This Isn't a Stretch: Quantifying Ball Acquisition Proficiency to Evaluate Fielders on Assisted Put-Outs

https://github.com/danielhocevar/SMT-Baseball-2023

Abstract. Player and ball tracking technologies such as Statcast [1] have greatly improved our ability to measure the defensive proficiency of fielders. However, certain positions have historically been harder to assess than others, necessitating continued development of new statistics. We propose that shortcomings in these evaluative frameworks come as a result of failing to acknowledge differences in the composition of each position's defensive touches. Current methods focus overwhelmingly on independent batted ball acquisition and throwing abilities, but limited efforts have been made to evaluate a player's ball-receiving ability from throws, aside from catchers. This is notable because **over 80%** of a first baseman's total defensive touches, for example, come from catching the ball. Therefore, this paper will introduce a ball capture metric intended to improve defensive analysis of first basemen and be used as a stepping stone to quantify credit assignment on collaborative plays between fielders. This will augment existing metrics that evaluate fielding performance and be an improved discriminator of defensive ability, allowing us to make more comprehensive assessments of player value.

Introduction

It's the bottom of the eighth inning and the away team has a two-run lead with runners on first and third base and one out. Hits have been hard to come by throughout the game and the tension is high. The pitcher delivers a fastball that is pounded on the ground at 103.2 mph - directly at the third baseman. He flips the ball to the second baseman, who steps on the bag and fires a throw to first base to complete what the scorecard would deem a seemingly routine 5-4-3 inning-ending double play. However, **this video replay** reveals that the ending of the play was anything *but* routine. These types of plays invoke awe-filled remarks such as "what a stretch by the first baseman". Such comments reveal our desire to place special value on the effort of a particular fielder in completing a play involving the interaction of multiple defensive teammates. However, these qualitative descriptions do little to provide a concrete administration of credit, and trying to establish a more quantitative framework to attribute credit to each fielder remains quite nebulous.

To address this question, we define **three types** of possible actions that can be performed by a fielder: acquiring a batted ball, throwing an acquired ball, and acquiring a thrown ball. A distinction is made between acquiring a batted ball versus a thrown ball because the former can happen at most once in a play sequence, while the latter, paired with a throw, can theoretically occur ad infinitum (and in the case of **some rundowns**, seem to)! When evaluating defensive proficiency, emphasis is placed primarily on a fielder's ability to make a play on a batted ball, assessed with metrics such as Defensive Runs Saved [2], the SABR Defensive Index [3], and Statcast's Fielding Run Value [4]. Collaborative defensive play sequences have received understandably less attention than these individual metrics due to their decreased frequency and increased difficulty to assess, headlined by the "assist" statistic [5], arm strength, and metrics evaluating double play conversion rates [6]. However, aside from catchers, very few attempts have been made to rate a fielder's ability to capture a throw, the third type of outlined defensive action. Potentially overlooked due to how integral catching a ball is to the sport of baseball, it is important to note that these types of plays make up the majority of multiple infield position's defensive event composition, illustrated in Table 1.

Table 1: Composition of defensive touches for infielders and outfielders, from the provided data

	Proportion of Defensive Touches (%)		
Player Position	Batted Ball Acquisition	Thrown Ball	Thrown Ball Acquisition
Center Fielder	61.44	37.38	1.18
First Baseman	13.46	4.65	81.89
Left Fielder	59.67	39.84	0.49
Right Fielder	60.86	38.74	0.41
Second Baseman	31.03	30.51	38.46
Shortstop	29.95	29.08	40.97
Third Baseman	44.59	33.16	22.24

Although these actions may not require on average the same level of concerted "effort" as acquiring a batted ball or throwing a baseball, examining caught throws could improve holistic fielder assessments, particularly at positions where they make up large proportions of their defensive touches, such as at first base. This coincides with observations by those who follow baseball that the aforementioned individual defensive statistics tend to fall short when evaluating first basemen in particular, contradicting both the "eye-test" and anecdotal reports from teammates and commentators [7]. Therefore, this project will investigate collaborative plays between infielders and more specifically look to quantify a first baseman's catching ability on assisted put-out attempts. Analogous to other "plus-minus" metrics such as Outs Above Average [8], our proposed Catches Above Average (`CAA`) metric is a flexible framework intended to be incorporated into existing paradigms that rely solely on batted ball acquisition metrics to measure defensive performance.

