



Swing Shift: A Mathematical Approach to Defensive Positioning in Baseball

Baseball Track
Paper ID 1548701

Abstract

Defensive repositioning strategies (shifts) have become more prevalent in Major League Baseball in recent years. In 2018, batters faced some form of the shift in 34% of their plate appearances [7]. Most teams employ a shift that overloads one side of the infield and adjusts the positioning of the outfield. In this work we describe a mathematical approach to the positioning of players over the entire field of play. The model uses historical data for individual batters, and it leaves open the possibility of fewer than four infielders. The model also incorporates risk penalties for positioning players too far from areas of the field in which extra-base hits are more likely. Our simulations show that an optimal positioning with three infielders lowered predicted batting average on balls in play (BABIP) by 5.9% for right-handers and by 10.3% for left-handers on average when compared to a standard four-infielder placement of players.

1. Introduction

Defensive repositioning strategies, or shifts, involve moving fielders from their traditional placements in the field to locate them where batters tend to hit the ball. Shifts against left-handed hitters usually involve moving the shortstop to the right side of second base and adjusting the third baseman to cover a range between the traditional shortstop position and second base. For a right-handed batter, the second baseman moves to overload the left side of the infield. The first baseman shifts slightly but still has the ability to cover first base. In each case the outfielders often shift to overload parts of the outfield as well.

In 2018, Major League Baseball teams employed a shift in 34% of plate appearances [7]. One player who frequently faces a shift is Freddie Freeman of the Atlanta Braves. The heat map shown in Figure 1 shows the locations of the balls that Freeman put in play during the 2018 season [1]. Over his career, Freeman has shown a tendency to pull the ball to the right side of the infield (doing so on 40% of plate appearances [5]). The dark areas of the heat map in Figure 1 show this for the 2018 season. Teams employ a shift against Freeman to exploit this tendency, and evidence suggests that the shift cost Freeman the National League batting title in 2018 [6]. More generally, there is evidence that the shift reduces batting averages on balls put in play (BABIP). Mike Petriello reported that in 2017, the league had a 0.276 BABIP against a shift and a 0.306 BABIP without a shift [4].

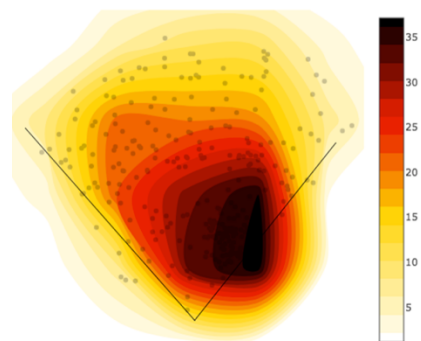


Figure 1: Heat map of balls put in play by left-handed hitter Freddie Freeman (2018 season) [1]. Darker colors denote higher hit density.



In this work, we propose a mathematical approach to positioning players based on two factors: (i) a particular batter's hit distribution, and (ii) the risk of not covering locations on the field where extra base hits are more likely to occur. This approach to positioning enhances the traditional shift by allowing seven fielders (all but the pitcher and catcher) to be placed in non-traditional positions.

2. Methods

We define an integer program to assign players to positions in the field. To do this we first divide the field into two regions. The first region is the area of the field within a 75-foot radius from home plate. We assume the pitcher and the catcher, who have fixed positions in the field, can cover balls hit into this first region. The second region is the rest of the field, and we partition this region into disks 5 ft in diameter representing possible locations to place the seven remaining fielders. In a field where the outfield wall is 400 ft from home plate (at all points on the wall), this division yields 5,021 possible disks. In our work we use Sun Trust Park, home field of the Atlanta Braves, as the example field, although we could adjust the dimensions to model any field. Using Sun Trust Park's field dimensions, there are 4,229 possible disks in which to position fielders since the outfield at Sun Trust Park does not extend to 400 ft in every direction. We set up binary decision variables x_{ij} as follows, where i (the fielder) varies from 1 to 7, and where j (the disk or field position) varies from 1 to 4,229:

$$x_{ij} = \begin{cases} 1, & \text{if fielder } i \text{ is assigned to disk } j \\ 0, & \text{otherwise.} \end{cases}$$

The seven fielders are not labeled or distinguished by any typical position names, and they are not restricted to certain typical areas of the field. Rather, they are seven players who can be placed anywhere necessary to improve defensive strategy.

