

Deconstructing the Rebound with Optical Tracking Data

Rajiv Maheswaran, Yu-Han Chang, Aaron Henehan and Samantha Danesis
University of Southern California
Los Angeles, CA, 90292, USA.
Email: maheswar@usc.edu

Abstract

This paper leverages STATS' SportVU Optical Tracking data to deconstruct several previously hidden aspects of rebounding. We are able to move beyond the outcome of who got the rebound to discover the non-linear relationship between shot location and its impact on offensive rebound rates, implications of the height of where rebounds are obtained, and estimates of where players should move in order to improve rebounding rates. We also leverage machine-learning methods to estimate the predictability of rebounding.

1 Introduction

Pat Summitt, coach of the Tennessee Lady Vols, has been quoted, perhaps apocryphally, as saying, "offense sells tickets, defense wins games, rebounding wins championships." [1,2,3] Rebounding is advocated as a fundamental characteristic of success by coaches [4] and is a key component of many advanced metrics evaluating both player and team performance [5,6]. While some may admit to the subjectivity of a credited assist, the fundamental nature of a rebound is less questioned. After all, the player credited with a rebound typically ended up with the ball in their hand. Aided by optical tracking data from STATS Inc.'s SportVU system [7], we employ a data-driven approach to discover previously hidden intricacies about what happens from the time a missed shot is released to the time it ends up in a player's hand.

2 Methodology

Our ability to understand any phenomena is limited by the observations we are able to obtain in a systematic manner. The individual human mind translates what it perceives into different hypotheses about the underlying physics of the phenomena and optimizes performance based on these factors. In basketball, to build a common scientific grounding to test hypotheses, we had been constrained by the observations found in box scores. Unfortunately, these observations are abstractions themselves with varying degrees of reliability ranging from the assist, shown to be highly subjective [8], to points, which are accurate in the sense that those credited with the points did almost always score them. Like points, rebounding assignment is generally highly accurate but the value associated with them is highly dependent on the context of how the event came to be. Metrics based on these observations propagate any uncertainty or false valuation associated with them.

In order to build a richer model of rebounding, we need information such as where was the shot taken, where did it go when it was missed, who was around, and what happened when it got there. In the language of Probabilistic Graphical Models or Dynamic Bayes Networks, we have two observable variables (i.e., field goal attempts and rebounds) that are parts of a temporal sequence involving several hidden variables (i.e., the locations of the ball and players between these two variables). Our goal is to identify the transitions to and from these hidden variables to extract their impact on both the strategic aspects and credit assignment for rebounding.

While other researchers have attempted to look at these phenomena leveraging game-charting [9], they haven't had the sample sizes to do larger-scale studies and have relied on subjective measures of context and placement. STATS Inc.'s SportVU optical tracking data records the $\{x,y\}$ positions of all players on the court at 25 frames per second annotated with relevant play-by-play events. Thus, we are able to augment any event of interest with spatial coordinates of all players. More significantly we are able to leverage data tracking $\{x,y,z\}$ positions of the ball, which allows us to calculate the height at which a particular event happens or the frame when a particular height is reached. In addition to augmenting events with positions of players and the ball, we can build data sets describing where the ball *could* have been rebounded as opposed to where it was rebounded. The data is obtained as a large collection of tuples for each of 14 entities (10 players, 3 referees and the ball) annotated with various events. We built a suite of filtering, annotation and visualization capabilities to enable us to understand this data. Examples of this are shown in Figure 1.

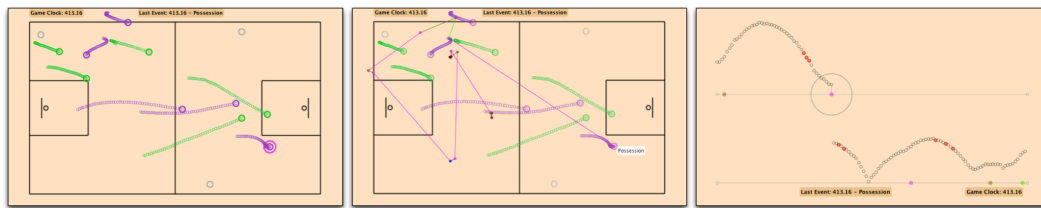


Figure 1. Visualizations of players moving over time (Left), annotations of events and locations (Center), and the height of the ball (Right) over time.

