Ashton Shortridge*, Kirk Goldsberry and Matthew Adams

Creating space to shoot: quantifying spatial relative field goal efficiency in basketball

Abstract: Every basketball player takes and makes a unique spatial array of shots. In recent years, technology to measure the coordinates of these constellations has made analysis of them possible, and the possibility exists for distinguishing between different shooters at a level of spatial detail finer than the entire basketball court. This paper addresses the challenge of characterizing and visualizing relative spatial shooting effectiveness in basketball by developing metrics to assess spatial variability in shooting. Several global and local measures are introduced and formal tests are proposed to enable the comparison of shooting effectiveness between players, groups of players, or other collections of shots. We propose an empirical Bayesian smoothing rate estimate that uses a novel local spatial neighborhood tailored for basketball shooting. These measures are evaluated using data from the 2011 to 2012 NBA basketball season in three distinct ways. First we contrast nonspatial and spatial shooting metrics for two players from that season and then extend the comparison to all players attempting at least 250 shots in that season, rating them in terms of shooting effectiveness. Second, we identify players shooting significantly better than the NBA average for their shot constellation, and formally compare shooting effectiveness of different players. Third, we demonstrate an approach to map spatial shooting effectiveness. In general, we conclude that these measures are relatively straightforward to calculate with the right input data, and they provide distinctive and useful information about relative shooting ability in basketball. We expect that spatially explicit basketball metrics will be useful additions to the sports analysis toolbox.

Keywords: basketball; empirical Bayes estimation; local spatial statistics; shot map.

DOI 10.1515/jqas-2013-0094

***Corresponding author: Ashton Shortridge**, Michigan State University – Geography, 673 Auditorium Rd, East Lansing, MI 48824, USA, Tel.: +517-432-3561, Fax: +517-432-1671, e-mail: ashton@msu.edu

Kirk Goldsberry: Harvard University – Institute for Quantitative Social Sciences, Cambridge, MA 02138, USA

Matthew Adams: CourtVision Analytics, New York, NY 10009, USA

1 Introduction

Perhaps the ultimate measure of success in basketball is effectiveness at scoring points. A team that scores more effectively than its opponent will win, while the team that generates assists or rebounds more effectively may very well not. Basketball statistics reflect the importance of field goals. The standard traditional measure of a player's contribution is points per game and its efficiency-minded relative, field goal percentage. Oliver (2003) promoted the use of effective field goal percentage, which accounts for the extra value of a three-point shot. In recent years, basketball has experienced a quantitative revolution, with possession-based metrics like points per shot or points per possession gaining increasing attention (Kubatko et al. 2007). Despite the current general interest in scoring metrics, relatively little academic attention has been paid to spatial variation in shooting effectiveness: the likelihood of a shot going in depends very much on where that shot was taken.

Every basketball player generates unique spatial constellations of shots. For example, some players are highly active mid-range shooters, while other players shoot almost exclusively in areas close to the basket. While players have long been qualitatively assigned to categories (e.g., “wing,” “post,” “spot shooter”) according to the general patterns of their offensive games, it is only recently that quantitative data could readily be collected to enable the mapping and analysis of these spatial constellations. This micro-geography of shooting effectiveness drives vital aspects of basketball. It forms the rationale for the three-point line, answers foundational questions, such as why the court is not twice as wide or long, and guides the bulk of basic offensive and defensive strategy. While the importance of spatial variation in shooting effectiveness seems so self-evident that we hesitate to belabor the point further, it is notable that no commonly reported modern basketball metrics explicitly account for this spatial variation. To illustrate just one consequence, these metrics fail to differentiate great pure midrange shooters from dominant post players who thrive by scoring close to the rim.

With the increasing collection of spatial shot data, several researchers have explored local shooting

metrics. Reich et al. (2006) divided the half-court into 121 zones and developed a spatial autoregressive model that accounted for a variety of game situations to explain a player's shooting ability. The model is capable of predicting where and how well a player will shoot under different situations (e.g., playing at home, with certain personnel on the floor, against a good shot-blocking team). Other zone-based analysis was published in online in the mid-2000s (82games.com 2005). More in the spirit of the present paper, Piette et al. (2010) proposed individual shooting and scoring metrics for NBA players that account for where the shots are taken. Unlike our paper, the metrics account solely for the distance from the basket for shots, rather than their absolute location on the court. A number of quantitative sports analysts have explored shot mapping in blog posts since 2008, when play-by-play records, including location information, were either scraped from ESPN web pages (Eli 2008) or became publicly available on the web for others to use (Parker 2008); several blog articles have mapped raw or kernel smoothed NBA shooting rates using this data (e.g., Bailey 2008; Eli 2008; ThinkBlueCrew 2010). Goldsberry (2012) employed fine spatial resolution shot maps to assess the spatial range and spread of NBA shooters. We use the same underlying approach for characterizing shot location, but our goal is to formally develop relative shooting proficiency metrics in a manner that accounts for location.

In this paper we introduce several measures of relative field goal effectiveness that explicitly account for spatial variability in scoring. These measures identify an expected point total for locations across the court, and contrast this expected point total with the actual points scored by a particular player. Critically, these metrics can be either calculated locally for each position on the floor, or they can be aggregated into single, global measures of relative shooting effectiveness. Local measures can be mapped to visualize a player's relative scoring effectiveness across the court, or subjected to further analysis. Much modern geographic spatial analysis research is effectively local (e.g., Anselin 1995; Lloyd 2011), reflecting the importance of understanding spatial variation in natural and human processes. Global measures on the other hand are convenient summaries of players' relative efficiencies.

