Enhancing NBA True Shot Charts: A Probabilistic Spatial Approach to Rebounding and Possession Value

Basketball
20251396

1. Introduction

It's the last five seconds of regulation, and your team trails by one point, but the ball is in your hands. Tick, tick. The defense is physical and smothering. Tick. You pull up for the game winning shot. Tick. Ball hits back rim and bounces out... BEEP! What happens next? Hopefully your big got the putback, otherwise, time to pray to the refs for a foul. Everything comes down to Plan B. Traditional shot charts, which focus solely on made baskets, fail to capture these crucial dynamics.

To overcome these shortcomings, recent research has introduced the concept of true shot charts incorporate expected free throw value, which has revealed an overvaluation of three-point shots in 'make-only' models [1]. Building on this progress by integrating the concepts of 'court realty' and hexmap rebound distributions, our paper introduces a probabilistic spatial model to propose a novel metric: Rebound Rate Over Expected (RROE) [2,3]. This model extends continuous true shot charting with expected offensive rebounded points-per-possession (PPP), providing additional game state context and offering a more comprehensive view of shot value in basketball analytics that accounts for secondary outcome point potential.

2. Data & Methods

This research is based on a 10-game sample of center of mass tracking and accompanying event labels, provided courtesy of the NBA, from the 2023-24 season [4]. Below is a figure breaking down the field goal attempts contained within.

Maheswaran, et al. employ the most similar approach in literature, leveraging Voronoi tessellation along with spatial probability distributions to deconstruct rebounding into three dimensions Positioning, Hustle, and Conversion [5]. While they successfully show how spatial dynamics adjust between shot time and rebound, our work focuses more directly on the Positioning base case and a player's ability to overcome expectations and impact their team with Hustle and Conversion.

2.1 Rebounding Metric
The first of two core techniques leveraged in the rebounding metric is rebound distribution hexmaping as implemented by Jacobs [3]. These plots embed the rebounding density across cells and illustrate the distinguishable patterns that result from missed shots originating in distinct regions of the court. Furthermore, these patterns are leveraged to project a team's expected PPP from second chance opportunities.

Next, we make use of Voronoi diagrams to create court realty plots of an NBA court, a method utilized by Cervone, Bornn and Goldsberry [2]. The guiding insight here is that space is more or less valuable depending on where it is and who is occupying it. A Voronoi plot is drawn with cell

borders equidistant between all players on the court. The resulting regions are 'owned' by the player contained in them. We similarly use Voronoi cells to assess the rebounding 'real estate' value the player was able to create at shot time.
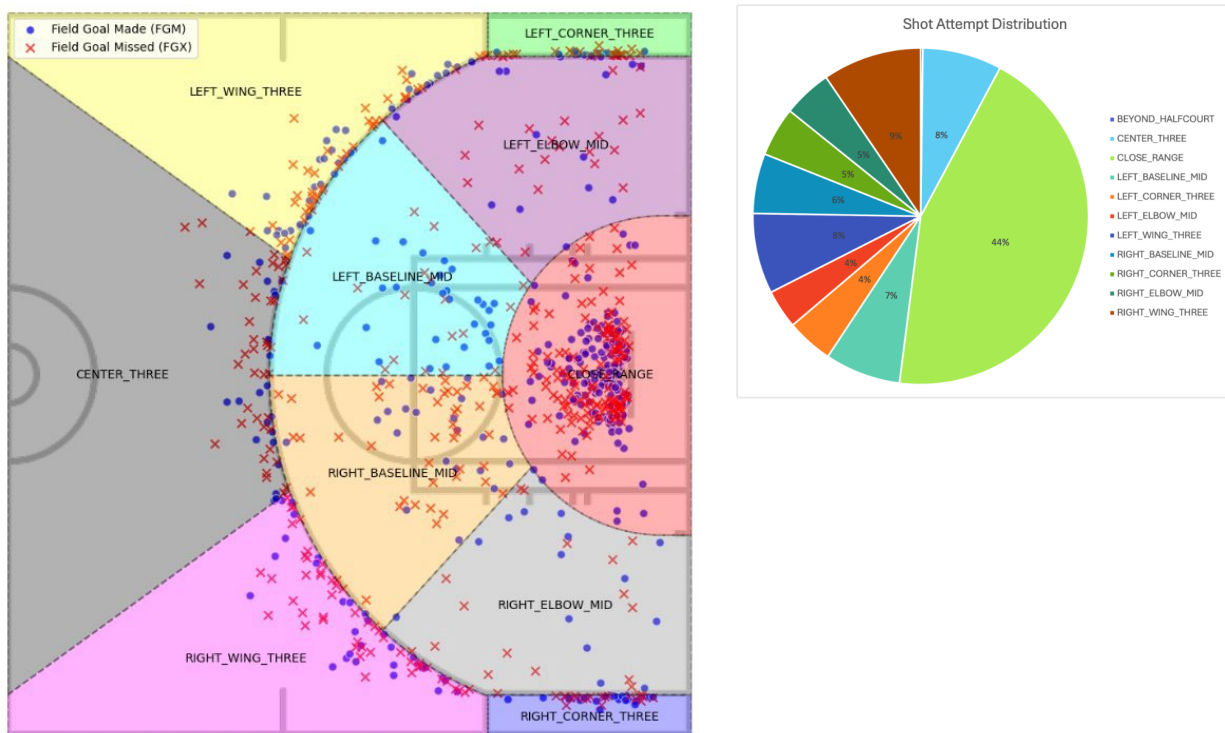
The attempt distribution reflects the modern philosophy of threes and layups, with *~78%* of attempts occurring either from close range or behind the arc.

