Dan Feldheim
Summary of Offensive Rebounding Prediction Model
10-24-2020

Using player tracking data provided by the OKC Thunder, I was able to create a 4-parameter logistic regression model capable of predicting whether a play will result in an offensive rebound with an accuracy of 83%, area under the ROC curve of 0.69, and logloss of 0.54. The workflow that yielded the model consisted of the following steps:

1. Examination of the data for skewness, kurtosis, and multicollinearity
2. Formulation of hypotheses and parameters to test (**Table 1**)
3. Importing data into Python and wrangling
4. Calculations
5. Iterative model testing, quality evaluation, and refinement (Python and R)

The parameters that were calculated and tested in the model are listed in **Table 1.** Most displayed p-values indicating their significance in fitting the data. However, many were highly correlated as judged by their variable inflation factors and were not observed to improve the fit. In order to distill the model down to its critical and non-overlapping components, various combinations of parameters were tested and their contributions to the fit assessed by the deviance residuals and ROC curve. This iterative refinement process revealed the most impactful inputs to be the AtRim offensive and AtRim defensive distance variation coefficients, and the number of top offensive rebounders in the play. The top offensive rebounders were selected from the player_rebound_data file based upon rebounds/chance, with a minimum number of minutes played imposed to ensure that players used sparingly or during garbage-time minutes were not included. The distance variation coefficient equals the standard deviation of a team's distance from the basket divided by the mean distance of the team from the basket. It is thus both a measure of the team's overall distance and spacing from the basket. The importance of the AtRim variation coefficient is not surprising as it leads to the obvious conclusion that having the offensive players close to the basket and the defensive players farther from the basket increases the odds of the offense obtaining a rebound. What is more intriguing is the model's suggestion that if the offensive players are positioned at different distances from the basket (large distance variation) and the defense is positioned at similar distances from the basket (small variation), the odds of the offense obtaining the rebound will increase. An additional surprise was how dominant the variation coefficients were in this model, even over the number of players in the restricted zone, player angle from the basket (which was not close to being significant), any of the AtShot data, and the number of offensive players positioned at a distance closer to the basket than the defense (boxing out).

The regression model suggests that for every 1 unit increase in distance variance the odds of securing an offensive rebound increase by nearly 40-fold, with all other factors held constant. A potentially interesting outcome of the model can be observed by focusing in on two offensive rebounders as illustrated in **Figure 1**, in which the circles represent player positions. Note that in comparing the orange circles to the blue circles, the mean player distance from the rim has not changed, but players 1 and 2 have moved to positions that have increased the overall variance by nearly 0.5 units. Thus, while players 1 and 2 in orange appear to be well-positioned to obtain the rebound, their distance ***variation*** from the rim is a relatively low 0.2. If players 1 and 2 moved to the positions shown in blue, their distance variance would increase to 0.7 and the team's odds of obtaining the rebound would increase by a factor of 6. To translate this concept to game strategy, suppose that player 1 is Steven Adams playing down low and player 2 is Andre Roberson coming in for the offensive rebound from his usual position in the corner. The team’s odds of obtaining the rebound would increase if Andre assumes a distance from the rim different from that of Steven’s as opposed to being equidistant from the rim with Steven. In addition to improving offensive rebounding via distance variation, the model indicated that for every top-ten offensive rebounder such as Andre that is included in the play, the odds of obtaining the offensive rebound increase by 10%.

Although this model hints at a role for variable distance in obtaining offensive rebounds, it is clear that it has failed to address all of the factors involved in offensive rebounding. The McFadden’s pseudo R^2 of just 8%, low recall (10%), and high logloss indicate that a small fraction of the parameters that are important in describing how rebounds are obtained have been considered. Despite the two different angle variables that were calculated and incorporated into the model, it is surprising (and somewhat concerning) that neither one contributed to the fit. Future refinements of the model would certainly need to focus on incorporating player angle as a way to improve the fit. Including shooter position and specific locations of each player position (guards, forwards, big men) may also yield improvements.

Email

It is far too premature to take this analysis to the coaching staff, but if the model were refined and vetted further, this is the type of email I would send to start a discussion. As the Thunder do not currently have a head coach, I inserted my top choice.

To: Coach Becky Hammon
From: Dan Feldheim, Thunder Analytics
Re: Increasing the odds of obtaining an offensive rebound

Dear Coach Hammon

I have analyzed over 300,000 plays in an attempt to gain insight into how to increase the odds of a team getting an offense rebound. Much of the model’s conclusions are obvious to you, for instance, the closer the offensive players are to the basket, the greater the odds of obtaining the rebound. However, a more nuanced prediction from the model is that the **variation** in player distance from the rim may also be important.

The figure below illustrates the importance of player distance variation. The players represented by the orange circles would appear to be positioned ideally to obtain an offensive rebound because they are both close to the basket. However, they are also the same distance from the rim (low distance variation), which actually decreases the rebounding odds vs. their positions indicated by the blue circles, in which they are located at different distances from the rim.

In terms of how this could influence game strategy, consider Steven Adams down low as player 1 and Andre Roberson as player 2 coming in for the offensive rebound from his usual position in the corner. Despite the fact that their average distance from the rim does not change in going from the orange to blue circles, the team’s odds of obtaining the rebound would increase if Andre assumes a distance from the rim different from that of Steven’s position from the rim.

The same predictions hold for the defense; that is, anything the offense can do to keep the defense farther from the rim AND at the same distance from the rim, would increase the offense's odds of obtaining the rebound.

