Mojica (9th lowest PRLA) as both players managed to be part of the highest and lowest PRLA list even though they only took 37 and 36 FGA respectively. In the case of J. Go, this means that he was able to make the most of his small number of FGA while it was the opposite for E. Mojica.

BN HR

The results of paired, weighted two one-sided T-test (TOST) on the local PRLA values of the players showed that at $p = 0.05$, the hypothesis that a player's PRLA is below zero could be rejected for 14 players including 8 from the top 10 list of Table 4.10 (A. Melecio at J. Lastimosa were not included); while the hypothesis that a player's PRLA is above zero could be rejected for 14 other players including 8 from the bottom 10 list of Table 4.10 (S. Ildefonso and R. Escoto were not included).

Table 4.10

UAAP Season 81 players with highest and lowest PRLA (FGA >= 28)

Highest (FGA >= 28)							
	Player	**Team**	**Pos**	**FGA**	**EPTS**	**PTS**	**PRLA**
1	J. Ahanmisi	ADU	SG	194	164	212	48
2	A. Tolentino	FEU	PF	117	104	127	23
3	A. Kouame	ADMU	C	132	141	161	20
4	B. Akhuetie	UP	C	177	192	212	20
5	Ju. Gomez de Liaño	UP	G	180	160	179	19
6	P. Orizu	FEU	C	51	56	74	18
7	A. Asistio	ADMU	G	89	81	98	17
8	A. Melecio	DLSU	SG	203	172	188	16
9	J. Go	DLSU	SG	37	34	48	14
10	J. Lastimosa	ADU	PG	81	72	86	14

Table 4.10 (cont.)

Lowest (FGA >= 28)							
	Player	**Team**	**Pos**	**FGA**	**EPTS**	**PTS**	**PRLA**
82	K. Zamora	UST	SG	76	66	39	-27
81	P. Manalang	UE	PG	135	126	103	-23
80	P. Sarr	ADU	C	146	131	108	-23
79	S. Ildefonso	NU	PF	90	85	65	-20
78	J. Parker	FEU	PG	94	84	64	-20
77	A. Wong	ADMU	G	47	41	22	-19
76	R. Escoto	FEU	PF	93	84	69	-15
75	W. Navarro	ADMU	F	66	58	43	-15
74	E. Mojica	ADU	SG	36	32	18	-14
73	J. Manzo	UP	PG	96	91	78	-13

Aside from the global SScE and PRLA values, other ways to summarize and describe shooting relative to the spatial distribution of field goal attempts include Spread and the different Range metrics introduced in 3.4.4. Table 4.11 shows the 5 players with the highest and lowest Spread among players with at least 100 FGA. Table 4.12 shows the 5 players with the highest and lowest Effective Range (ERNG) and Total Effective Range % (TERNG) for the same group of players. Table 4.13 shows the 5 players with the highest and lowest Net Effective Range (NERNG) and Player Effective Range % (PERNG) also for the same group of players.

As expected, the players with the highest number of FGA also had the highest Spread (the player with the 4th most FGA, D. Ildefonso, was ranked 6th in Spread) as shown in Table 4.11. The five players on the top Spread list took a lot of shots and took them from a lot of locations on the court. This is also supported by the popular opinion about these players. All of them are considered scorers who can take the outside shot, drive to the basket, or pull-up from mid-range. In contrast, the list of players with the least Spread includes B. Akhuetie and P. Sarr who were ranked 7th and 13th respectively in terms of their number of FGA. Their inclusion on this list, as with that of A. Kouame, can be attributed to their position and style of play. All three players are Centers who take a majority of their shots near the basket. Figure 4.20 shows the correlation between FGA and Spread. There was almost perfect linear correlation between FGA and Spread (Pearson correlation coefficient *$r = 0.94$, p-value at near zero value*) aside from some outlying players labeled on the graph.

Table 4.11

UAAP Season 81 players (FGA >= 100) who have the highest and lowest number of shooting cells with at least 1 FGA (Spread).

Highest (FGA >= 100)					
	Player	**Team**	**Pos**	**FGA**	**Spread**
1	A. Pasaol	UE	F	294 (1st)	151
2	R. Subido	UST	PG	197 (3rd)	144
3	J. Ahanmisi	ADU	SG	194 (5th)	143
4	A. Melecio	DLSU	G	203 (2nd)	142
5	Ju. Gomez de Liaño	UP	G	180 (6th)	124

Lowest (FGA >= 100)					
	Player	**Team**	**Pos**	**FGA**	**Spread**
24	A. Kouame	ADMU	C	132 (17th)	51
23	B. Akhuetie	UP	C	177 (7th)	54
22	C. Cansino	UST	SG	115 (22nd)	65
21	Z. Huang	UST	F	101 (24th)	71
20	P. Sarr	ADU	C	146 (13th)	74

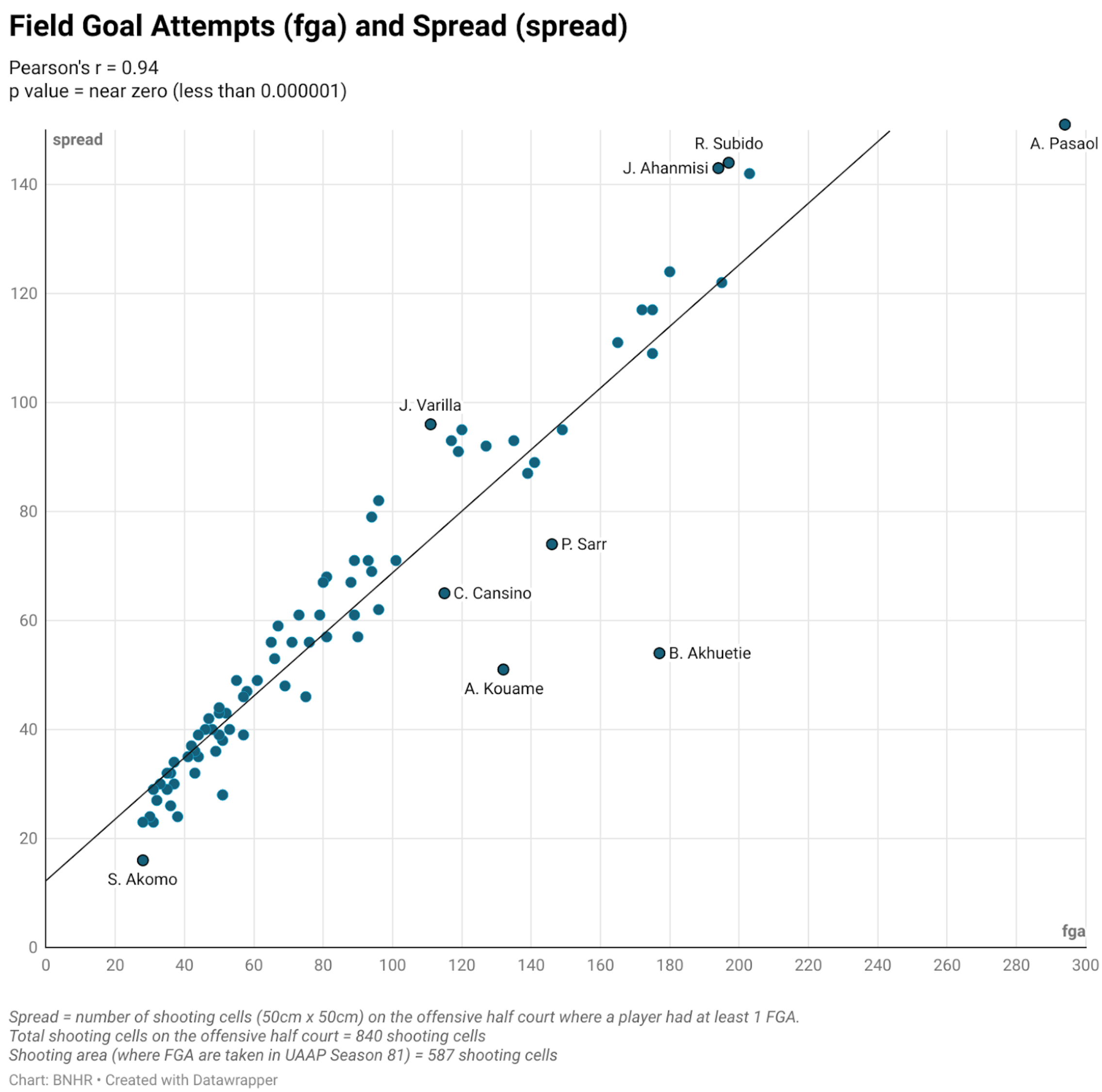


Figure 4.20

Field Goal Attempts (FGA) and Spread for UAAP Season 81. Selected outlying players are labeled.

As with FGA and Spread, most of the players with the highest and lowest Spread also had the highest and lowest Effective Range and Total Effective Range % metrics as shown in Table 4.12. Four of the five players with the highest Spread are included in the list of players with the five highest ERNG/TERNG. R. Subido, ranked 2nd in Spread, is ranked 6th in ERNG/TERNG. Meanwhile, all the players with the lowest Spread also had the lowest ERNG/TERNG. The assumption is that if a player takes field goals at a lot of areas on the court then the chance of him being effective at more areas on the court increases. This assumption is validated in Figure 4.21 which shows that there was linear correlation between Spread and TERNG (Pearson correlation coefficient *r = 0.95, p-value at near zero value*). Outlying players are once again labeled in the graph.

BN HR

Table 4.12

UAAP Season 81 players (FGA >= 100) who have the highest and lowest number of shooting cells where they score more than expected (SScE > 0); ERNG = count of cells where SScE > 0; TERNG = ERNG / number of cells in the scoring area (587) .

Highest (FGA >= 100)						
	Player	**Team**	**Pos**	**Spread**	**ERNG**	**TERNG**
1	J. Ahanmisi	ADU	SG	143 (3rd)	72	12.27
2	A. Melecio	DLSU	PG	142 (4th)	57	9.71
3	A. Pasaol	UE	F	151 (1st)	56	9.54
4	Ju. Gomez de Liaño	UP	G	124 (5th)	52	8.86
5	D. Ildefonso	NU	SF	122 (6th)	49	8.35
5	J. Clemente	NU	SG	111 (9th)	49	8.35

Lowest (FGA >= 100)						
	Player	**Team**	**Pos**	**Spread**	**ERNG**	**TERNG**
24	B. Akhuetie	UP	C	54 (23rd)	26	4.43
23	C. Cansino	UST	PG	65 (22nd)	27	4.60
22	P. Manalang	UE	PG	93 (15th)	27	4.69
21	P. Sarr	ADU	C	74 (20th)	28	4.77
20	A. Kouame	ADMU	C	51 (24th)	29	4.94
20	Z. Huang	UST	PF	71 (21st)	29	4.94

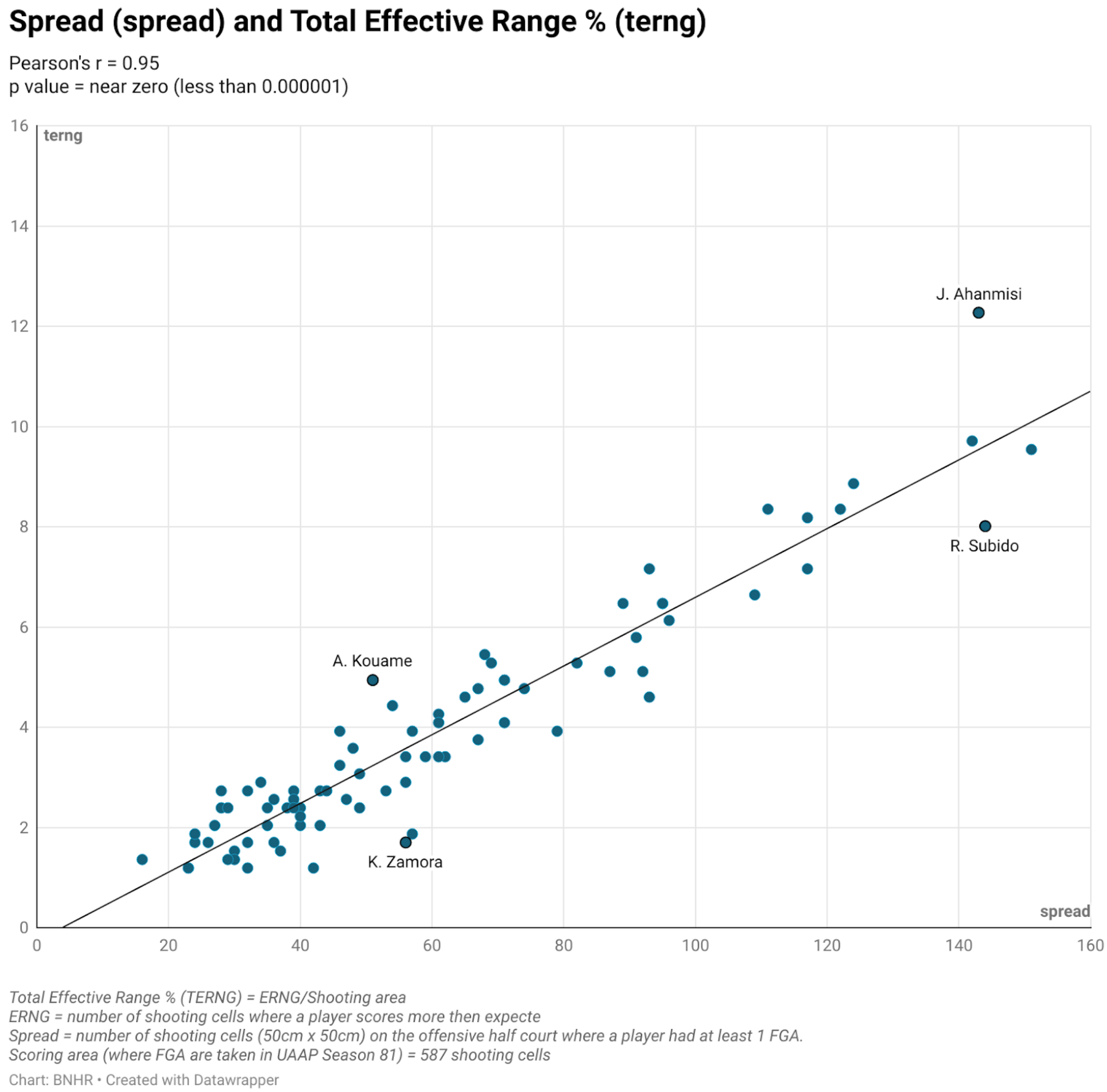


Figure 4.21

Spread and Total Effective Range % (TERNG) for UAAP Season 81. Selected outlying players are labeled.

