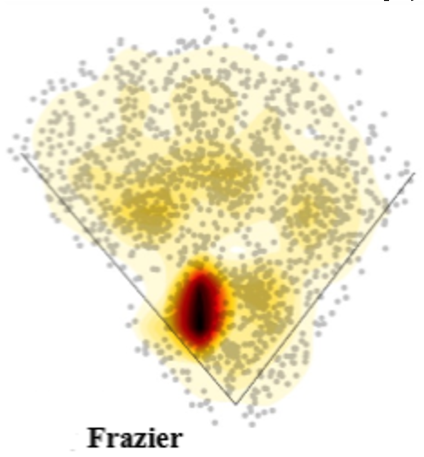
Investigating Optimal MLB Defensive Shifts

FURMAN UNIVERSITY

Noah Johnson, Rosie Kim, Patrick Lyle, Will Przedpelski, Devon Waskiewicz,
Dr. Liz Bouzarth[1], Dr. Ben Grannan[2], Dr. John Harris[1], and Dr. Kevin Hutson[1]
[1]Department of Mathematics and [2] Department of Business and Accounting
Furman University, Greenville, South Carolina

Introduction

Imagine you are the manager of a baseball team, and you see the following heat map of balls that Pittsburgh Pirate Todd Frazier has previously hit. Where would you position your fielders to best be able to cover his heat map, as shown below?



Heat map of non-home-run balls hit in play by right-handed Todd Frazier from 2014-2018. Data Source: Baseball Savant

To answer this question, we look at strategic defensive repositioning of a team's fielders to cover a player's strengths, commonly called a shift. If the positions are changed enough from the traditional player positions, then it can be compared to the original to determine whether the batter is defended better than he normally would be.

Model

We are working to improve the model of Bouzarth et al. 2021, primarily in terms of accuracy and ease of use. The researchers created an integer programming model which reads in the hit distribution of a batter as input and then outputs optimized positioning for each of the seven moveable players (which excludes the pitcher and catcher) to defend most optimally on the field.

The objective of this problem is to find the maximum field coverage while balancing the need to cover a batter's hit distribution with the risk of leaving parts of the field where extra base hits most commonly occur for this player. The model ensures that every 8x8-foot circle patch on the field has a coverage score less than or equal to 1 (that is, each patch is not overly crowded), that players cannot block each other's line of sight, that at least one player is within a reasonable range of first base, that there are at least three players in the infield, and that exactly seven players are placed.

The model of Bouzarth et al. outputs the diagram on the left to display where the seven movable players should stand when batter Todd Frazier is at the plate.

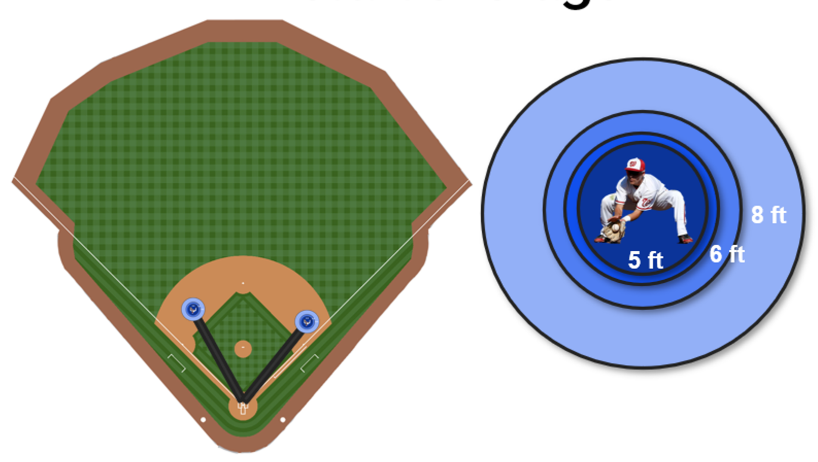
Changes To Model and Results

In an effort to increase the accuracy of the model, we made the following changes to the model provided by Bouzarth et al. 2021.