2.2 Continuous Shot Chart

The basis for our shot charting comes from the generalized additive model approach used by Ehrlich and Sanders to create true, continuous shot charts. To accomplish this, they incorporate the secondary impacts of missed shots by augmenting base PPP with an expected free throw value based on foul propensity in different shot regions. They plot the value model topographically on a half-court to show the 'peaks' and 'valleys' in true shot value. The major finding is team's actually take too many threes, as these attempts are less likely to generate free throw value, resulting in a "three-point premium" since the 2018-19 NBA season.
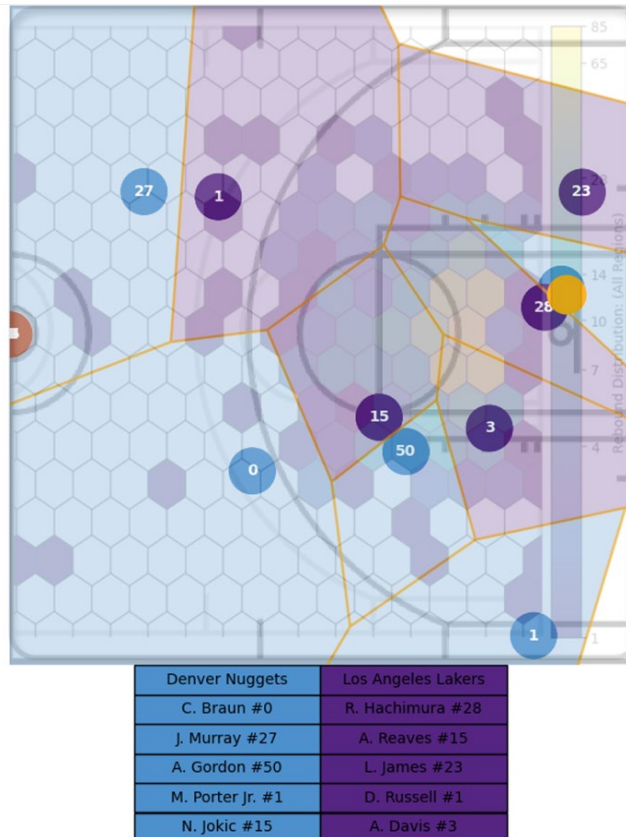
3. Design & Implementation

This project is written in Python and all of the code can be found here: github.com/dkstephanos/TrueImpactShotCharts. Due to the small sample size, the overall rebound distribution across shot regions was used to inform the model. With more data, future work could look at the impacts of shot region specific distributions on expected rebounds.

3.1 Rebounding Model & Metrics

The rebound model is a spatial model that leverages the rebound distribution and Voronoi court realty plots. At shot time, each player is given a percent chance of securing the rebound by assigning ownership to hexagons in the rebound distribution that their Voronoi polygon covers. This is done for each shot attempt using the real locations of the players on-court. When aggregated, a player's actual rebounds are compared to his projections to create the RROE metric.