Our research focuses specifically on *missed field goals*. In the data set, we were able to extract over 11,000 such events. To address the noise inherent in the data, we eliminate any event that associates an event with a location outside the court of play and any shot attempts that are taken beyond 30 feet of the hoop. We filter the data further based on the specific questions at hand as described below. We employ a variety of data-driven visualization and machine-learning methods focused on extracting new understanding about rebounding.

3 Rebound Location

We start by understanding the spatial location distribution of where rebounds are credited. This will help inform us of any biases that we may have due to our observations of where rebounds occur. STATS SportVU technology uses a combination of the distance of the ball to various players as well as the box score credit of the rebound to identify the most likely point at which a rebound occurred. In Figure 2, we show the spatial distribution of rebounds. We also show the offensive rebound percentage in each sector of the half court, shown as both a table and a heatmap. We only display results for sectors with over 50 events.

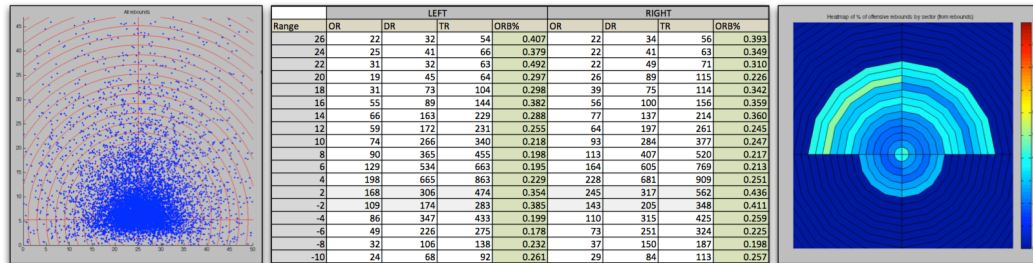


Figure 2. Spatial distribution of rebound locations (Left), A table of offensive and defensive rebounds by sector (Center) where the range denotes the upper limit of a two-foot sector in either the left or right half of the court [negative numbers for ranges indicate locations between the hoop] and the baseline, and a heatmap of offensive rebound rates (Right).

We see that most rebounds occur close to the hoop, but that rebound locations do spread over the entire court with over 100 instances between 24-26 feet from the basket. If we calculate the offensive rebound rates as a function of the sectors, i.e., the probability of an offensive rebound conditioned on the location of the rebound, we see that within two feet of the hoop, there is a high chance that the rebound is an offensive rebound (40%). From 2-10 feet, the offense is unlikely to have been the rebounder (22%) and from 10 feet onwards there is an increasing likelihood of an offensive rebound as a function of distance, reaching 40% again in the 24-26 foot sectors. These probabilities align with general expectations that offensive rebounds happen close by (tip ins) or further away from the basket. Unfortunately, rebound locations are the end of a sequence of connected probabilistic events and not useful for strategic purposes. To gain strategic value, we must go to where the chain of events begins, namely the field goal attempt.

4 Shot Location

One of the first strategic aspects of an offense is the type and location of the field goal attempts they try to create with their plays. Typically these plays are constructed to maximize the expected points per attempt. However, the location of the field goal attempt is not independent of the likelihood that one can recover a miss. While many intuitively understand this phenomenon, we have only vague notions such as “long shots lead to long rebounds” to characterize these effects. By conditioning the rebound likelihoods on the shot locations, we can now get explicit characterization of the relationship between shot location and offensive rebound percentage. In Figure 3, we show the spatial distribution of all shots, and the offensive rebound percentage as a function of each sector of the court where shots were taken. This is shown as both a table and a heatmap. Again the display is limited to sectors with over 50 events.