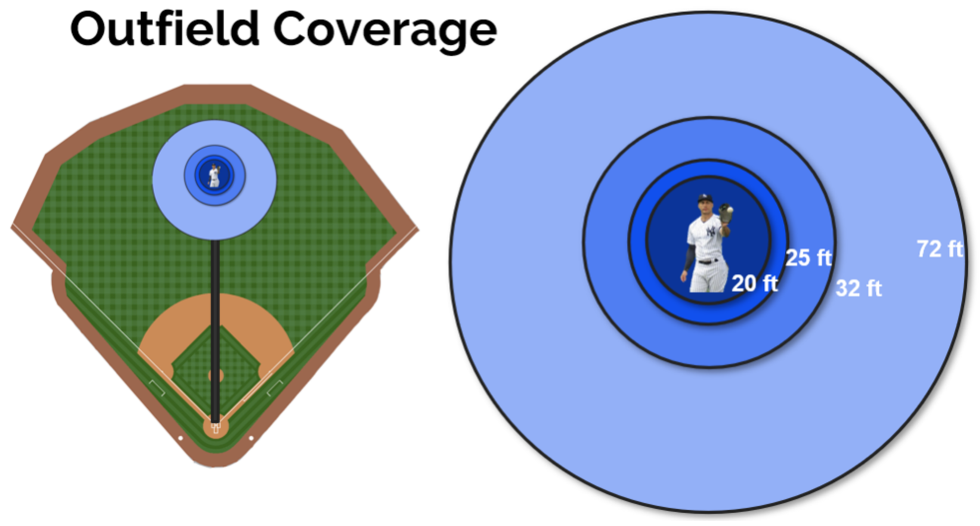
Fielder Acceleration: We individualized player speed by accounting for unique acceleration and reaction time values for each defender. We used the idea of outfielder jump, the distance, in feet, a fielder can cover in the correct direction in the first three seconds after the ball is hit, using data compiled from Baseball Savant, Major League Baseball's statistics repository. After three seconds, we used sprint speed to help estimate coverage area because we assume maximum speed is reached after initial acceleration. For infielders, jump data was not present, so we approximated coverage based on their sprint speeds. When implemented, these changes made coverage zones smaller but also more dynamic and individualized to specific teams of fielders.

Ball Speed Coverage Groupings: We created four groups of batted ball horizontal exit velocities into which each batted ball a player has on record falls. This is more robust than the previous metric, since each ball would naturally have a different time interval during which a fielder can make a play. Using Baseball Savant's batted ball exit velocity data, we divided each batted ball into four equal-frequency bins and used the median of each bin to represent a less-skewed representation of the average exit velocity of that particular type of ball (slow balls in the first bin, fast balls in the fourth bin, etc.). The slower the ball, the larger the coverage zone in the model becomes, as each fielder would have more time to react to a slow ball than to a fast ball.