Denver Nuggets	Los Angeles Lakers
C. Braun #0	R. Hachimura #28
J. Murray #27	A. Reaves #15
A. Gordon #50	L. James #23
M. Porter Jr. #1	D. Russell #1
N. Jokic #15	A. Davis #3

The figure above illustrates the overlay of spatial plots for an individual shot attempt. The hexmap used in the overlay represents the full distribution off all collected rebounds in the sample.

Once these projections are created, we augment the average PPP for shots following an offensive rebound and add that to the expected free throw value based on the foul rate for the shot location. This final augmented PPP is used to generate the continuous true shot charts.

3.2 Continuous Shot Chart

To inform the continuous shot charts, we first calculate the points generated on each possession. Next, we create the true points metric by extending the points produced with the value add from made free throws on shot attempts [1].

We then propose our own metric, true impact points produced (TIPP), which embeds the potential points derived from offensive rebounds to account for all scoring secondary outcomes. We leverage the rebound model detailed in section 3.1 to do this and use the player estimates to provide a

team's percent chance of securing an offensive rebound per shot. These percentages are summed together as expected points by multiplying them by the average PPP of shots following a miss.

To remove bias, we calculate this value directly from the dataset and find it to be *1.102*. The resulting point estimates are then fed into the topographical mapping algorithm. In the map, higher level elevation corresponds to greater PPP projections for shot attempts in that location, and is colored red. Lower elevation regions are blue.

In an effort to overcome the small sample size and smooth out the visualizations, null values were filled with a percent of the mean, based on distance from basket (with a value adjustment for 3-point attempts), and Gaussian filtering was applied to the final result.

3.3 RROE
Traditional rebounding statistics track raw counts across offensive and defensive possessions. To account for opportunity, play-by-play data is used to extend raw total rebounds into an aggregate rebound rate (REB Chance%) [6]. This is done by dividing a player's real rebounds by the number of rebounding opportunities that occurred while that player was on the court. Using our rebounding model as a baseline for expectations, we propose Rebound Rate Over Expected as an extension by substituting the real rebounds collected with the net in real and expected.

We define RROE per player as such:

$$RROE = \sum_{k=0}^{n} \frac{(R_k - E_k)}{n} \tag{1}$$

Where *n* is the number of rebound opportunities, R_k is either 0 or 1 depending on if the player actually got the rebound (1) or did not (0), and E_k is defined as:

$$E_k = \sum_{c \in H} P(r \mid x,y) \cdot I(V_c(p) > 0.5) \tag{2}$$

Where *H* is the set of hex cells on the court, *P(r | x,y)* is the probability that a rebound lands in the hex cell, and *$V_c(p)$* is the percent coverage of that player's Voronoi cell with the hex cell. This means that on every rebound opportunity, each player on the court will receive an expected rebound score. As a result, a value adding rebound is one that is secured with a low estimate, thus maximizing the numerator in the metric. A rebounder who frequently secures the board but who also begins opportunities with significant real estate and thus a high estimate will have smaller signals in comparison.

4. Results

4.1 Rebounding Metric
First, we look from the perspective of team defensive rebounding, and find that the model shows strong predictive ability in picking the rebounding team, demonstrating a Brier score of *0.0767.*

Next, we created a receiver-operator characteristic curve and found an area under curve of *0.98*, as shown below. This is fundamentally a comparison of the true and false positive rates, and scores close to 1 suggest a well discerning model.
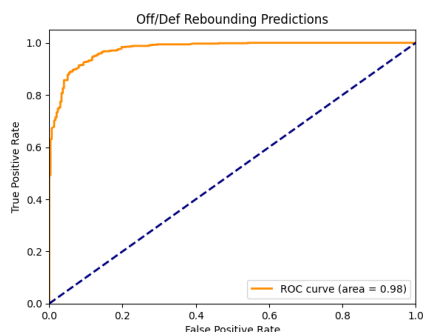
On a player level, our average expected rebound rate of *10.19%* was close to the real rate of *9.97%*, with a net-overestimation of *2.19%*. This suggests our predictions are relatively stable around the mean. One potential explanation is team rebounding. A team rebound occurs when a shot goes out of bounds, or other scenarios in which no individual player can be credited, since our model does not project rebounds outside the bounds of the court.

