Optimal Fielding Position in Baseball

Kian Kurokawa and Ian Hughes

University of Puget Sound

August 23, 2022

Abstract

In this paper, we create a model that finds the optimal position for an individual outfielder in the game of baseball. First, we simulate an individual hitter's batted ball data. This batted ball data tells us where the ball was hit and its time of flight. Then, we model this fielder's position over time. We can then find a fielder's optimal initial position given that we know how many outs the fielder can record at each starting point. We then add two other outfielders to find each of the three outfielder's optimal position. After solving for each outfielders optimal positioning there is an additional out gained per 10 batted balls, which is practically significant in the game of Baseball.

1 Introduction

Outs are extremely valuable in the game of baseball. Because of this, fielders' positioning is of great importance for the team playing defense; a split-second could be the difference between a hit and an out. For those unfamiliar with the game of baseball, an explanation of the rules relevant to this analysis are provided in the Appendix.

First we find the optimal positioning of a single outfielder and use this as a baseline for explanation and analysis. From here we extend the model to include the other two outfielders in order to maximize the number of outs the defense can get. Traditionally, fielders are positioned as shown in Figure [1].

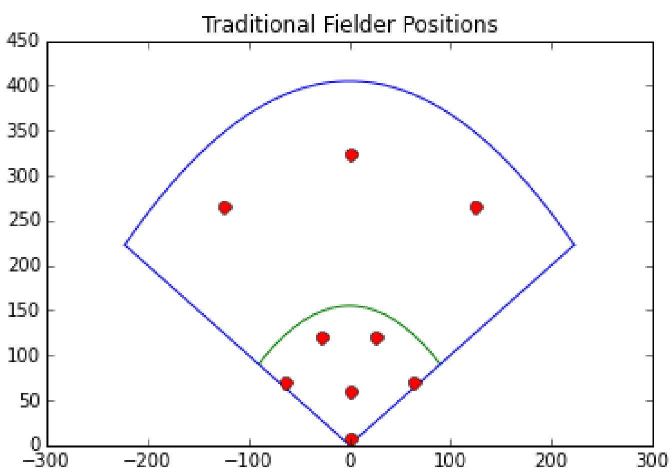

Figure 1

This traditional fielding, however, can be improved upon if we know the tendencies of different hitters. We would expect hitters to develop certain patterns after they have a reasonable number of at bats, say over 200. Because of this, a fielder's position should change slightly depending on the hitter.

Our model generates simulated batted ball data for offensive players based on their most recent full season of play. This gives a location and time of flight for 200 batted balls. We then combine this with a position versus time model for the outfielder. Combining the simulated batted ball data and the fielder position versus time model allows us to find the positioning on the field that will maximize the number of outs that an outfielder can get.

2 Simulating Batted Ball Data

While we used Houston Astro's star Jose Altuve in the year 2015 in our model, any major league player with over 200 at bats can be used in our simulation. Initially, we planned on using batted ball data that included location and flight time. Because this data is unavailable, we chose to simulate our own given a player's hitting tendencies. The launch angles and velocities for the balls Altuve hit in the 2015 season were available, as they are for many players. There are three general areas of the outfield, left right and center, in baseball. The percentage of times Altuve hit to each of these three fields is also available and will help us find the on-field angle of a randomly hit ball.

2.1 Distance and Hang Time of a Batted Ball

Using a simple differential equation with initial velocity and launch angle, we can propagate a ball's location forward through time. The equation and method we used to do this is located in the Appendix.

To begin, we center normal distributions using the mean and standard deviation of Altuve's 2015 exit velocity and launch angle data. We can then sample from these distributions for each batted ball and plug these samples into our differential equation to find the distance and time of flight for each individual sample.

2.2 On-Field Angle

Given that a baseball field is 90 degrees, we designate each field section, left, center, and right, 30 degrees. So, left field encompasses the angles from 135 to 105, center has the angles from 105 to 75, and right field has the angles from 75 to 45. First we use a random variable to choose what field our batter hits to. We do this using the player's actual left, center, and right field percentages. Our random variable is in-line with these percentages, or in the long-run sampling from this random variable will give Altuve hitting to left field 45.3% of the time.

Altuve's Distribution of Left, Center, and Right Field Batted Balls		
Left	Center	Right
45.3%	35.5%	19.1%

Then, after a field is chosen we designate what direction the ball is hit using the normal distribution (Figure [3]) centered approximately 15 degrees away from the respective fields left endpoint.

