Analyzing Arm Strength’s Impact on Minor-League Position Player Promotion

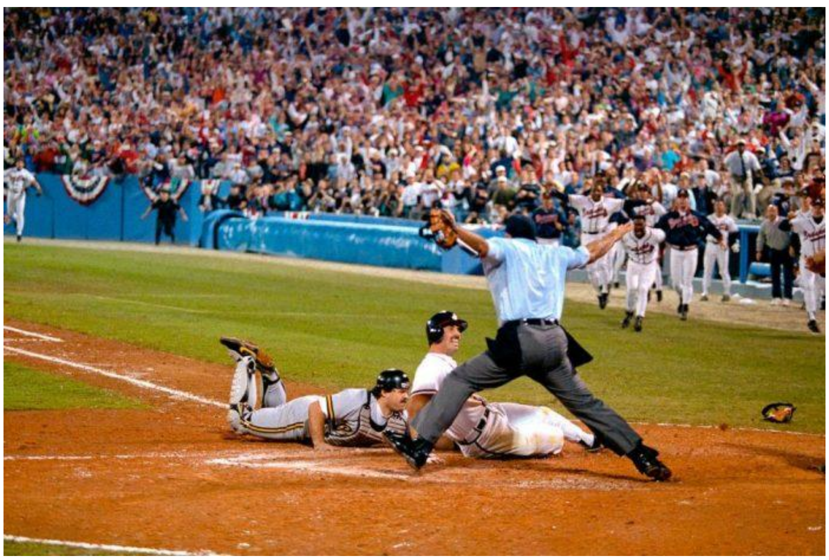
https://github.com/DakotaOlson/arm-strength-analysis

Abstract

Arm strength is considered one of the five tools that make a position player a useful contributor to their team, but how much does the skill factor into the real minor league management decisions an organization makes? This analysis examines the impact arm strength had on position players being promoted in an organization’s minor league system. Estimated throwing speeds of 1,601 throws made by position players in the same organization were calculated using baseball position data collected during the throw with a simulation and optimization approach. In this analysis, the maximum estimated throwing speed was determined and used to represent the arm strength of 117 players in the same organization by position, year, and minor-league level to determine the impact arm strength had on promotion. Results indicated that arm strength had little to no correlation with a position player being promoted or demoted regardless of position as many players with the strongest arm strengths remained at the same minor-league level or appeared to have left the organization, while several players with the weakest arm strengths at their positions were promoted. This indicates that despite being an important skill and one of the five tools, the arm strength skill is being overlooked when determining which players to promote in the organization being analyzed.

I. Introduction

It's October 14th, 1992 at Atlanta-Fulton County Stadium just before the stroke of midnight. Francisco Cabrera of the home team Atlanta Braves stands at the plate opposite Pittsburgh Pirates' closer Stan Belinda. The Pirates are leading 2-1 with two outs, but the bases are loaded in the bottom of the ninth-inning of Game Seven of the 1992 NLCS. Cabrera hits a line drive into shallow left field that is fielded by Barry Bonds as David Justice scores the tying run. Braves' third-base coach Jimy Williams waves the slow-running Sid Bream home as Bonds throws to catcher Mike LaValliere. Bream nears home plate as the throw arrives offline and bounces in front of LaValliere toward the first-base line. Bream slides toward home plate as LaValliere receives the ball and dives trying to make a tag that will keep the Pirates' season alive, but Bream reaches home barely before LaValliere can make the tag and is called safe [R1].



The Braves had completed the ninth-inning comeback to win their second straight NL pennant and advanced to the 1992 World Series. The Pirates had come up short for the third consecutive year, would lose Bonds and starter Doug Drabek to free agency, and failed to return to the playoffs for 21 years. If Bonds' throw to LaValliere was more accurate or had more speed, LaValliere would have made the tag that sent this game into extra innings and perhaps the Pirates would have ended up as the dominant team in the NL throughout the 1990s. Overlooked skills can make a significant difference on a team's path to winning a championship. I will be analyzing one of the factors that could have altered the outcome of this game: fielder arm strength. To enhance this analysis, I will examine arm strength by fielder position and how it impacts minor-league position player promotion.

II. Data

SMT provided each of the three datasets utilized. All three datasets contain information collected from ninety-seven games over four years ranging from Low-A to Double-A in the home parks of teams in the same organization. Due to data anonymization that occurred, the years represented in the datasets range from 1900 to 1903. The games are distributed across the minor-league levels as shown in Table 1.