We see that the best location to attempt a field goal to maximize the chance of an offensive rebound is close to the basket. Shots attempted within 6 feet of the basket and missed are recovered at a 36% rate. From 6-10 feet, the rate becomes 28%. There is a significant change in the 10-22 foot range, where missed shots are only rebounded at a 21.5% rate. There is also an interesting transition at the three-point line where rebound rates outside 22 feet jump up to 25.5%.

It is not clear using heuristic reasoning that shots close to the basket would yield high offensive rebounds. One might assume that longer shots that are missed will return the ball to a region where

offensive players can corral them or conversely that close shots are good because the shooter is close to the possible rebound location. We verify that it is the latter effect and quantify its exact value. Furthermore, we identify that mid-range shots lead to poor offensive rebound rates. It is interesting to note the sharp increase in rebound rate for shots past the three-point line. One hypothesis is that three-point attempts may lead to greater run-outs, which remove a potential defensive rebounder; but this needs to be tested. It remains an open question as to why, but the effect is clear.

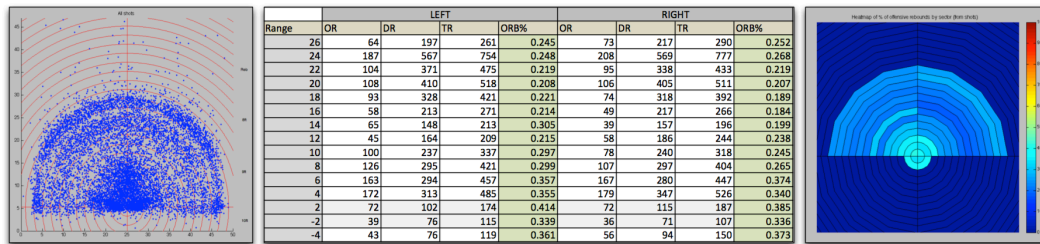


Figure 3. Spatial distribution of shots locations (Left), A table of offensive and defensive rebounds by sector (Center) where the range denotes the upper limit of a two foot sector in either the left or right half of the court; negative numbers for ranges indicate locations between the hoop and the baseline, and a heatmap of offensive rebound rates (Right).

We note that offensive rebound rates decrease as a function of shot distance with a sharp jump at the three-point line. This is very similar to effective field goal percentage as a function of shot distance. This result implies that mid-range shots are even worse than previously characterized due to their effects on offensive rebound rates. Strategically, teams have even more reason to eschew mid-range shots for shots closer to the basket or three-pointers.

5 Rebound Height

One of the interesting aspects of the data set is the ability to identify the height of the ball when the rebound was credited. To get this information, we developed tools that can tell when the ball descends below a particular height level after it hits the rim and bounces. This was used to enable analysis of where the ball *could* have been rebounded if players were positioned differently. Certain shots, e.g., air balls and shots that were rebounded above the rim, would not yield these markers. We thus constrained the set of shots to those shots which bounced upwards after hitting the rim and then descended below 11 feet. This filtered set has 9122 events and is used to generate Figure 4.

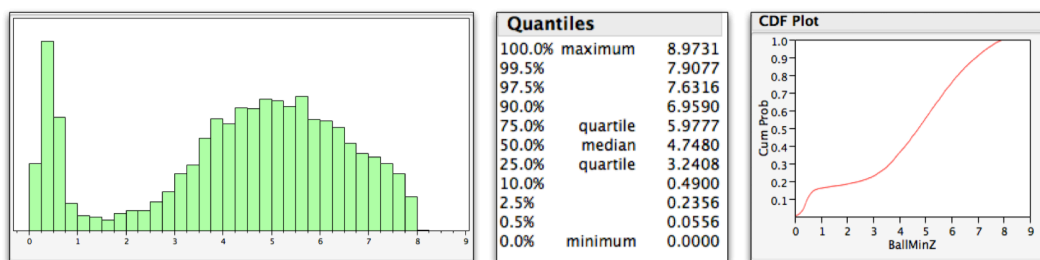


Figure 4. The distribution of the height of the ball when it is rebounded (Left), Samples of various values of the cumulative distribution function (Center) and the complete cumulative distribution function of rebound height (Right).