Data

The data used for this project came from 97 anonymized Minor League Baseball (MiLB) games across three seasons. Part of this data came in the form of sequential game events coded by player position and event type and ordered by timestamp. Specialized functions were created to translate these game codes into human-readable play-by-play descriptions, providing contextual information to augment animated visualizations created using the 3D (x,y,z) ball coordinate data and 2D (x,y) player coordinates, hyperlinked in Figure 1.

```
play_by_play("1903_32_TeamNB_TeamA1", 2) #game_id, play_number
```

```
[1] "pitch thrown by pitcher, ball hit into play by batter,
ball acquired by right fielder, end of play."
```

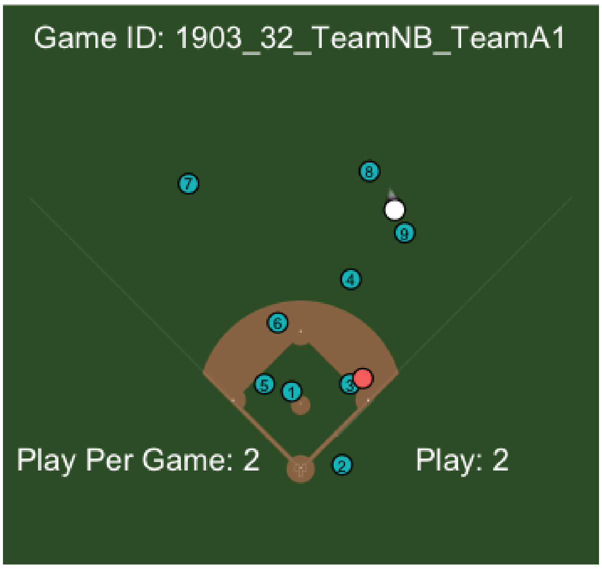

Fig. 1: **Example play animation.**

Each player was also provided with a unique identification number such that those on the field at the time of an event and their positions (whether in the field or on which base) could be obtained. However, numerous gaps in this information that made it unrealistic to accurately derive the game state - the number of outs in an inning at the advent of an event. This information could have been used to proxy for *leverage* and improve the quality of evaluation for the players included in this analysis. To this end, player identification was solely used to distinguish and rank first basemen, and the majority of the analysis utilized the 3D ball location coordinates and metrics that could be calculated with these measures. A robust process (described in the **analysis code**) was used to identify throws to first base and the final dataset contained 1918 throws.

Methodology

To quantify catch difficulty, a geometric approach was taken. Consider a first baseman waiting to receive a throw from the second baseman, just as Vladimir Guerrero Jr. was from Ernie Clement in the example play. Vlad is initially square to Clement and sets a target with his glove, and Clement is tasked with putting a throw in that general vicinity, analogous to a pitcher aiming at a catcher's glove. Standing straight up, Vlad has a bounded 3 dimensional range of motion that can be achieved by moving only his arm, and if Clement's throw is outside of this initial range, Guerrero will have to bend over, jump, or stretch/lean to increase the volume of this range. The intuition follows that the larger a first baseman needs to make his viable surface volume in order to catch a throw, the worse the throw is. This is analogous to a strike zone, except that it expands as the first baseman makes more adjustments to be able to corral an incoming throw. Therefore, after rotating the surface so that the angle between the first baseman and fielder is zero (that is, they are square to each other), we describe the "catchability" of Clement's throw by both its **(x,y,z) coordinates** as it intersects with the surface and the **size** of this surface at the time of intersection. So then, a throw that intersects with a small surface volume can be thought of as a "good throw", irrespective of whether or not it is caught, simply because it is bounded by a spatial domain deemed reachable by the first baseman. Clement's throw, on the other hand, would have a very large surface scale factor due to how much Guerrero had to stretch in order to make the catch, giving Vlad a lower catch probability to overcome. We call this probability Expected Catch Rate (`xCR`). For the `xCR` model, an ellipsoid (formula in Appendix) was used as the 3D surface to predict the outcome of whether a throw was caught or not. The proposed features were:

- Ellipsoid scale factor
- The (x,y,z) coordinates of the throw as it intersected the ellipsoid

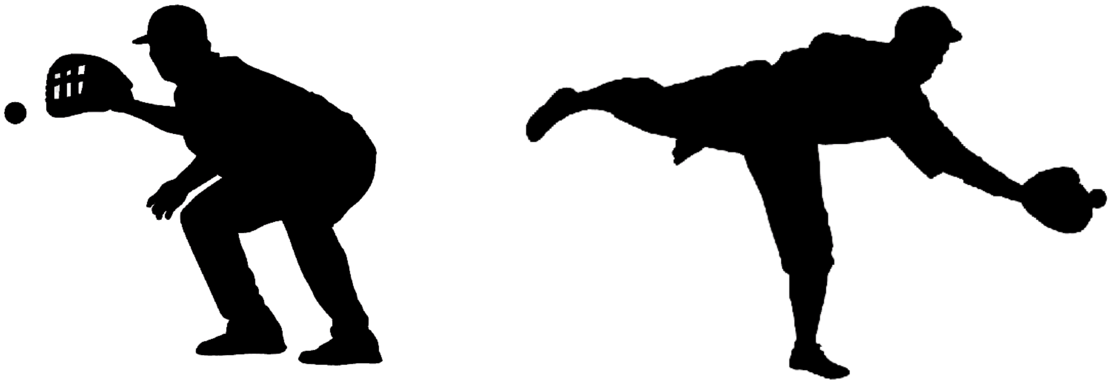

Fig. 2: The fielder on the left doesn't have to move very much to catch the incoming throw, so the necessary surface range is relatively small. The fielder on the right, however, has to lean their body in an awkward position and stretch out their arm to make the catch, so their required catch surface is much larger.

- Whether the ball bounced
- Throw Specifications: velocity, distance, and duration
- Dummy variables to account for stadium specific z-coordinates

Neither the throw specification nor the stadium identifying variables were statistically significant in our baseline logistic regression model, so we used the remaining variables in the final catch model and tested 4 binary classification models: logistic regression, Naïve Bayes, random forest, and XGBoost. We used a 60/40 train/test split, growing 100 trees for the random forest and 10 rounds of boosting for the XGBoost with no additional tuning. The models had nearly identical values of test precision, recall, and F1 score, so we selected XGBoost since it had the largest test AUC (0.812). `xCR` values are visualized in Figures 3 and 4.