By themselves, ERNG/TERNG only tell us how many locations on the court a player scores more than expected. A player with high ERNG/TERNG is someone who scores effectively at different locations on the court but it does not tell us whether or not a player is an effective scorer based on his field goal attempt distribution. For that, we need to look at NERNG and PERNG. Compared to ERNG/TERNG, Figure 4.22 shows that there was some negative correlation, although not perfect, between Spread and NERNG (Pearson correlation coefficient *r = -0.58, p-value at near zero value*). This means that players who took field goals at more locations were effective at less locations. In Table 4.13, the three players with the lowest Spread are in the top 5 list of NERNG while the two players with the highest Spread are in the bottom 5 list of NERNG. Of note is J. Ahanmisi who had the third highest but still managed to get a positive NERNG (+1, 2nd). This means that even though he took FGA in a lot of areas on the court, he was still able to score better than expected in more than half of those areas.

Meanwhile, Figure 4.23 shows that there is no evidence to support the conclusion that there was correlation between Spread and PERNG (Pearson correlation coefficient *r = 0.07, p-value = 0.51*). This suggests that taking shots at more locations on the court does not necessarily equate to being a more efficient or less efficient scorer.

Table 4.13

UAAP Season 81 players (FGA >= 100) who have the highest and lowest proportion of shooting cells where they score more than expected (SScE > 0); NERNG = ERNG - count of cells where SScE <= 0; PERNG = ERNG / Spread .

Highest (FGA >= 100)						
	Player	**Team**	**Pos**	**Spread**	**NERNG**	**PERNG**
1	A. Kouame	ADMU	C	51 (24th)	7	56.86
2	J. Ahanmisi	ADU	SG	143 (3rd)	1	50.35
3	B. Akhuetie	UP	C	54 (23rd)	-2	48.15
4	A. Tolentino	FEU	PF	93 (14th)	-9	45.16
5	C. Cansino	UST	PG	65 (22nd)	-11	41.54

Lowest (FGA >= 100)						
	Player	**Team**	**Pos**	**Spread**	**NERNG**	**PERNG**
24	R. Subido	UST	PG	144 (2nd)	-50	32.64
23	P. Manalang	UE	PG	93 (15th)	-39	29.03
22	A. Pasaol	UE	F	151 (1st)	-39	37.09
21	M. Lee	UST	PG	117 (8th)	-33	35.90
20	A. Caracut	DLSU	PG	92 (16th)	-32	32.61

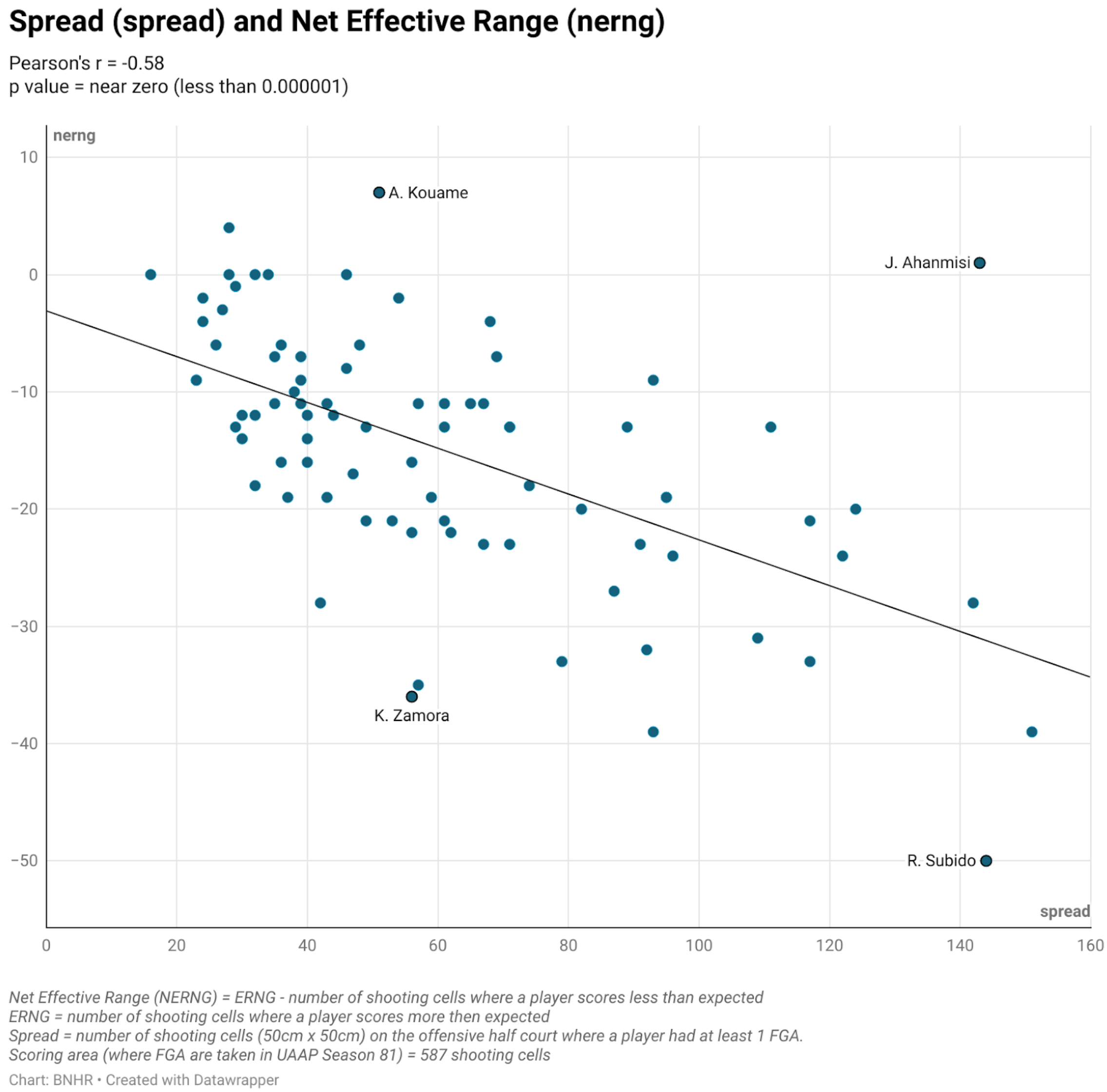


Figure 4.22

Spread and Net Effective Range (NERNG) for UAAP Season 81. Selected outlying players are labeled.

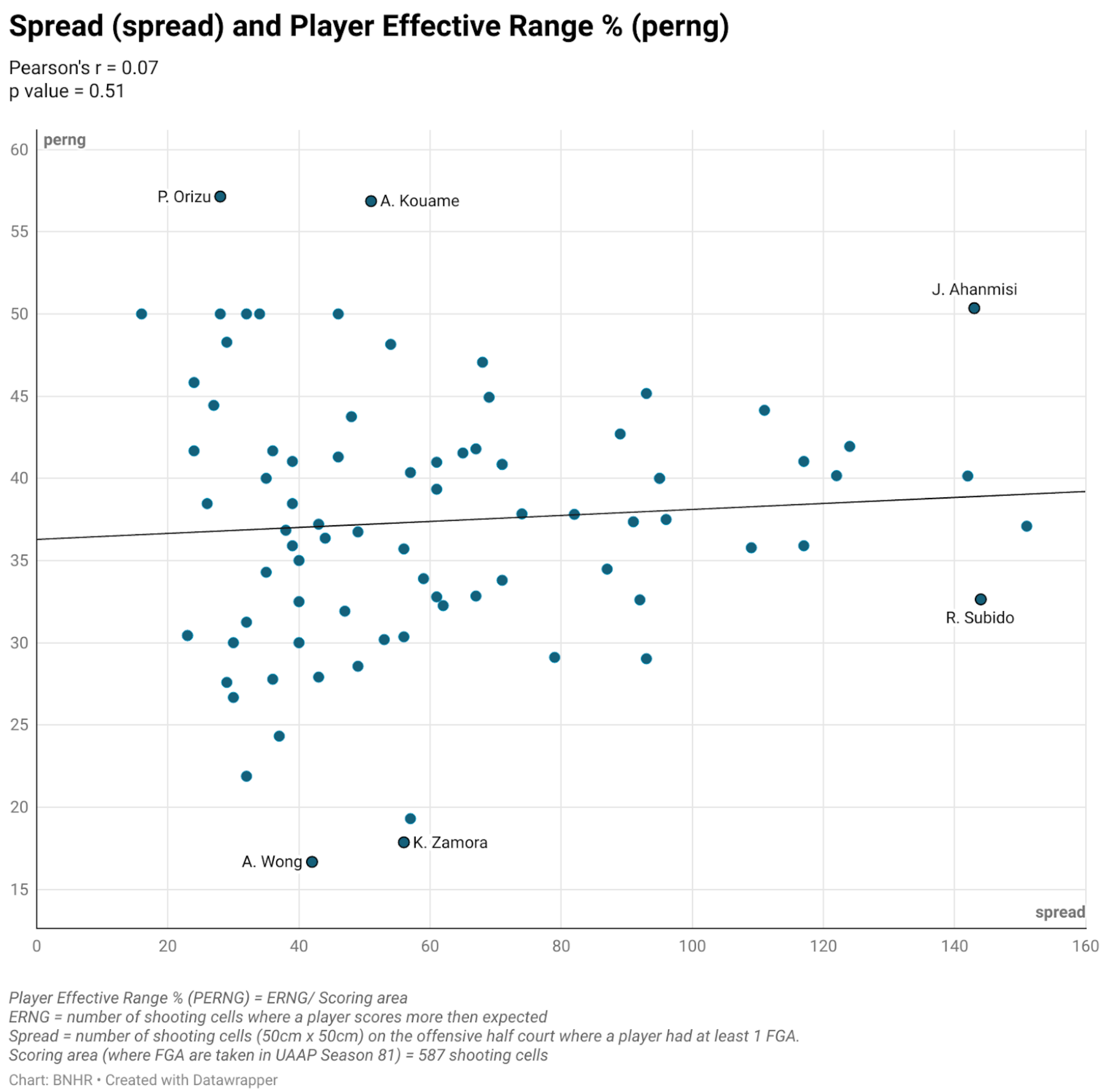


Figure 4.23

Spread and Player Effective Range % (PERNG) for UAAP Season 81. Selected outlying players are labeled.

4.6.1 Was J. Ahanmisi the best shooter/scorer during UAAP Season 81?

The case of J. Ahanmisi is an interesting one.

J. Ahanmisi's statistics for UAAP Season 81 are shown in Table 4.14. Of all the players with FGA >= 28, he had the 5th lowest EPPA (78th total) but the 10th highest PPA and the 4th highest SScE. He is one of only two players (FGA >= 28) with a confidence interval for their SScE that's entirely above zero. For players with FGA >= 100, he ranked last (24th) in EPPA but 3rd in PPA and 1st in SScE. He has the highest PRLA and the 3rd highest points scored even though he only had the 6th highest EPTS.

What this means is that even though he had one of the most difficult and toughest field goal attempt distributions (evidenced by his low EPPA), he still managed to score beyond what was expected with every shot he took. He was the only player with FGA >=100 who had an absolute value of SScE greater than 0.2 and a confidence interval that's entirely above zero. All the other players with FGA >= 100 scored more or less as expected based on their field goal distribution and had confidence intervals for their SScE that spanned zero. What's even more impressive is that J. Ahanmisi was able to do this both in terms of the volume and spatial distribution of his field goal attempts. He had 194 FGA (5th) and a 143 Spread (3rd) but still managed to be efficient with his shot. For all players with FGA >= 100, he ranked 3rd in PPA (1st among non-centers), 1st in Effective Range and Total Effective Range %, and 2nd in Net Effective Range and Player Effective Range % (1st among non-centers). In fact, he was the only player ranked in the top 3 of PPA, Spread, Effective Range, Net Effective Range, Player Effective Range %, and Total Effective Range %. This means that he took a lot of field goals, he took them from a lot of different locations on the court, and more often than not he was able to score more effectively than anyone else taking the same shot.

Table 4.15 shows a comparison between J. Ahanmisi and select players. W. Comboy and A. Caracut were identified as players with similar shooting habits in Section 4.4. A. Melecio had similar FGA, EPPA, Spread, ERNG, and TERNG values while B. Akhuetie had similar PPA, PTS, NERNG, and PERNG values. In terms of the spatial metrics introduced in the study, J. Ahanmisi has the best statistics among all the players while in terms of conventional statistics (FG%, eFG, PPA), he beats everyone except for B. Akhuetie.