Distribution of Games by Minor-League Level					
Level	Home Team	1900	1901	1902	1903
Double-A	TeamB	9	9	9	9
High-A	TeamA3	0	9	9	9
Single-A	TeamA2	0	0	9	9
Low-A	TeamA1	0	0	7	9

Table 1: Number of games included for each year by minor-league level.

The ‘game_events’ dataset includes information about the ball events (e.g., pitch, hit, throw) that occurred during each play and contains 103,444 entries and seven fields. The fields ‘event_code’ and ‘player_position’ contain IDs that represent an event that occurred during each play and the position of the player involved with the event respectively. These fields were used to find entries where a fielded baseball was thrown and which position made the throw. A ‘DataFrame’ [A1] was created that contained the 3,698 entries from ‘game_events’ where a fielded baseball was thrown. One entry had to be removed as the throw was recorded to have happened after an ‘end_of_play’ event which prevented it from being used to calculate the estimated speed of the throw.

The ‘game_info’ dataset includes information about the players on the field for each play and contains 26,315 entries and twenty fields. The ‘player_id’ who made each throw was identified using the field corresponding to the position who made the throw in ‘game_events’, and this value was added to the ‘DataFrame’. The fields ‘home_team’ and ‘top_bottom_inning’ were also added to the ‘DataFrame’ to later determine if a player was promoted and filter out players not in the organization respectively. One limitation identified was that 176 throws were missing ‘game_info’ entries. Of these, 136 plays’ ‘game_info’ entries were inferred as the entries for both the plays before and after the missing play were in the same half-inning and there were no defensive substitutions. The other forty throws were omitted.

The ‘ball_position’ dataset includes information about the position of the baseball throughout each play and contains 785,041 entries and six fields. Three fields, one for the position in feet for each dimension in three-dimensional space, were used to calculate the displacement of the baseball during the throw. An assumed value for air density had to be used while calculating the effect of air resistance as ‘ball_position_z’ represents the height of the baseball to the field and not sea level.

III. Methodology

The estimated speed was calculated for each of the 3,657 throws in a ‘DataFrame’ using the following approach:

1. The ball positions measured throughout the trajectory of each throw were found.
2. Ball positions that should have been recorded before or after each throw (e.g., bouncing, rolling, being fielded off the ground) were removed. There were 109 throws where this left a maximum of one measured ball position during the throw and an estimated speed couldn’t be calculated.
3. Using kinematic equations [A3], the initial horizontal (ground [A4]) and vertical velocities of each throw was estimated with the respective displacements between the two ball positions closest to the time of the throw. For these estimates, it was assumed there was no air resistance and the only force acting on the baseball was gravity [A5].
4. The trajectory of each throw was simulated using the estimated velocities with gravitational and drag forces being applied [A9]. This was done as the acceleration caused by drag force is proportional to the current velocity of the baseball. The acceleration caused by drag force [A8] was calculated using the following assumptions: the mass of the baseball is $5\frac{1}{8}\ ounces$ [R2], the cross-sectional area of the baseball was calculated using a circumference of $9\frac{1}{8}\ inches$ [R2][A6], the air density is $0.002377\ slug/ft^3$ [R3][A7], and the drag coefficient is 0.345 which is explained below.
5. The difference between the simulated trajectory and the real trajectory was calculated using Euclidean distance [A10].
6. ‘SciPy’ [A11] was used to repeat steps four and five trying new velocity values until the Euclidean distance between the trajectories was minimized, yielding more accurate velocity estimates.
7. Using the optimized estimated initial horizontal and vertical velocities, the estimated throwing speed was calculated in miles per hour.

The drag coefficient of an authentic MLB baseball at 80-mph ranges from 0.29 to 0.44 [R4]. Of the values simulated, a drag coefficient of 0.345 generated estimated throwing speed results with the smallest average Euclidean distance between trajectories for each throw and was used as the assumed value. Estimated throwing speeds with an average Euclidean distance between trajectories greater than three feet were removed from the 'DataFrame'. There were fifty-one such throws.

