Optimizing Defensive Alignments in Major League Baseball

Jiyue (Tom) Yang, Donghun Cho, Joe Datz, Julia Foust, Giaco Gentile, Lucas Gregoire, Connor Horan, Jacob Kefalos, Hugh McMurray, Nathaniel Palmer, Troy Reinhardt, Aaron Reuter, Jeffrey Wheeler (Faculty)

Math 1103 - BIG Problems, Dept. of Mathematics, University of Pittsburgh

University of Pittsburgh

Carnegie Mellon University

The Problem

Can optimal defensive alignments be determined from a set of batting data? The **Pittsburgh Pirates** posed this problem to us and provided data and guidance to inform our solution. Our goal was to **determine optimal player positioning** against each batter to minimize the expected number of runs allowed.

Background

Sabermetrics & The Shift **Saber**metrics is a term coined by Bill James, in homage to the **S**ociety for **A**merican **B**aseball **R**esearch, to describe the use of data analytics in baseball. In 1946, the first Shift, a strategic realignment of the defense to one side of the field, was used against Ted Williams. In 2009, Dan Fox, the Director of Baseball Informatics for the Pittsburgh Pirates, encouraged Shifting techniques that significantly increased defensive efficiency and, ultimately, brought the Shift to the MLB.

For this project, we worked with Dan Fox and his colleagues at the Pirates to create a method to optimize the defensive alignment against a given batter.

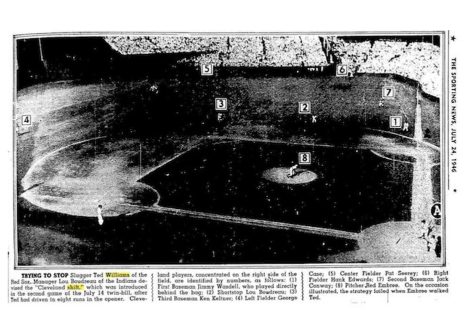
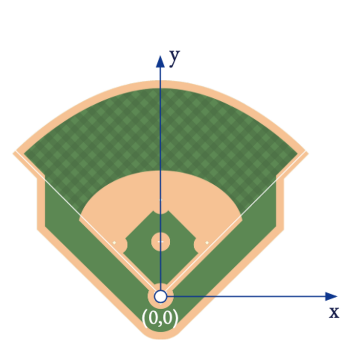


Figure 1: First Shift against Ted Williams by Lou Boudreau and the Cleveland Indians

Figure 2: Dan Fox, featured in the Tribune-Review ('Meet the man who built Pirates' analytics department,' Travis Sawchik) and ESPN ('How the Pirates got defensive,' David Schoenfield)

Data

- 231,025 data points
- Some key categories: bat side, player and ball positions



Idea: Machine Learning

Clustering We cluster data (ball positions) and try to optimize defensive alignments based on these clusters. K-Means is one algorithm that can solve our clustering problem.

Problems

- No constraints
- Can't handle data bias
- Overfitting

Attempt 1: Quadratic Optimization with Linear Constraints

Objective Function Minimize the distance between the predicted ball position when caught (6 ft in the air) and the total distance between players.

$$\min \sum_{i=3} (\hat{x} - x_i)^2 + (\hat{y} - y_i)^2$$
$$x_{pos} = (v_0 \cos\theta_v \cos\theta)(t_{total})$$
$$y_{pos} = (v_0 \cos\theta_v \sin\theta)(t_{total})$$

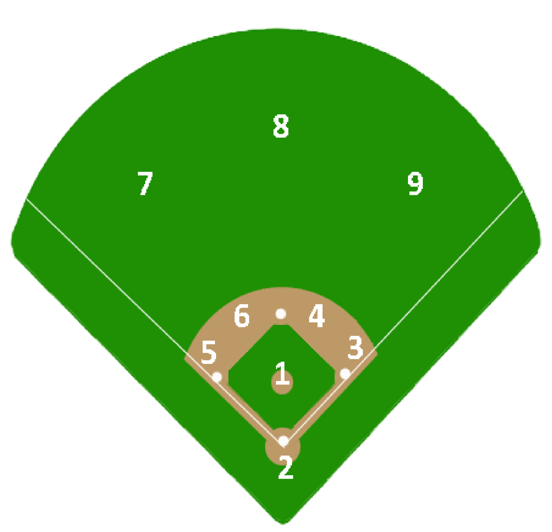
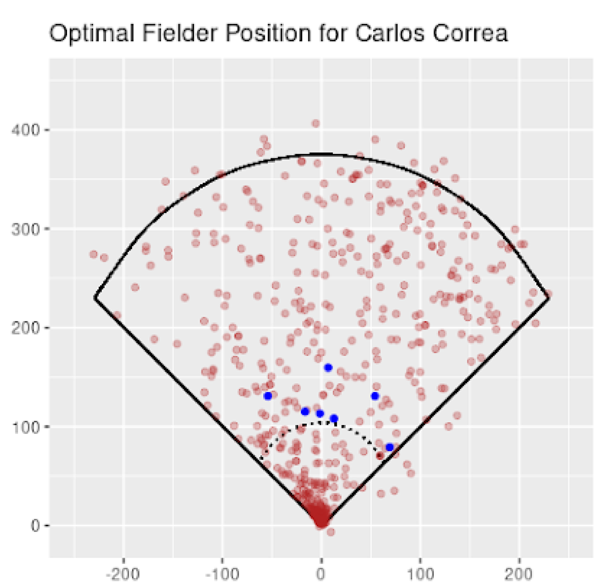


Figure 3: Label of each fielding position

Constraints

- One player near 2nd and 3rd base at all times
- 1st baseman in box near 1st base
- Players between foul lines
- Custom constraints for individual park outfields

Results



Problem The players are clustered around second base. We need separate analysis for infield and outfield players.