To define the objective function and constraints, we need to create coverage zones for each of the possible fielding positions. In determining coverage zones, we assume that every fielder has the same ability, although, in practice, this could be adjusted for each defensive player. Our work assumes that a fielder can move at 25 ft/sec to field a ball hit at home plate with an exit velocity of 125 ft/sec. An outfielder would have approximately two seconds to react to and field a line drive under these assumptions. An infielder would have about one second to field a ball in a side-to-side motion but less time to react to a ball while moving forward. To quantify coverage, we define a $4,229 \times 4,229$ matrix P with entry p_{ij} representing the coverage score for a fielder at location i to field a ball at location j on the field. The coverage scores range from 0 to 1 and decrease while moving away from position i .

In determining the p_{ij} values, we add detail by distinguishing between outfield and infield placements. Specifically, all points outside of a 210-foot radius of home plate are categorized as the outfield. For a player in the outfield, we assign the highest coverage score for all positions j within a 40-foot radius around position i . For positions j outside of the maximal coverage radius but within an 80-foot radius, we apply a constant decrease in coverage based on the distance between locations i and j :



$$p_{ij} = \begin{cases} 1, & \text{if } d(i, j) \leq 40 \\ 1 - \frac{d(i, j) - 40}{40}, & \text{if } 40 < d(i, j) < 80 \\ 0, & \text{if } d(i, j) \geq 80, \end{cases} \quad (1)$$

where $d(i, j)$ represents the straight-line (Euclidean) distance between locations i and j . Note that coverage zones for players positioned in the outfield are circles, indicating symmetric coverage in all directions. Figure 2a shows an outfield coverage zone. Darker (lighter) shading indicates larger (smaller) p_{ij} values.

Coverage zones for players placed in the infield are not entirely circular in shape. This is because of the realistic feature of our model that an infielder has more coverage when moving backwards and side-to-side to field a ball than when moving forward. Additionally, fielders placed nearer to home plate have a smaller range of coverage than ones placed further from home plate. To determine infield coverage scores p_{ij} , we begin by letting r be the distance in feet of position i from home plate. The maximum distance an infielder could field a ball while moving laterally in either direction is $\frac{r}{5}$. The denominator comes from

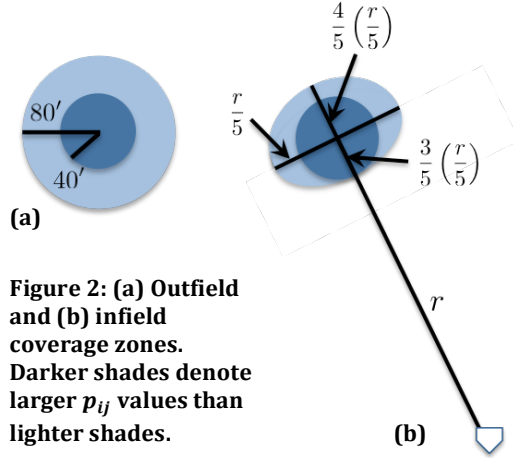


Figure 2: (a) Outfield and (b) infield coverage zones. Darker shades denote larger p_{ij} values than lighter shades.