Outlier analysis was performed on the remaining 3,497 estimated throwing speeds. The distribution of estimated throwing speed was determined to not be normal [A12], which is understandable as a fielder isn't always trying to throw the baseball at maximum speed. Because the distribution wasn't normal, the IQR method for outlier detection [A13] was used to remove ninety-four throws. Each of these throws had low-estimated speeds, so the effect on this analysis was inconsequential. Throws made by pitchers were removed as this analysis focused on position players. This left 3,221 estimated throwing speeds ranging from 27.9 to 102.8-mph made by 450 players (Figure 1).

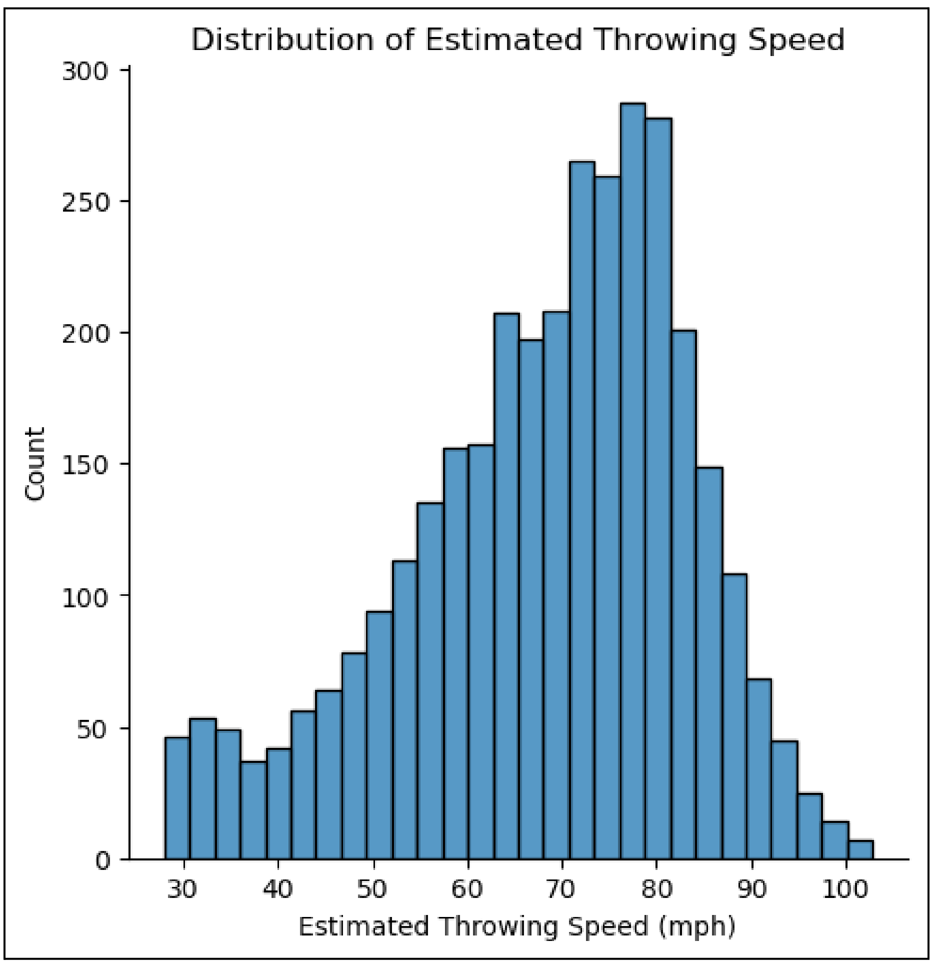


Figure 1

Figure 2: Maximum estimated throwing speed for each player at each position they fielded over a box and whisker plot [A14].

Maximum estimated throwing speed was chosen to represent arm strength as these are minor-league players and the maximum value can be considered as indicative of a player's arm strength potential (Figure 2). Throws made by players outside the organization were filtered out using the field 'top_bottom_inning' as the home team always fields in the top half-inning. This reduced the number of throws to 1,601 made by 117 players (Figure 3).