Among these shots, we see that almost all were rebounded below 8 feet. To put that in context, the standing reach of almost all players drafted over the last decade is above 8 feet [10]. The fact that the ball is rebounded below this height indicates that the rebounder was not in the ideal position to get the ball, i.e., they had to move to get the rebound and in theory someone else could have gotten it if they had made it to that point sooner. We note that 15% of rebounds had a height of less than 1 foot indicating, due to sampling rates, that it hit the floor. A strategic shift that would obtain a fraction of these rebounds could have a significant impact on offensive rebound rates.

6 Potential Rebound Location

The ability to calculate where the ball went below a particular height enables us to figure out where the ball *could* have been rebounded if players were positioned in a different manner. We choose 8 feet as the height where the ball could have been rebounded, based on the information in Figure 4 and the standing reach data. We calculate the location of the ball when it crosses its threshold and display this information in Figure 5.

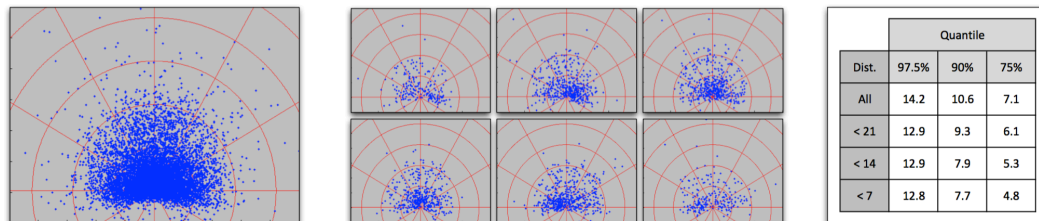


Figure 5. Distribution of the ball locations where it descends below 8 feet for all shots (Left), distributions of locations where the ball descends below 8 feet conditioned on shots from 21-28 feet in 30 degree sectors (Center) and samples of the CDF of ball distance from hoop when the ball descended below 8 feet, conditioned on the distance of the shot location.

One of the interesting results is that there is not discernable bias in terms of direction based on shot location. There are theories that advocate that corner threes tend to be rebounded on the long side of the shot. In the center figure of Figure 5, we show the scatter plots of the 8-foot crossings of shots taken from 21-28 feet, i.e., three pointers from the six sectors of 30 degrees each. We propose that it is not possible to discern which sector is associated with which scatterplot by inspection thus indicating that there is a consistent uniform distribution of rebound angle and distance. We can verify this by looking at the means and entropies of the angles and distances in the 8-foot crossing locations.

A second important result is that the ball reaches the 8-foot crossing relatively close to the basket. The sample CDF shows almost all shots fall within 14 feet of the basket (97.5% of all shots fall within 14.2 feet). This radius reduces dramatically as the shot distance decreases. What this implies for strategy is that if rebounding is a concern, players should move much closer to the basket after a shot is taken. The angle is not significant as it is essentially uniformly random.

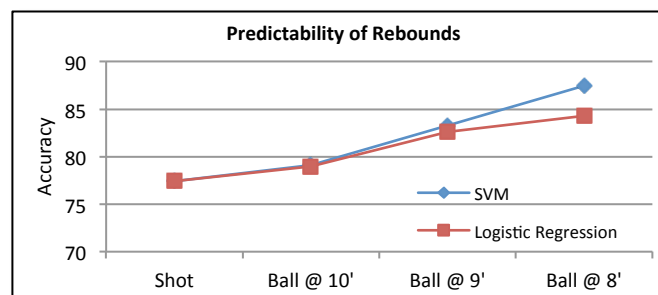
7 Predictability of Rebounds

Thus far, we have examined the factors that contribute to a team successfully rebounding the ball. Armed with our knowledge of these factors, we can also ask the question: given information about the players and the ball at a particular point in time after a shot is taken, can we predict with high

accuracy which team will get the rebound? We might expect that the predictability is least at the moment the ball leaves the shooter's hands, and that it is greatest just before the eventual rebounder grabs the ball. Using a suite of *machine learning* algorithms that learn from the existing data and build predictive models using that data, we show that this is indeed true. We show the results using logistic regression and support vector machines below.

