Pace and Space: An Alternative Measure of NBA Shooting Prowess

Abstract— **Given the evolution of the game of basketball over decades, we now live in an era in which reliable spacing, especially by means of 3 point shooting, is essential to the success of NBA offense. While it is simple to evaluate a players shooting prowess based on his field goal percentages, this methodology fails to take into account many factors that dictate the likelihood of success prior to a shot attempt. This means field goal statistics alone make for a very noisy way to evaluate the a players true skill.**

In a game with no defenders, every shot would be taken from the location with highest associated shot success probability – namely, right under the basket. In a real game, however, driving to the rim every play is not possible due to obstructing defenders. Instead, we postulate that players travel toward a nearby location where they are more effective shooters – players can be seen as following a gradient ascent method. Furthermore, we hypothesize that good shooters also move to avoid opposing defenders and free up space. To test our hypothesis, we analyzed a sample of over 100,000 attempted shots using SportVU data for over 600 games from the 2015/16 season. This data was released to the public in 2016 by the NBA. The SportVU data provides all players movements leading up to a shot, which we used to extract positions for our gradient ascent as well as locations of the nearest defender to a shooter. Using our analysis, we are able to discover underrated shooters who utilize movements to get to better shooting locations such as Marco Belinelli. Our analysis also affirms the greatness of LeBron James: he is able to navigate to his favorite shooting locations and separate with ease from defenders who guard him.

I. INTRODUCTION

With the growth of basketball as a sport, there have been various styles that have evolved as well. More recently, with the Golden State Warriors winning 3 of the past 4 seasons in the NBA, questions naturally arise as to how they have established such dominance. Immediately, one can observe the explosive scoring prowess of multiple players. The Warriors find themselves with highly accurate shooters in terms of field goal percentages (FG%). This includes Stephen Curry, Klay Thompson, and more recently, Kevin Durant, with FG%'s of 49.5%, 48.8%, and 51.6%, respectively, for the 2017-2018 season. Even more surprising are these players' 3 point percentages (3PT%) for the past season, which are 42.3%, 44.0%, and 41.9%, respectively, for the same season [1]. The accumulation of multiple accurate shooters naturally tends to more wins, or in the Warriors' case, championships.

Our work served to dive deeper into understanding the importance of good offense in the NBA. While we could have simply analyzed at FG% or 3PT% to infer the necessity of shooting prowess, this methodology takes little consideration of the context of shots. A multitude of factors - namely, the other team's defenders - create a significant obstacle for shooters. As such, a significant amount of insight is lost in judging a player solely by their FG%. Other related works have recognized this as well, and focused on other contexts such as the need for teamwork in passing for offensive prowess [2]. Other work has considered positional data to show the depth hidden behind such relatively simple statistics, especially in terms of rebounds [3].

Our work chooses to focus on two specific aspects of the game context surrounding shots using positional data. First, in section II, we discuss the specific data source used and the required processing. This leads into section III, where we discuss our gradient ascent method, and section IV, where we discuss defensive coverage. We then compare the two metrics in section V and conclude in section VI.

II. DATA USED

Data were taken from SportVU data, which provides locations for each player and the ball 25 times per second per game [4], which the NBA made public in 2016. We used a mirror to the data from GitHub for over 600 games, as well as data files that had already extracted over 100,000 attempted shots [5]. The SportVU data organization made it such that each of these shots were classified as specific *events*. For each event, there were a set of *moments*, which gave the specific player and ball locations at each time stamp. The sheer size of the SportVU data made it difficult to process it all, so we chose to sample 2.5% of the moments per event. From these data, we could approximately trace out the paths of the players leading up to a shot. Furthermore, we only look at **jump shots** throughout this paper and not layups or dunks.

These data were useful for our two pronged approach in the analysis. Being able to approximately trace the shooter's path allows for our analysis in section III to look at the shot probabilities. Further, knowing the positions of all 10 players means that we had a reliable way of assessing the defense's proximity. This aspect of the data is specifically addressed in section IV.

III. GRADIENT ASCENT

Our first metric to measure the effectiveness of shooters will involve gradient ascent. Specifically, using the SportUV data, we first created a surface over the basketball court that represented the probability of making a $jumpshot$ at various locations along the court. We then analyzed players who have taken more than 100 shots and determined how they moved to a better position with respect to the surface when they took a shot. This allowed us to categorize which shooters are effective in their movements. Our methodology is presented in section III-A and our results for this metric are presented in section III-B.