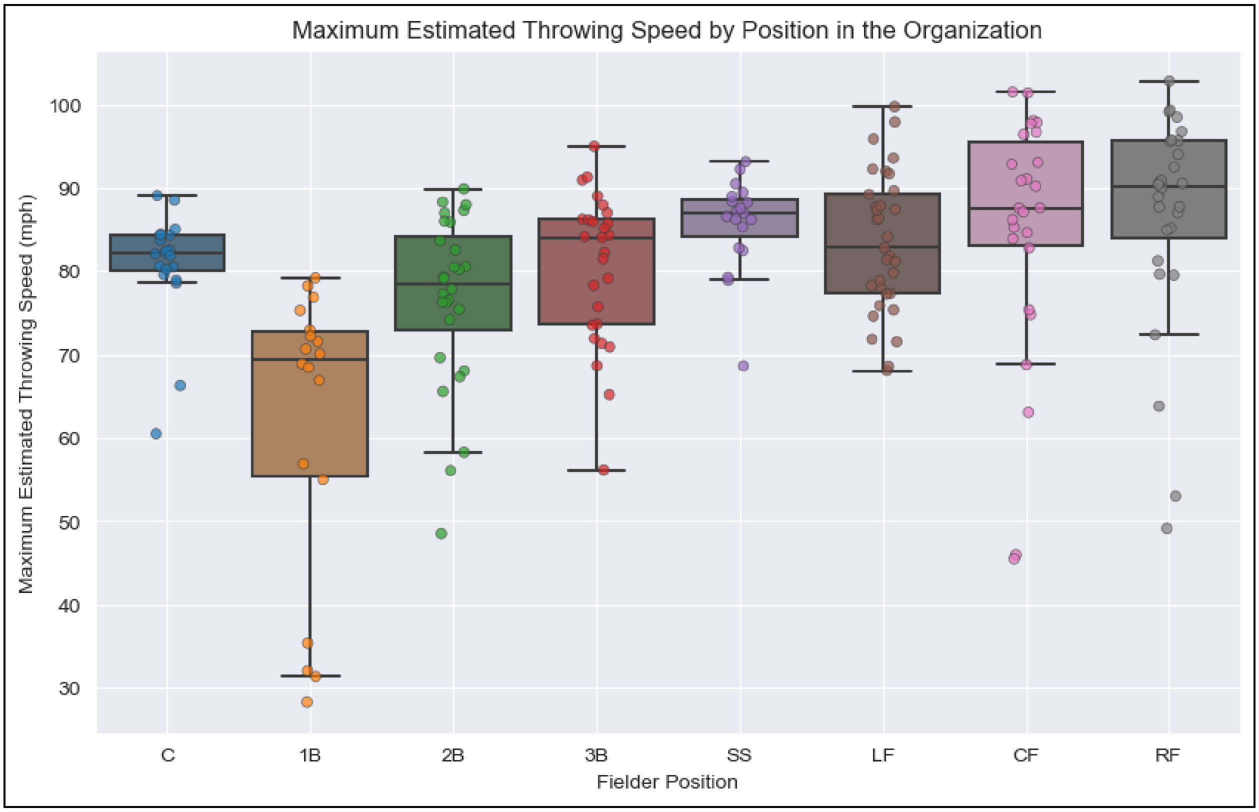


Figure 3: Maximum estimated throwing speed for each player in the organization at each position they fielded over a box and whisker plot [A14].

IV. Discussion

Estimated throwing speed was calculated for 1,601 throws made by position players in the same organization and the maximum throwing speed for each player by position, year, and minor-league level was found to analyze how arm strength impacts promotion. From 1900 to 1901 no promotions (or demotions) occurred. This was expected as 1900 only included Double-A data. From 1901 to 1902 no promotions occurred between years, however, during 1902 one player was promoted and two players were demoted (Figure 4).