To illustrate their potential, these statistics are used to assess field goal efficiency of National Basketball Association (NBA) players from the 2011 to 2012 regular season. We use spatial data identifying the locations of both missed and made field goals for that season. Results are contrasted with non-spatial scoring metrics. Finally, we

conclude with some thoughts on the utility of spatially-explicit measures for basketball.

2 Spatial framework and background rate estimation

An NBA basketball court is a rectangular surface 94 ft long and 50 ft wide. Aside from desperation heaves at the end of the timed playing periods, nearly all shots taken at a basket are made in the half-court closest to that basket. Our present interest is to develop a data model to represent the constellation of shots taken within a half-court by one or more players over the course of a season. Consider a Cartesian coordinate system with an origin in one corner of the half-court, the x-axis running along the baseline and the y-axis along the sideline. Any location on the court can then be specified by a coordinate pair; an infinite number of such locations lie within its boundaries. In Geographic Information Science, similar conceptual spaces are discretized using a regular tessellation into a finite set of equally-sized square elements in the interests of computational tractability – as if a basketball court were surfaced with square tiles. The elements could be structured as squares in a vector data structure, or as cells in a raster structure. Both structures are widely used in geographic information systems (GIS) applications. The number of field goal attempts and points scored is then stored for each cell across the half-court. A key characteristic is the spatial resolution of the elements; the finer the resolution, the greater the precision of the constellation, but the greater the data collection accuracy must be.

We maintain that shooting effectiveness by a particular player or group of players from a particular cell i should be measured against the background shooting effectiveness from that cell by a larger sample of representative shots, for example, all shots taken from that cell during an NBA season. An intuitive approach is to estimate that background effectiveness for cell i by the raw success rate:

$$r_i = \frac{FGM_i}{FGA_i} \quad (1)$$

where FGA_i and FGM_i are the field goals attempted and made at cell i , respectively. Use of the raw rate as an estimate of the actual background shooting success rate is problematic, since rates from cells with small numbers of attempts are unlikely to be well-estimated; indeed, a map of raw rates is spatially noisy. We adopt an empirical Bayes (EB) rate estimator, an approach popularized

in spatial analysis of disease risk (Langford 1994; Bailey and Gatrell 1995) and widely used in spatial epidemiology and health geography research today (Waller and Gotway 2004; Foley et al. 2009; Fahrmeir and Kneib 2011). This approach develops an estimate of the true rate at each location as a weighted combination of a prior probability distribution function and the local raw rate. The weight, also called the shrinkage factor, controls the contribution of the raw rate. In the basic case, the prior probabilities are simply the overall distribution of rates, and the EB estimated rate will lie between the local raw rate and the global average. In the basketball context, we make the reasonable assumption that a better prior would utilize only locations about the same distance from the basket, and relatively close to the location being estimated.

Specifically, for a location i positioned d feet from the basket, where $d < 30$ (about 85% of all shots taken during the 2011–2012 season), the prior distribution includes those cells satisfying both of the following conditions:

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

For locations 30 feet or farther from the basket it is necessary to expand the local neighborhood for the prior distribution, as many cells at long distances have few or no attempts. This is done in a linear fashion with distance:

- *equidistant from the basket*: within 1.2 feet of d , plus 6 inches for every foot increase past 30 from the basket
- *close*: < 5 feet from i , plus 1 foot for every foot increase past 30 from the basket

In practice, the definition above is effective at flexibly modifying the prior to account for shooting frequency. 90% of locations have a local neighborhood in excess of 20 cells, with at least 100 shots taken. Only 1% have fewer than 13 neighboring cells and fewer than 54 shots taken. The prior means and variances for weights calculation were estimated via method of moments: the prior mean for a location i is the neighborhood average $\hat{\gamma} = \sum \text{makes}_j / \sum \text{shots}_j$ for all cells j in the local neighborhood, and the prior variance is a weighted local sample variance (Clayton and Kaldor 1987; Marshall 1991):

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

where n_j is the number of shots taken in cell j , r_j is the raw field goal rate observed in cell j , and $\bar{n}_j$ is the sample mean of the number of shots taken within all neighborhood cells j . The weighting factor for shrinking the contribution of the prior mean for each cell i is:

$$\hat{W}_i = \frac{\hat{\varphi}_i}{(\hat{\varphi}_i + \hat{\gamma}_i / n_i)} \quad (3)$$

where n_i is the number of shot attempts from i . The empirical Bayesian rate for each location i is a weighted combination of the local raw field goal rate r_i and the Bayesian local average $\hat{\gamma}_i$:

$$\hat{\theta}_i = \hat{W}_i r_i + (1 - \hat{W}_i) \hat{\gamma}_i \quad (4)$$

Weights close to 1 will result in a rate estimate similar to the raw rate for that location, while weights close to zero will result in a rate matching the Bayesian local average. Locations with relatively small numbers of attempted shots, or in local neighborhoods with low variance, will have small weights.