Remaining gaps could be slimmed by incorporating player biometrics or rebounding skills into the Voronoi cell plot process, but it is interesting to note the predictivity of the model while only taking into account relative positioning of players. This could be a natural by-product of court realty, as players are always attempting to optimize their controlled space, and a player's size and ability factor heavily into these attempts.

4.2 Continuous Shot Chart

Our primary interest in generating the topographical shot charts is to compare trends in our point estimate model to PPP and similar value add models.

4.2.1 Comparing Point Estimate Models

First, we look at the expected value of shot attempts across regions and metrics. While true points and true impact points are both additive models, the margins are different across shot region. This is due to fluctuations in foul and offensive rebound propensity when accounting for different shot attempts.
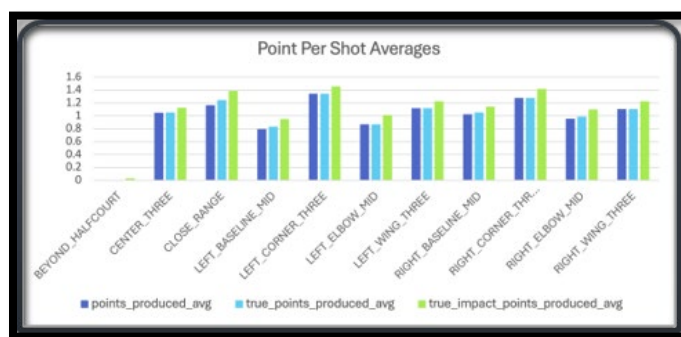
The spatial nature of the model also reduces the risk of overfitting with the small sample. Additionally, when compared to traditional points produced and the free throw augmented true points produced metric, our TIPP metric shows greater sensitivity to the dataset, as there are more miss examples in which no foul occurs.

4.2.2 Comparing Topographical Maps

We generated continuous shot maps for raw points produced, as well as the true points and true impact points models, below. Red indicates higher expected value in terms of PPP.
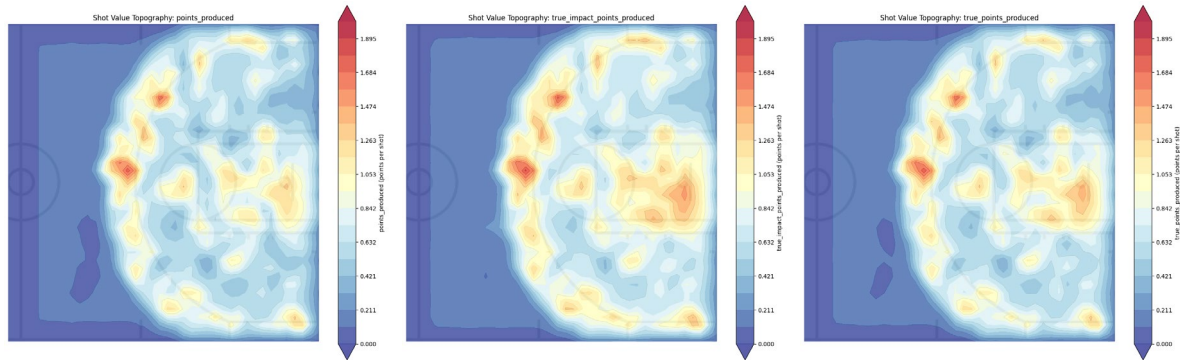

Some of the trends observed by Ehrlich and Sanders [1] bear out, such as increased elevation around the painted area, however, given the small number of mid-range attempts, there is little noticeable movement in these areas, and thus it is difficult to respond to their claim of over-indexing on 3pt attempts at the expense long twos.

A noticeable similarity between the true impact model and true points is the relative value added to 3pt misses. Since TIPP is an additive model, all the point estimates are naturally higher, but the uniformly higher elevation around the 3pt line in the true impact model suggests offensive rebounding is having an impact here. Longer 3pt shots typically result in longer rebounds [7], so this suggests that teams are finding value by crashing the boards in these scenarios, potentially *further* motivating an increase in 3pt attempts.

4.2.3 Plotting Net Gains

Finally, we plotted the net gains in our true impact model compared to actual PPP. We find some moderate gains around the 3pt line, but the primary hotspot is just to the right of the rim, with other high elevation real-estate concentrated in the paint and the dunker's spot. This means that shots in these areas are much more likely to derive value from secondary outcomes such as fouls and offensive rebounds. The right-leaning skew is likely attributable to the general right-hand dominance of players in the NBA.