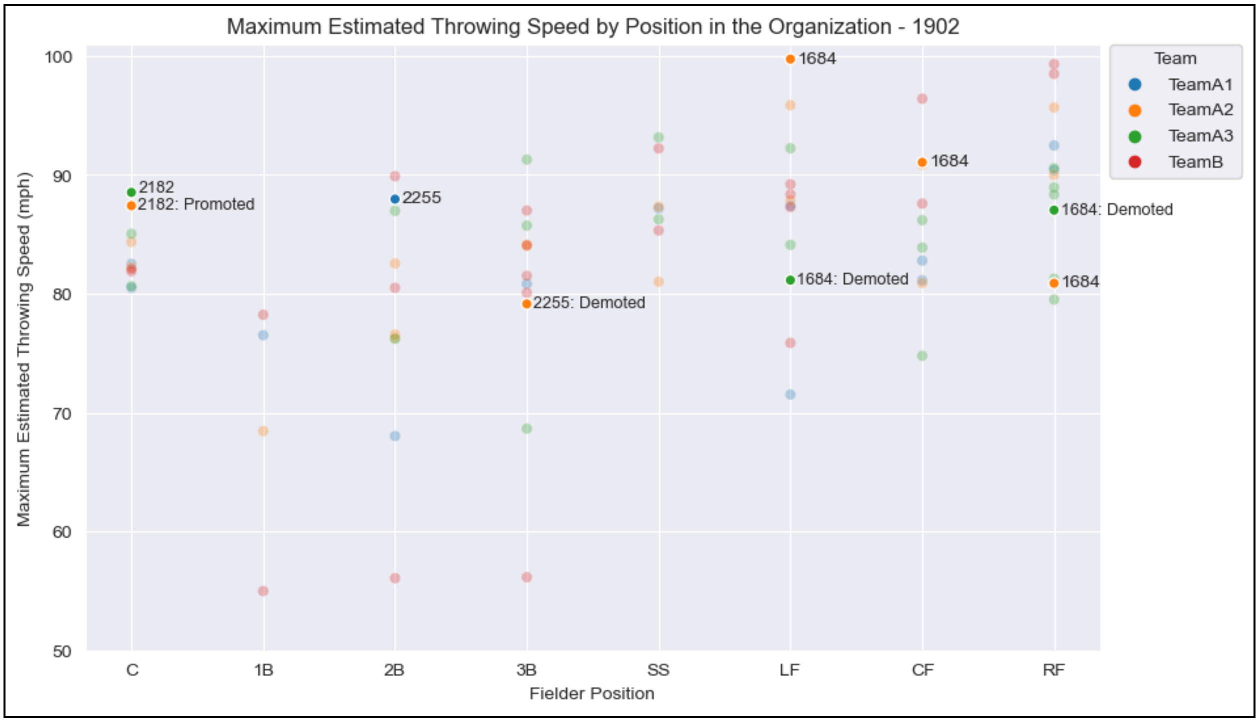


Figure 4: Maximum estimated throwing speed for each player in the organization at each position they fielded during 1902 with promoted/demoted players highlighted.

Player ‘2182’ is a catcher who was promoted from Single-A to High-A and had the strongest arm strength of any catcher in the organization during 1902. Player ‘2255’ is a third baseman who was demoted from Single-A to Low-A and demonstrated the third-weakest arm strength of any third baseman in the organization during 1902. It should be noted that following the demotion, ‘2255’ transitioned to being a second baseman and demonstrated the second-strongest arm strength of any second baseman in the organization during 1902. Whether this increase led to ‘2255’ being promoted following 1902 will be investigated. Player ‘1684’ is an outfielder who was demoted from High-A to Single-A and demonstrated a slightly below-average arm strength for an outfielder in the organization during 1902 (87.0-mph compared to the average 88.4-mph). Following the demotion, ‘1684’ recorded the strongest estimated throwing speed of any throw made by a player in the organization during 1902 with a 99.7-mph throw from left field. Again, whether ‘1684’ was promoted following 1902 will be investigated.

From 1902 to 1903, sixteen promotions or demotions occurred, excluding promotions during 1902. The first notable aspect was that none of ‘2182’, ‘2255’, and ‘1684’ appeared in 1903 data indicating the players were either promoted past Double-A, injured, or left the organization. Between 1902 and 1903, eight players were promoted (Figure 5 and Table 2).