The development of an EB estimated rate with this novel local neighborhood prior has a number of advantages over spatial kernel smoothing methods: first, it explicitly accounts for distance from the basket, and therefore does not include cells in the neighborhood that are substantially closer to or farther from the hoop; second, it flexibly modifies the local neighborhood to ensure sufficient observations for robust neighborhood rate estimation; and third, it provides more adjustment to raw rates in locations where fewer shots are attempted, and maintains local but meaningful departures from the neighborhood rate in locations supported by a high local number of shots. The particular neighborhood parameter settings produce a spatially compact neighborhood around each tile; results were very similar to those using alternative small-neighborhood parameter settings. We note that alternative Bayesian estimators, including fully Bayesian approaches, may be effective at modeling both local rate means and variances, and will return to this point in the discussion.

3 New measures of spatial shooting efficiency

Consider a spatial surface overlying a half-court representing the background local probability of hitting a shot. While this surface can not be directly observed, its nature can be inferred using a large sample of shots (e.g., with local EB derived from attempts by all NBA players during the 2011–2012 regular season) and a set of attempts and outcomes of interest k taken on that surface. k might be an individual player's contribution during the regular season, but other sets might also be evaluated: that player's performance against playoff-bound teams, or in the

fourth quarter, or against particular defenders. k could also represent groups of players: a franchise's starting five, or all Eastern Conference forwards. In this paper we simply let k refer to the shots taken by individual players for the 2011–2012 NBA season. The expected local points scored by k shooting from spatial cell i within the half-court is estimated as:

$$ELPTS_{ik} = \hat{\theta}_i \times FGA_{ik} \times PPB_i \quad (5)$$

where $\hat{\theta}_i$ is the EB rate estimate for cell i , FGA_{ik} is the number of shots taken from i by k , and PPB_i is the points per basket a shot from i is worth. The result is the number of points we would expect k to score from i if k scores at the EB estimated rate for that location. A summary measure characterizing the average difficulty of k 's constellation of shot attempts is the expected points per shot (EPPS), estimated as:

$$EPPS_k = \sum_{i=1}^N ELPTS_{ik} / FGA_k \quad (6)$$

where N is the number of cells i in which k has attempted one or more field goals and FGA_k is the total number of field goals (from anywhere on the court) attempted by k .

A more familiar shot-based metric: k 's actual points per shot, excluding those from free throws, is defined as:

$$PPS_k = PTS_k / FGA_k \quad (7)$$

Where PTS_k are the total points scored from field goals by k . By combining Equations 6 and 7, we introduce a spatial efficiency metric, spatial shooting effectiveness (SSE):

$$SSE_k = PPS_k - EPPS_k \quad (8)$$

This value, measured in points, indicates the difference between k 's expected and actual points per shot. A value of 0 indicates that k is scoring at the EB estimated rate. A value considerably larger than 0 indicates that k is scoring more effectively for that constellation of shots, while a negative value indicates that k is scoring less effectively. Finally, a measure of output, points over league average (POLA), can be constructed:

$$POLA_k = SSE_k \times FGA_k = PTS_k - \sum_{i=1}^N ELPTS_{ik} \quad (9)$$

POLA indicates the number of points from field goals that k scored relative to the EB estimate of the expected number of points for k 's shot constellation. A value of 0 indicates an average output, a value of 65 indicates that k scored 65 more points than expected, while a value of -12 indicates that k scored 12 fewer points than expected.

Local measures may be valuable as well; such measures identify shooting proficiency from any, or every, specific location on the court, and can therefore be mapped and analyzed for spatial pattern. We suggest two measures. The first is a local points per shot metric for cell i :

$$LPPS_{ik} = PTS_{ik} / FGA_{ik} \quad (10)$$

where PTS_{ik} is the number of points scored by k from i . $LPPS_{ik}$ is a disaggregated version of Equation 7 that identifies locations where k tends to score more or less efficiently. This measure may range from 0 to 3; unlike LISA metrics such as Getis G which measure local proportions of the global statistic, a mean weighted on local FGA will return k 's overall PPS. Second, local point difference can also be calculated for location i :

$$LPD_{ik} = (PTS_{ik} - ELPTS_{ik}) / FGA_{ik} \quad (11)$$

If k shoots better than the population average from location i , this number will be above zero; if it is below zero, k has shot worse than the average. The units are easy to understand, as they are simply points per shot. The third local metric we consider here is a measure of local shooting effectiveness:

$$LSSE_{ik} = LPPS_{ik} - (\hat{\theta}_i \times PPB_i) \quad (12)$$

which measures the difference between actual and expected points per shot for each location i . $LSSE_{ik}$, $LPPS_{ik}$ and LPD_{ik} are local statistics because they are calculated for every tile i on the half-court; they can be mapped to visualize spatial variation in scoring efficiency.

The ability to assess the significance of the departure of these metrics from one another, or from league averages, is clearly important. We have done that in two ways: departure of actual from expected points, and direct comparison of SSE or POLA between two players k . Each is explained in turn in the following paragraphs.