A. METHODOLOGY

We first removed all shots except *jumpshots* from the given SportUV data. We combined these shots to form a surface over the basketball court where the value of the surface at a particular location represents the percentage of shots that went in from that particular location. To create this surface, we first binned the court into a square mesh with side length a tuneable parameter which we set to 1 feet. We then bucketed all the jump shots according to the square they landed on to get a value for each square which represented the probability of making a jump shot from the particular square. Smoothly interpolating this data results in the left half of Figure 1.

After arriving at our probability surface, we looked at players in our data set who had taken more than 100 shots. Then for each such player and for each jump shot the player took, we looked at the location of the player approximately 15 seconds prior to the shot and the final location of the player when they took the shot. We then measured the value of these two locations and then averaged this value across all of the shots the player had taken. Fundamentally, the final value measures the amount of 'gradient ascent' the player has done in route to taking his shot. We call this the player's **gradient score**. Essentially, the gradient score represents an increase in the probability of making a shot due to the player's movements. For example, consider the surface show in Figure 1. If a player moves from a light red region to a white region before taking a shot, this would signify that the player has increased his chances of making a shot by around $5\% - 10\%$. Therefore, if the gradient score is large, this signifies that the player is good at moving to better locations on the court. The distribution of this value across all players is show in the right half of Figure 2.

B. RESULTS

The ten players with the highest gradient score and the ten players with the lowest gradient scores (out of all the players that had taken more than 100 shots in our data), are shown in Tables I and II respectively. Our analysis shows that the players in Table I are good at moving to a better location right before taking a shot. As stated in section III-A, this is measured by looking at the increase in the value of the surface in Figure 1. However, we have to note that this surface was created by taking into account **all** the jumpshots taken. Hence, this explains why some of the well known shooters, such as Stephen Curry, do not have the highest gradient score. These exceptional shooters have other unique talents that are not reflective in the average player such as the ability to shoot over very close defenders. This highlights one flaw in our metric which is that we should be looking at surfaces **conditioning** on each individual player. However, we did not have sufficient data for this purpose.

This also explains the appearance of certain players in the table of players with the worst gradient scores. For example, the catch-and-shoot style of a player such as Klay Thompson does not involve a lot of moving to get to a better location (but his movements are key in him finding open space). In fact, our methodology speaks even more to Klay Thompson's prowess as a shooter, as the intelligence of movements are scored by the overall league's shooting percentages in each zone. Per our metric, Klay Thompson does not move to better areas to shoot but his FG% and 3PT% show that he simply does not need to. As such, our results here must be taken in context of other statistics to tell the full story.

However, a great advantage of our methodology is identifying players who are not well known to be especially gifted at shooting but have very intelligent movements which aids and enhances their shooting abilities such as Marco Belinelli. We believe that this aspect of the methodology would be very helpful in identifying promising young talent who are still very raw with respect to their shooting but have high basketball IQ and are able to get to more optimal shooting locations in the basketball court.

1. Jimmy Butler: 6.38 %	6. Kentavious Caldwell-Pope 2.82 %
2. LaMarcus Aldridge: 5.74 %	7. CJ Miles 2.67 %
3. Marc Gasol: 5.27 %	8. Stephen Curry: 2.61 %
4. Zach LaVine: 4.66 %	9. Serge Ibaka: 2.59 %
5. Marco Belinelli: 3.93 %	10. LeBron James: 2.44 %

TABLE I: Table of the players with the highest gradient score. The values show the average percentage increase in the player's shot going in based on his movement.

1. Klay Thompson: -6.87 %	6. Jamal Crawford: -3.93 %
2. JJ Redick: -5.52 %	7. Kristaps Porzingis: -3.83 %
3. Rudy Gay: -5.18 %	8. Trevor Ariza: -3.60 %
4. Gordon Hayward: -4.18 %	9. Rajon Rondo: -3.52 %
5. Emmanuel Mudiay: -4.10 %	10. Eric Gordon: -3.24 %

TABLE II: Table of the players with the lowest gradient score. The values show the average percentage decrease in the player's shot going in based on his movement.

IV. DEFENSIVE COVERAGE

Our second metric to measure the effectiveness of shooters will involve consideration of the opposing team's defenders. Specifically, using the SportUV data, we can measure the distance from a shooter to the defenders on the opposing team. Our methodology and defensive coverage metric are presented in section IV-A and our results for this metric are presented in section IV-B.