BN HR

Table 4.14

Statistics and rank of J. Ahanmisi in UAAP Season 81

J. Ahanmisi (ADU) *UAAP Season 81*		**Rank (FG >= 28)** **82 players**	**Rank (FG >= 100)** **24 players**
FGA	194	5th	5th
EPPA	0.845	78th	24th
PPA	1.093	10th	3rd
SScE	0.247 (0.427, 0.068)	4th	1st
EPTS	164	6th	6th
PTS	212	3rd	3rd
PRLA	48	1st	1st
Spread	143	3rd	3rd
Effective Range	72	1st	1st
Net Effective Range	1	3rd	2nd
Player Effective Range %	50.35	3rd	2nd
Total Effective Range %	12.27	1st	1st

Table 4.15

Statistics of J. Ahanmisi and select players in UAAP Season 81

Player	Pos	FGA	EPPA	PPA	EPTS	PTS	SScE	PRLA
J. Ahanmisi	SG	194	0.845	1.093	164	212	0.247	48
W. Comboy	G	119	0.874	0.857	104	102	-0.017	-2
A. Caracut	PG	127	0.819	0.877	111	104	-0.059	-7
A. Melecio	SG	203	0.847	0.926	172	188	0.078	16
B. Akhuetie	C	177	1.084	1.198	192	212	0.114	20

Player	Pos	Spread	ERANGE	NERANGE	PERANGE	TERANGE
J. Ahanmisi	SG	143	72	1	50.35	12.27
W. Comboy	G	91	34	-23	37.36	5.79
A. Caracut	PG	92	30	-32	32.61	5.11
A. Melecio	SG	142	57	-28	40.14	9.71
B. Akhuetie	C	54	26	-2	48.15	4.43

Player	Pos	2P	2PA	2P%	3P	3PA	3P%	FG%	eFG
J. Ahanmisi	SG	46	99	46.5	40	95	42.1	44.3	54.6
W. Comboy	G	24	61	39.3	18	58	31.0	35.3	42.9
A. Caracut	PG	34	84	40.5	12	43	27.9	36.2	40.9
A. Melecio	SG	46	111	41.4	32	92	34.8	38.4	46.3
B. Akhuetie	C	106	174	59.9	0	3	0.0	59.9	59.9

BN HR

The global spatial metrics computed in the study support the conclusion that J. Ahanmisi was a good, effective, and efficient shooter and scorer—*if not the best, most effective, and most efficient one*—during UAAP Season 81 and the conventional statistics also seem to confirm this. However, these global spatial metrics and conventional statistics do not provide context about where on the court J. Ahanmisi was able to perform better or worse than the other players nor does it give us a spatial comparison of their shooting performance. By mapping the local values of SScE and PRLA, we can look at the spatial distribution of these metrics to see where J. Ahanmisi was able to differentiate himself from the other players.

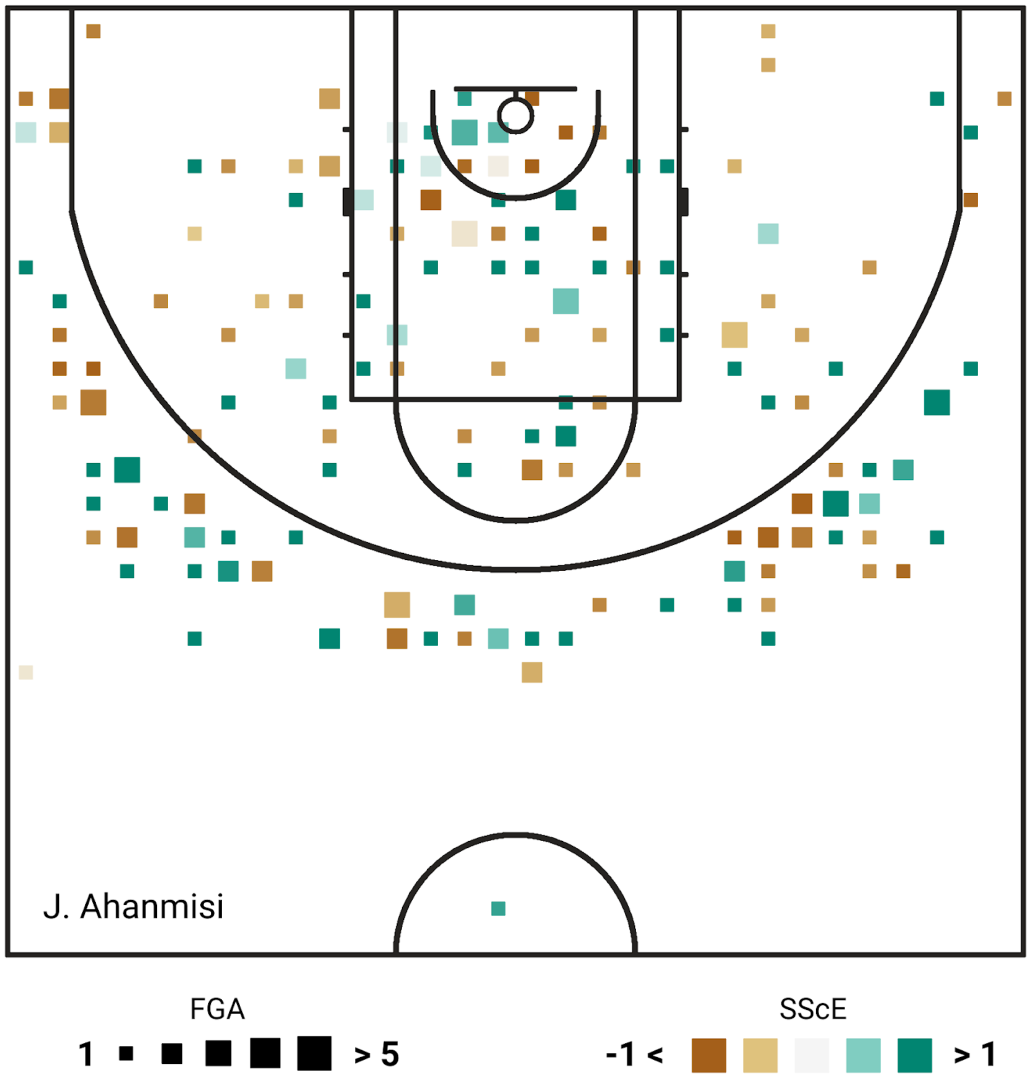
Figure 4.24 shows the local SScE maps of J. Ahanmisi, W. Comboy, A. Caracut, A. Melecio, and B. Akhuetie while Figure 4.25 shows the local PRLA of the same players.

J. Ahanmisi's SScE and PRLA maps show that he took shots at almost all areas on the court—at the rim, paint, mid-range, and three-pointers. This is similar to his neighbors/players with similar shooting habits—W. Comboy and A. Caracut—and also similar to fellow SG A. Melecio. What separates J. Ahanmisi from these three players is that he was an effective scorer at all these court areas. Near the basket (distance to basket <= 2m), J. Ahanmisi had a 1.13 PPA, 0.09 SScE, and 1.94 PRLA which was better than W. Comboy (0.69 PPA, -0.43 SScE, -11.38 PRLA), A. Caracut (0.94 PPA, -0.18 SScE, -5.97 PRLA), and A. Melecio (1.00 PPA, -0.03 SScE, -1.32 PRLA). For three-pointers, J. Ahanmisi had a 1.30 PPA, 0.41 SScE, and 38.86 PRLA which was also better than W. Comboy (0.93 PPA, 0.08 SScE, 4.74), A. Caracut (0.84 PPA, -0.03 SScE, -1.37 PRLA), and A. Melecio (1.05 PPA, 0.19 SScE, 17.72 PRLA). Even for mid-range shots (distance between 3m and 6m from the basket), J. Ahanmisi was still able to score above average with a 0.78 PPA, 0.08 SScE, and 4.04 PRLA which was on par with W. Comboy (0.81 PPA, 0.12 SScE, 3.31 PRLA) but higher than A. Caracut (0.60 PPA, -0.07 SScE, -2.07 PRLA) and A. Melecio (0.64 PPA, -0.03 SScE, -1.216 PRLA).

The difference between J. Ahanmisi and B. Akhuetie is apparent in their SScE and PRLA maps. Where J. Ahanmisi took field goals at almost all areas on the court, B. Akhuetie's field goals were more concentrated near the basket. B. Akhuetie took 147 FGA at locations <= 2m from the basket. This made up 83% of all his FGA. In comparison, J. Ahanmisi only took 30 FGA from this area which comprised just 15% of his total FGA. However, B. Akhuetie was very effective in that small area near the basket where he took his shots. For locations <= 2m from the basket, B. Akhuetie had a 1.28 PPA, 0.127 SScE, and 18.62 PRLA; all of which were higher than J. Ahanmisi for the same area (1.13 PPA, 0.09 SScE, 1.94 PRLA). B. Akhuetie was also effective in the mid-range albeit in a small sample size. Of the 13 FGA he took between 3m to 6m from the basket, he had a 0.77 PPA, 0.13 SScE, and 1.65 PRLA which was similar to the statistics of J. Ahanmisi (0.78 PPA, 0.08 SScE, 4.04 PRLA) from the same area. It should be noted that J. Ahanmisi took almost four times more shots from the area (49 FGA) than B. Akhuetie. For three-pointers, B. Akhuetie was an afterthought in Season 81. He only took 3 attempts and missed all of them for 0

PPA, -0.81 SScE, and -2.419 PRLA. Put another way, B. Akhuetie actually cost his team almost 1 point every time he took a three pointer in Season 81. This was in stark contrast with J. Ahanmisi's performance from deep that season.

If B. Akhuetie made his bread near the basket, J. Ahanmisi made his beyond the arc (or from locations 6.5 meters and more from the basket). J. Ahanmisi's 0.41 SScE for field goals taken 6.5 meters away or more from the basket meant that for every attempt he took at that distance, he was scoring almost half a point more than the average shooter for that season. He took 94 FGA from that area which was almost half of his total FGA and managed a 38.86 PRLA. He wasn't just a three-point specialist either. Of the 100 FGA he took within 6.5 meters from the basket, he had 0.90 PPA, 0.09 SScE, and 9.1 PRLA. Nevertheless, even if he missed all of his shots within 6.5 meters from the basket, he would still have a total of 29.76 PRLA and 0.153 SScE which would still rank him 1st and 3rd in those metrics respectively. Think about that for a second.


J. Ahanmisi
FGA
1
> 5
SScE
-1 <
> 1

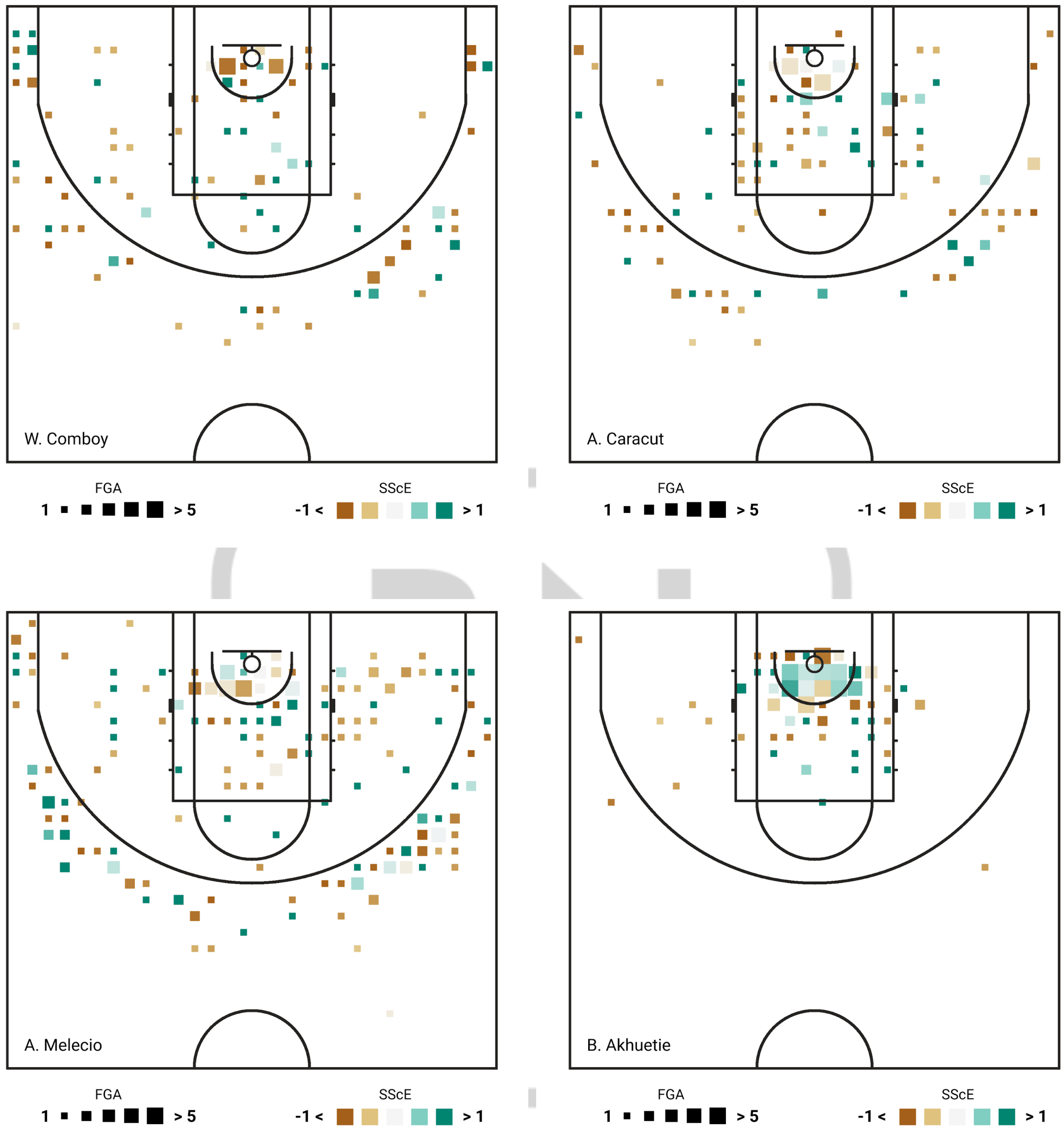
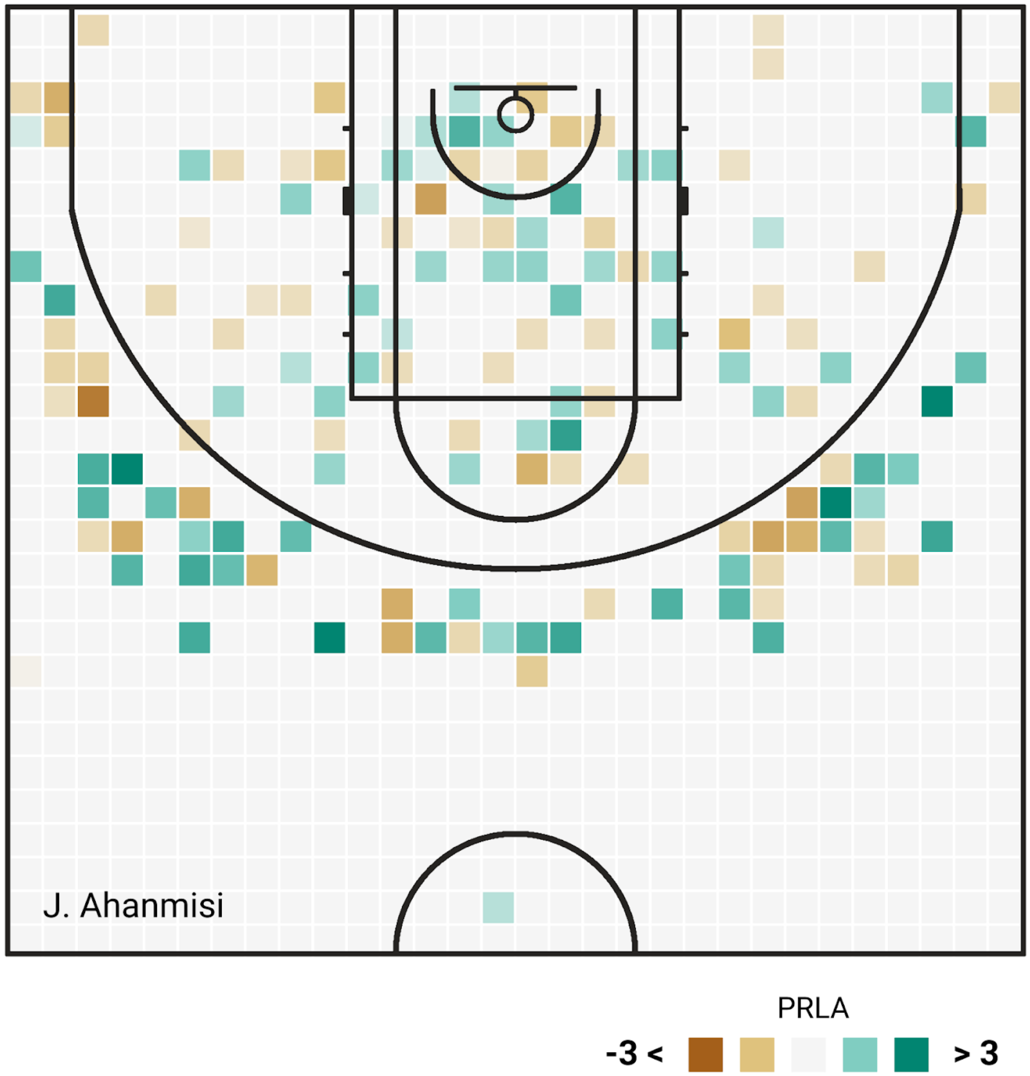