Both shooting metrics, POLA and SSE, measure the difference between expected (based on EB rate estimates for a set of cells comprising the group shooting constellation) and actual points scored. A useful question is whether the observed points differ significantly from the expected points. We use weighted, paired t-tests to evaluate whether a group's observed-expected point total differs from 0; observations are cells on the court from which a player has taken shots. Weights are the number of shots taken within that cell. Our approach follows that described by Bland and Kerry (1998). Consider a set of $LSSE_{ik}$ values and a corresponding set of FGA_{ik} values. Then the weighted mean SSE_{ik} is:

$$SSE_k = \frac{\sum_{i=1}^N FGA_{ik} \times LSSE_{ik}}{\sum_{i=1}^N FGA_{ik}} \quad (13)$$

The weighted variance is:

$$s^2 = \frac{WSS_k - (N \times SSE_k^2)}{(N-1)} \quad (14)$$

where WSS_k , the weighted sum of squares, is:

$$WSS_k = \frac{\sum_{i=1}^N FGA_{ik} \times LSSE_{ik}^2}{\sum_{i=1}^N FGA_{ik} / N} \quad (15)$$

A t-test can then be calculated that accounts for the different numbers of shots taken in each location. A second useful question to consider is whether Group A's POLA or SSE is significantly larger than Group B's. Again a weighted t-test is used, though this is can not be paired, as the group constellations can not be assumed to be identical. Weights for POLA or SSE for each cell are the shots taken from that cell by the group.

4 Case study: the 2011–2012 regular season

To illustrate the calculation and utility of the new metrics, and to quantify the influence of space on shooting efficiency in the NBA, we analyzed a database recording every shot taken during the 2011–2012 NBA regular season. Data were obtained from ESPN.com; shot location data were available

on a game by game basis in publicly available text xml files, though no information was provided concerning collection methodology or accuracy. This database includes Cartesian coordinates for over 141,000 field goal attempts, and detailed attribute information including who took the shot and whether or not the attempt resulted in a made basket. All analysis was conducted in R v. 3.0.2 (R Core Team 2013) and used the classInt (Bivand 2013), RColorBrewer (Neuwirth 2011), sp (Bivand et al. 2013), and weights (Pasek 2012) libraries. To construct a robust estimate of league-wide local field goal probability, we used EB rate estimation, as documented in Section 2. The 2011–2012 raw NBA field goal rate is mapped in Figure 1A, while the EB rate inferred surface is shown in Figure 1B. It is visually apparent that the EB rate surface is much smoother than the raw rate, as much of the noise present in the raw rates has been removed. Raw rates with small numbers of shots, typically in locations far from the basket, were affected substantially: EB estimated rates were as much as 0.94 below the raw rate and as much as 0.42 above the raw rate. However the mean rate difference (EB–raw) was 0.003, with half the rate differences falling between -0.035 and 0.057 . Distinct spatial patterns are apparent in the EB estimated rate surface. Rates are high very close to the basket, but fall as distance increases to about 15 feet. There is an arc at about this distance with higher rates; this arc is wider at the free throw line and on the left side. At distances >20 feet the smoothed rates resume their descent, with generally extremely low rates more than 30 feet from the basket. Along the three point line several higher spots emerge: in the corners and on the right elbow. Shots taken from behind the backboard are low-percentage, particularly on

Figure 1 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.

the figure's left: shots from these locations evidently must be attempted with one's left hand, which may explain the discrepancy. The EB field goal rate surface is not left-right symmetric, though the broad pattern is persistent.

We generate local shooting percentage for individual players and calculate LPPS, LPD, and LSSE, as described above. Each local metric is calculated for each 1-square foot in which shots were taken, and summaries of their values across the court were collected to provide single, global shooting metrics for each player (SSE and POLA). Finally, graphics showing local shooting ability are constructed for selected players. The utility of the introduced metrics, particularly SSE and POLA, is evaluated in three case studies. The first contrasts the shooting performance of two NBA players during the 2011–2012 season in detail; the second is a league-wide analysis to determine the most and least effective shooters; the third offers an assessment of the spatial shooting ability of two players. Throughout these studies the spatial metrics are compared with non-spatial measures of shooting ability.

We first illustrate the new shooting evaluations by comparing two NBA players: Dwight Howard and Steve Nash. During the 2011–2012 NBA season Steve Nash took 554 shots and Dwight Howard took 721 shots. They were very different players, with very distinctive spatial shooting tendencies. As a center, almost all of Dwight Howard's shots were taken close to the basket; as a point guard, Steve Nash's shots were more scattered, generally occurring much further from the basket. The traditional metric to compare these

player's shooting effectiveness is field goal percentage (FG%); Howard was second in the league in FG%, shooting 57.3%, while Nash was 11th, at 53.2%. Using effective FG%, which accounts for three point shooting, Nash ranked third at 58.1%, while Howard, who attempted no three point shots, ranked fourth at 57.3%. Points per shot, an efficiency metric, could not distinguish between these shooters, as both scored at a rate of 1.16 points per shot.

All of these measures neglect to account for the striking spatial differences between Howard and Nash's shooting tendencies. Figure 2A illustrates the expected points per shot surface, based on the 2011–2012 EB estimated field goal rate for the league as a whole. This rate is multiplied by the points per shot earned from each location on the court to produce the surface in the figure. Cells intersected by the three point line are adjusted by the proportion of three point shots vs. two point shots attempted. This map reflects the same spatial patterns observed in Figure 1B, but with the impact of the shot value factored in. The shot constellations of Howard and Nash are overlain on this map in Figure 2B. It is visually apparent that almost all of Howard's shots were in areas of where the league has relatively high EPPS values; this was not the case with Nash. To quantify this striking visual contrast, spatial shooting metrics were calculated for both players.