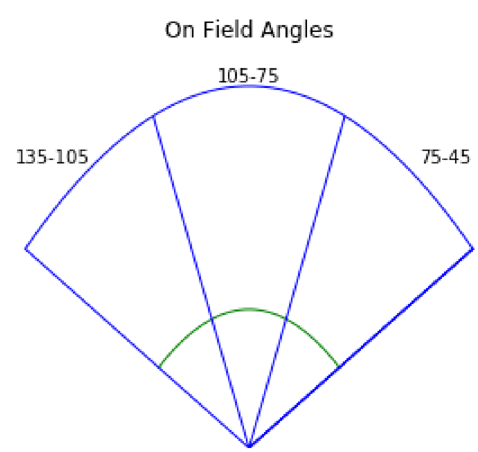

Figure 2

So, for center field, we would start at the leftmost angle of 105 degrees and then sample our degree distribution. This then tells us how many degrees the ball will be from the left endpoint of 105 degrees. As shown in Figure [3] we have a higher probability of the ball being hit to dead center or 15 degrees away from the leftmost angle of center field.

Degrees from Leftmost Edge of Chosen Field

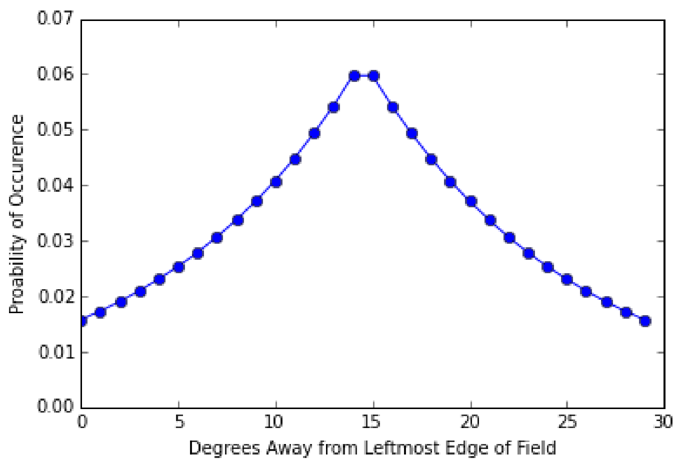

Figure 3

Combining these two factors gives us one of 90 on-field angles that the ball can go to. Adding this to the distance and time of flight variables give us an exact coordinate position for a batted ball.

3 Modeling the Fielder

Although we used Lorenzo Cain to model our fielder, any major league outfielder, given their max velocity and acceleration, could be used in our simulation. We decided to make the assumption that the fielder will take a straight path from his initial starting position to the landing spot of the batted ball.

We wanted to find the initial position that maximized the total number of balls caught in our model. To do this we set up 9 discrete starting positions for our fielder. Relative to our initial starting position, we shifted the fielder 50 feet left or right, 50 feet forward or back, or a combination of the two as shown in Figure [4].

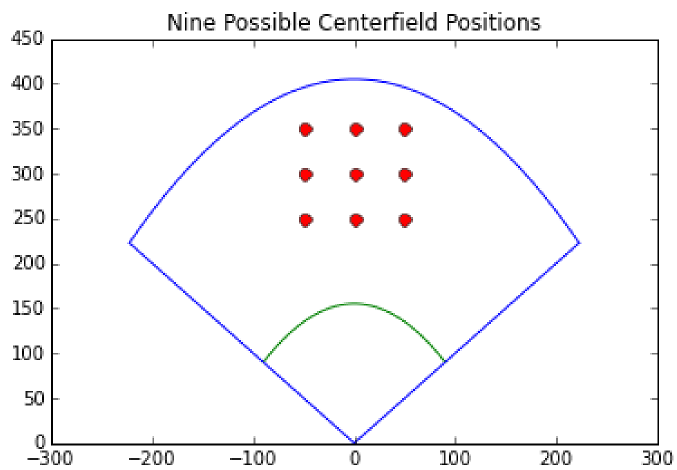

Figure 4

Using the distance formula we calculated each batted balls distance away from our fielder. Since each batted ball had an associated hang time we determined that the fielder, given a function of position based on time, caught the ball if he could run the distance faster than a batted ball's hang time.

3.1 Modeling Fielder Movement

To model our player movement we obtained data based on the players top speed and acceleration. Given these numbers we approximated the distance traveled given time using formulas from introductory physics. The formula can be found in the appendix under Fielder Movement.

We decided that the fielder's range is the same in all directions so in discrete time the fielder can cover a circular area based on the radius as seen in Figure [5].