Figure 4.24

Local SScE map of J. Ahanmisi and select players. The size of the pixel indicates the number of field goal attempts in the cell; the color indicates how effective a player is at scoring in the cell, positive values = more effective scoring (Local SScE).


J. Ahanmisi
PRLA
-3 <
> 3

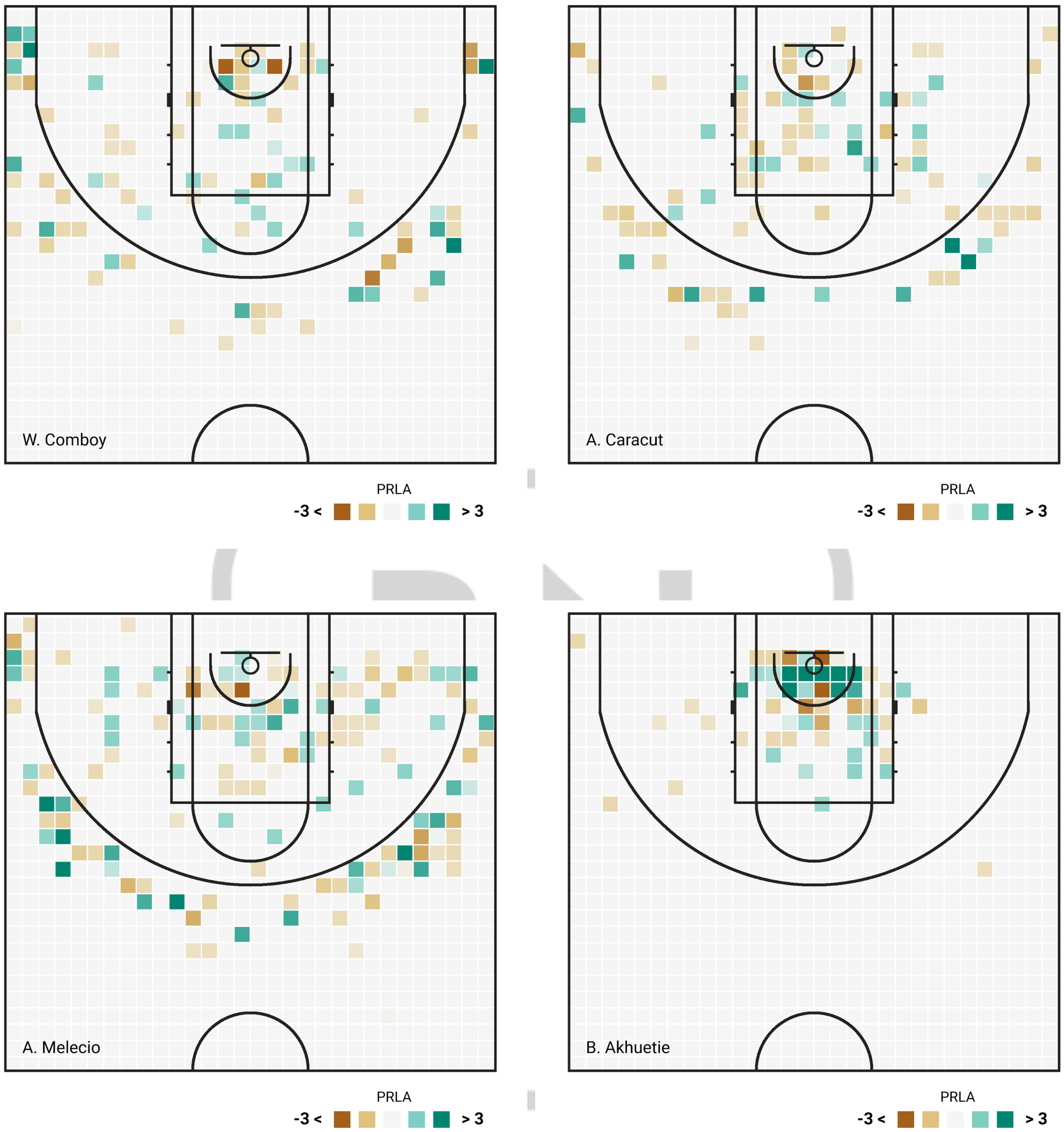


Figure 4.25

Local PRLA map of J. Ahanmisi and select players. The color indicates the number of points a player scores above or below the league average, positive values = more effective scoring (Local PRLA).

Using an unpaired weighted t-test, Ahanmisi's total SScE was found to be significantly higher than 46 of the other 81 players with FGA >= 28 at $p = 0.05$. For players with FGA >= 100 and at $p = 0.05$, Ahanmisi's SScE was found to be significantly higher than all players except for the 9 players below him on the list and #15 C. Cansino.

His total SScE was significantly higher than W. Comboy (*t-statistic=1.82, p-value=0.03*) and A. Caracut (*t-statistic=2.32 p-value=0.01*). However, comparing his total SScE with A. Melecio (*t-statistic=1.37 p-value=0.09)* and B. Akhuetie (*t-statistic=0.85 p-value=0.20)* found no support for the hypothesis that J. Ahanmisi's SScE was significantly higher than the two.

Ahanmisi's SScE for field goals near the basket was significantly higher than Comboy (*t-statistic=2.67, p-value=0.01*) but there was no support for the hypothesis that it was significantly higher than the other three players. For his SScE on three-pointers, it was found to be significantly higher than A. Caracut (*t-statistic=1.66, p-value=0.05*) and B. Akhuetie (*t-statistic=1.63, p-value=0.05*).

4.7 Team analysis

Aside from player performance, the metrics introduced in the study can also be used to study team performance.

4.7.1 UAAP Season 81 Finals Matchup - UP vs ADMU: A Good Offense vs The Best Defense

The UAAP Season 81 Finals provides an interesting case for team-wide analysis. Aside from the fact that it was the first time in 32 years that the UP Fighting Maroons made it to the Finals and they were going up against the defending champions in the ADMU Blue Eagles, this matchup featured one of the better offenses of Season 81 (UP) against probably the best defense of Season 81 (ADMU).

Table 4.16 shows the SScE and oppSScE metrics of the eight UAAP teams while table 4.17 shows their PRLA and oppPRLA metrics.

UP, Season 81's runner up, had a positive SScE and led the league in EPPA and PPA. This indicates that UP took field goal attempts at areas on the court where it was easy to score and even converted those attempts at an above-average rate. This is a testament to just how effective UP's offense was that season. Only two teams had a positive SScE—the aforementioned UP and FEU. However, when the 95% confidence intervals are included, the SScE intervals of all eight teams overlap. Seven teams had confidence intervals that span zero—only UST didn't. UST's confidence interval was entirely below zero. Using an unpaired t-test, UST's SScE was found to be significantly lower than the FEU (*t-statistic = 2.20, p-value = 0.01*) and UP (*t-statistic = 2.00, p-value = 0.02*). FEU and UP were the top 2 teams in terms of SScE and both also made it to the Final 4 in Season 81. UP also had an SScE that was significantly higher than UE (*t-statistic = 1.79, p-value = 0.04*)—Season 81's cellar-dweller. Interestingly, the SScE of Season 81's champion, ADMU, was not found to be significantly higher than any of the other teams' SScE. This was not the case for oppSScE.

The metric oppSScE is equal to the SScE of a team's opponents. It can be used to show how effective opponents score against a team. A positive value means opponents are scoring more than expected—i.e. the team has poor defense or is bad at preventing the opposing team from scoring—while negative values indicate the opposite. The lower the value of oppSScE, the better. For Season 81, five teams had negative oppSScE and three of these teams had a 95% confidence interval entirely below zero. One of these three teams is ADMU that led the league in oppSScE and oppPPA even though their oppEPPA was just average. This indicates that ADMU had a really good defense that prevented opposing teams from scoring even at areas where it was normally easy to score from. An unpaired t-test of ADMU's SScE showed that it was significantly lower than all of the other teams except ADU and DLSU. On the other hand, UP had a good oppEPPA but had poor oppPPA and the worst oppSScE. It also had an oppSScE that was significantly higher than four of the other seven teams. This indicates the opposite of what happened with ADMU. In the case of UP, they were unable to prevent opposing teams from scoring even at areas where it was normally difficult for teams to score.

The same pattern emerged for PRLA and oppPRLA. ADMU had average PRLA but dominant oppPRLA indicative of a good offense and great defense while UP had great PRLA but poor oppPRLA indicative of a great offense and bad defense.

How about the Spread and Range metrics? Tables 4.18 shows the Spread and Range metrics of the UAAP teams and their opponents in Season 81. In terms of offense, ADMU had good Spread but were at the bottom in terms of Effective Range which suggests that even though ADMU attempted field goals at different areas on the court, they were only actually effective in a few of them (around 38% of their Spread or 20% of the total scoring area). UP fared a bit better than ADMU. What's interesting with UP is that they had a really small Spread (276 shooting cells) which contributed to their low ERNG and TERNG. However, they were able to score effectively at a slightly higher percentage of their Spread than ADMU (PERNG - UP: 40%, ADMU: 38%).

The difference between UP and ADMU is more apparent in their opponents' Spread and Range metrics. ADMU's opponents had the second lowest Spread and the lowest ERNG, NERNG, PERNG, and TERNG. Whether this was because of ADMU's defense or the offensive tactics of their opponents, this suggests that not only was the area where ADMU's opponents attempted field goals small, ADMU was also able to successfully limit the number of locations where their opponents scored effectively. ADMU opponents managed to score above what's expected in just ⅓ of their Spread and ⅙ of the total scoring area. In contrast, UP was ranked near the bottom for the Spread and Range metrics. Compared to ADMU's opponents, UP's opponents took field goals at more locations on the court (323 Spread vs 297 for ADMU) and scored better than expected at more of these locations (41.8% PERNG vs 33.0% for ADMU; 23.0% TERNG vs 16.7% for ADMU).

Table 4.16

Expected and actual points per attempt for UAAP teams and their opponents in Season 81.

Team	Season 81 outcome	EPPA [rank]	PPA [rank]	SScE 95% CI [rank]
ADMU	Champion	0.924 [2]	0.920 [3]	-0.004 (0.064, -0.071) [3]
ADU	Final 4	0.896 [8]	0.875 [5]	-0.022 (0.046, -0.089)[5]
DLSU	eliminated	0.900 [7]	0.892 [4]	-0.009 (0.060, -0.077) [4]
FEU	Final 4	0.917 [4]	0.952 [2]	0.035 (0.107, -0.037) [1]
NU	eliminated	0.907 [5]	0.852 [7]	-0.054 (0.011, -0.12) [6]
UE	eliminated	0.922 [3]	0.855 [6]	-0.066 (0.004, -0.136) [7]
UP	Finals	0.962 [1]	0.989 [1]	0.027 (0.102, -0.047) [2]
UST	eliminated	0.902 [6]	0.824 [8]	-0.078 (-0.007, -0.15) [8]

Table 4.16 (cont)

Team	Season 81 outcome	oppEPPA [rank]	oppPPA [rank]	oppSScE 95% CI [rank]
ADMU	Champion	0.916 [4]	0.815 [1]	-0.101 (-0.033, -0.170) [1]
ADU	Final 4	0.920 [5]	0.838 [3]	-0.082 (-0.015, -0.148) [2]
DLSU	eliminated	0.907 [3]	0.832 [2]	-0.075 (-0.008, -0.142) [3]
FEU	Final 4	0.898 [1]	0.882 [4]	-0.017 (0.054, -0.087) [5]
NU	eliminated	0.926 [7]	0.895 [5]	-0.032 (0.032, -0.096) [4]
UE	eliminated	0.924 [6]	0.972 [8]	0.049 (0.121, -0.023) [7]
UP	Finals	0.901 [2]	0.960 [6]	0.059 (0.130, -0.012) [8]
UST	eliminated	0.933 [8]	0.965 [7]	0.032 (0.099, -0.035) [6]

Table 4.17

Expected and actual points for UAAP teams and their opponents in Season 81.