Steve Nash's 554 attempts result in 520.9 expected points, an average of 0.94 expected points per shot (EPPS); Dwight Howard's 721 attempts result in 727.2 expected points, an average of 1.01 EPPS. By itself, this analysis

Figure 2 (A) Expected points per shot (EPPS) map for the 2011–2012 NBA regular season, obtained by multiplying the rate estimates in Figure 1B by the local points per field goal; (B) Shot maps for Steve Nash (orange diamonds) and Dwight Howard (blue x's) overlain on a map of expected points per shot.

simply reveals that one should expect fewer points per shot from Nash than Howard, given their respective shot constellations. Next we compare these estimated point totals with actual points accrued by the shooter as a means to assess their shooting efficiency relative to their cohort of NBA peers. Steve Nash's shots actually resulted in 645 points, an average of 1.16 points per shot (PPS). His spatial shooting effectiveness (SSE) was 0.23 points per shot, meaning that he scored over a fifth of a point per shot more than would be expected from his shot constellation. Nash accrued 124.1 points more than expected (POLA), indicating that he shot much better than an average NBA shooter from these locations. Dwight Howard's 721 shots actually resulted in 832 points, an average of 1.15 PPS; scoring 104.9 POLA. His SSE was 0.15 points per shot, a full third less than Nash's. Interestingly, both shooters accumulated points at nearly the same rate: 1.16 points per shot; however, Nash's PPS might be considered more impressive in the sense that it is much larger than his EPPS value of 0.94. However, a weighted difference of means t-test of each player's LSSE distributions resulted in a t-statistic of 1, with an associated *p*-value of 0.32, indicating no support for the hypothesis that Nash's SSE is larger than Howard's.

For the league-wide analysis, we calculated PPS, EPPS, and SSE values for the 250 individual players that attempted at least 250 field goals during the 2011–2012 NBA regular season. Table 1 includes summary statistics for these three key variables. Players' points per shot distribution (free throws excluded) ranges more widely than EPPS, but has a similar mean. Spatial shooting effectiveness (SSE) ranges over more than half a point per shot, indicating considerable individual variability in scoring efficiency, even while accounting for each player's spatial constellation of shots. That said, substantial variability in points scored is explained by spatial position: the standard deviation of total points from field goals for this group of players is 232, while the standard deviation of the difference in total points from expected points is just 47.

Table 2 reports the players with the ten highest and ten lowest EPPS values in the NBA during the 2011–2012

regular season (team and position data from <http://www.nbastuffer.com>). Players with the highest EPPS values have constellations with the highest field goal percentage, centered in the highest PPS areas in Figure 2A. Players with the lowest EPPS values are taking more shots in the low-percentage areas of that figure; for example, Kobe Bryant is popularly known for taking difficult shots; his position in the table appears to support this. Table 3 indicates players with the highest and lowest points per shot averages, excluding free throws. This is a non-spatial measure of shooting efficiency.

Taken individually, Tables 2 and 3 indicate the expected and actual scoring effectiveness of individual players' shooting constellations. These metrics are combined to identify spatial shooting effectiveness (SSE) as a global measure of relative shooting prowess. Table 4 shows the ten best and worst SSE values in the NBA during the 2011–2012 season. Steve Novak scored 0.32 points per shot better than the league average from his shooting constellation, while Tony Douglas scored 0.21 points less per shot from his. The narrow range of SSE is noteworthy: just one percent of high-frequency NBA shooters are able to score more than two-tenths of a point more per shot than the NBA average for their constellations, while no player is able to remain high-frequency (>250 FGA) while shooting worse than 0.21 points per shot below the average for their constellations. Table 4 also prints 95% confidence intervals for weighted paired t-statistics evaluating whether the observed SSE is >0 (for players with positive SSE) or <0 (for players with negative SSE). Perhaps unsurprisingly, the ten highest and lowest SSEs had confidence intervals that did not include zero. Out of the 250 players evaluated, the SSE confidence intervals of 187 spanned zero, 38 players had a positive SSE with 95% confidence intervals above zero, while just 25 players had a negative SSE with confidence intervals entirely below zero. Intervals are relatively wide: those of the top 10 players all overlap, as do those of the bottom 10 players.

Figure 3 illustrates the relationship between EPPS and PPS for all players in the 2011–2012 season with at least 250 field goal attempts. These two components comprise the spatial shooting effectiveness (SSE) metric, and this scatterplot provides some insight into how players end up where they do. A player with an average points per shot and an average shooting constellation will end up in the middle (Jameer Nelson claims that spot on the graph). Players who take shots from higher percentage locations on the floor will be at the top of the graph, while those who hit shots at a higher percentage will be on the right. The SSE gradient runs from right to left on a diagonal. By inspection, the three players with the lowest SSE scores

Table 1 2011–2012 NBA regular season shooting statistics for 250 players.

	PPS	EPPS	SSE
Maximum value	1.37	1.21	0.32
Mean value	0.99	0.98	0.01
Minimum value	0.74	0.86	−0.22
Standard deviation	0.09	0.06	0.09

Table 2 2011–2012 NBA players with highest and lowest EPPS values.