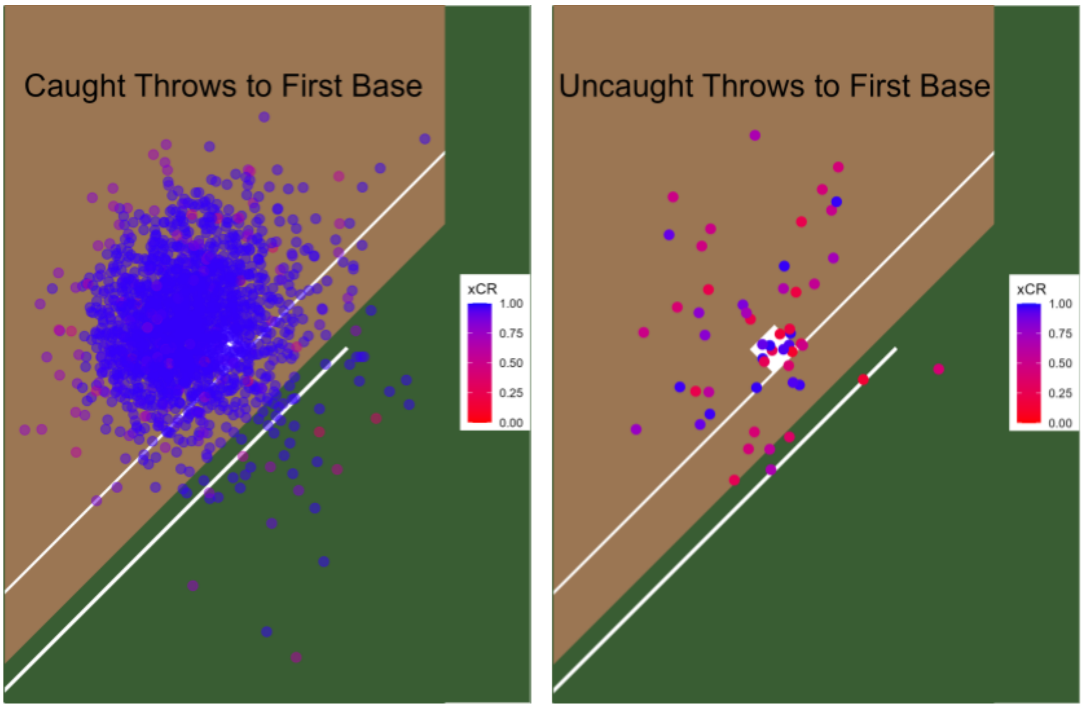

Fig. 3: Spray charts of caught and uncaught throws to first base, colored by catch probability (xCR). They align with our intuition of the relationship between the catch difficulty and spatial proximity to first base.

With a catch probability now assigned to every throw, we can now quantify the catching ability of a first baseman based on how well they receive throws of varying difficulty. We do so by taking the cumulative sum of the difference between the catch status (0 for uncaught, 1 for caught) and the `xCR` for each player. This

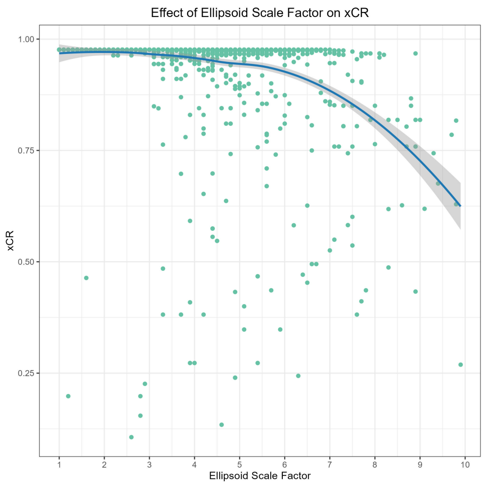

Fig. 4: xCR decreases as the ellipsoid center gets further from first base, which aligns with common schools of thought. Additionally, xCR is significantly lower for throws that bounce prior to reaching first base.

Table 2: Top and bottom five first baseman, ranked by Catches Above Average.

Rank	Player ID	CAA
1	1918	4.30
2	2526	4.17
3	2227	3.02
4	2480	2.32
5	8183	2.10
...	...	...
78	1980	-0.80
79	2774	-1.21
80	9218	-1.25
81	4202	-1.27
82	5524	-2.03

sum is Catches Above Average (`CAA`), a leaderboard of which is given in Table 2. An example comparison illustrating how to interpret the metric is given in Figure 5.