dividing an average player speed of 25 ft/sec by a standard baseball exit velocity of 125 ft/sec. We set the coverage boundary of the player moving towards the ball to be $\frac{3}{5}$ of this lateral coverage and the coverage boundary moving backwards to be $\frac{4}{5}$ of this lateral coverage, indicating that a player can react better moving side-to-side to field a ball than moving forward or backward. The infield coverage zones are shown in Figure 2b. Additionally, for any field position j that is within $\frac{3}{5}$ of $\frac{r}{5}$ ft of player position i , we set the coverage score $p_{ij} = 1$. For all other infield locations (light shading in Figure 2b), we again use a constant decrease in coverage as a function of distance between i and j and set

$$p_{ij} = \begin{cases} 1, & \text{if } d(i, j) \leq \frac{3r}{25} \\ 1 - \frac{d(i, j) - \frac{3r}{25}}{\frac{4r}{25}}, & \text{if } j \text{ is in the lightly shaded region} \\ 0, & \text{if } j \text{ is outside the shaded area.} \end{cases} \quad (2)$$

Our goal is to place fielders in the field based on a batter's tendency to hit to certain locations. So, using data from baseballsavant.mlb.com from the 2014-2018 Major League Baseball seasons, we tabulate positions on the field where balls hit in play from different batters were fielded. As an example, the heat maps in Figures 3a and 3b show the fielding locations of balls put in play by left-handed Anthony Rizzo and right-handed Todd Frazier during the 2018 season, respectively [2]. Dark areas of black and red represent locations where a large number of balls were fielded. Light areas of yellow and brown are locations of less frequency for balls put in play. Areas of white indicate low frequency areas.

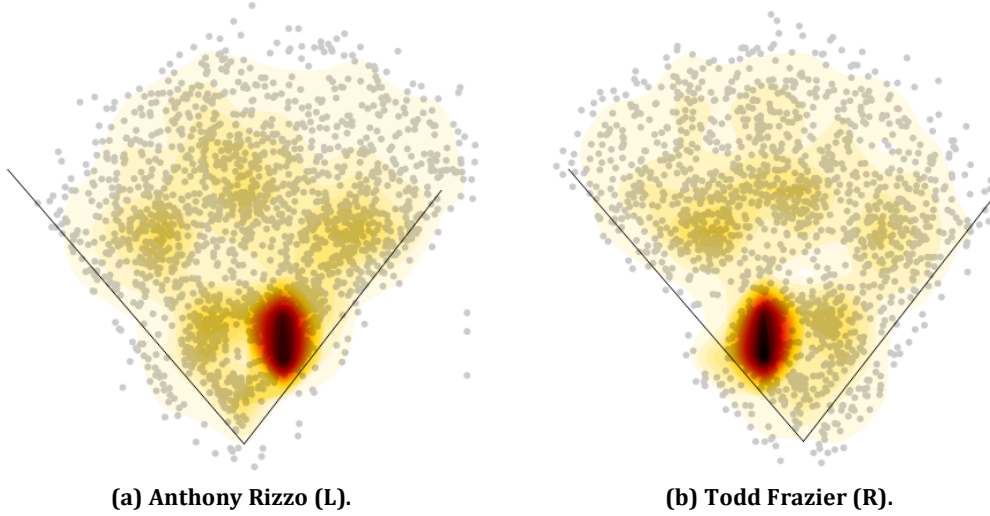


Figure 3: Heat map of balls in play by (a) left-handed Anthony Rizzo and (b) right-handed Todd Frazier [2]. Darker colors denote higher hit density.

Based on the historical data available for a given batter, we define an intensity rating w_i , for each position i on the field. This represents a normalized frequency with which the batter hits the ball to that location. However, placing fielders totally based on this intensity calculation leaves the possibility of being vulnerable to balls hit to the extremes of the field, such as around the foul lines and warning track of the field. So we introduce a risk vector, \vec{r} , which essentially adds extra intensity to each batter's intensity vector \vec{w} in the areas around the foul lines and locations more than 325 ft from home plate. The final integer programming formulation for the player assignment problem (PAP) is given below. A discussion of the objective function and constraints follows the formulation.