Highest				Lowest			
Player	Team	Pos	EPPS	Player	Team	Pos	EPPS
1. Tyson Chandler	NYK	C	1.21	250. Patrick Patterson	HOU	PF	0.86
2. Nikola Pecovic	MIN	C	1.14	249. Elston Brand	PHI	PG	0.86
3. DeAndre Jordan	LAC	C	1.14	248. Josh Howard	UTA	SF	0.86
4. Luc Mbah a Moute	MIL	SF	1.11	247. Al Jefferson	UTA	C	0.86
5. Kenneth Faried	DEN	SF	1.11	246. Kobe Bryant	LAL	SG	0.88
6. Greg Monroe	DET	PF	1.11	245. Marc Gasol	MEM	C	0.88
7. Gerald Wallace	NJN	SF	1.09	244. Kevin Garnett	BOS	PF	0.88
8. Zaza Pachulia	ATL	C	1.09	243. Jarrett Jack	NOH	PG	0.88
9. Tiago Splitter	SAS	C	1.08	242. Dirk Nowitski	DAL	PF	0.88
10. Joakim Noah	CHI	C	1.08	241. Luis Scola	HOU	PG	0.88

Table 3 2011–2012 NBA players with highest and lowest PPS values (free throws excluded).

Highest				Lowest			
Player	Team	Pos	PPS	Player	Team	Pos	PPS
1. Tyson Chandler	NYK	C	1.37	250. Tyrus Thomas	CHA	PF	0.74
2. Steve Novak	NYK	SF	1.35	249. Toney Douglas	NYK	PG	0.74
3. DeAndre Jordan	LAC	C	1.26	248. Lamar Odom	DAL	PF	0.79
4. Manu Ginobili	SAS	SG	1.24	247. Ricky Rubio	MIN	PG	0.80
5. Tiago Splitter	SAS	C	1.24	246. Corey Maggette	CHA	SF	0.81
6. Brandon Rush	GSW	SG	1.22	245. Tyler Hansbrough	IND	PF	0.82
7. Matt Bonner	SAS	PF	1.18	244. Kemba Walker	CHA	PG	0.83
8. Kenneth Faried	DEN	SF	1.17	243. Josh Howard	UTA	SF	0.83
9. James Harden	OKC	SG	1.17	242. Norris Cole	MIA	PG	0.84
10. Stephen Curry	GSW	PG	1.17	241. Johan Petro	NJN	C	0.84

Table 4 2011–2012 NBA players with highest and lowest SSE values and accompanying 95% confidence intervals.

Highest			Lowest		
Player	Team	SSE (95% CI)	Player	Team	SSE (95% CI)
1. Steve Novak	NYK	0.32 (0.17, 0.47)	250. Toney Douglas	NYK	−0.21 (−0.35, −0.09)
2. Stephen Curry	GSW	0.24 (0.09, 0.38)	249. Lamar Odom	DAL	−0.20 (−0.33, −0.08)
3. Steve Nash	PHX	0.23 (0.13, 0.33)	248. Tristan Thompson	CLE	−0.20 (−0.29, −0.10)
4. Manu Ginobili	SAS	0.20 (0.05, 0.35)	247. Ricky Rubio	MIN	−0.17 (−0.28, −0.06)
5. Matt Bonner	SAS	0.18 (0.04, 0.33)	246. Corey Maggette	CHA	−0.16 (−0.27, −0.05)
6. Brandon Rush	GSW	0.18 (0.06, 0.30)	245. Alonzo Gee	CLE	−0.15 (−0.25, −0.06)
7. Gerald Green	NJN	0.17 (0.03, 0.30)	244. Tyrus Thomas	CHA	−0.15 (−0.26, −0.04)
8. Kyle Korver	CHI	0.16 (0.03, 0.29)	243. Derrick Williams	MIN	−0.15 (−0.25, −0.05)
9. Tyson Chandler	NYK	0.16 (0.05, 0.26)	242. Norris Cole	MIA	−0.14 (−0.25, −0.04)
10. Tiago Splitter	SAS	0.16 (0.04, 0.27)	241. Marcus Camby	HOU	−0.14 (−0.27, −0.02)

have very different shooting profiles. Lamar Odom's shot constellation of his EPPS was of average difficulty, but he scored at a low rate on those shots. In contrast, Toney Douglas, the lowest-ranking SSE shooter, had a relatively difficult shot constellation, while Tristan Thompson scored at a higher rate per shot, but his EPPS was relatively high.

Similar patterns are evident at the high end of the SSE gradient. Steve Novak had the best SSE value on the strength of a very high points per shot metric over a relatively high EPPS score. Curry and Nash trailed him, as both scored fewer points per shot on much lower EPPS shooting constellations. Tyson Chandler had the highest point per shot

Figure 3 Points per shot vs expected points per shot, 2011–2012 NBA season. Selected outlying players are labeled. Diagonal lines are SSE contours.

Figure 4 Spatial shooting effectiveness vs. effective FG percentage, 2011–2012 NBA season. Selected outlying players are labeled.

score of the season, but because his expected points per shot were quite high, his SSE ranks only ninth.

Steve Novak's SSE was tested against those of each other player using the weighted t-test described in Section 3. Novak's league-leading SSE was significantly above the SSE of all but 10 players: the 9 below him on the SSE top 10 and the 11th-ranked SSE player, Jason Smith.

An important question for any new metric is whether its information content is largely captured by an existing metric. For SSE, a natural comparison is with effective field goal percentage. The correlation between these two metrics in this study is 0.80: substantial but not perfectly linear; their relationship is depicted in Figure 4. EFG by itself is not able to distinguish between players like Nikola Pekovic and Steve Nash, who have similar EFG percentages (58.1% vs. 56.4%). However, their spatial shooting effectiveness is quite different, with Nash being one of the best in the NBA and Pekovic being a below-average shooter for his constellation (0.23 vs. -0.01). A comparison of their SSE distributions with the weighted t-test revealed that Nash's SSE was significantly greater (p -Value 0.002).