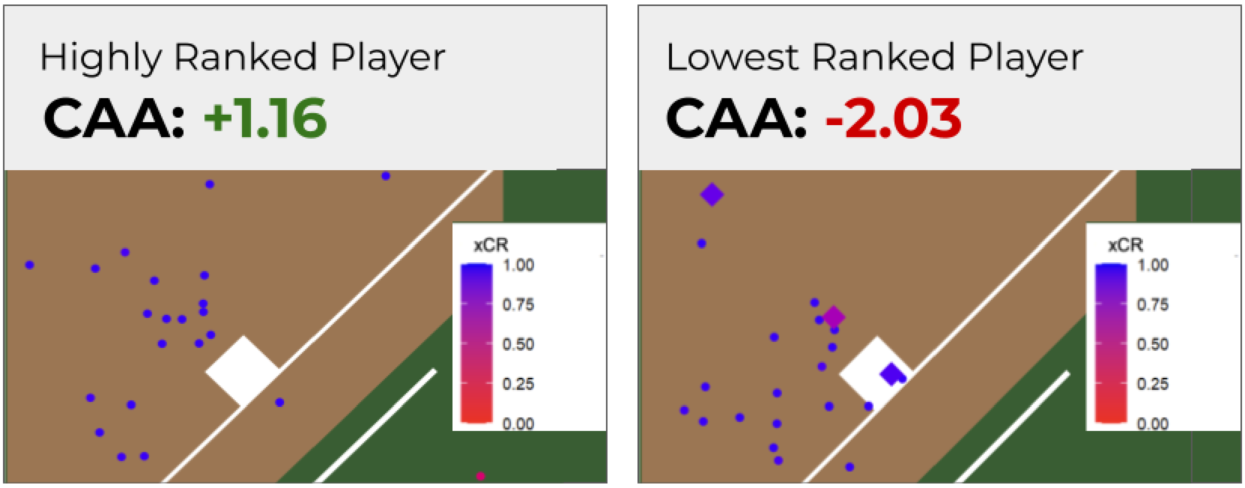

Fig. 5: Sample size matters when comparing players using CAA, so the two players chosen for this example each received 22 throws. The player on the left caught every throw, the most notable being one with a 38% catch probability. However, they are not higher on the CAA leaderboard because all of the other throws they received had xCR values greater than 96%, a testament to the throw quality from their teammates. Conversely, the player on the right had 3 missed catches at 70%, 90% and 95% catch probabilities. Since they also received high quality throws from their teammates, they are penalized harshly for their misses.

Discussion

The meta-analytic criteria [9] of **discrimination** and **independence** were employed to determine whether `CAA` provides significant "unique and reliable information". While robust quantitative analysis requires access to more multi-season data than available to us, we conducted simplified technical and semantic evaluations of our metric and its computation method using these indicators. According to meta analytics, metrics evaluating individual players should be able to adequately distinguish between players based on their performance. Figures 6 and 7 compare the distribution and variance of `CAA` to other existing metrics to evaluate its discriminatory power, and there is evidence that it is a stronger discriminator of player performance than standard Fielding Percentage and possesses similar discriminatory power to Outs Above Average. We use a semantic approach with the Clement-to-Guerrero example play to illustrate the independence of `CAA`. Here, Outs Above Average would give Vladimir Guerrero Jr. no credit at all, despite the fact that he made a spectacular stretch to complete the final out of the inning. Our framework identifies the fact that he caught an incoming throw with a low xCR, and boosts his `CAA` score accordingly, improving his overall defensive rating. This example demonstrates where `CAA` provides valuable information over existing metrics.