PAP: maximize $(\alpha \vec{r} + P \vec{w}) \vec{x}$
subject to

$$\sum_{j=1}^{4,229} x_{ij} = 1 \text{ for } 1 \leq i \leq 7 \quad (3)$$

$$P^T \vec{x} \leq 1.5 \quad (4)$$

$$\vec{f} \cdot \vec{x} \geq 1 \quad (5)$$

$$\vec{N} \cdot \vec{x} = 3 \text{ or } 4 \quad (6)$$

$$|\theta_j - \theta_l| \geq 4^\circ x_{ij} x_{kl} \text{ for } 1 \leq i, k \leq 7; 1 \leq j, l \leq 4,229 \quad (7)$$

$$x_{ij} \in \{0,1\} \text{ for } 1 \leq i \leq 7; 1 \leq j \leq 4,229 \quad (8)$$

The goal is to position fielders in a way that maximizes coverage of a batter's intensity ratings and the risk areas. The objective function contains a parameter α to allow the modeler to adjust the impact of the risk vector \vec{r} relative to the intensity ratings \vec{w} . For batters who have shown abilities to adjust to shifts, α can be set higher to avoid the risk of extra-base hits. For this study, we set $\alpha = 0.0004$ for each batter, which we determined through experimentation.

The constraints in (3) require that each of the non-pitcher/non-catcher fielders are placed in the field. Constraint (4) ensures that no area of the field is over-covered. The right-hand side value of 1.5 is adjustable and is set high in this work to ensure that areas with high intensity have large coverage.

Constraint (5) ensures that one fielder is placed within a range, defined by the vector \vec{f} , of locations for which a player could reasonably reach first base ahead of a runner. This ensures that someone is close enough to first base to be able to receive throws there.

Constraint (6) concerns the placement of players in an infield area defined by the vector \vec{N} . This area consists of all points in the field that are both within 180 ft of home plate and within 145 ft of first base, as shown in Figure 4a. These are locations at which fielders could realistically field a ground ball and make a throw to first base to get an average runner out. This region could be adjusted based on individual batters. Our model ensures that there are three or four fielders placed in this area.

The constraints in (7) ensure that no infielder will be placed within the same line of sight to home plate as another infielder. Here θ_j and θ_l represent the angles (in degrees) between the first base line and disk locations j and l , respectively, using home plate as the common vertex (see Figure 4b). The constraints guarantee that if $x_{ij} = 1$ and $x_{kl} = 1$, then angle difference between θ_j and θ_l needs to be at least 4° .

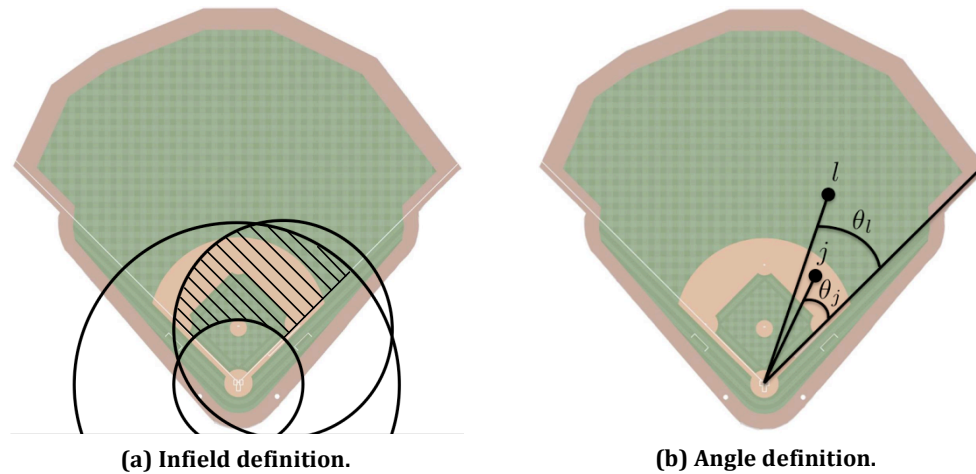


Figure 4: Visual representation of (a) the infield definition to identify positions from which fielders can throw someone out at first base (striped region) and (b) angle definitions used in determining line of sight constraints.