Shot Value Topography: net_value
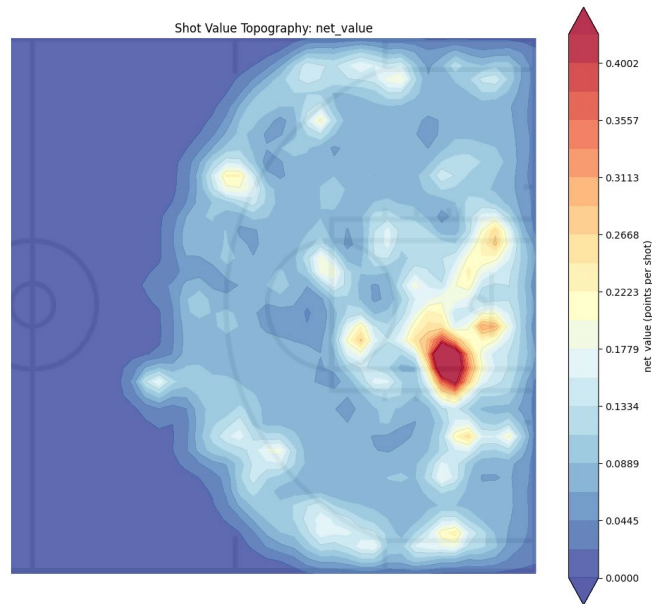
While it is hard to be conclusive with a small sample, there is a trend towards strong gains around the rim and the relative gains on most long-distance shots. In general, this seems to support a shot-on-goal prioritization strategy, as the net impact of expected points due to secondary outcomes spans between 0.1-0.4 PPP in the most frequently attempted regions, which is a strategically significant margin.

A possible explanation for the relative difference in observed rebound rate between close range shots and longer attempts, in particular 3pt shots. The average missed 2pt attempt in the dataset was corralled by the offense *32.23%* of the time, compared to *28.54%* for 3pt misses. This divide grows even more apparent when your compare close range shots (*38.73%*), to the unimpressive *21.0%* on mid-range misfires. Our model is likely picking up on this dynamic in the visualization, and giving a greater expected offensive rebound points estimate for threes and layups. This may indicate somewhat of an inverse effect to that observed by Ehrlich and Sanders, who found that the true points model showed an over-indexing on 3pt attempts when adjusting for expected free-throw value. What we can say for sure is that within our 10-game sample, teams were able to derive more potential value from second chance opportunities originating from a three than a mid-range attempt.

4.3 Rebound Rate Over Expected

One important thing to note is what this metric tells us about rebounding value. Our prediction is based upon court realty owned by a player at shot time, and thus a player performing beyond expectation in our model is successfully collecting rebounds that fall outside of their 'real estate'. We should then expect to see players that provide rebounding value with their athleticism and crashing ability to stand out in these results, whereas a player who excels at boxing out and creating space before the shot attempt might underperform relative to the model's expectations. This does not necessarily mean that one of these players is a better overall rebounder than the other, but there is a distinction in how these types of players are contributing on the boards.

4.3.1 Comparing to Raw Predictions

We first compared the RROE to the raw predicted rebounds totals. Since RROE is an opportunity adjusted metric, we see a lack of correlation with the raw prediction numbers. This suggests that our metric is capable of discerning value created by both high- and low- volume rebounders, as show in the figure below.
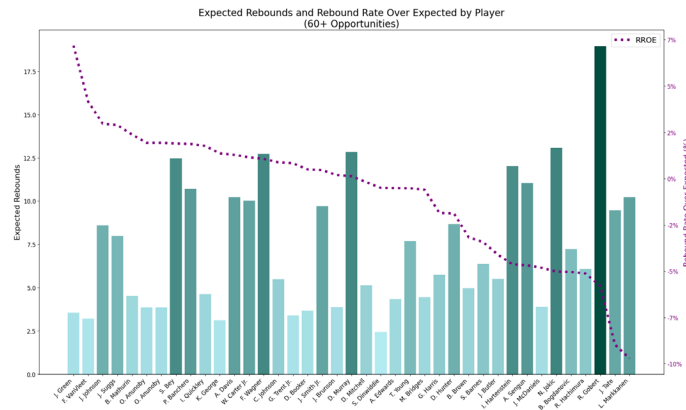

Overall, the net range in observed RROE runs from *–10%* to *+7%*. This means on the far ends we are projecting a player to collect on average *8.5%* more or fewer of the opportunities available to them. Considering the league leader in non-adjusted rebound rate for 2023-24 was Jusuf Nurkić at *22.8%*, this amounts to a *37.28%* fluctuation from max value.