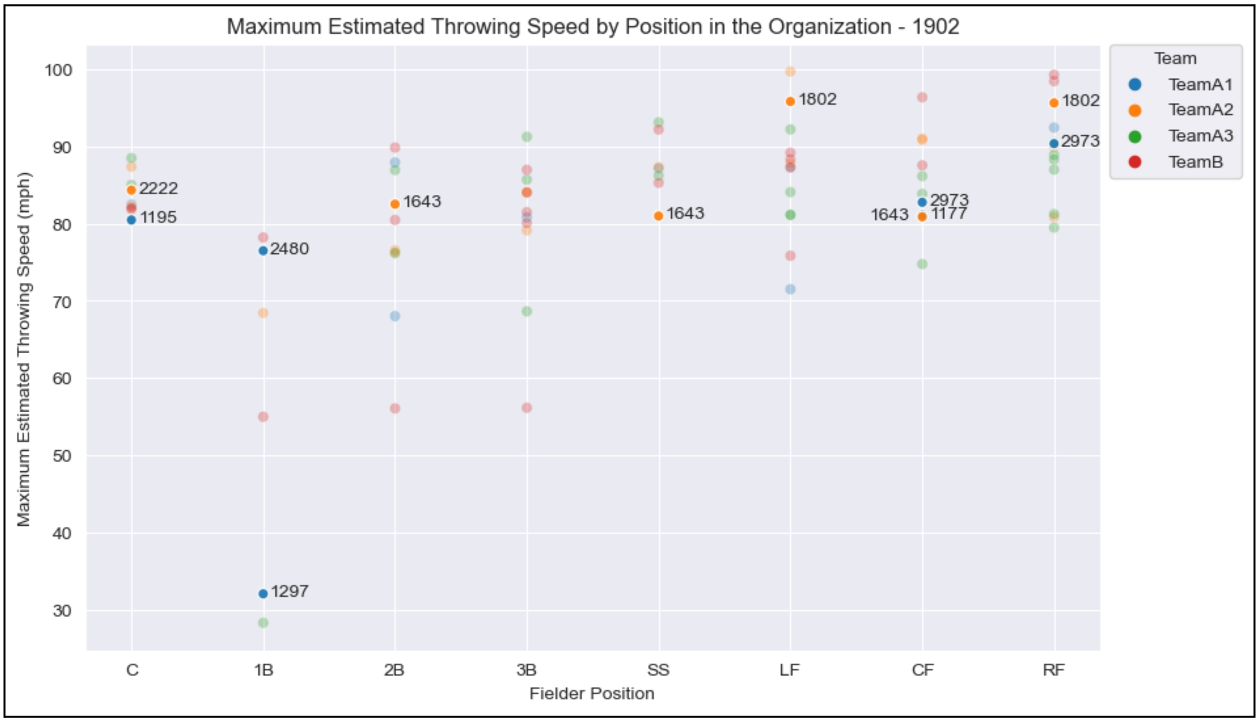


Figure 5: Maximum estimated throwing speed for each player in the organization at each position they fielded during 1902 with players promoted after the season highlighted.

Players Promoted/Demoted After 1902 or During 1903				
Player	Position(s)	Year	Max Est. Throwing Speed	Change
2222	C	1902	84.4 mph	Promoted
1195	C	1902	80.5 mph	Promoted
2480	1B	1902	76.5 mph	Promoted
1297	1B	1902	32.0 mph	Promoted
1643	2B, SS, CF	1902	82.5 mph (2B)	Promoted
1802	LF, RF	1902	95.9 mph (LF)	Promoted
2973	CF, RF	1902	90.4 mph (RF)	Promoted
1177	CF	1902	81.1 mph	Promoted
1240	C	1903	77.8 mph	Promoted
2382	1B, 2B, 3B	1903	82.3 mph (3B)	Promoted → Demoted
2053	2B, LF	1903	82.8 mph (LF)	Promoted
2137	2B	1903	57.4 mph	Promoted
1725	3B, OF	1903	73.5 mph (3B)	Promoted
2148	SS	1903	86.9 mph	Demoted
1181	SS	1903	77.5 mph	Promoted
2837	RF	1903	94.0 mph	Promoted

Table 2: ‘Max Est. Throwing Speed’ is from before the promotion/demotion occurred* and players who played multiple positions have their strongest throw listed. (*Excluding ‘2382’)

During 1903, six players were promoted, one player was demoted, and one player was promoted and then demoted (Table 2 and Figure 6).