Visual Representation of Fielder Range

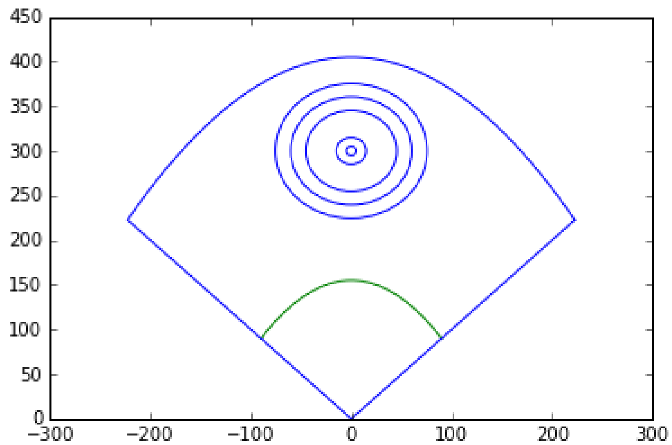

Figure 5

You can see that the radial ranges each fielder can cover is not uniform because we must take acceleration, which is non-constant over time, into account.

4 Results and Possible Extensions

Out of the 9 initial positions we found that moving the fielder in and right by 50 ft would be the optimum position to maximize outs. As seen in Figure [6] the green dots represent the landing area of the batted balls while the blue star shows the optimum.

Visual Representation of Fielder Range

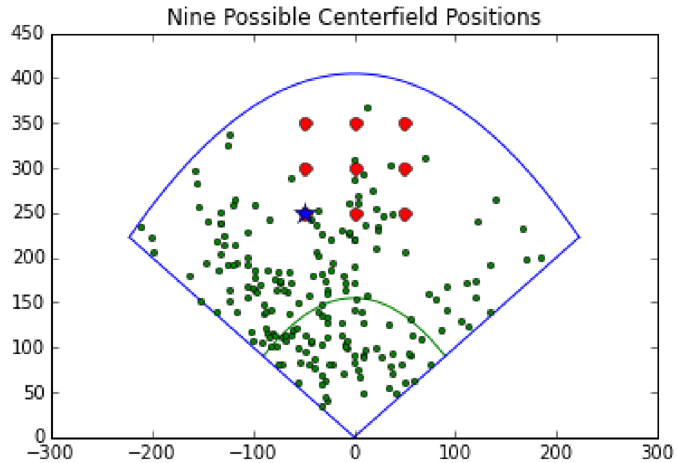

Figure 6

Comparing our results we found that optimal positioning yielded an average of 9.36 more outs than traditional positioning for 50 simulations. This would result in a drop of 0.00468 in BA (batting average) on average.

We must take note that these results are in the case of only one outfielder. We suspect that optimal positioning for the other positions would result in an increase of the number of outs gained.

5 Complete Outfield Model Results

Visual Representation of Fielder Range

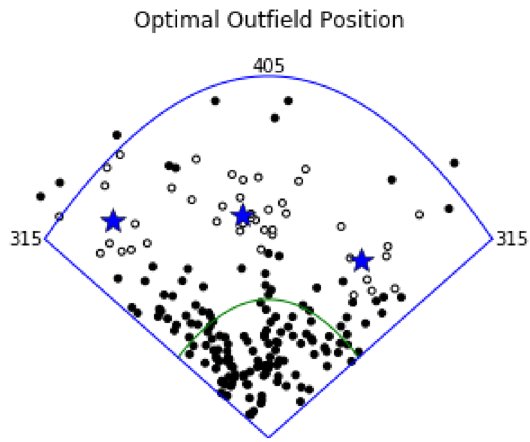

Figure 7

After running the simple model with one outfielder we wanted to know how optimal positioning was altered when we added two other outfielders. Mainly, our question was how many more outs could be gained when all outfielders' positioning was optimized?

In the game of baseball center field has priority over the other outfielders on balls in the air. After finding the optimal position for the center fielder we took all balls that he caught out of the simulated hit data. We then took the left fielder and found his optimal position given that the center fielder would catch all the balls within his range. We then did the same process to the right fielder given that the center fielder and left fielder were in their optimal positioning.

In our enhanced model we expanded the number and area of initial starting positions our fielders could start at. We used a 25 $\times$ 25 grid where each square equaled 4 $\times$ 4 feet giving us a total area of 1000 square feet.

As seen in Figure [7], we can see that the blue stars represent the optimal positioning for each outfielder. The graph is also color coded where black dots represent balls that are not caught and clear dots are balls that are caught.

Comparing our results we found that optimal positioning yielded an average of 20.8 more outs than traditional positioning for 50 simulations. This would result in a drop of 0.100 in BA on average. To give the reader an idea of how important this is consider that for every 10 balls hit, optimum positioning would gain the team 1 out. Over a season of 162 games and 500 AB, the outs gained would be extremely significant.