3. Results

Using 2019 data obtained from baseballsavant.mlb.com, we created a hit distribution for a variety of left-handed and right-handed batters. We simulated 10,000 balls in play by a sample of batters against three different positioning strategies. The first strategy was a standard positioning. To determine this standard coverage we ran the optimization program with a weight vector that averaged the weight vectors of ten randomly-selected left-handed and ten randomly-selected right-handed batters constrained with four infielders. The standard coverage positioning is shown in Figure 5, and we note that it is similar to a traditional positioning of players in the field. The second strategy was an optimal positioning as determined by the PAP program where the weight vectors were determined using 2014-2018 batter-specific hit data and where we constrained the positioning to have three infielders in constraint (6). The third positioning strategy had the same conditions as the second except that we constrained the number of infielders to be four.

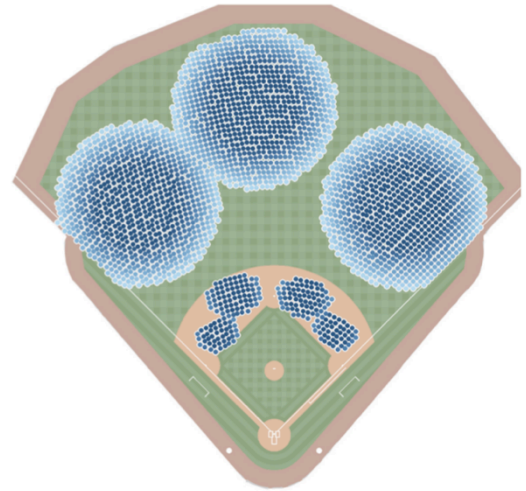


Figure 5: Standard positioning.

Figure 6 reproduces the heat maps of left-handed Anthony Rizzo and right-handed Todd Frazier from Figure 3. The figure also shows the optimal positioning of fielders by the second (three-infielder) analytical method. The figure shows positioning of multiple infielders in areas of red and black on the heat map of these batters. For Rizzo, the three infielders are placed on the first-base side of the infield while another fielder is placed more in a rover position in short left field to handle balls hit to the third base side of the infield. For Frazier, only two infielders are positioned in the area of red and black since there must be an infielder to cover first base. Note the fourth infielder plays as a rover in short center field to cover the dense area of yellow and brown in these areas.

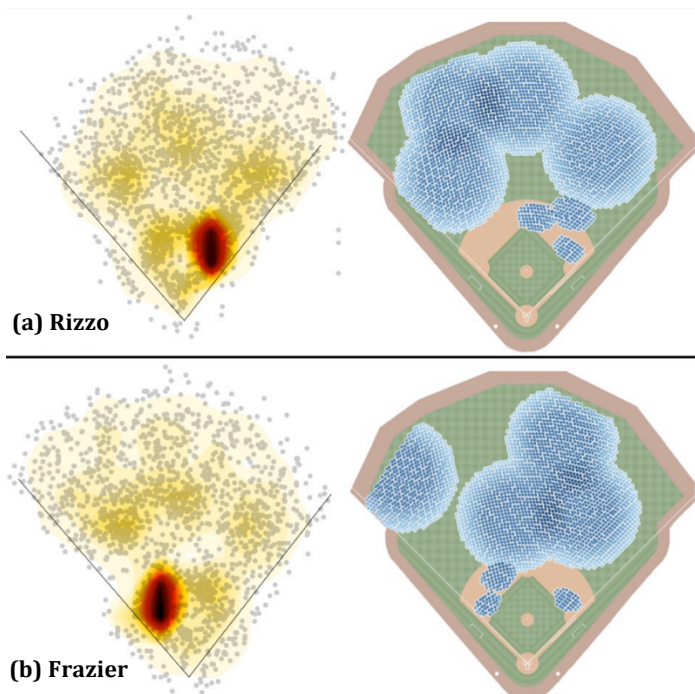
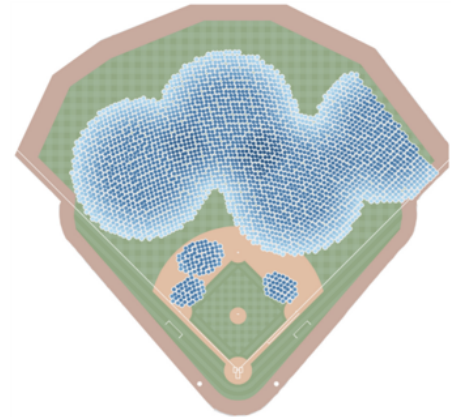


