Vacuums and Stone Hands

An Exploration of First Baseman Receiving

Anonymous Author(s)

Link to GitHub Repo:
https://github.com/nicholson2208/smt-data-challenge

Abstract

In an era where the game is increasingly quantified, the receiving aspect of first baseman defense has largely been ignored. While particularly skilled (or lacking) play contributes to a fielder's reputation, even advanced defensive metrics fail to explicitly consider the role of the first baseman in the assist. In this analysis, I describe what makes an out at first from the perspective of the first baseman. I find that bounced and offline throws are less likely to be outs, and that some players seem to be better at fielding these than others, lending credence to the conventional wisdom that some possess this skill. While the data are insufficient to compute full player-level first baseman receiving rankings, I demonstrate a viable framework for its evaluation. Finally, I discuss the role that first baseman receiving can take in player development and valuation, and suggest future extensions of this approach.

SMT Data Challenge
August 2023

Vacuums and Stone Hands: An Exploration of First Baseman Receiving

Anonymous Author(s)

Figure 1: Rhys Hoskins stretches to shorten the throw to first for an out in the bottom of the third inning of 2022 World Series Game 6. Without the stretch, Astros catcher Martín Maldonado likely would have been safe, but Hoskins gets no credit on this play.

ABSTRACT

In an era where the game is increasingly quantified, the receiving aspect of first baseman defense has largely been ignored. While particularly skilled (or lacking) play contributes to a fielder's reputation, even advanced defensive metrics fail to explicitly consider the role of the first baseman in the assist. In this analysis, I describe what makes an out at first from the perspective of the first baseman. I find that bounced and offline throws are less likely to be outs, and that some players seem to be better at fielding these than others, lending credence to the conventional wisdom that some possess this skill. While the data are insufficient to compute full player-level first baseman receiving rankings, I demonstrate a viable framework for its evaluation. Finally, I discuss the role that first baseman receiving can take in player development and valuation, and suggest future extensions of this approach.

1 INTRODUCTION

Most assists are fairly routine, yet for how commonly the first baseman receives throws from other fielders, little public information exists on this phenomenon. Anecdotally, some first baseman have the reputation of being "vacuums," yet the first baseman receive is often overlooked – left as an "additional wrinkle" in the current set of player-tracking enabled advanced statistics [10].

Consider contrasting plays in the 2022 World Series by Phillies first baseman Rhys Hoskins. In Game 6, Astros catcher Martín Maldonado grounded out softly to third. On paper, this reads as a one of a potential dozen mundane-looking 5-3 putouts in a game.

The mundane takes on new meaning with attention toward this play's subtleties. While Maldonado was not a burner down the line, it took a long run, an in-line throw, and Hoskins's outstretched frame to beat the runner for the out.

Contrast this play with another from the same Series (Figure 2). Astros first baseman Yuli Gurriel chopped a ball to short. Hit harder than Maldonado's, the shortstop had time for the throw. The throw was on-line, but skipped, which handcuffed Hoskins. He then watched the ball trickle away and the runner on third score. Instead of recording an inning-ending 6-3 putout, the shortstop was charged with an error.

Even with plentiful statistics, an obvious, but unmeasured participant exists: the first baseman himself. In the first case, Hoskins saved an infield hit by shortening the throw distance, while in the

Figure 2: Rhys Hoskins is unable to scoop a short-hopped throw to first in Game 2 of the 2022 World Series. Instead of making the third out of the inning, a run was able to score. While certainly a tough play, Hoskins gets no blame and the shortstop was charged with an error.

second, he got eaten up by a short hop. Yet, Hoskins gets no credit or blame in either case.

We can add additional context and understanding to the game by exploring the receiving end of an assists to first. In this work, I ask:

RQ: What makes an out from the perspective of the first baseman?

This article is organized as follows: First, I discuss contemporary fielding stats, other attempts to quantify first baseman receiving, and how they can be improved. Using player-tracking data from SMT, I then impute outs and provide a description of "first baseman receives" and describe high and low likelihood plays. Finally, I discuss this work's implications for player development and valuation, and suggest directions for future work in fielding analytics.

2 BACKGROUND

Defensive has lagged behind pitch and hitting analytics due in part to its uniquely complex dependencies. Though many defensive metrics exist, they often ignore these complications.

2.1 Defensive Metrics

The foundational defensive metric is fielding percentage [1]. Because most plays are routine, the leaders in fielding percentage are typically first baseman, limiting its usefulness for comparison across position groups. Its underlying assumption that opportunity is an important factor is strong, but its utility is limited because it assumes that all plays are equally difficult.