4.3.2 Grouping by Position

Considering that players court realty is largely defined by their on-court positions at any given point in time, it makes sense to analyze these players in their corresponding groups. For this, we chose to categorize into: bigs, wings, and guards.

Bigs typically operate closer to the rim and thus have higher raw estimate and real rebounding numbers, while wings and guards tend to occupy areas outside of the bulk of the rebounding distribution. As a result, deriving value is different for a big, who needs to win a physical match around the basket, than to a perimeter player, who creates value by crashing into those high propensity areas. This is similar to the idea of a player winning in the Positional dimension or the Hustle dimension, as defined by Maheswaran, et al. [5].

One thing that stands out across the positional spectrum is the success of athletically dynamic players, with the top individual performers being two former big-men teammates, Jarret Allen and Nic Claxton, and two former dunk contest participants, Aaron Gordon and Jalen Green. Moving forward, RROE could be used to isolate these types of players in draft assessment or team building to find players who are making the most of their rebounding opportunities or align strategically with the tendencies of other team members. A full breakdown of player RROE by position is found in the figure below.

Net Expected Rebounds & Rebound Rate Above Expected by Position
(min 50+ Opportunities)
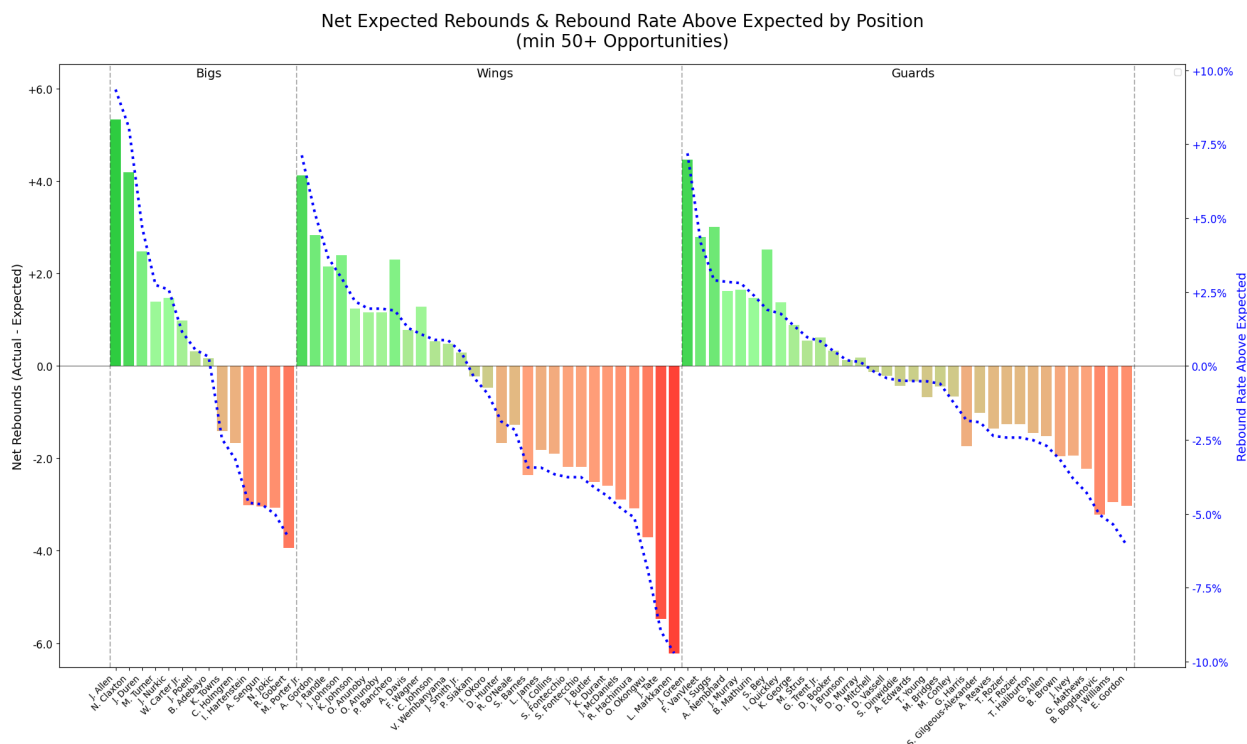
4.3.3 The "Brook Lopez Effect"