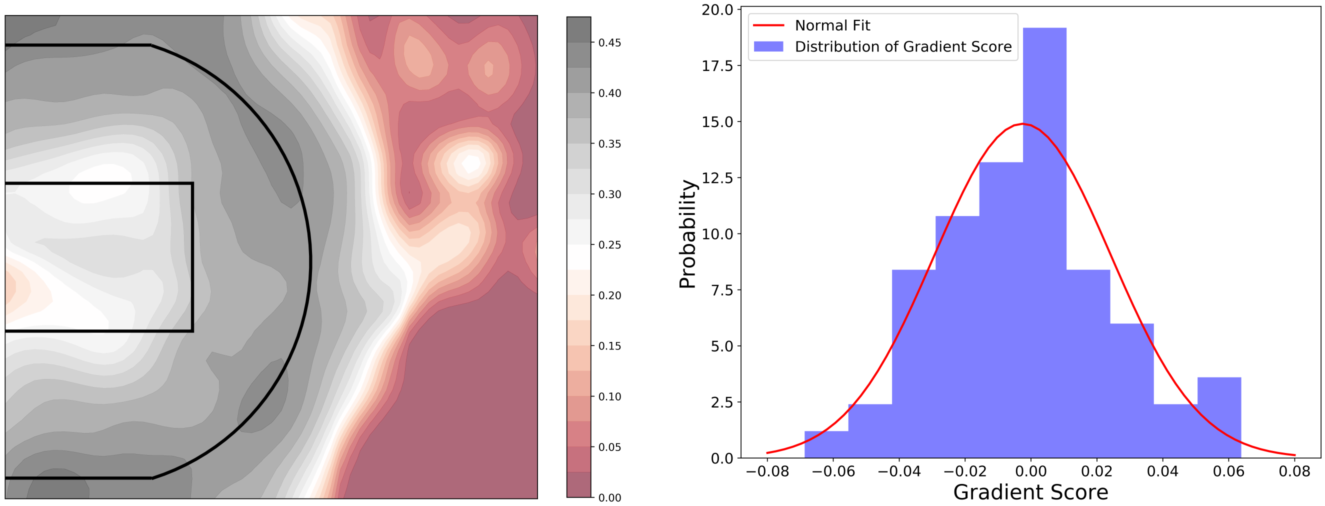

Fig. 1: Left Figure: Surface over one half of a basketball court representing the probability of a shot going in when shot from a particular location. Note that we are only taking into account jump shots. The color represents probability with a legend present next to the figure. Right Figure: Distribution of gradient scores across all players that we have the data and restricted to the players that have taken more than 100 shots. A normal distribution fit is also shown.

A. *METHODOLOGY*

We define **defensive coverage** as the total separation between a shooter and the nearest opposing defender over the course of some period of time (approximately 15 seconds) prior to a jump shot attempt.

We calculated this metric using the following assumptions:

1) The defender nearest to the shooter possesses the greatest influence on the accuracy and comfort of a jump shot attempt apart from the inherent shooting prowess of the shooter.
2) Variance in arm length of an opposing defender is negligible and irrelevant to our definition of defensive coverage.

Let **D** represent the distance between the shooter and the nearest opposing defender at a given moment of time **t**. **D** is calculated in terms of the coordinates provided by the SportVU data for each moment prior to a jump shot attempt. Let (X, Y) represent the x,y coordinates of the shooter and $(\bar{x}, \bar{y})$ represent the x, y coordinates of the nearest defender. Therefore, the Pythagorean theorem can be used to calculate **D** for a given moment:

$$D = \sqrt{(X - \bar{x})^2 + (Y - \bar{y})^2}.$$

For each moment preceding a jump shot, a value of **D**. Therefore, we can think of **D** as a function of t as the shooter is taking his jump shot.

Given these determinations, a plot of **D** against **t** was constructed for each jump shot in the SportVU dataset. The area below this plot was then calculated and deemed the total Defensive Coverage (feet seconds) for the jump shot attempt (see Figure 2). However, given the variability in the total time that defensive coverage occurred on a shot-to-shot basis, this total Defensive Coverage was then divided by the total time in seconds to yield a Normalized Defensive Coverage (feet) for the jump shot attempt. This normalization allowed for direct comparison between all jump shot attempts in our data set. Ideally, a shooter would want this Normalized Defensive Coverage to be as high as possible since that would be indicative of adequate separation from the nearest defender, promoting a sense of comfort when attempting a jump shot.

B. *RESULTS*

After binning Normalized Defensive Coverages for both Makes and Misses, we find that there exists no significant difference between the mean Normalized Defensive Coverage for Makes and Misses among all players. Both histograms can be well-approximated by a Gaussian (normal) distribution as shown in Figure 3.

Furthermore, we ranked the mean Normalized Defensive Coverages for jump shot Makes in order from highest-to-lowest and lowest-to-highest to conceptualize which players are the best and worst at generating separation between themselves and their corresponding defenders while being able to accrue points from these jump shot attempts. Table III and IV both respectively list the players that can individually be clustered into the top 10 of these two groups as well as their mean Normalized Defensive Coverage for their jump shot Makes.