Figure 6: (a) Anthony Rizzo and (b) Todd Frazier balls in play heat maps (left) [2] and optimal positioning with three infielders (right).

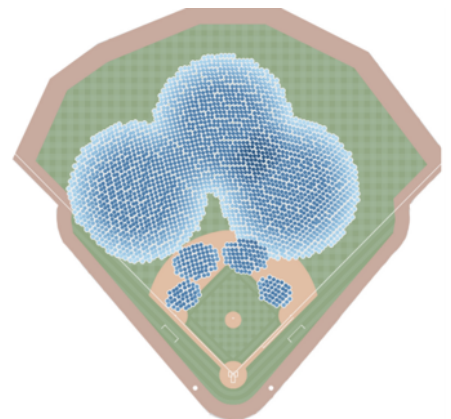
Figures 7a and 7b show the different positioning of fielders with three infielders and with four infielders against right-handed batter Edward Encarnacion. Note that the placement of three outfielders is similar in each figure. In the case of four outfielders, the fourth is placed in deep right field. We conjecture this is due to the risk vector setting a player placement to cover extra base hits.

Our simulations show that an optimal positioning with three infielders lowered predicted batting average on balls in play (BABIP) by 5.9% for right-handers and by 10.3% for left-handers on average when compared to a standard four-infielder placement of players. Table 1 shows the simulated change in BABIP when comparing traditional fielder placement to an optimal positioning with three infielders. With the exception of Lorenzo Cain, the simulations show that most players see a decrease in their BABIP statistics, indicating that the repositioning strategies have a positive defensive effect. Cain's change in BABIP is small, and we conjecture the cause may be that certain players are able to adjust their batting approach between seasons.

Batters have already adjusted their hitting approaches to defensive repositioning [7]. Hitters have had to adapt by either trying to hit the ball towards where a non-shifted shortstop would play or trying to lift the ball over the top of the shift. Figure 8 shows heat maps for Bryce Harper for the 2014-2015 and 2018-2019 seasons. In the 2018-2019 seasons there is a dark red area in the short right field that doesn't appear earlier in Harper's career, indicating a tendency for Harper to try to lift the ball over the infield. For our integer programming model to be most effective, the most recent information should be used (ideally teams run this model at least before every series, if not daily, to get best results). In fact, when we ran the model for Harper with the positioning being based on data from July 2018 - June 2019 and then tested it on the data from the remainder of the 2019 season, the model decreased his BABIP by 9.2% which is a slight improvement from 8.77%. Using more recent data could improve our positioning since it adds more



(a) Three infielders.



(b) Four infielders.

Figure 7: Positioning for Edward Encarnacion with (a) three and (b) four infielders.