Team	Season 81 outcome	EPTS [rank]	PTS [rank]	PRLA [rank]
ADMU	Champion	925 [1]	921 [1]	-4 [3]
ADU	Final 4	879 [3]	858 [4]	-21 [5]
DLSU	eliminated	863 [5]	855 [5]	-8 [4]
FEU	Final 4	853 [7]	886 [3]	33 [1]
NU	eliminated	877 [4]	824 [6]	-53 [6]
UE	eliminated	852 [8]	772 [8]	-60 [7]
UP	Finals	890 [2]	915 [2]	25 [1]
UST	eliminated	854 [6]	780 [7]	-74 [8]

Table 4.17 (cont)

Team	Season 81 outcome	oppEPTS [rank]	oppPTS [rank]	oppPRLA [rank]
ADMU	Champion	816 [2]	726 [1]	-90 [1]
ADU	Final 4	862 [4]	785 [2]	-77 [2]
DLSU	eliminated	906 [6]	831 [3]	- 73 [3]
FEU	Final 4	883 [5]	867 [5]	-16 [5]
NU	eliminated	940 [8]	908 [7]	-32 [4]
UE	eliminated	833 [3]	877 [6[	44 [7]
UP	Finals	798 [1]	851 [4]	53 [8]
UST	eliminated	934 [7]	966 [8]	32 [6]

Table 4.18

Number and proportion of the scoring area where the UAAP teams and their opponents scored more than expected in Season 81.

Team	Season 81	Spread [rank]	ERNG [rank]	NERNG [rank]	PERNG [rank]	TERNG [rank]
ADMU	Champion	323 [3]	122 [t6]	-79 [8]	37.8% [8]	20.8% [t6]
ADU	Final 4	315 [6]	139 [2]	-36 [1]	44.1% [1]	23.7% [2]
DLSU	eliminated	310 [7]	136 [3]	-38 [2]	43.9% [2]	23.2% [3]
FEU	Final 4	330 [1]	143 [1]	-44 [3]	43.3% [3]	24.4% [1]
NU	eliminated	319 [5]	126 [4]	-66 [5]	39.5% [5]	21.5% [4]
UE	eliminated	325 [2]	124 [5]	-77 [6]	38.2% [6]	21.1% [5]
UP	Finals	276 [8]	110 [8]	-56 [4]	39.9% [4]	18.7% [8]
UST	eliminated	322 [4]	122 [t6]	-78 [7]	37.9% [7]	20.8% [t6]

Table 4.18 (cont)

Team	Season 81 outcome	opp Spread [rank]	opp ERNG [rank]	opp NERNG [rank]	opp PERNG [rank]	opp TERNG [rank]
ADMU	Champion	297 [2]	98 [1]	-101 [1]	33.0% [1]	16.7% [1]
ADU	Final 4	324 [7]	116 [2]	-92 [2]	35.8% [2]	19.8% [2]
DLSU	eliminated	312 [4]	126 [t3]	-60 [4]	40.4% [4]	21.5% [t3]
FEU	Final 4	347 [8]	145 [8]	-57 [5]	41.8% [5]	24.7% [8]
NU	eliminated	317 [5]	126 [t3]	-64 [3]	39.7% [3]	21.5% [t3]
UE	eliminated	310 [3]	135 [t6]	-40 [8]	43.5% [8]	23.0% [t6]
UP	Finals	323 [6]	135 [t6]	-53 [6]	41.8% [6]	23.0% [t6]
UST	eliminated	296 [1]	127 [5]	-42 [7]	42.9% [7]	21.6% [5]

The global metrics indicate that UP had a slightly better offense than ADMU while ADMU was the better defensive team—if not the best defensive team of Season 81.

A look at the local SScE, PRLA, oppSScE, and oppPRLA maps of both teams in Figures 4.26, 4.27, 4.28, and 4.29 provides better insight as to where the differences between the two teams can be found.

Both teams were quite proficient near the basket (<= 2m) as evidenced by the large green colored cells around that area in the SScE and PRLA maps. UP took half of their field goals from this area (460 FGA) and found success with 1.18 PPA, 0.062 SScE, and 28.72 PRLA. ADMU only took 36% of their total field goals from this range

(359 FGA) but actually performed better than UP with 1.23 PPA, 0.114 SScE, and 41.039 PRLA. This isn't surprising since both teams employed dominant big men (e.g. B. Akhuetie for UP and A. Kouame for ADMU) and athletic players who can drive to the basket (e.g. Ju. Gomez de Liaño for UP and T. Ravena for ADMU). Both teams struggled from mid-range (between 3m to 6m from the basket) but UP managed to hold a slight advantage over ADMU with the former having a 0.60 PPA, -0.062 SScE, and -7.64 PRLA while the latter had 0.58 PPA, -0.093 SScE, and -12.93 PRLA. Surprisingly, ADMU performed poorly near and beyond the three point line (>= 6.5 meters from the basket) even though those shots accounted for 42% of their total field goals. They took 417 three-pointers or an average just shy of 30 attempts per game. In that area, ADMU only managed a 0.83 PPA, -0.043 SScE, and -18.01 PRLA. UP only averaged 20 three-pointers per game, one of the lowest for all the teams, but performed better than ADMU with 0.88 PPA, 0.009 SScE, and 2.41 PRLA. Had ADMU been better at converting their three-pointers, they probably would have been more dominant in Season 81.

In terms of defense, the local oppSScE and oppPRLA maps showcase the dominance of ADMU especially at areas near the basket and beyond the three-point line. There aren't many green-colored shooting cells that indicate scoring above expected that can be found in these areas. In fact, most of them are brown in color indicating that ADMU's opponents scored worse than expected in these locations. This is significant because the teams in Season 81 took around 70% of their field goals and scored a majority of their points from these two areas. ADMU's defense was solid near the basket and only allowed 1.04 PPA, -0.041 SScE, and -14.19 PRLA from their opponents. Opposing teams attempted about the same number of field goals as ADMU (344 vs 359) from this area but ADMU was able to outscore their opponents by almost 0.2 points for every shot taken near the basket. They were even more dominant defending three-pointers where they limited opponents to a measly 0.71 PPA, -0.179 SScE, and -54.50 PRLA. ADMU's opponents scored almost 1 point less than expected for every 6 three pointers they attempted. Three-point shooters who favored the right side of the court had a hard time against ADMU evidenced by the large number of brown cells along the three-point line on the right side of the court in the oppSScE and oppPRLA maps. Overall, ADMU's opponents were ineffective near or far from the basket, only managing a 0.88 PPA, -0.107 SScE, and -69.40 PRLA from those areas. For comparison, the opponents of the other seven UAAP teams had an average of 0.98 PPA, -0.006 SscE, and -4.44 PRLA from the same area.

Defense was not the strongest suit of UP for Season 81. Among the eight teams, they had the worst oppSScE and oppPRLA in all locations except mid-range shots (3m to 6m from the basket) which was their lone bright spot on defense. They actually led the league in defending that area, only allowing 0.55 PPA, -0.132 SScE, and -21.7 PRLA from opponents. Unfortunately, opposing teams only attempted 17% of their shots from that area against UP. In all other areas, UP got torched—particularly near the basket and from three-point range. UP was horrendous at defending these two areas. Near the basket, they allowed their opponents to get 1.18 PPA, 0.075 SScE, and 23.3 PRLA which was equivalent to UP's own performance in that area. This means that even though UP was effective on offense near the basket, their defense allowed their opponents to be equally effective against them. UP also gave up a league-worst 1.02 PPA, 0.167 SScE, and 52.2 PRLA on opposing three-pointers. UP and ADMU's opponents attempted around the same number of three-pointers (305 for ADMU opponents, 312 for UP opponents) but where ADMU's opponents were scoring 1 less point than expected for every six shots, UP's opponents were scoring 1 more point than expected for the same number of shots. UP's opponents took 67% of their field goals from near the basket or from three-point range where they manhandled UP to the tune of 1.10 PPA, 0.121 SScE, and 75.5 PRLA. Compared to ADMU, UP allowed almost a quarter of a point more for every shot taken in these two areas. All of these suggest that UP had significantly worse defense than ADMU and that they were unable to prevent their opponents from scoring against them especially in the areas that mattered most—near the basket and three-pointers.

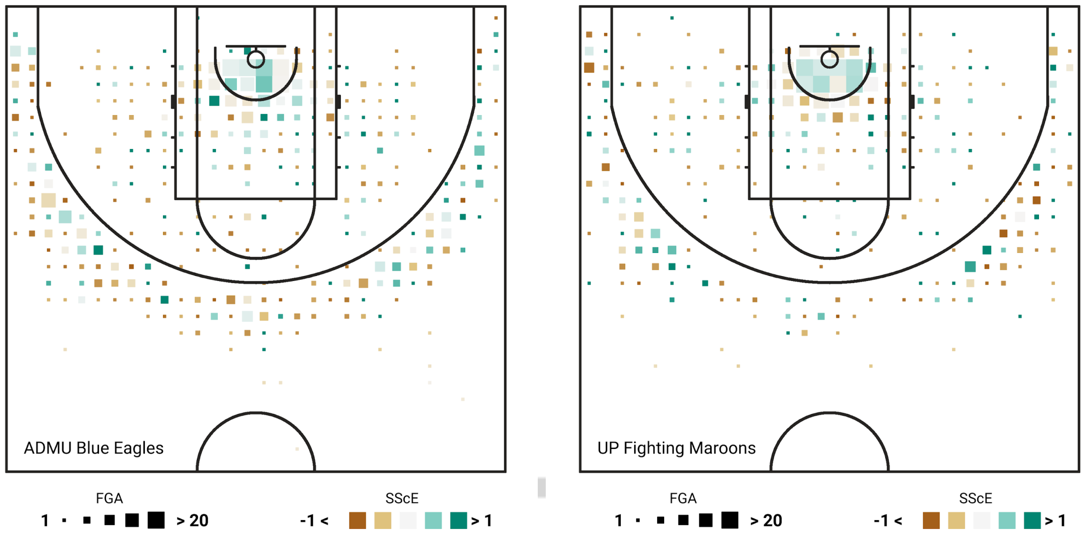


Figure 4.26

Local SScE maps of ADMU and UP in Season 81.

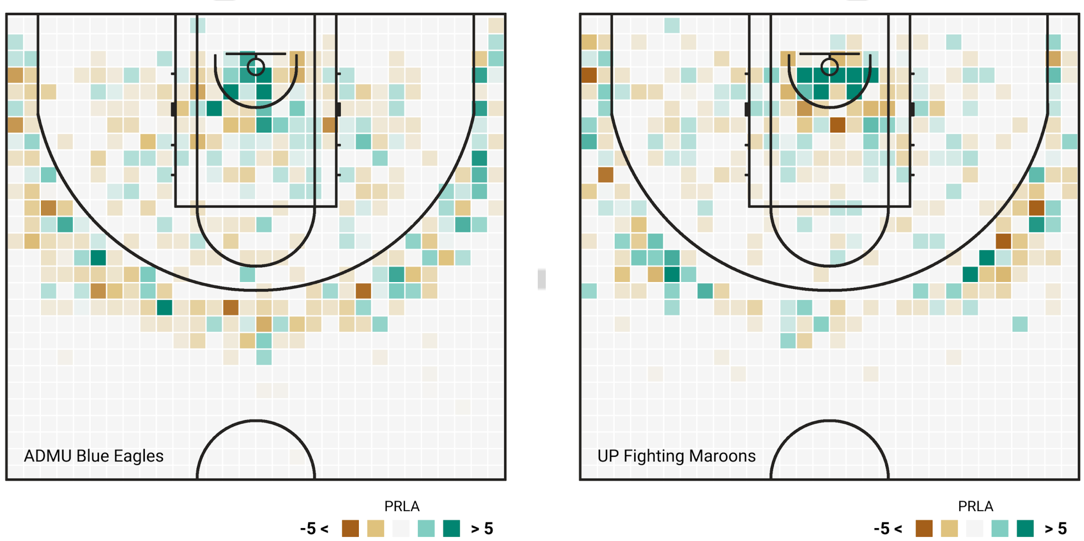


Figure 4.27

Local PRLA maps of ADMU and UP in Season 81.

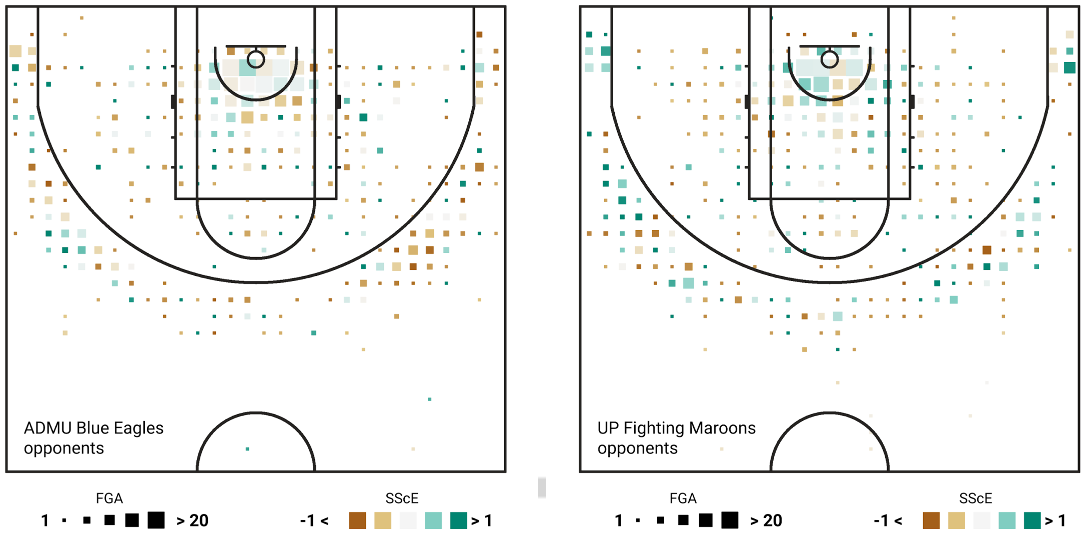


Figure 4.28

Local oppSScE maps of ADMU and UP in Season 81.

