Optimal Fielder Positioning Model

By Clinton Hausman, Michael Shames, and Bradley Waddell
Objective
Optimize the positioning of fielders based on a given hitter’s batted ball data

- Shifts are becoming increasingly more common in MLB
  - Started with extreme hitters
  - Now, many types of shifts are implemented
  - Most hitters should have an optimal shift, with varying severities

Source: Baseball Info Solutions
Set-up: Segmentation

- Used Baseball Savant’s hc_x, hc_y
- Separated field into 96 segments (right to left, front to back)
  - 12 partitions by angle from home plate
  - 8 partitions by distance from home plate
- All data and results assume no runners are on base
New Goal

Pick the 7 “segments” with the highest aggregate contribution score, maximizing the defensive contribution to the team for a given batter
The Spatial Batted Ball Information

- What hc_x and hc_y convey

\[ d = \sqrt{d_x^2 + d_y^2} \]
\[ \theta = 45^\circ + \tan^{-1}(d_x/d_y) \]
Component 1: Batted Ball Density

- Calculated percentage of balls hit in each segment from the partitioning
- Used gradient color scale to highlight the density as a heatmap
Component 2: Severity Scores
Fielder-independent measurement

- How ‘bad’ is a batted ball for the defense?
- Based on Exit Velocity and Launch Angle
  - Fit polynomial models to the relationship between EV, LA and hit rate
- Compiled via multiplication
Component 2: Severity Scores
Fielder-independent measurement
Component 2: Severity Scores
Component 3: Play Distance Matrix

- Fielder-dependent measure
- Broadly describe fielder’s range
  - More distance between segments -> smaller probability of out being recorded
- Measures compiled in 96x96 matrices
Component 4: Relative Angle Matrix

- Used catch probability article from MLB.com as reference point
  - Average speed of fielders was approximately one foot per second slower running back than forward or to the side
- Measures compiled in 96x96 matrices

Source: MLB.com
Fielder Optimization

- How do we maximize the total contribution to the defense of a batter’s balls in play?
- Consider a fielder in segment $i$--what is their contribution in segment $j$?
- Iteratively picked the 7 segments with highest contribution scores for the highest total fielding team contribution

\[ \text{contribution score}_{ij} = (\rho_i + s_i)(d_{ij} + a_{ij}) \]
Results

Trout Contribution Heat Map With Optimal Fielder Positions
Results

Yelich Contribution Heat Map With Optimal Fielder Positioning

hc_x
hc_y

contribution_score

5000
7500
2500
Results
Limitations and Assumptions

- Limited data
  - Exit Velocity, Launch Angle since 2015
  - hc_x, hc_y since 2014
  - Location of fielders is unavailable
- Outcome bias
  - All outcome data is fielder-dependent
- Optimization bias toward infield
- No throwing constraints
- hc_x and hc_y were scaled
- Assume same location for every ball in a given segment
Future Modifications and Improvements

- Optimize positioning for more types of hitters
- Fielder location data
  - Would allow us to consider more than just fielder-independent measures
- Account for difference between catching a fly ball and fielding a ground ball more precisely
- Account for fielder skill levels instead of weighting each fielder’s skills equally
Acknowledgements

- Thank you to Andy Andres and Morris Greenberg for their invaluable guidance throughout this project.
Appendix

\[
contribution \ score_{ij} = (\rho_i + s_i)(d_{ij} + a_{ij})
\]

- \( c_{ij} \) = contribution score of a fielder in segment \( i \) to a batted ball in segment \( j \)
- \( s_i \) = average severity score of a batted ball in segment \( i \)
- \( a_{ij} \) = angle at which a fielder in segment \( i \) must run in order to field a ball in segment \( j \)
- \( d_{ij} \) = linear distance between the centerpoint of segment \( i \) and segment \( j \)
- \( p_i \) = density score of segment \( i \)
Appendix

Visual representation of graph