6 Model Extensions and Conclusion

A possible extension to our model would be to minimize the number of expected bases on each hit, instead of minimizing the number of hits. In this extension each ball would get an expected value of bases based on the direction, velocity, and time the ball is in the air. We would then find the new optimum positioning to minimize bases.

We also did not factor in the deviations of the fielders path to the ball which would result in a decrease in range, which complicates the simple differential equations we used to model fielder movement.

In conclusion we found that optimal positioning, given a certain players hitting tendencies, can significantly affect the number of outs gained over a season. With a season consisting of 162 games and approximately 500 at bats optimal positioning gains a team a significant number of valuable outs and, as a result, decreases the runs a team has given up and also increases wins.

Appendix

Distance and Hang Time Equation[1]

We first had to compute the x and y components of the batted balls velocity. Before we could do this we had to convert the launch angle into radians and velocity into ft/s. After doing this we can find the x-components velocity by taking the value of the angles cosine multiplied by the velocity. We found the y-components velocity similarly except we used the angles sin value.

$$v = \text{velocity (mph)}$$

$$a = \text{launch angle (degrees)}$$

$$vx = \text{x-component of velocity (ft/s)}$$

$$vy = \text{y-component of velocity (ft/s)}$$

$$vx = v * 1.467 * \cos(\tfrac{\pi}{180} * a)$$

$$vy = v * 1.467 * \sin(\tfrac{\pi}{180} * a)$$

We then took into account the balls acceleration and how it would affect the distance the ball travels. To do this we found the x and y components of the balls acceleration.

$$g = 32.2 \text{ (ft/s) gravity}$$

$$c = 0.0000835 \text{ (ft/s) drag constant}$$

$$ax = -c * v * vx * g$$

$$ay = -c * v * vy * g - g$$

While the balls y position was greater than 0 we used python to update the x and y components of velocity at small discrete time steps. We then used these updates in velocity to find the balls x and y position. When y equaled 0 we added the discrete time steps and returned the x position and time traveled.

$$dt = 0.1 \text{ (seconds)}$$

$$vx = vx + ax * dt$$

$$vy = vy + ay * dt$$

$$x = x + vx * dt$$

$$y = y + vy * dt$$

$$t = t + dt$$

Fielder Movement

We decided to model fielder movement based on player tracking data and human movement models. Intuitively we thought that Lorenzo Cain will not reach top velocity instantly and that his top acceleration would hold be until he reaches max velocity. Accessing player tracking data we found that Lorenzo Cain's top velocity and acceleration was $30ft/s$ and $15.1ft/s^2$ respectively. To model this we initialized acceleration as $15.1ft/s^2$ and decreased it as time went on until our fielder hit his max velocity. We assumed that when he hits max velocity acceleration would be equal to 0.

Using Python we updated the values of acceleration, velocity, and distance using small discrete time steps. We used an if statement to set acceleration equal to 0 when top velocity was reached and kept track of how far our fielder ran using our updated values. Given a certain time we could approximate the distance a fielder ran and whether or not he could catch a ball so many feet away in a certain time.

Initializing Variables

$xt = 0$ (distance traveled)

$t = 0$ (time)

$dt = .01$ (discrete time step)

$v_0 = 0$ (initial velocity)

$a_0 = 15.1$ (initial acceleration)

Python Code

$t = t + dt$ (updating time by dt)

$a = a_0 - 0.2 * t$ (decreasing acceleration as t goes on)

$v = v_0 + a * t$ (updating velocity)

if $v > 30$: (if player reaches top velocity)

$v = 30$

$xt = xt + v * dt$ (updating distance)

Rules of the Game[2]

Baseball is a bat-and-ball game played between two teams of nine players each who take turns batting and fielding.

The batting team attempts to score runs by hitting a ball that is thrown by the pitcher with a bat swung by the batter, then running counter-clockwise around a series of four bases: first, second, third, and home plate. A run is scored when a player advances around the bases and returns to home plate.

Players on the batting team take turns hitting against the pitcher of the fielding team, which tries to prevent runs by getting hitters out in any of several ways. A player on the batting team who reaches a base safely can later attempt to advance to subsequent bases during teammates' turns batting, such as on a hit or by other means. The teams switch between batting and fielding whenever the fielding team records three outs. One turn batting for both teams, beginning with the visiting team, constitutes an inning. A game comprises nine innings, and the team with the greater number of runs at the end of the game wins. Baseball is the only major team sport in America with no game clock, although almost all games end in the ninth inning.

References

[1] Wyers, Colin. "Hang Time." The Hardball Times. N.p., 6 July 2009. Web. 13 May 2016.

[2] "Baseball." Wikipedia. Wikimedia Foundation, n.d. Web. 13 May 2016.