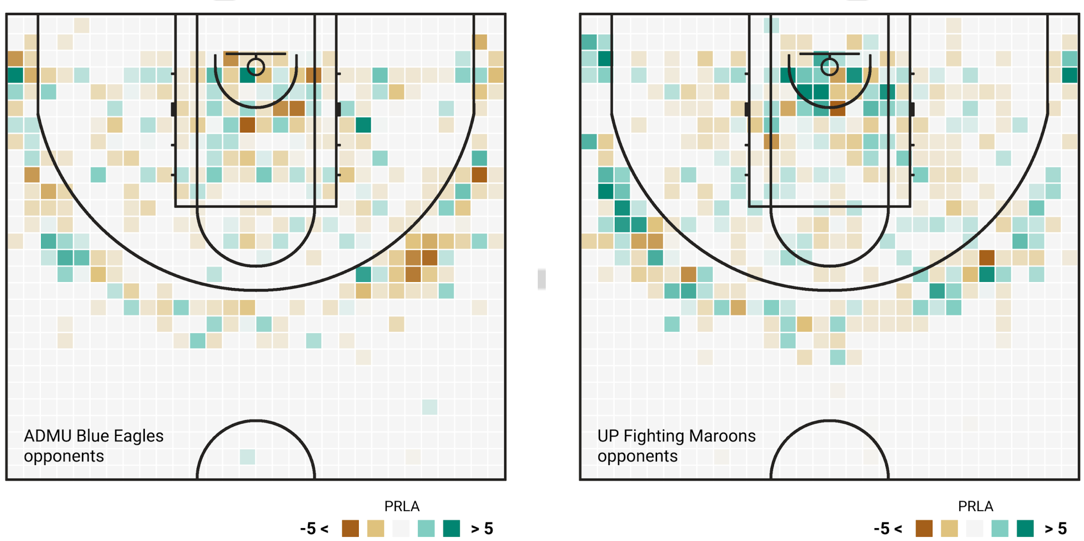


Figure 4.29

Local oppPRLA maps of ADMU and UP in Season 81.

4.8 Metric Analysis

4.8.1 Correlation of SScE and PRLA with EFG, FGA, PTS

Computing the correlation of SScE and PRLA with conventional statistics such as EFG, FGA, and PTS provides a good opportunity to determine if another metric captures the same information as SScE and PRLA (and vice versa) as well as to identify whether or not there is evidence to suggest that the changes in SScE and PRLA values can be explained by another metric.

Figure 4.25 shows the correlation between SScE and Effective Field Goal % (EFG) while Figure 4.23 shows the correlation between PRLA and EFG. In terms of SScE and EFG, there was a near perfect positive correlation between the two (Pearson correlation coefficient $r = 0.94$, p at near zero value) indicating that SScE can work equally well as EFG for a global or summary metric of shooting performance. A player with high SScE had high EFG and vice versa. In addition to this, SScE also has a local component computed per shooting cell that can be used to differentiate players based on where they shoot or score from. A correlation between the local SScE and a local computation of EFG is something that can be looked at in future studies. In terms of the correlation between PRLA and EFG, there was substantial correlation between them (Pearson correlation coefficient $r = 0.80$, p at near zero value) but not as perfectly linear as SScE and EFG.

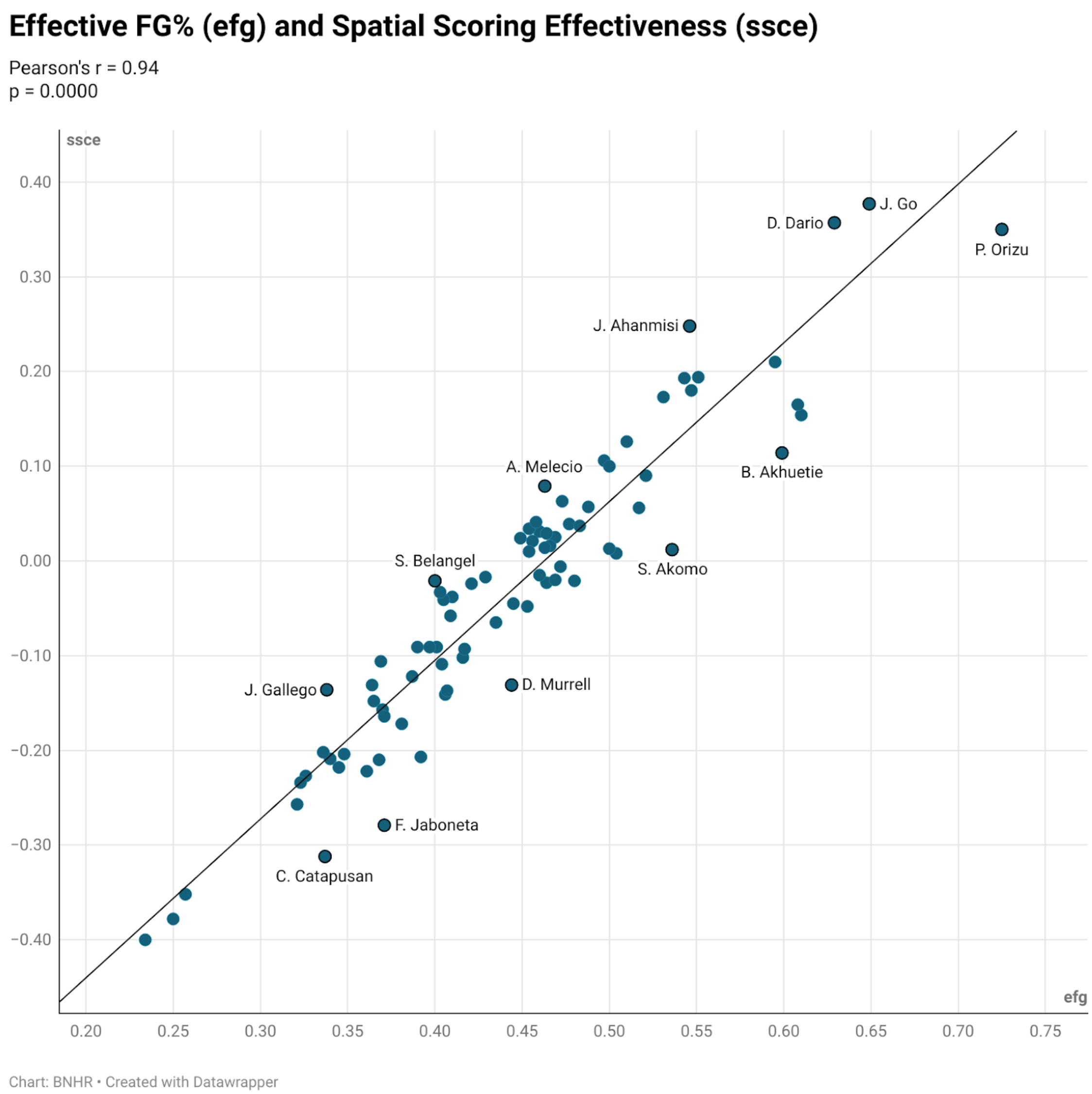


Figure 4.25

Effective FG% vs Spatial Scoring Effectiveness for UAAP Season 81. Selected outlying players are highlighted.

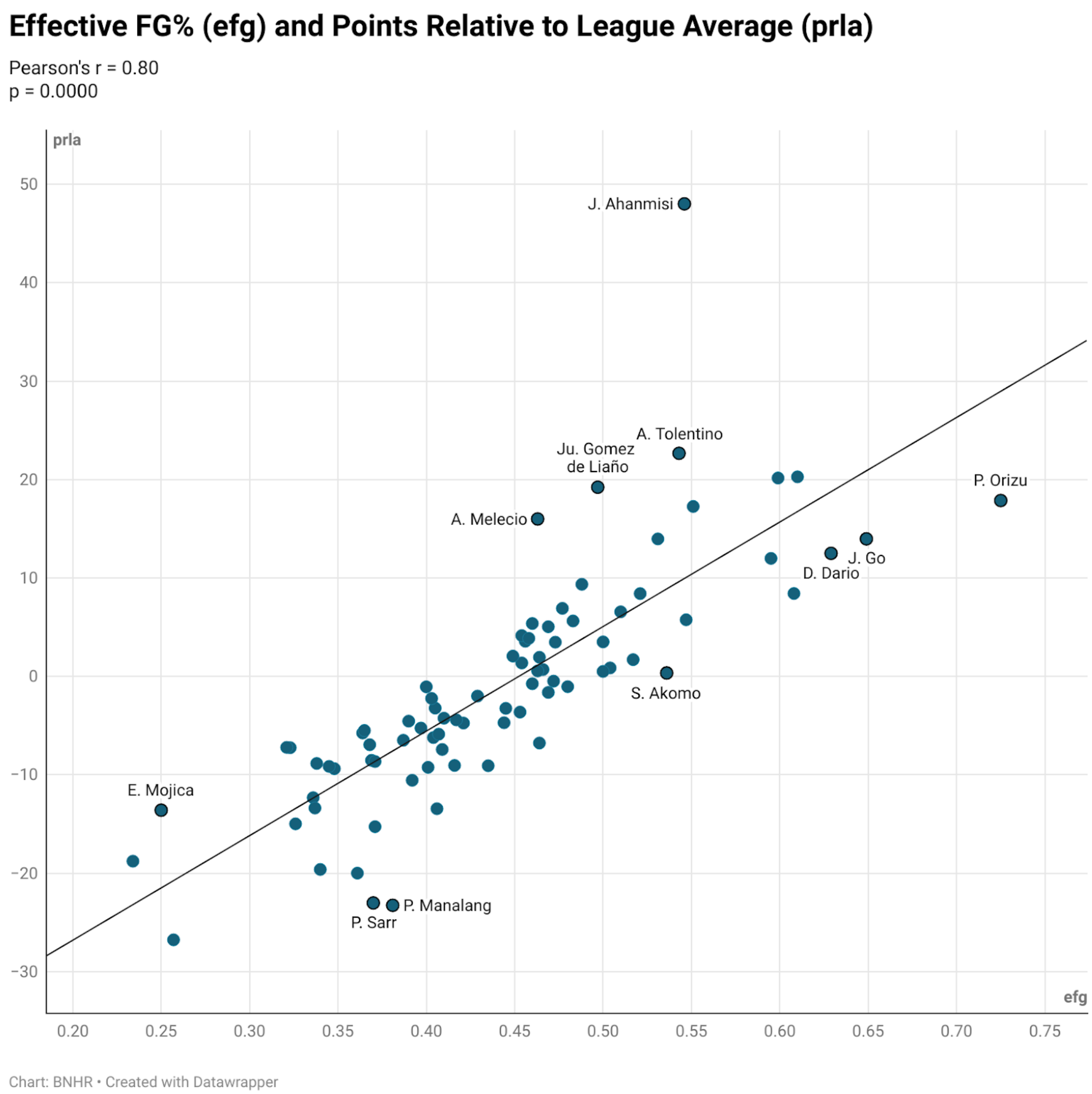


Figure 4.23

Effective FG% vs Points Relative to League Average for UAAP Season 81. Selected outlying players are highlighted

The second thing looked at was whether or not SScE and PRLA are correlated with FGA or the number of field goal attempts by a player. Does taking more FGA mean that a player will have a higher or lower SScE and PRLA values? Figure 4.24 shows the correlation between SScE and FGA while Figure 4.25 shows the correlation between PRLA and FGA. In both instances, the Pearson correlation coefficient r was low (0.20 for SScE and FGA, 0.27 for PRLA and FGA) suggesting weak, if not minimal, linear correlation between the two statistics and FGA. Put another way, the players who took a high number of field goals aren't always the ones with the highest SScE or PRLA. In fact, we can see in Figure 4.25 that the player with the most number of FGA (A. Pasaol) actually has a negative PRLA value.

Figure 4.24 displays one of the observations and intuitions previously made about how the range of SScE values become narrower as players take more field goal attempts indicating that as a player takes more shots, it becomes increasingly less likely for him to score more or less than what is expected. In the figure, a triangular pattern appears with a wide base parallel to the y-axis that seems to converge to y=0 as we move along the x-axis (i.e. as more FGA are taken). Only one player is not included in this triangle—J. Ahanmisi.