In regards to players who make shots and rank among

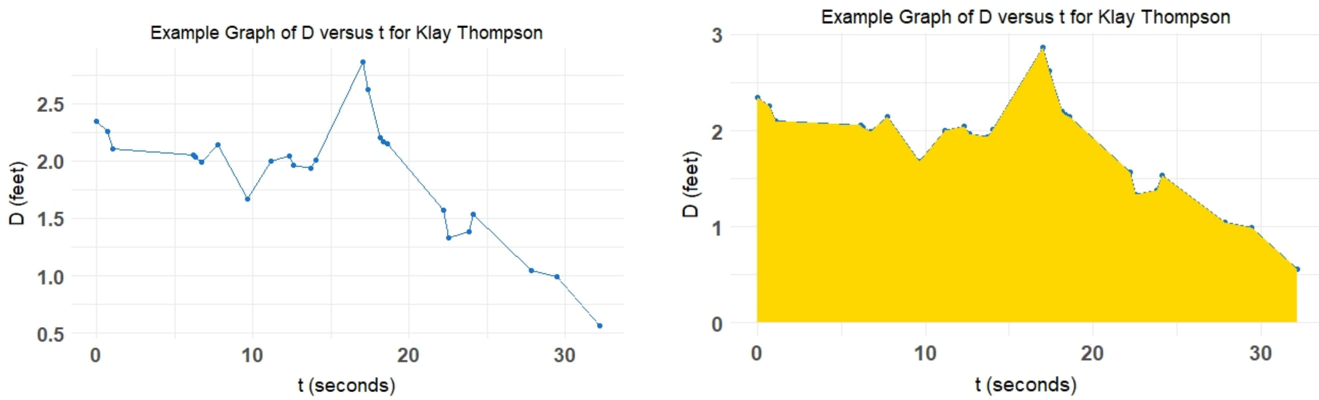

Fig. 2: Left Figure: An example plot of D versus t for the moments that precede a mid-range shot attempt by Klay Thompson of the Golden State Warriors when facing the Los Angeles Lakers at home on January 14, 2016. Right Figure: The area under the curve of the plot was assessed as the total Defensive Coverage for the jump shot attempt.

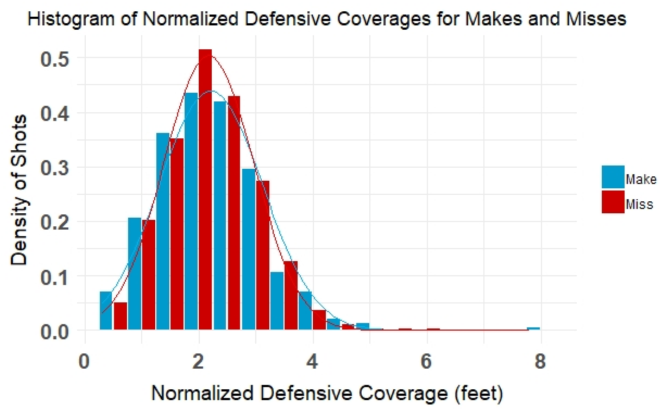

Fig. 3: A histogram showing the distribution of Normalized Defensive Coverages for both Makes and Misses. The solid lines represent Gaussian fits to the data with peaks centered at 2.21 ± 0.04 feet for Makes and 2.17 ± 0.02 feet for Misses.

1. Russell Westbrook: 7.80	6. Enes Kanter 4.39
2. LeBron James: 5.24	7. Kentavious Caldwell-Pope 4.32
3. Trevor Ariza: 4.96	8. Ish Smith: 4.31
4. D'Angelo Russell: 4.90	9. Blake Griffin: 4.25
5. Rudy Gay: 4.44	10. Hassan Whiteside: 4.24

TABLE III: Table of the players with the highest mean Normalized Defensive Coverages for shots they make. These players are the best at maintaining separation between themselves and their corresponding defenders on the opposing team.

1. Ersan Ilyasova: 0.32	6. Serge Ibaka: 0.50
2. Kevin Durant: 0.37	7. Ryan Anderson: 0.53
3. Kawhi Leonard: 0.38	8. Stanley Johnson: 0.56
4. Rudy Gay: 0.39	9. Paul Gasol: 0.57
5. Aaron Afflalo: 0.41	10. Jahlil Okafor: 0.57

TABLE IV: Table of the players with the lowest mean Normalized Defensive Coverages for shots they make. These players are the worst at maintaining separation between themselves and their corresponding defenders on the opposing team.

the highest mean Normalized Defensive Coverages, these players are extremely athletic and capable of maintaining ample separation between themselves and their defenders. For example, Russell Westbrook and LeBron James are widely-regarded as two of the most athletic players in the NBA that can play on the perimeter. Not to mention, players like Rudy Gay and Blake Griffin are also extremely athletic inside the arc.