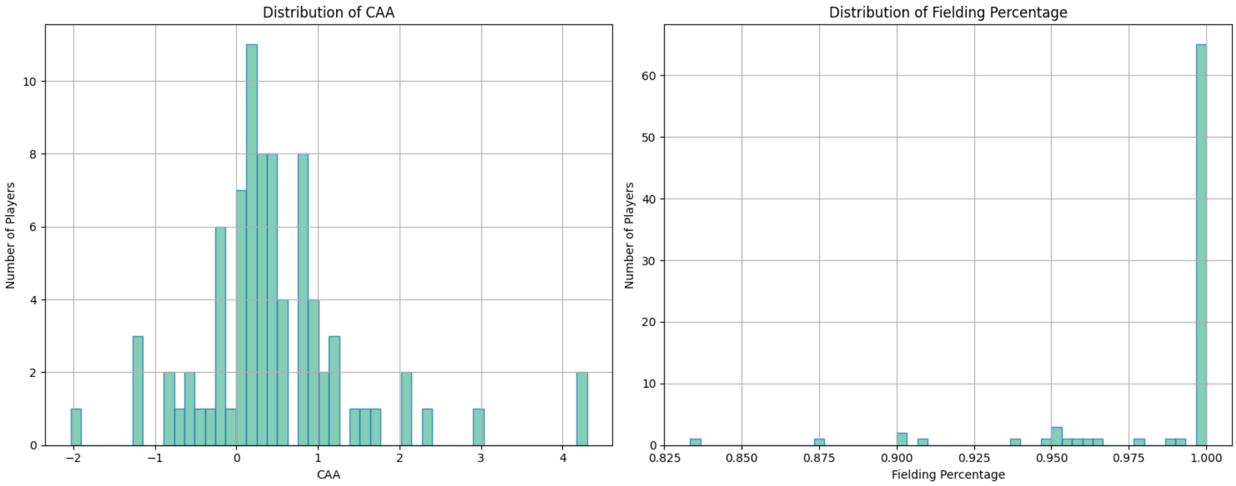

Fig. 6: CAA has a wider distribution and greater variance than Fielding Percentage, indicating that it is a stronger discriminator of player performance. Here, Fielding Percentage is computed on the same data as CAA with a simple classifier that assesses an uncaught ball with an xCR greater than 0.7 as an error.

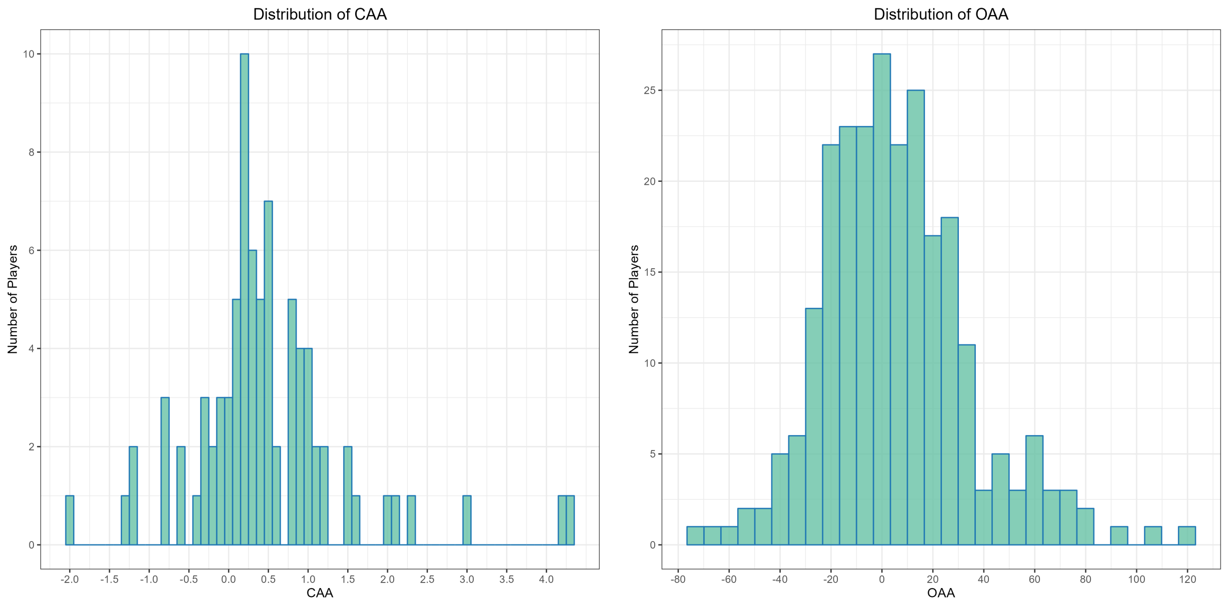

Fig. 7: The distributions of Catches Above Average and MLB's Outs Above Average [10] have striking similarities, even with widely different sample sizes. Both have a mean close to zero and a slight right skew that likely arises from survival bias. Both metrics are cumulative, so better players will get more opportunities to make defensive plays, pulling the distributions rightward.

Conclusion

In this paper, we have proposed a new approach to evaluating the defensive prowess of first-basemen. Motivated by the fact that the overwhelming majority of a first basemen's defensive touches come from receiving put-out throws, we introduce a novel model for grading the difficulty of catching a put-out throw called Expected Catch Rate (`xCR`). Using this model, we sum the difference between `xCR` and actual outcome across all plays in which a player is involved to produce a final metric we call Catches Above Average (`CAA`). `CAA` proves to be a much stronger discriminatory metric than standard Fielding Percentage, and provides new insight to player evaluation overlooked by other popular metrics such as Outs Above Average. Future application of our framework could extend to credit assignment problems. For example, `xCR` could be used to determine who "deserves" to be charged with an error on a defensive misplay, or alternatively help reveal whether a completed play came as a result of a good throw or a good reception by the thrower's intended target, adjusted by the difficulty of acquiring a batted ball, if applicable.