Figure 5 shows maps of POLA for two players, Kevin Garnett and Greg Monroe. The players are superficially similar shooters using nonspatial metrics: Garnett scored 785 points from the field with a 50.5 EFG%, and he averaged 1.01 points per shot. Monroe scored 814 points with a 52.1 EFG% and 1.04 PPS. From the maps though, it is

clear that they have different constellations and markedly different success relative to the NBA average from those constellations. Garnett's points above league average was 101.2 points, while Monroe's was -48.7 . A weighted t-test of the difference of these player's POLA scores resulted in a t -value of 10.5 and a near-zero p -Value: Garnett scored significantly more effectively from his constellation than Monroe did from his. This map reveals more about the source of Greg Monroe's shooting difficulties: his local POLA in the cell directly under the basket was -48 . Monroe was effectively an average shooter from the rest of his constellation, but at the basket he hit 85 of 148 shots (57%) at 1.15 PPS. The EB field goal percentage for that cell is 74%, with an EPPS of 1.47. Monroe apparently had trouble scoring effectively right at the basket in comparison to the rest of the league.

5 Discussion and Conclusion

Spatial shooting effectiveness (SSE) and points above league average (POLA) are new measures to evaluate basketball shooting; they compare and contrast shooters using a spatially explicit perspective that accounts for the individual shot constellation and the relative performance of all other shooters from those spatial locations.

Figure 5 Maps of points above league average (POLA): (A) Kevin Garnett; (B) Greg Monroe, 2011–2012 NBA season. Square size indicates number of shots taken from that location, while hue indicates POLA value for that spot. Garnett’s summed POLA value was 101.2 points, while Monroe’s was -48.7 .

They take advantage of the recent availability of location-specific shooting data. A substantial contribution of this paper is the formalization of both global and spatially local versions of these metrics, along with formal tests enabling the evaluation of the significance of derived metrics, either against the league average for a particular constellation or between different players or groups of shots. Another contribution is the application of an empirical Bayesian smoothed rate estimator using a novel local neighborhood for obtaining background average field goal percentage. These shooting metrics can be used to evaluate effectiveness of individuals, teams, or other groupings against various populations, including league or division averages or subsets such as playoff-bound teams, all forwards, or shots taken in the fourth quarter.

To illustrate these measures we evaluated the shooting performance of every player in the 2011–2012 NBA regular season with at least 250 field goal attempts. We conducted a detailed comparison of two players and also summarized league-wide shooting effectiveness of individuals. The detailed comparison demonstrated how standard metrics like points per shot can fall short: both Nash and Howard netted 1.16 points per shot (not including free throws), but Nash’s SSE was higher than Howard’s, indicating he was better at scoring from his spots on the floor than others shooting from the same positions. The league-wide comparisons showed that SSE and POLA are useful and distinctive measures of ability that provide different insights than nonspatial shooting metrics such

as effective field goal percentage. The comparison of Kevin Garnett to Greg Monroe also showed the utility of spatial metrics to identify significant differences in the spatial variation and effectiveness of different players, and provided insights to why those differences might be present.

This paper demonstrated relatively straightforward approaches to estimate spatially varying background shooting rates and to evaluate the departure of interesting subsets from those rates. Alternative, fully Bayesian methods may provide more complete methods to model shooting proficiency while accounting for spatial autocorrelation, spatially varying numbers of shots attempted, and other factors (Waller and Gotway 2004; Best et al. 2005). For example, Reich et al. (2006) divided the court into 121 zones and used a spatial CAR model to take advantage of covariances between neighboring zones for smoothing, while also capturing covariate relationships. Bayesian methods can also provide local estimates of model uncertainty, thereby enabling more robust comparisons than those presented in this paper. In addition, we used player’s raw shooting rates to develop the metrics. Individual constellations typically have many tiles with only a few shots (Figure 5), for which raw rates can be unreliable. Local EB rate smoothing of individual shooters (or other subsets k of interest) may provide more robust estimates (Marshall 1991).

We anticipate several ways in which the introduced metrics can shed useful insight. First, global measures of shot selection and effectiveness can be compared across

different players and groups of players, illuminating new ways to employ quantitative techniques in sports analysis. Spatial patterns of shot selection and efficiency probably reflect offensive strategies and their responses to defensive adjustments; these metrics may offer an opportunity to isolate those patterns. Second, formally defined metrics can be embedded in more sophisticated models of basketball scoring such as those in Reich et al. (2006). This may enable us to study how shot effectiveness is affected by the game environment, moving beyond spatial pattern analysis to process modeling. Finally, local metrics detailing the spatial patterns of shooting effectiveness – shooting effectiveness maps for particular players or teammates, perhaps in different game contexts, or contested vs. not contested shots – might usefully be compared, perhaps through map similarity analysis (White 2006; Foody 2007). With ongoing expansion of the collection of spatial data about athletic contests with SportVU cameras and other technology (Plafke 2013), we suspect that this line of analysis is just beginning, and that methods developed in the spatial sciences will prove to be highly valuable additions to the quantitative sports analyst.

Acknowledgments: The authors thank the anonymous reviewers whose perspectives substantially improved this paper.