However, when it comes to those players who exhibit the lowest mean Normalized Defensive Coverages, it is very likely that these players are extremely talented at scoring even when guarded heavily. For example, Kevin Durant and Kawhi Leonard are two isolation-heavy perimeter players who can simply overpower their defenders and generate simple offense. Durant, in particular, is very well-known for his ability to shoot effortlessly over the most reputable perimeter defenders in the NBA, such as Paul George and Jimmy Butler.

V. COMPARING THE TWO METRICS

In this section we will discuss our two metrics as well as their implications. First, as briefly discussed in section III-B, our gradient score metric shows the movement tendencies of the various players and how this affects their shooting. We discovered an example with Klay Thompson, who has a poor gradient score, but this speaks more to his catch and shoot style and accuracy. We also saw players such as Marco Belinelli, who makes intelligent movements towards areas of higher shooting accuracy. With the defensive coverage metric, we saw players such as Russell Westbrook and LeBron James with high scores, which means they both are adept at finding areas where they are comfortable shooting. On the other hand, we see that players such as Kevin Durant and Kawhi Leonard score worse in terms of keeping adequate spacing from other defenders, but are very capable scorers as well.

When assessing the two metrics together, we see that LeBron James is capable of both intelligent movements and adequate spacing for his shots. This agrees with context in the NBA, given LeBron James' overall prowess as a player. In this way, we see that the two metrics highlight two different aspects of a player's shooting ability. While both metrics are reliant on positional data, the gradient score concentrates on the area of the court while the defensive coverage concentrates on the relative positions of defenders in relation to shooters. Both are clearly important in shot selection, so neither can be considered in a vacuum.

It is also important to consider the strengths and weaknesses of each metric. The gradient score metric contextualizes shots in terms of movements to areas of overall higher shooting accuracy. By rewarding players who make more intelligent movements, we can identify those who can potentially be relied on for more shooting. However, this accuracy is calculated by considering the league as a whole, and as we saw earlier, can be misinterpreted, such as in the case of Klay Thompson. As for the defensive coverage metric, we can find players who are more adept at finding openings in opposing defense, and also more likely to make shots. However, as in the case of Kevin Durant or Kawhi Leonard, this is not an impediment to their shooting and speaks even more to their athletic ability to still make shots.

Interestingly, these two metrics can account for each others' weaknesses because they catch points that can be misinterpreted. For example, the gradient score method does not consider that players may be able to shoot with high accuracy from more obscure areas of the court, but the defensive coverage method can show that they are doing this simply because they have already spotted weaknesses in defense. Similarly, a player with a low defensive coverage score could just have tendencies to move themselves to higher accuracy areas and have defenders move to the same areas as well.

VI. CONCLUSION

In this paper we developed two very different metrics to compare the effectiveness of NBA shooters. The first metric used a gradient ascent approach to determine players who use intelligent movement to get to better shooting locations on the court while the second metric looked at separation between a shooter and defenders to determine athletic players who are able to separate from nearby defenders and create space to shoot. The ability to move to good locations in addition to the ability to create space are valuable assets in the modern NBA and our analysis confirms that well regarded players, such as Lebron James, are able to do both of these tasks effectively.

However, there are anomalies to our analysis. For example, we disregard the vast array of different playing styles that very elite players posses. For example, tall shooters such as Kevin Durant do not need to create extra space for shooting that others players may utilize. Nevertheless, our analysis is still useful in reaffirming the greatness of players such as Lebron James in addition to finding highly underrated players who excel at moving to optimal shooting locations and creating space to shoot. In this way, we see the synergy between the two seemingly different metrics to tell a fuller story about the shooting prowess of specific players.

REFERENCES

[1] Basketball Reference. https://www.basketball-reference.com/
[2] Sliz, Bradley A (2016). An Investigation of Three-point Shooting through an Analysis of NBA Player Tracking Data. https://arxiv.org/pdf/1703.07030.pdf
[3] Rajiv Maheswaran, Yu-Han Chang, Jeff Su, Sheldon Kwok, Tal Levy, Adam Wexler, Noel Hollingsworth (2014). The Three Dimensions of Rebounding, MIT Sloan Sports Analytics Conference.
[4] NBA Advanced Stats. https://stats.nba.com/players/drives/
[5] https://github.com/sealneaward/nba-movement-data/tree/master/data