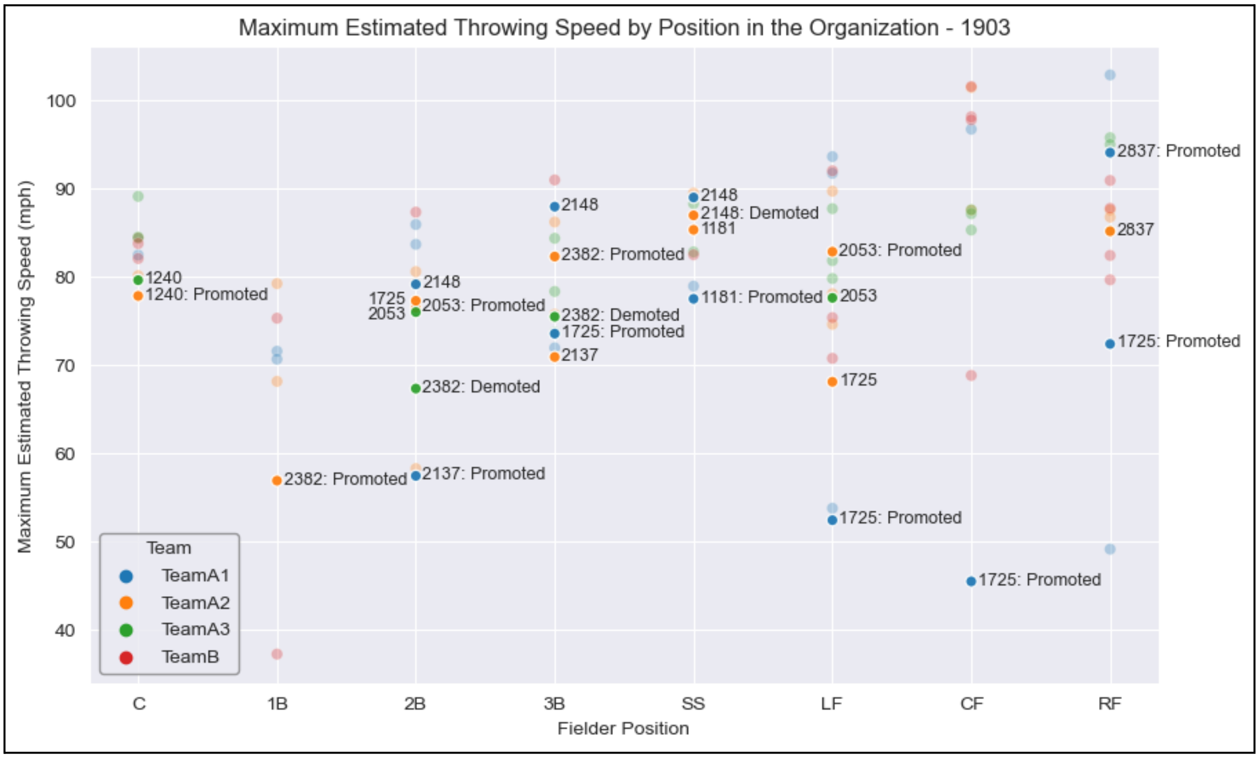


Figure 6: Maximum estimated throwing speed for each player in the organization at each position they fielded during 1903 with promoted/demoted players highlighted.

As shown in Figures 5 and 6, there appears to be little correlation between a player's arm strength and being promoted. For players promoted between 1902 and 1903, Figure 5 shows the catcher with the weakest arm strength ('1195') and the shortstop with the weakest arm strength ('1643') were all promoted while most of the players with the strongest arm strengths at those positions weren't promoted or appeared to no longer be part of the organization ('2182', '1684'). For players promoted or demoted during 1903, Figure 6 shows again that the catcher with the weakest arm strength ('1240'), the shortstop with the weakest arm strength ('1181'), and the outfielder with the weakest arm strength ('1725') were all promoted. Additionally, '2148' had one of the strongest arm strengths of shortstops and was demoted. There does appear to be a slight correlation between outfielders with 90.0-mph or better arm strengths getting promoted ('1802', '2973', '2837'), but for other positions, there appears to be little or no correlation.

It can be concluded that this organization isn't valuing the five-tool skill of arm strength when deciding who to promote or demote regardless of position. Catchers '2182' (88.5-mph) and '2222' (89.0-mph) had arm strengths nearly 10-mph better than average catchers. '2182' appeared to leave the organization after 1902, and '2222' remained in High-A throughout the 1903 season. Right fielder '1008' had the strongest arm strength of any player (regardless of whether they were in the organization) with a maximum estimated throwing speed of 102.8-mph, yet remained in Low-A throughout 1903. '1684' had the strongest arm strength of any left fielder in the organization and also appeared to leave the organization after 1902. These are potentially lost runs saved.

V. Improvements

I would suggest several enhancements for improvement. The first would be a larger sample size. No player in the organization in the dataset was recorded to have thrown a fielded ball more than forty times, and an increased sample size would allow for a more accurate estimate of a player's "true" arm strength potential. Additionally, since a fielder isn't always trying to throw a baseball at maximum speed it would be best to collect data outside of a game. The simulation engine for a thrown baseball's trajectory could be improved by accounting for lift force which requires knowing the spin of a thrown ball to get a more accurate estimate of throwing speed. This analysis could be followed up by considering other factors that determine the fielding ability of a player, such as accuracy and running speed to determine if other fielding ability factors are more determinant in deciding if a player gets promoted.