Player (RH)	BABIP Change	Player (LH)	BABIP Change
Altuve	-7.14%	Bruce	-13.68%
Arenado	-6.36%	Carpenter	-9.91%
Cain	+1.92%	Freeman	-11.82%
Encarnacion	-8.93%	Gallo	-5.13%
Frazier	-6.25%	Gordon	-9.73%
Longoria	-8.93%	Harper	-8.77%
Martinez	-3.85%	Heyward	-11.72%
Posey	-5.56%	Kiermaier	-9.91%
Pujols	-9.91%	Rizzo	-13.51%
Trout	-4.46%	Votto	-9.01%

Table 1: Percentage change in BABIP after repositioning.



weight to the places where a batter is currently hitting the ball rather than where he hit the ball four years ago.

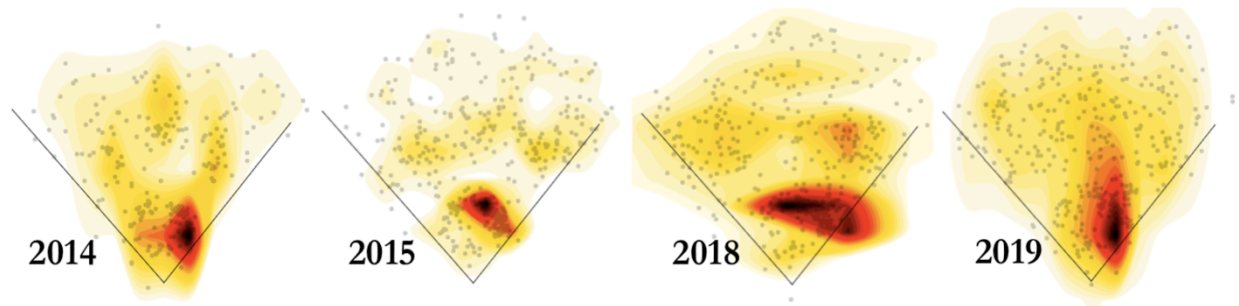


Figure 8: Heat maps for Bryce Harper from the 2014, 2015, 2018, and 2019 seasons [3].

4. Conclusion

Our research shows that using a mathematical programming approach to positioning defensive players in the field based on a balance between the normal areas a batter hits the ball with the penalty of leaving too much of the field uncovered can lead to significant reductions in players' BABIP. Further, by weighting more recent information, the model can adjust to changing hitting strategies of batters. While encouraging, we consider this work a proof of concept that warrants further investigation. Specifically, since our data indicates where balls put in play were fielded, we should examine similar data where balls are originally hit. Further, we could enhance this work by considering situational positioning based on pitch counts and runners on base. We could also address changes that must occur in pitching strategy as a result of repositioning the defense, since pitchers may be uncomfortable with certain pitches given where the defense is placed behind them. Finally, as mentioned earlier, there are ways to customize this work based on specific skills of fielders and different exit velocities of batters. With these enhancements, major league teams can strategically shift their defense to reduce the impact of opposing batters.



References

- [1] www.baseballsavant.com, Accessed on July 26, 2019.
- [2] www.baseballsavant.com, Accessed on August 5, 2019.
- [3] www.baseballsavant.com, Accessed on November 25, 2019.
- [4] M. Petriello. "9 Things You Need to Know About the Shift," mlb.com/news/9-things-you-need-to-know-about-the-shift-c276706888.
- [5] T. Poe. "Freeman's 'Shift This' Narrative Lacks Context," thesportsdaily.com/2019/03/31/freemans-shift-this-narrative-lacks-context-wow11/, March 31, 2019.
- [6] M. Salfino. "Hitters Hurt Most by the Shift," wsj.com/articles/the-hitters-hurt-most-by-the-shift-11553256970, March 22, 2019.
- [7] T. Sawchik. "Don't Worry, MLB-Hitters Are Killing The Shift On Their Own," fivethirtyeight.com/features/dont-worry-mlb-hitters-are-killing-the-shift-on-their-own/, January 17, 2019.