Future Work

While the initial results of our analysis are quite promising, there is still much room for extension. Practically, information on player heights and wingspans could potentially be valuable to the catch probability model, helping to control for a player's reach. Optimizing the target shape used to measure throw accuracy could be more reflective of a player's range of motion and including parabolic throw trajectories as a model parameter may improve the catch probability model. A major assumption of our `CAA` framework is that the first baseman is stationary at the time of receiving the ball, thereby presenting a fixed target for their teammate to throw at. While this is reasonable for many plays at first base, it doesn't capture plays happening at other bases nearly as well because there is a higher chance that the fielder covering the bag will be in motion, such as when turning a double play. Therefore, developing this metric for other infielders would have to account for this and the disproportionately smaller sample size of throws that non-first basemen receive. Finally, `CAA` could play a role in quantifying the "trust" between two fielders and help answer the bigger-picture question of whether there is a causal relationship between the receiving ability of a team's first baseman and the willingness of their infielders to attempt difficult throws. This could be further extended to evaluate an infielder's decision-making and assess how influential catching ability is compared to other factors that affect the propensity to make a throw, such as game state and the speed of a runner.

Acknowledgements

This work was made possible through collaboration within the University of Toronto Sports Analytics Student Group (**UTSPAN**). We also thank Dr. Meredith Wills for her support throughout this project.

Appendix

The formula for the ellipsoid is as follows, where $\Rightarrow$ signifies the returned scale factor based on each case:

$$Ellipsoid = \begin{cases} (\frac{x}{\alpha})^2 + (\frac{y}{\alpha})^2 + (\frac{z-4}{\frac{4}{3}\alpha})^2 \leq 1; \quad 1 \leq \alpha \leq 12.5 \times \frac{3}{4} \Rightarrow \alpha \\ (\frac{x}{\alpha})^2 + (\frac{y}{\alpha})^2 + (\frac{z-4}{12.5})^2 \leq 1; \quad 12.5 < \alpha \leq 13.5 \Rightarrow \alpha \\ \vdots \\ \Rightarrow -1; \quad \alpha > 13.5 \\ \Rightarrow 1; \quad 0 < \alpha < 1 \end{cases}$$

References

1. Major League Baseball. Statcast. `https://www.mlb.com/glossary/statcast`. Accessed: 2023-08-26.
2. Major League Baseball. Defensive runs saved (drs). `https://www.mlb.com/glossary/advanced-stats/defensive-runs-saved`. Accessed: 2023-08-26.
3. Society for American Baseball Research. Sabr defensive index. `https://sabr.org/sdi`. Accessed: 2023-08-26.
4. Baseball Savant. Fielding run value leaderboard. `https://baseballsavant.mlb.com/leaderboard/fielding-run-value/`. Accessed: 2023-08-26.
5. Major League Baseball. Assist (a). `https://www.mlb.com/glossary/standard-stats/assist/`. Accessed: 2023-08-26.
6. Baseball Prospectus. Double place percentage. `https://legacy.baseballprospectus.com/glossary/index.php?mode=viewstat&stat=281/`. Accessed: 2023-08-26.
7. Daniel R. Epstein. Best of bp: 1,000 grounders: What we still don't know about first base defense. `https://www.baseballprospectus.com/news/article/71706/best-of-bp-1000-grounders-what-we-still-dont-know-about-first-base-defense/`. Accessed: 2023-08-28.
8. Tom Tango. History of fielding. `http://tangotiger.com/images/uploads/History_of_the_Fielding.pdf/`. Accessed: 2023-08-26.
9. Alexander M. Franks, Alexander D'Amour, Daniel Cervone, and Luke Bornn. Meta-analytics: tools for understanding the statistical properties of sports metrics. *Journal of Quantitative Analysis in Sports*, 12(4):151–165, 2016.
10. Baseball Savant. Outs above average. `https://baseballsavant.mlb.com/leaderboard/outs_above_average/`. Accessed: 2023-08-23.