Please contact me if you have any questions or comments about this evaluation.
Sincerely,
Dan

Table 1. Parameters calculated and hypotheses tested in a logistic regression model for offensive rebounds. Parameters shaded blue had the largest impact on the model and were used to fit the data.

Category	Parameter	Hypothesis for Increasing the Odds of an Offensive Rebound
Player Spacing	Player distance from basket	Rebound odds increase as distance to basket decreases
	Player angle from basket	Rebound odds increase as angle increases toward 90° from baseline
	Offense and defense team distance variation coefficient	VC = STDEV/Mean Distance; Increased variation from basket and distances close to basket increase rebounding odds
	Team angle variation coefficient	Increased angle variation increases odds
	Number of players in restricted zone	More offensive players vs. defensive players close to the basket increase odds
	Average distance between offensive players	Spreading out the offense increases odds
	Average distance between defensive players	Confining defense increases odds of an offensive rebound
	Median distance between offensive and defensive players	Creating distance between offensive and defensive players increases odds
Relative Player Position	Number of players boxing out	Being closer to the basket than the closest defensive player increases odds
Offensive Rebounding Skill Level	Presence of selected individual top rebounders	The presence of certain highly efficient rebounders increases the odds
	Number of top rebounders in game during the play	Having any one of the top rebounding players in the play increases odds
Shot Type	Jump Shot, Layup, Free Throw, or Other	Certain shot types may be more amenable to rebounding by the offense

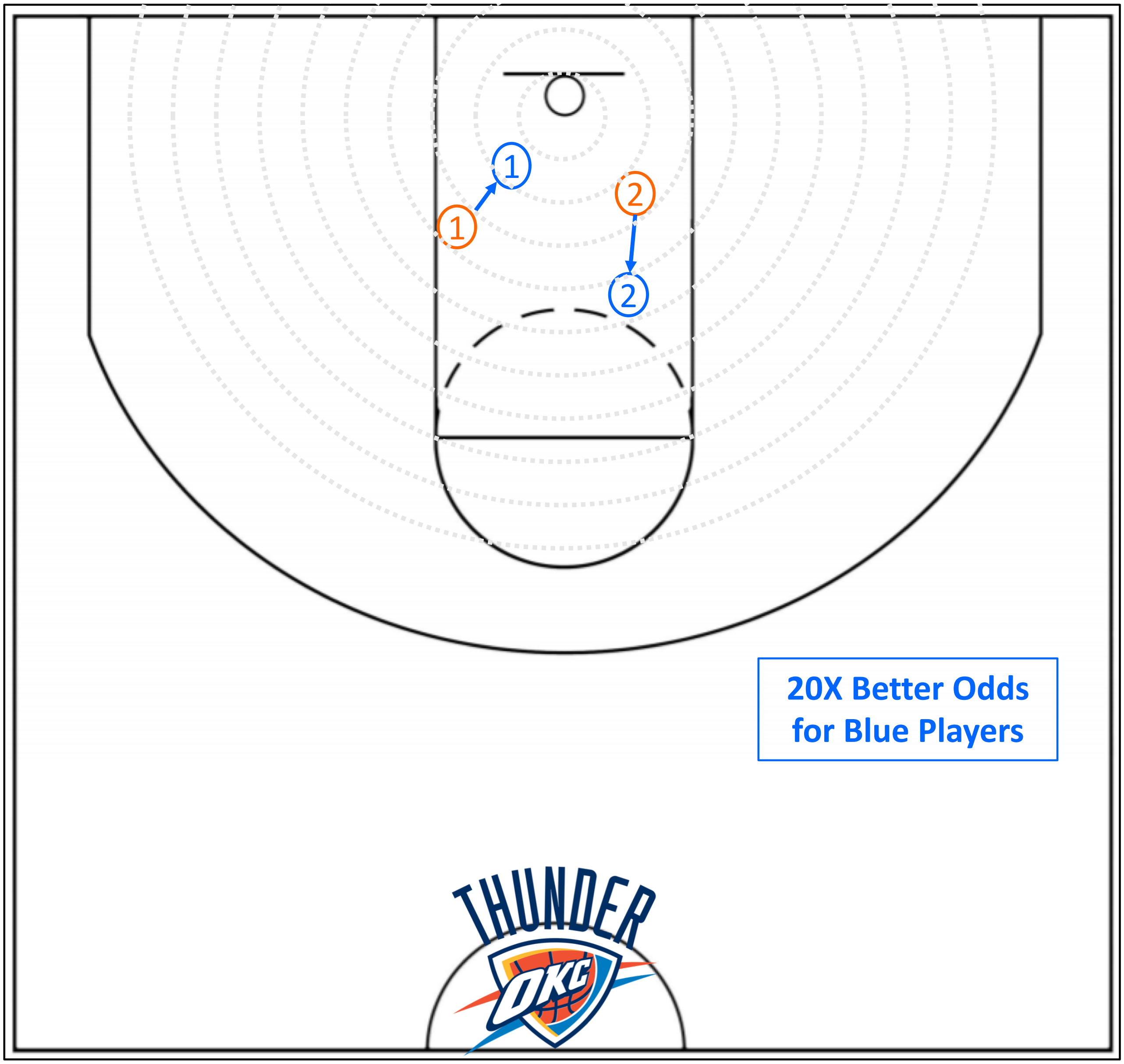

Figure 1. Illustration of the contribution of the variance in player distance from the rim to the odds of obtaining an offensive rebound. In comparing the orange circles to the blue circles, the mean player distance from the rim has not changed, but players 1 and 2 have moved to positions that have increased their overall variance by nearly 0.5 units. This increases the odds of obtaining the rebound by a factor of 20. The grey rings indicate distance, with each ring representing 2 ft.