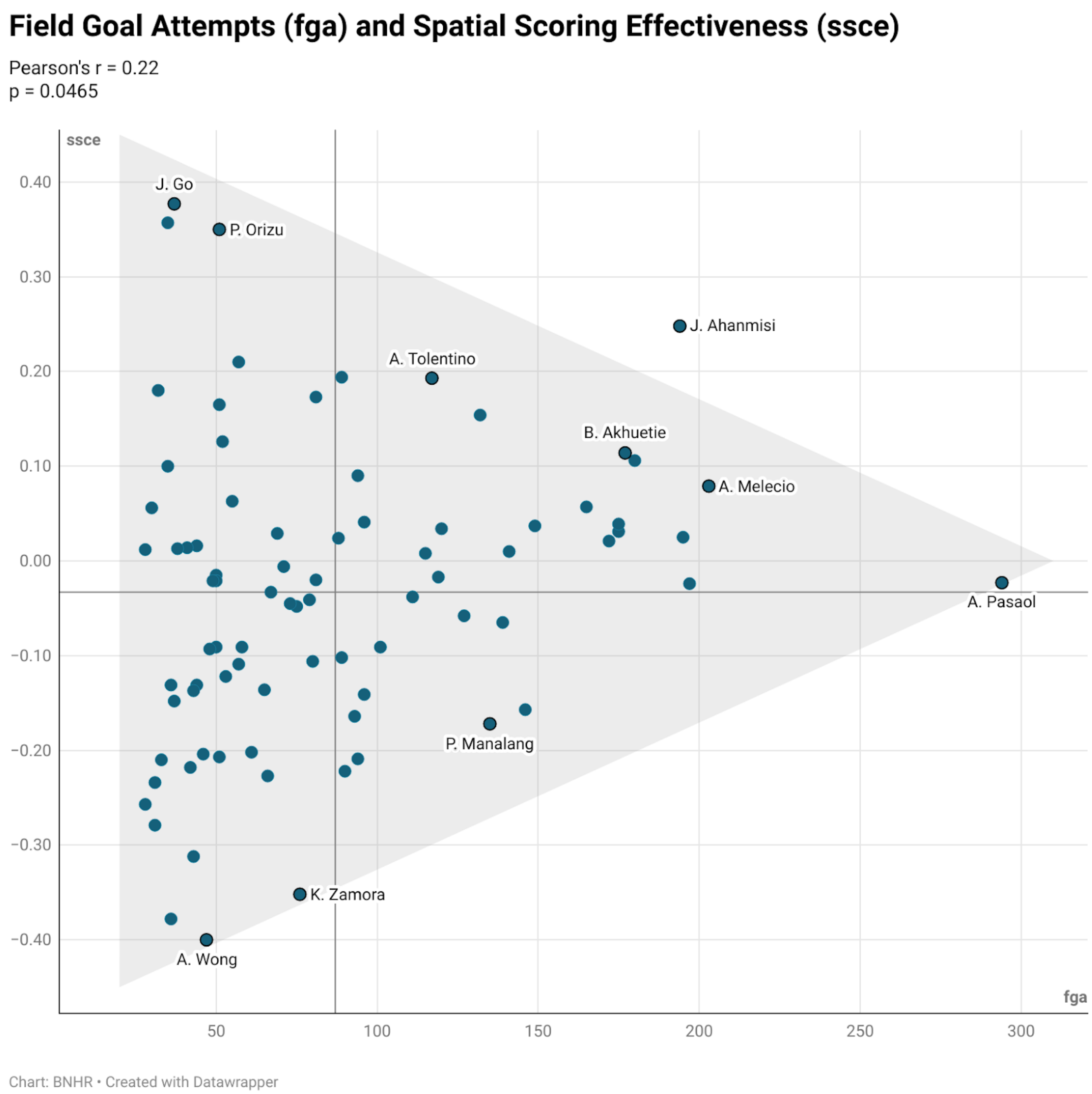


Figure 4.24

Field Goal Attempts vs Spatial Scoring Effectiveness for UAAP Season 81. Selected outlying players are highlighted.

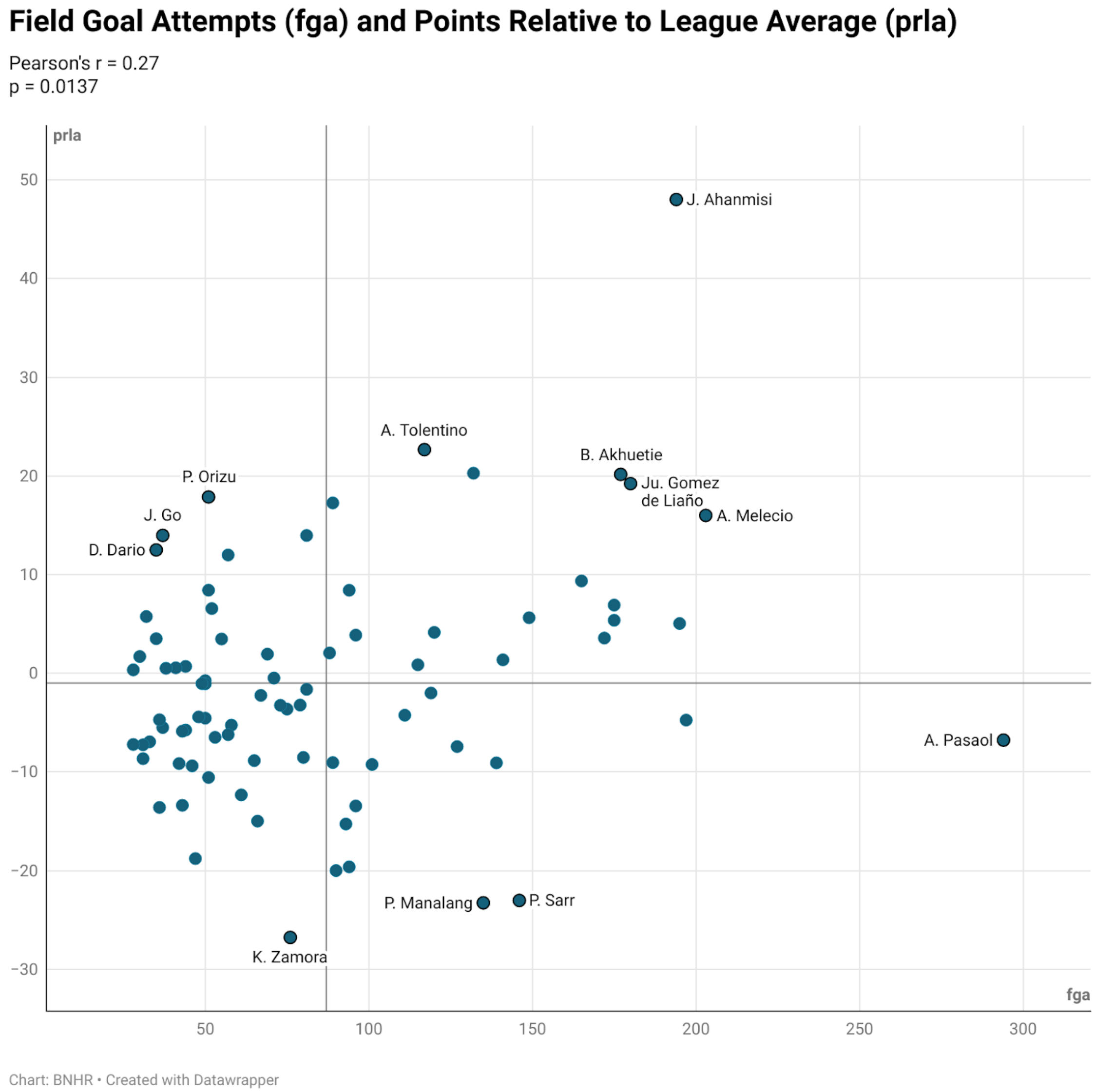


Figure 4.25

Field Goal Attempts vs Points Relative to League Average for UAAP Season 81. Selected outlying players are highlighted

So shooting more field goals does not mean better SScE and PRLA values but how about points? Does scoring more points mean that a player will have better SScE and PRLA? Figures 4.26 and 4.27 show the correlation of SScE and points scored and the correlation of PRLA and points scored, respectively. In both graphs, we see that there was some linear correlation (Pearson correlation coefficient r = 0.40, p = 0.0002 for SScE vs PTS; Pearson correlation coefficient r = 0.47, p at near zero value for PRLA vs PTS). It's not as strong as the correlation of the two metrics with EFG but it's also not as weak as their correlation to FGA. In fact, there was a general trend in both graphs where the values of the metrics do increase as the number of points also increase. This pattern was more apparent between PRLA and PTS than SScE and PTS. An interesting case in both graphs is A. Pasaol who had the most FGA in Season 81 but only had average SScE and PRLA values.

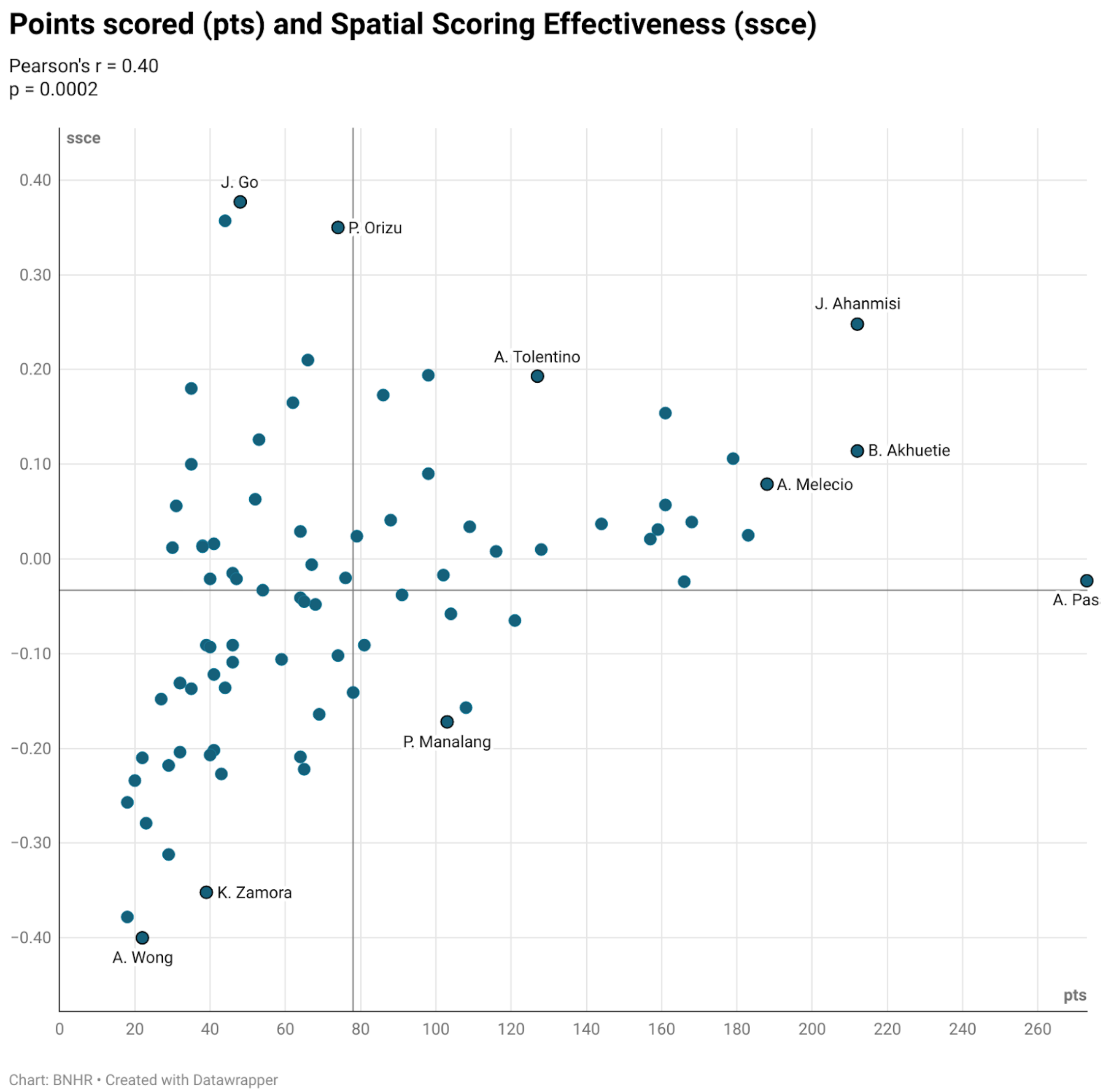


Figure 4.26

Points Scored vs Spatial Scoring Effectiveness for UAAP Season 81. Selected outlying players are highlighted.

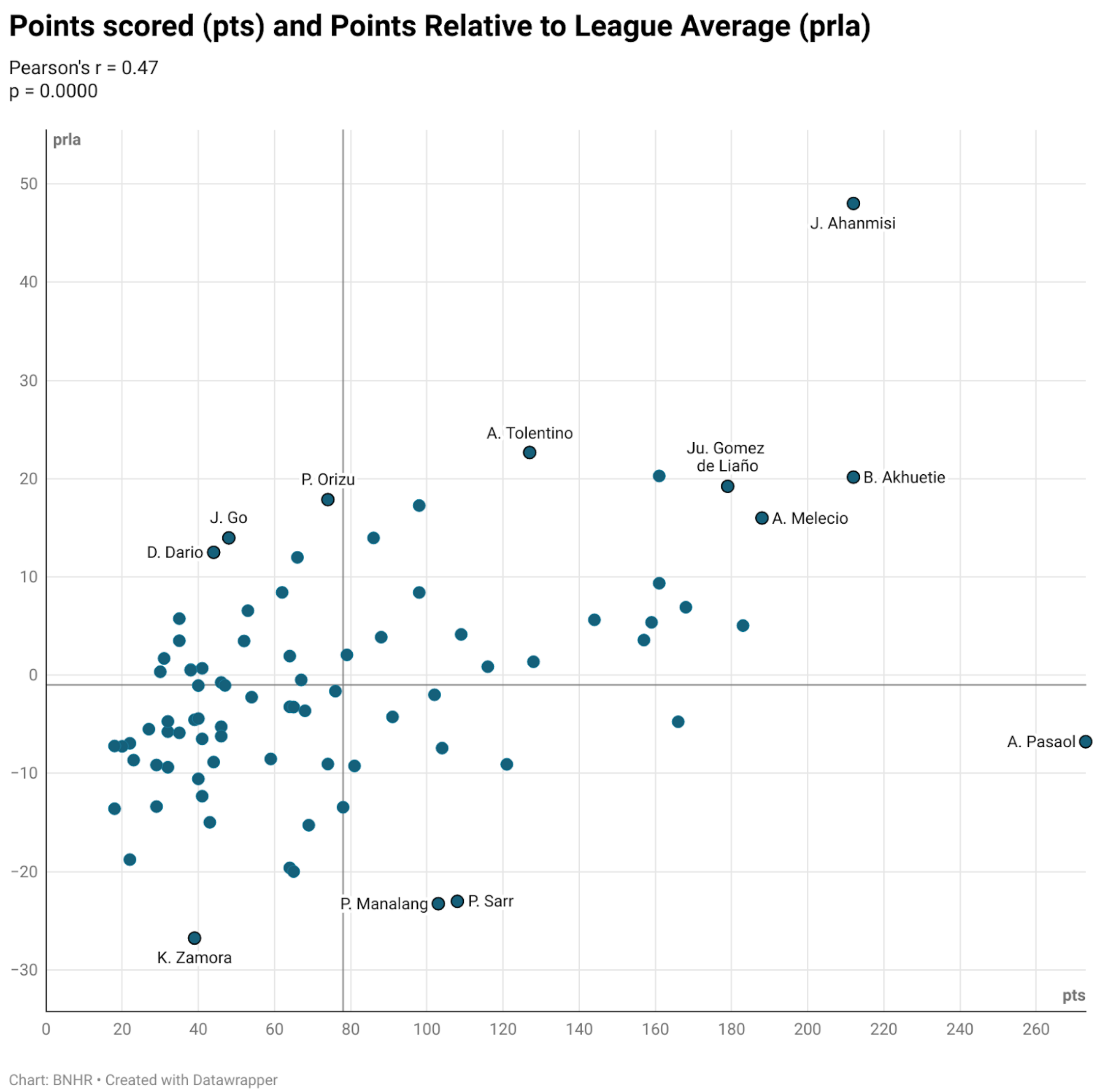


Figure 4.27

Points Scored vs Points Relative to League Average for UAAP Season 81. Selected outlying players are highlighted

5. Summary and Recommendations

In the beginning, this study set out to accomplish the following: (1) to generate and share a spatial field goal dataset of the UAAP, (2) to use the said dataset to spatially characterize field goals and find players with similar shooting habits, (3) to create "spatially-aware" metrics of shooting that incorporate the field goal distribution of teams and players, and (4) to show that spatial analysis and visualization are applicable in the study of shooting and scoring in Philippine basketball. This study has achieved all these objectives.