VI. Acknowledgements

I would like to thank my colleague Olivia Falk (University of Minnesota - Class of 2022 - B.S. Physics and Astrophysics) who provided support on how to properly include drag force while simulating the trajectory of a thrown baseball.

VII. References

[R1] "92 NLCS, Gm 7 PIT@ATL: Bream Beats Bonds' Throw." *YouTube*, uploaded by MLB, 16 Aug. 2013, www.youtube.com/watch?v=FgjIVvEQo_o. Accessed 14 Aug. 2023.

[R2] *Official Baseball Rules.* 2023 ed., Major League Baseball, 2023.

[R3] Stephen Corda, Introduction to Aerospace Engineering with a Flight Test Perspective, John Wiley & Sons, 2017.

[R4] Albert, Jim et al. "Report of the Committee Studying Home Run Rates in Major League Baseball." 2018.

VIII. Appendix

[A1] DataFrame: A two-dimensional labeled data structure with columns of potentially different types (like a spreadsheet) included in pandas.

[A2] pandas: A Python programming language library used for data manipulation and analysis.

[A3] Kinematic Equation: $v_i = \frac{d}{\Delta t} - \frac{a * \Delta t}{2}$, where v_i is initial velocity, d is displacement (change in position), Δt is elapsed time (end time subtracted by start time), and a is acceleration.

[A4] Horizontal (Ground) Displacement: $d_{horz} = \sqrt{(x - x_i)^2 + (y - y_i)^2}$, where d_{horz} is displacement in the horizontal direction, x is 'ball_position_x', x_i is 'ball_position_x' at the time of throw, y is 'ball_position_y', and y_i is 'ball_position_y' at the time of throw

[A5] Gravitational Acceleration: $32.174\ ft/s^2$.

[A6] Cross-Sectional Area: $A = \pi * r^2$, where A is the cross-sectional area and r is the radius.

[A7] Air Density: Air density is $0.002377\ slug/ft^3 \Leftrightarrow 1.225\ kg/m^3$ at standard sea-level conditions.

[A8] Drag Force: $F_{D(horz/vert)} = \frac{1}{2} * \rho * C_D * A * v * v_{(horz/vert)}$, where $F_{D(horz/vert)}$ is the force caused by drag in the horizontal or vertical direction in poundal, ρ is the air density, C_D is the drag coefficient, A is the cross-sectional area, v is velocity relative to the air, and $v_{(horz/vert)}$ is the velocity in the horizontal or vertical direction.

[A9] Acceleration Components: $a_{horz} =- \frac{F_{D,horz}}{m}$ and $a_{vert} =- g - \frac{F_{D,vert}}{m}$, where $a_{(horz/vert)}$ is the acceleration in the horizontal or vertical direction, g is gravitational acceleration, $F_{D(horz/vert)}$ is drag force in the horizontal or vertical direction, and m is the mass.

[A10] Euclidean Distance: $d(p, q) = \sqrt{(q_{horz} - p_{horz})^2 + (q_{vert} - p_{vert})^2}$, where p and q are points with coordinates (p_{horz}, p_{vert}) and (q_{horz}, q_{vert}) respectively.

[A11] SciPy: A Python library used for scientific and statistical computing that includes modules for optimization.

[A12] Normal Distribution: A probability distribution where the mean (average) and median (50th percentile) are expected to be about the same and when graphed by frequency a bell-shaped curve is observed. As fielders aren't always trying to throw the baseball at maximum speed this creates a negatively skewed distribution where the mean is less than the median.

[A13] IQR Method: $IQR = Q_3 - Q_1$, where IQR is the interquartile range, Q_3 is the third quartile (75th percentile), and Q_1 is the first quartile (25th percentile). Any data points that are over the upper limit $Q_3 + 1.5 * IQR$ or below the lower limit $Q_1 + 1.5 * IQR$ are considered outliers.

[A14] Box and Whisker Plot: A chart that includes a box that contains fifty percent of the data with the lower edge being the first quartile, the upper edge being the third quartile and a line for the median. Additionally, the chart contains 'whiskers' which are the minimum and maximum values after excluding outliers according to the IQR method. The included box and whisker plots contain outliers as outliers in this case are determined by maximum estimated throwing speed by fielder position rather than when outliers were removed earlier in comparison to all throws.