One particular player-type stands out as undervalued in this model: the team-rebounder. Brook Lopez of the Milwaukee Bucks. In the 2023-24 season, Lopez's defensive rebounding percentage was a paltry 13%. This would place him in the bottom 10th percentile of rebounding big men, yet the Bucks finished third overall in defensive rebounding [8]. Clearly, Lopez is doing something to help his team in these situations despite his lack of traditional counting stats.

Two potential exemplars of the "Brook Lopez Effect" stand out in our results: Nikola Jokić and Alperen Sengun. Both bigs perform poorly in our model, suggesting they do not collect a lot of rebounds outside of their controlled area at shot time. Interestingly, the top performing wing and guard respectively are Jokić's and Sengun's teammates: Aaron Gordon and Jalen Green. This makes sense because team rebounding is zero-sum. If you don't secure the board someone else will, and if that someone is your teammate you created space for, they essentially siphoned that rebound from you. It is difficult to say to what degree Gordon and Green are creating value by crashing and how much their bigs contribute to this by creating space, but it is a dynamic worthy of future exploration.

The NBA has an advanced stat that is supposed to account for these scenarios: deferred rebound chances. These are effectively the scenario we described, where a rebound falls near a player and that player defers to their teammate. Jokić, Sengun and Rudy Gobert, who all performed poorly in our model, were in the top 10 in deferred rebound chances for the 2023-24 season [6]. A counterexample is Jarret Allen, who both led our model in RROE and finished last season 10th in deferred rebound chances, suggesting it is possible to excel in both metrics.

5. Conclusion

This research uses the concept of court realty to propose a powerful new rebounding model. We use this model to both extend cutting edge continuous shot charts, providing further insight into shot selection in the NBA, and to introduce a Rebound Rate Over Expectation metric to capture dynamic rebounders who excel at getting hard-to-reach boards. The predictivity of the model results and the trends within player positions suggest an insightful new way of characterizing derived value from rebounding.

Future work could expand the tracking data set, allowing for analysis of expected rebounds across an entire season, focus on the idiosyncrasies of offensive versus defensive rebounding, or explore how dynamics shift in the playoffs. Additionally, player biometrics and ability could be introduced as model priors to give better context to on-court dynamics, and additional tracking stats like contested or deferred rebound opportunities could be used to augment initial RROE for a more comprehensive rebounding metric.

References

1. Ehrlich, J & Sanders, S. (2024). **Estimating NBA Team Shot Selection Efficiency from Aggregations of True Continuous Shot Charts: A Generalized Additive Model Approach**. Sloan Sports Analytics Conference. Retrieved from https://www.sloansportsconference.com/research-papers/estimating-nba-team-shot-selection-efficiency-from-aggregations-of-true-continuous-shot-charts-a-generalized-additive-model-approach
2. Cervone, D., Bornn, L., & Goldsberry, K. (2016). **NBA Court Realty**. Retrieved from http://www.lukebornn.com/papers/cervone_ssac_2016.pdf
3. Jacobs, J. (2019). **Extending Possessions: A Geometric Distribution Approach to Basketball Possessions**. Squared2020. Retrieved from https://squared2020.com/2019/11/27/extending-possessions-geometric-distribution
4. NBA. (2024). Proprietary Player Tracking Data. Retrieved from NBA internal database.
5. Maheswaran, R., Chang, Y.-H., Su, J., Kwok, S., Levy, T., Wexler, A., & Hollingsworth, N. (2014). *The Three Dimensions of Rebounding*. MIT Sloan Sports Analytics Conference.
6. NBA Advanced Stats. (2024). NBA.com/Stats - Official NBA Statistics and Advanced Analytics. https://www.nba.com/stats
7. Jacobs, J (2019). **Offensive Crashing: Using Data to Understand and Optimize Offensive Rebounding**. Squared2020. Retrieved from https://squared2020.com/2019/11/18/offensive-crashing
8. Sports Reference LLC. (2024). Basketball-Reference.com - Basketball Statistics and History. https://www.basketball-reference.com