5.1 Summary and conclusion

First, the study generated a field goal dataset of UAAP Season 81 from online shot-chart data and showed that it was possible to have usable spatial field goal data even without the use of player tracking systems. This dataset is released under an open license at: https://github.com/benhur07b/ms-thesis-spatial-analysis-shooting-philippine-basketball/tree/main/data in order to encourage and support other studies that want to spatially analyze shooting and scoring in the UAAP, build on the current study, and expand the body of work about the use of spatial analysis and visualization in Philippine basketball. Of course the accuracy of the field goal locations from the online shot-charts could not be validated nor compared with those from camera tracking systems because the latter data does not currently exist.

Second, the study was able to characterize the field goals for UAAP Season 81 by finding patterns in the field goal distributions of players using Non-negative Matrix Factorization (NMF) instead of simply dividing the court arbitrarily. Interestingly enough, even though there are countless ways to divide the court into shooting areas, the study found that the field goals during UAAP Season 81 can be characterized by just five areas or shot-types: the restricted area/at-rim, key and wing three-pointers, mid-range + some paint, left-block + some paint, and corner three pointers + some wing three-pointers. As a consequence of using NMF for computing the court areas (i.e. spatial basis vectors) of field goals, the study was also able to generate the frequency at which individual players took shots at these shooting areas. These shooting frequencies were used to show that it was possible to identify players with similar shooting habits by applying a Nearest Neighbor analysis.

Third, several new metrics were introduced in the study based on the premise that there exists a spatial surface overlying the offensive half-court which represents the background local values of the number of points scored per attempt at that location on the court and that although these background values cannot be directly observed, their values can be inferred from a large enough sample in order to estimate the expected points per attempt that should be scored at a specific location on the court. The expected points per attempt at each location obtained by applying Empirical Bayes (EB) estimation and a prior distribution that includes nearby and equidistant locations resulted in a map that was less spatially noisy than the raw points per attempt. The metrics Spatial Scoring Effectiveness (SScE) and Points Relative to League Average (PRLA) were computed by comparing the expected points per attempt (or points) scored at each court location with the actual points per attempt (or points) scored. SScE and PRLA have several advantages over conventional statistics: (1) they explicitly incorporate the spatial nature of shooting which can be used to differentiate players with similar conventional statistics, (2) they have local values that can be mapped to show the spatial distribution of SScE and PRLA which is useful if we want to focus on analyzing or improving a player or team's performance at a specific area on the court, and (3) confidence intervals for the EB-estimated PPA, SScE, and PRLA can be computed for each shooting cell or set of shooting cells which is useful when comparing a player/team's expected and actual performance or a player/team's performance with that of another. Aside from SScE and PRLA, new Range metrics based on the SScE were introduced—e.g. Effective Range, Net Effective Range, Player/Team Effective Range %, and Total Effective Range %—that can serve as summary values for comparing teams and players based on how large of an area on the court they score better than expected.

Lastly, the application of the metrics introduced in the study to UAAP Season 81 showed how spatial analysis and visualization can be used to compare, contrast, and present player and team performance. Additionally, these metrics were able to highlight and provide insight on interesting cases during Season 81. J. Ahanmisi is one example. Although conventional statistics tend to agree that Ahanmisi had a good shooting/scoring season, they failed to give justice to just how good his shooting performance was for that season. Spatial metrics were able to uncover this and put it in full display. They showed that even though Ahanmisi's conventional shooting statistics were similar to other players, he actually had one of the most difficult field goal distributions during the season—evidenced by his low EPPA—yet he was one of only a handful of players who managed to significantly score beyond what was expected as shown by his SScE and PRLA statistics. In addition to this, the study found that Ahanmisi not only scored significantly higher than what was expected of his field goal distribution, he also performed significantly better than a majority of the players that season. Even more impressive was that he was able to do this consistently while being one of the top players with the most number of field goal attempts and most number of field goal locations as shown by his Spread and Range metrics. Similarly, the case of ADMU and UP—the two teams who met in the Finals of Season 81—demonstrated how the metrics introduced in the study can be used for team-wide analysis of both offense and defense. It showed the locations on the court where the two teams were effective at scoring or where they were effective at limiting their opponents from scoring. The analysis of both teams found that UP had an effective yet limited offense that was focused on shots near the basket based on their Spread, Range metrics, SScE, and PRLA. However, their defense also allowed their opponents to score effectively, especially near the basket and from the three-point area. In contrast, ADMU's offense was average. The difference between their expected and actual points scored was not found to be significant. The same was also true when comparing their offensive performance against that of the other teams. What the study found, however, was the sheer dominance of ADMU's defense at limiting the effectiveness of opposing teams from scoring from the field. The spatial metrics introduced in the study showed that ADMU led the league both in preventing opponents from scoring effectively (oppSScE and oppPRLA) and in limiting the number of locations on the court where opponents score effectively against them (RNG, NERNG, PERNG, TERNG). They also found that ADMU's opponents scored significantly lower than expected based on the spatial distribution of their field goals and that the difference in the expected and actual performance of ADMU's opponents was significantly larger than that of the opponents of a majority of the teams. It's not a stretch to say that for ADMU in UAAP Season 81, the coach's adage of "Defense wins championships." rang true.

The application of spatial analysis and visualization of shooting in basketball was able to extract and display information absent from conventional statistics, provide new insight into player and team performance, and challenge some assumptions that we have about shooting/scoring while also validating others.

5.2 Recommendations for future work

In terms of the data, it is hoped that in the future, player tracking systems could provide better and larger datasets that can be used in studying Philippine basketball spatially. A larger dataset could provide more information and better estimates about how shooting habits and performance change over time. Data covering multiple seasons will not only allow comparisons of player and team performance over time but also enable the creation of models based on the spatial data that can be used to predict future performance of players and teams. Having data from the professional basketball leagues like the PBA and MPBL would also enable studies about how and if shooting and scoring ability in the collegiate leagues translate to the pros—i.e. do players with good spatial shooting metrics in college remain that way when they turn professional, are spatial shooting metrics in college a good indicator that a player will be a good shooter/scorer in the professional leagues. With more complete data, multivariate and hierarchical analysis of shooting and scoring will also be possible so that contextual information such as the time of the game, the deficit in the score, and the presence of a defender can be accounted for. These are just some of the things that are possible if more spatial basketball data becomes available in the Philippines.

Other clustering and classification algorithms can also be looked into for determining groups of similar players. This can include building models that can classify players into shooting/scoring archetypes based on their shooting tendencies and spatial metrics. This could then lead to a possible study to identify what kinds of player archetypes make up a successful offense or team.

For the estimation of the expected points and points per attempt at different areas on the court, fully-Bayesian methods can be explored aside from other local value estimators. The size of the shooting cells used can also be changed to see if there were spatial variations in shooting that were not captured by the 50cm x 50cm cells. The results of simpler methods such as kernel smoothing can be compared to those of Empirical Bayes or fully-Bayesian estimation.

Spatial Scoring Effectiveness, Points Relative to League Average, and the different Range metrics can be adapted in other studies to further analyze and visualize shooting/scoring. Examples include computing for the average distance of player/team shooting cells, left-right splits to determine if a player/team favors or is better at one side of the court over the other, comparing the performance of different on-court lineups, and the analysis of players based on their position, height, etc.

Bibliography

[1] Bailey, N., Bhuwalka, K., Lee, H., & Zhong, T. (2018), Understanding Features of Successful 3 Point Shots in the NBA, *Massachusetts Institute of Technology, MIT*. Available at: http://www.mit.edu/~nbailey/files/3ptreport.pdf

[2] Bornn L, Daly-Grafstein D. (2019). Using in-game shot trajectories to better understand defensive impact in the NBA. arXiv preprint arXiv:1905.00822

[3] Brunet J-P., et al. (2004) Metagenes and molecular pattern discovery using matrix factorization. *Proc. Natl Acad. Sci. USA, 2004, vol. 101* (pg. 4164-4169). DOI:10.1073/pnas.0308531101.

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

[5] Cervone, D., Bornn, L., & Goldsberry, K. (2016a). NBA Court Realty. *10th MIT Sloan Sports Analytics Conference*, 2016.

[6] Daly-Grafstein D, & Bornn L. (2019). Rao-Blackwellizing field goal percentage, *Journal of Quantitative Analysis in Sports 15:85–95*.

[7] D'Amour, A., Cervone, D., Bornn, L., & Goldsberry, K. (2015). Move or Die: How Ball Movement Creates Open Shots in the NBA. *9th MIT Sloan Sports Analytics Conference*, 2015.

[8] Franks, A., Miller, A., Bornn, L., & Goldsberry, K. (2015). Characterizing the spatial structure of defensive skill in professional basketball. *Annals of Applied Statistics 9 (2015), no. 1,* 94-121. DOI:10.1214/14-AOAS799.

[9] Frigyesi, A., & Höglund, M. (2008). Non-negative matrix factorization for the analysis of complex gene expression data: identification of clinically relevant tumor subtypes. *Cancer informatics*, 6, 275–292. DOI:10.4137/cin.s606.

[10] Gaujoux, R. & Seoighe, C. (2010). A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* 2010, 11:367. http://www.biomedcentral.com/1471-2105/11/367.

[11] Goldsberry, K. (2012) CourtVision: New Visual and Spatial Analytics for the NBA. *MIT Sloan Sports Analytics Conference*, 2012.

[12] Goldsberry, K., & Weiss, E. (2013), The Dwight Effect: A New Ensemble of Interior Defense Analytics for the NBA, *MIT Sloan Sports Analytics Conference*, 2013.

[13] Goldsberry, K. (2019). *Sprawlball: a visual tour of the new era of the NBA*. Boston: Houghton Mifflin Harcourt.

[14] Harris, C. R. et al. (2020). Array programming with NumPy, Nature, 585, 357–362, DOI:10.1038/s41586-020-2649-2.

[15] Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, 9, 90-95, DOI:10.1109/MCSE.2007.55

[16] Hutchins, L., Murphy, S., Singh, P., & Graber, J. (2008). Position-Dependent Motif Characterization Using Nonnegative Matrix Factorization. *Bioinformatics* (Oxford, England). 24. 2684-90. DOI:10.1093/bioinformatics/btn526.

[17] Jiao, J., Hu, G., & Jan, Y. (2019). A Bayesian Marked Spatial Point Processes Model for Basketball Shot Chart. https://arxiv.org/abs/1908.05745v2.

[18] Lee, D. & Seung, S. (1999), Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[19] Maheswaran, R., Chang, Y.H., Henehan, A., & Danesis, S. (2012). Deconstructing the Rebound with Optical Tracking Data. *MIT Sloan Sports Analytics Conference, 2012.*

[20] Maheswaran, R., et al. (2014). The Three Dimensions of Rebounding. *MIT Sloan Sports Analytics Conference*, 2014.

[21] Marty R. (2018). High-resolution shot capture reveals systematic biases and an improved method for shooter evaluation, *In Proceedings of the 2018 MIT Sloan Sports Analytics Conference.*

[22] McKinney, W. (2010). Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference, 51-56. DOI:10.25080/Majora-92bf1922-00a

[23] Miller, A., Bomn, L., Adams, R., & Goldsberry, K. (2014). Factorized point process intensities: A spatial analysis of professional basketball. In *31st International Conference on Machine Learning, ICML 2014* (pp. 398-414). (*31st International Conference on Machine Learning, ICML 2014; Vol. 1*). International Machine Learning Society (IMLS).

[24] Murillo M.A.S. (2019, July 22). Sports analytics in the Moneyball era. BusinessWorld. https://www.bworldonline.com/sports-analytics-in-the-moneyball-era/

[25] Pedregosa, F. et al. (2011). Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825-2830. URL http://jmlr.org/papers/v12/pedregosa11a.html

[26] Pintor, B., & Cataniag, N. (2014). CourtVisionPH: A System for the Extraction of Field Goal Attempt Locations from Broadcast Basketball Videos and Spatial Analysis of Shooting. *Undergraduate Research. University of the Philippines College of Engineering.*

[27] Python Software Foundation (2021). Python. URL https://www.python.org/.

[28] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[29] RStudio Team (2021). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

[30] Rolland, G., Vuillemot, R., Bos, W., & Rivière, N. (2020). Characterization of Space and Time-Dependence of 3-Point Shots in Basketball. *MIT Sloan Sports Analytics Conference*, 2020.

[31] Sandholtz, N., Mortensen, J., & Bornn, L. (2019). Measuring Spatial Allocative Efficiency in Basketball. https://arxiv.org/abs/1912.05129v1.

[32] Virtanen, P. et al. (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17(3), 261-272. DOI:doi.org/10.1038/s41592-019-0686-2.

[33] Shortridge, A., Goldsberry, K., & Adams, M. (2014). Creating space to shoot: quantifying spatial relative field goal efficiency in basketball. *Journal of Quantitative Analysis in Sports, 10(3),* 303-313. DOI:doi.org/10.1515/jqas-2013-0094.

[34] Socamos, K. (2018, January 16). *PBA News: Alaska analytics coach Layug on....* spin.ph.

BN HR

https://www.spin.ph/basketball/pba/more-than-just-numbers-as-analytics-coach-paolo-layug-tries-to-help-alaska-relive-lost-glory

[35] Tenner, Z., & Franks A. (2020). Modeling Player and Team Performance in Basketball.