Infield Coverage



Outfield Coverage

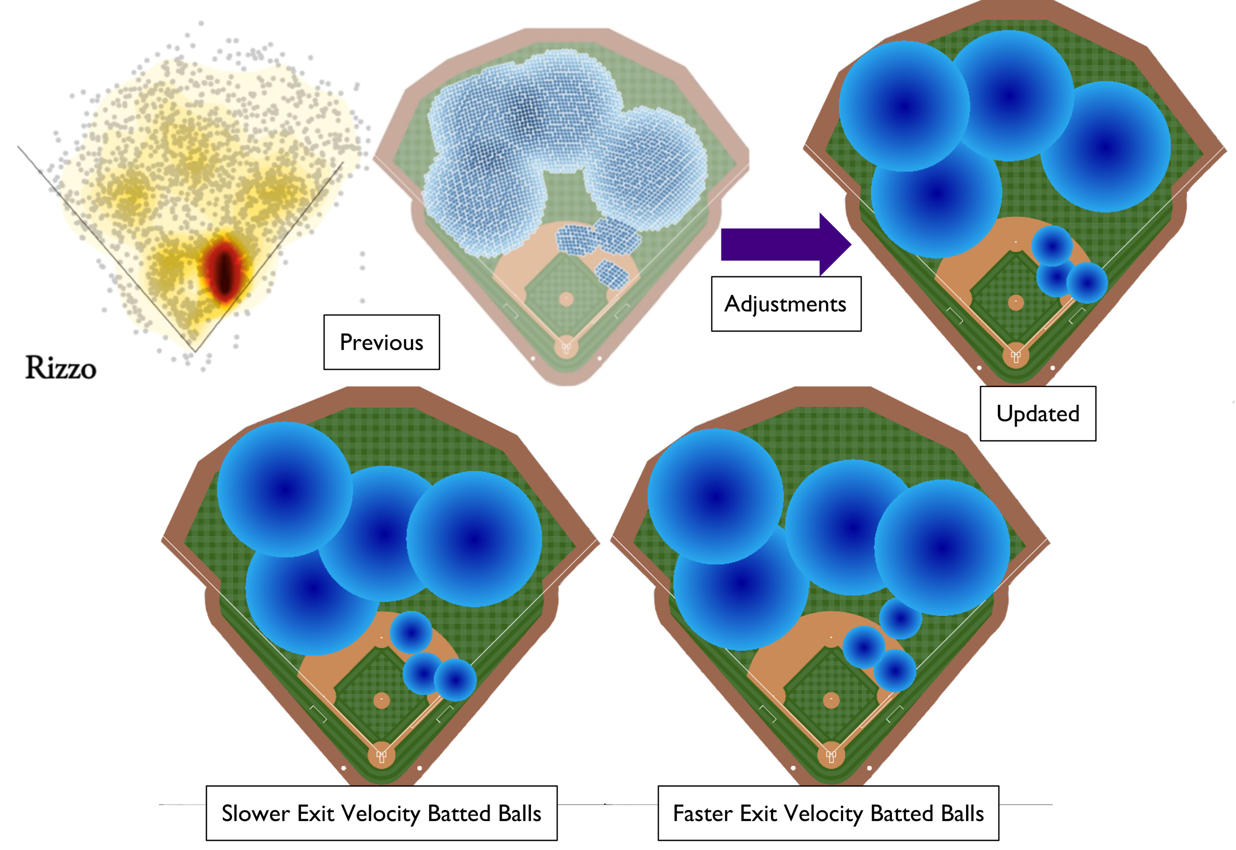


The above picture depicts an example coverage circle for an infielder and an example coverage circle for an outfielder, and the associated approximate distances a player of average speed can cover in these positions. The lightest blue circle represents how far this player can move and still field a slow ball, while the darkest blue circle represents how far this player can move and still feel a fast ball.

Coverage Patches: We increased the diameter of the coverage patches, which represent the fundamental unit of player positioning, from five feet to eight feet. This decreased run time of the computationally complex integer program without negatively impacting accuracy because the player's location need not be exact; as long as the player stands in an approximately similar position, the model remains valid.

Results Compared

One of our main goals was to improve the realism of the previous model, which we attempted to do with the changes discussed in the previous section. Below shows the heat map of Chicago Cub Anthony Rizzo from 2014-2018 and then compares the output of the previous model and our refined model. Further, we have also provided two heat maps corresponding to optimal shifts on Rizzo on slow and fast balls.



Future Research

While we worked to improve the previous model, some realism remains to be achieved. Namely, running the model on any of the 30 home ballparks in MLB would provide park-specific shifts, such as accounting for the Green Monster in left field at Fenway Park in Boston. Some of the model results with too many specified parameters (only slow balls, only curveball pitches thrown, etc.) create unnatural positionings that must be fixed before this model is finalized; this can be potentially predicted using machine learning based on players with sufficient data in these situations. Another possible improvement is the implementation of a situation index that takes the specific game circumstances, which lead into situational shifts when the count is 3-1 (favorable for the hitter) vs 0-2 (favorable for the pitcher), or if there is a runner is scoring position, etc.

Acknowledgements

We would like to thank the Furman University Mathematics Department and our research mentors for helping us in this research, and Drs. Bouzarth, Grannan, Harris, and Hutson for their continued guidance and support throughout the summer that made this project possible. We would also like to thank student peers Andrew Hartley and Ella Morton for generously allowing us to continue their research and assisting us at points along the way.