References

- 82games.com 2005. NBA player shot zones. 82Games.com. <http://www.82games.com/shotzones.htm>. October 10, 2005. Last checked 5/18/2014.
- Anselin, L. 1995. "Local Indicators of Spatial Association – LISA." *Geographical Analysis* 27(2):93–115.
- Bailey, J. R. 2008. "NBA Shot Location Visualization." *JasonRBailey.com*. <http://jasonrbailey.com/?p=127>. November 9, 2008. Last checked 5/18/2014.
- Bailey, T. C. and A. C. Gatrell. 1995. *Interactive Spatial Data Analysis*. Harlow, UK: Longman Scientific & Technical, 413 p.
- Best, N., S. Richardson, and A. Thompson. 2005. "A Comparison of Bayesian Spatial Models for Disease Mapping." *Statistical Methods in Medical Research* 14:35–59.
- Bivand, R. S. 2013. *classInt: Choose univariate class intervals*. R package version 0.1-21. <http://CRAN.R-project.org/package=classInt>. Last checked 5/18/2014.
- Bivand, R. S., E. Pebesma, and V. Gomez-Rubio. 2013. *Applied Spatial Data Analysis with R, Second edition*. NY: Springer.
- Bland, J. M. and S. M. Kerry. 1998. "Weighted Comparison of Means." *BMJ* 316(7125):129.
- Clayton, D. and J. Kaldor. 1987. "Empirical Bayes Estimates of Age-Standardized Relative Risks for use in Disease Mapping." *Biometrics* 43(3):671–681.
- Eli 2008. "Where Players Take and Make Shots." *CountTheBasket.com*. <http://www.countthebasket.com/blog/2008/03/27/where-players-take-and-make-shots/>. March 27, 2008. Last checked 5/18/2014.
- Fahrmeir, L. and T. Kneib. 2011. *Bayesian Smoothing and Regression for Longitudinal, Spatial and Event History Data*. New York: Oxford University Press, 512 p.
- Foley, R., M. C. Charlton, and A. S. Fotheringham. 2009. "GIS in Health and Social Care Planning." In: *Handbook of Theoretical and Quantitative Geography*. UNIL-FGSE-Workshop series. 22 p.
- Foody, G. M. 2007. "Map Comparison in GIS." *Progress in Physical Geography* 31(4):439–445.
- Goldsberry, K. 2012. "CourtVision: New Visual and Spatial Analytics for the NBA." *MIT Sloan Sports Analytics Conference* March 2–3, 2012.
- Kubatko, J., D. Oliver, K. Pelton, and D. T. Rosenbaum. 2007. "A Starting Point for Analyzing Basketball Statistics." *Journal of Quantitative Analysis in Sports* 3(3), Article 1.
- Langford, I. H. 1994. "Using Empirical Bayes Estimates in the Geographical Analysis of Disease Risk." *Area* 26(2):142–149.
- Lloyd, C. D. 2011. *Local Models for Spatial Analysis, 2nd Edition*. Boca Raton, Florida: CRC Press. 336 p.
- Marshall, R. J. 1991. "Mapping Disease and Mortality Rates Using Empirical Bayes Estimators." *J. Royal Statistical Society, Series C* 40(2):283–294.
- Neuwirth, E. 2011. *RColorBrewer: ColorBrewer palettes*. R package version 1.0-5. <http://CRAN.R-project.org/package=RColorBrewer>. Last checked 5/18/2014.
- Oliver, D. 2003. *Basketball on Paper: Rules and Tools for Performance Analysis*. Dulles, Virginia: Potomac Books.
- Parker, R. J. 2008. "Tracking the 2008 NBA Playoffs: What the Data Represents." *BasketballGeek.com*. <http://www.basketballgeek.com/2008/08/30/tracking-the-2008-nba-playoffs-what-the-data-represents/>. August 30, 2008. Last checked 5/18/2014.
- Pasek, J. 2012. "Weights: Weighting and Weighted Statistics." R package v. 0.75. <http://CRAN.R-project.org/package=weights>. Last checked 5/18/2014.
- Piette, J., S. Anand, and K. Zhang. 2010. "Scoring and Shooting Abilities of NBA Players." *Journal of Quantitative Analysis in Sports* 6(1), Article 1.
- Plafke, J. 2013. "How the NBA's SportVU Ball and Player Tracking Tech Changes the Face of Sports." *ExtremeTech.com*. <http://www.extremetech.com/extreme/169775-how-the-nbas-sportvu-ball-and-player-tracking-tech-changes-the-face-of-sports>. October 29, 2013. Last checked 5/18/2014.
- R Core Team 2013. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Reich, B., J. Hodges, B. Carlin, and A. Reich. 2006. "A Spatial Analysis of Basketball Shot Chart Data." *The American Statistician* 60(1):3–12.
- ThinkBlueCrew 2010. "A First Look at Shot Location Visualizations." *ThinkBlueCrew.blogspot.com*. <http://thinkbluecrew.blogspot.com/2010/07/first-look-at-shot-location.html>, July 31, 2010. Last checked 5/18/2014.
- Waller, L. A. and C. A. Gotway. 2004. *Applied Spatial Statistics for Public Health Data*. Hoboken, New Jersey: John Wiley. 520 p.
- White, R. 2006. "Pattern Based Map Comparisons." *Journal of Geographical Systems* 8:145–164.