Attempt 2: Quadratic Optimization with Quadratic Constraints

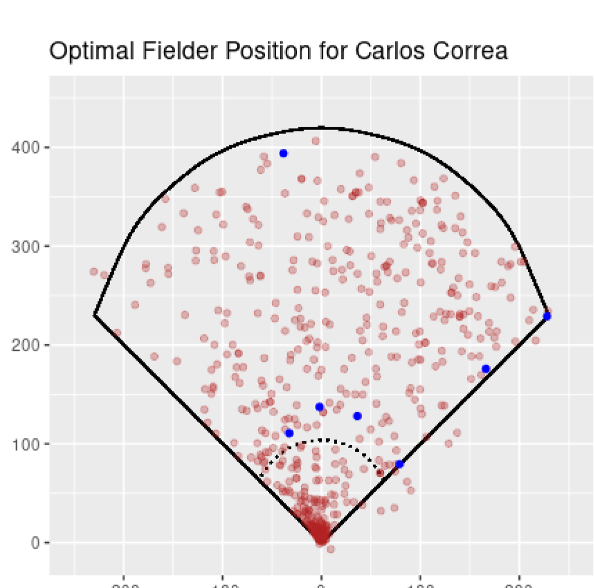
Objective Function Weight ball position by run values. Split infield and outfield players to avoid converging to one single position.

$$\min_{X_k, Y_k} \sum_j \{ \sum_{k=3}^{6} (1 - i_a) R_i[(x_j - X_k)^2 + (y_j - Y_k)^2]$$
$$+ \sum_{k=7}^{9} i_a R_i[(x_j - X_k)^2 + (y_j - Y_j)^2]\}$$

Constraints

- Restrict outfield players moving forward to infield positions

Results



Problem We have two outfielders lining at the foul line.

Attempt 3: Two-part Probability Simulator

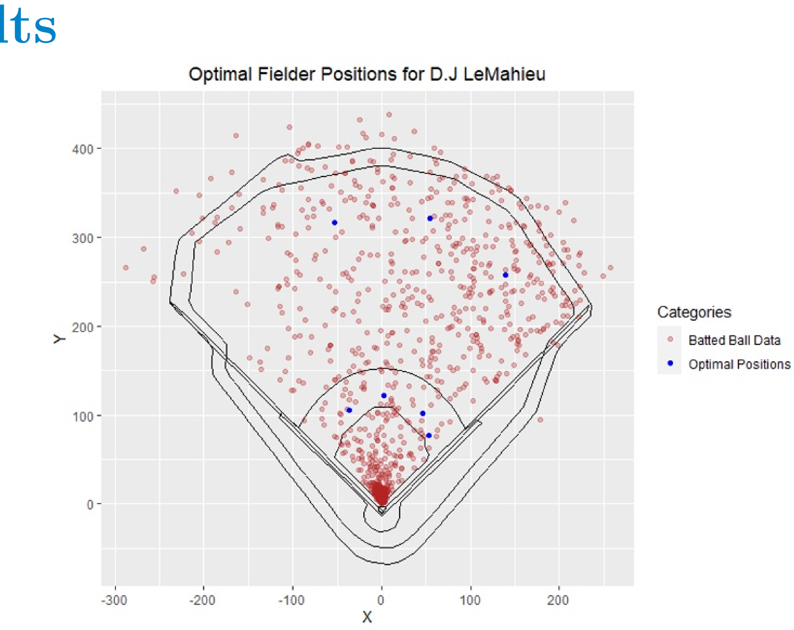
We try to find the arrangement that maximizes the probability of getting an out using 2 steps:

1. P(Ballpos X, Ballpos Y| Batter)

$$f(x, y) = \sum_{i=1}^{n} \frac{1}{2n\pi} e^{(x-x_i)^2 + (y-y_i)^2}$$

2. P(Out | Batted Ball Position, Player Positions)

Results



Conclusions

Our attempted approaches included machine learning and quadratic optimization with various constraints. Ultimately, we found that a **two-part probability simulator** created the best, most realistic defensive alignments.

Technology Used/Skills Learned

Sabermetrics Techniques • R • SQL • PCA • K-Means
Parzen Windows • Optimization • Probability Simulators

Acknowledgements

- Department of Mathematics, University of Pittsburgh
- Dr. Jeffrey Wheeler
- Dan Fox, Director of Baseball Informatics, Pittsburgh Pirates, & Colleagues at the Pirates
- Dr. Sam Ventura, Pittsburgh Penguins & CMU
- Dr. Andy Walsh, Health Monitoring Systems
- Joseph 'Nico' Gabriel, Pitt School of Pharmacy
- Allegheny Mountain Section of the MAA