Clearly, the predictability of the rebound increases as the ball gets closer to being rebounded. While the predictability starts out near 77% when the ball leaves the offensive shooter's hands, the probability that we can predict the correct team who will make the rebound increases to 87.5% by the time the ball has fallen back down to the 8-foot level, which is the altitude at which most players can begin to grab and control the ball. The support vector machine (SVM) is a state-of-the art machine learning technique that has been shown to work well in a wide variety of settings, with the important trait of avoiding over-fitting [11]. For this data, we used an implementation of an SVM provided by the WEKA machine-learning program [12]. The SVM is able to outperform simple logistic regression by several percentage points in terms of accuracy of its predictions. The graph below shows performance of an SVM with radial basis function (RBF) kernels, which enables the SVM to separate between offensive and defensive rebounds using a hyper-plane in a higher-dimensional space created using this non-linear kernel. Hence the SVM can create non-linear class separators when the solution is projected back down to the original search space.

Currently 12.5% of data-points that are not predictable, i.e., our predictions were incorrect 12.5% of the time. This may be due to inherent randomness or omitting a relevant feature. We expect that as we develop richer variables that better describe our data, we will be able to push the limits of predictability higher.



7 Summary and Future Work

By leveraging a data-driven approach on optical tracking data, we are able to discover several insights about the nature of rebounding: (1) the effect of shot location on rebound rate that exacerbates the ineffectiveness of mid-range shots and improves the significance of three-pointers and close shots, (2) the lack of significant directional bias in rebounds, and (3) quantification of the radius of the potential rebounds, showing that it is clustered closer to the basket than expected. We are in the process of calculating additional metrics such as the value of boxing out, and evaluating the movement patterns of players after the shot to find additional insights into rebounding. Ultimately, the data could allow us to find a more appropriate rebounding measure, i.e., the marginal improvement to team rebound rate due to individual or collective actions, because now we can understand what could have happened as opposed to only what did happen.

Acknowledgments. We would like to thank STATS, Inc. for sharing their data and for their help in curating it. In particular, we would like to thank Brian Kopp, Mitch Tanney and Nicole DeFord.

References

- [1] <http://www.basketball-tips-and-training.com/basketball-quotes.html>
- [2] <http://www.championshipcoachesnetwork.com/public/444.cfm>
- [3] http://sports.espn.go.com/nw/ncaatourney07/columns/story?columnist=hays_graham&id=2824894
- [4] <http://www.coachlikeapro.com/basketball-coaching-quotes.html>
- [5] “Calculating PER”, <http://www.basketball-reference.com/about/per.html>
- [6] Berri, David J., Martin B. Schmidt, and Stacey L. Brook. 2006. *The Wages of Wins: Taking Measure of the Many Myths in Modern Sport*. Stanford University Press.
- [7] STATS SportsVu, <http://www.spovu.com/>
- [8] “The NBA's Most Misleading Number”, Wall Street Journal, April 1, 2009. <http://online.wsj.com/article/SB123855027541776617.html>
- [9] “Game Charting Insight: Contested Rebounds”, <http://82games.com/rebounds.htm>
- [10] Draft Express Pre-Draft Measurements, <http://www.draftexpress.com/nba-pre-draft-measurements/?page=&year=All&source=All&sort2=DESC&draft=100&pos=0&sort=6>
- [11] Pattern Recognition and Machine Learning, Christopher Bishop, Springer 2006.
- [12] WEKA, <http://www.cs.waikato.ac.nz/ml/weka/>