A next step in fielding analysis is premised on weighting opportunities differently. Metrics like Ultimate Zone Rating (UZR) [6] and Defensive Runs Saved (DRS) [3] both assign some value for play difficulty based on where and how hard a ball is hit in relation to a fielder's typical position. "Typical" is the operative word – neither accounts for a fielder's actual position on a play, which can lead to inaccuracies in cases where fielders are abnormally positioned. Even so, these metrics are an improvement for non-first-base-playing fielders over fielding percentage. However, they both assign all credit for a play to the assisting player, while ignoring the receiver. Good and bad throws that end in outs are treated identically.

Player-tracking data enables more precise defensive metrics. The current-state-of-the-art in public fielding analysis, Infield Defensive Outs Above Average (OAA), considers the likelihood of a fielder reaching a hit ball and the likelihood of a throw beating the runner for an out, among others [10]. While an improvement for most fielders, this does nothing to account for the most frequent contribution of a first baseman. By ignoring the last portion of the play, receiving the ball for an out at first, it similarly ignores this player's role in converting an out.

2.2 First Baseman Receiving Ability

Recognizing that "first baseman receives" are a gap in typical defensive metrics, some have explored the phenomenon. Following a matched player methodology, Lichtman [5] showed that that differences in first baseman receiving ability exist across several dimensions, including height (taller is better) and handedness (lefties are better). While promising, this methodology is limited by lack of available data, and describes this phenomenon in aggregate. Addressing this with a more direct approach, Epstein examined first basemen receiving ability [4] by manually tagging 1,000 catches, stretches, scoops on bounced throws, and found systemic differences between players and a cost to poor receiving performance.

Now with detailed player- and ball-tracking data, we can marry these approaches and revisit these analyses more directly.

3 DATA

To explore first baseman receiving, I make use of the following data from SMT: "player_pos" and "ball_pos" tables, which contain position data, the "game_event" table, which contains time stamped labels for events like throws or catches, and the "game_info," which contains information on game state. Noticeably missing from "game_info" is information on outs. This anonymization proved a hurdle to the analysis, and required extensive preprocessing.

3.1 Aligning Ball and Player Data

Players and the ball appeared significantly misaligned (See Figure 3(a)) in seven of the 97 total games. Assuming that the position of the ball should be within arm's reach of a player when the ball is acquired, I calculated the time difference between when the ball and player occupy the same space, for each instance of a ball pickup at the play-level. I then shifted the ball timestamps by this time difference (See Figure 3(b)). If unaddressed, this issue would significantly skew computed values relying on position data.

3.2 Imputing Outs, Excluding Data, Computing Features

In order to study first baseman defense, I first imputed the number of outs made on a given play in each half inning by using a recursive strategy inspired by "Sudoku solvers" (See Appendix A). The errors in this process revealed a number of subtle data issues that naive approaches may have missed: namely, missing plays from "game_info" sequences, and disagreement in the presence of baserunners between "game_info" and "player_pos". To address

(Click to play. Open in Adobe Acrobat or use the link in the caption.)

[a] [b]

Figure 3: (a) shows a misaligned play (Play ID 124, from Game 1903_01_TeamNE_TeamA2), notice the second basemen is not near the ball or bag when the turn is made. (b) shows the same play with the ball and player timestamps aligned as described in 3.1. Click here to view these plays online.

this, I track data agreement at a play-level and only consider plays where the full half innings are present and in agreement.

I further exclude 16 games (of 97) where no trusted outs or not outs were found due to these discrepancies. Double plays are materially different from typical putouts in both the kinds of throws made and how the first baseman sets up. As such, I removed them from the dataset to get a purer sense of a first baseman's receiving ability.

I additionally computed a number of features (See Figure 4, Appendix B) and used these to select plays that involve a throw towards the first baseman. In total, I ended with 484 outs at first base, and 78 not outs.

Figure 4: A grounder to second with a sample of the features computed for each throw to first overlaid. The thick blue line is the distance of between the ball at release and the bag. The blue arrow represents the velocity vector of the throw and the blue arc is the x-y angle of the throw away from the bag. The yellow arrow is the x-y-z position of the ball when it is received in relation to the bag, and thick yellow line is the projection of that vector in the direction the ball's release point. The gray arrow is the velocity of the runner. See Appendix B for more.

4 LIMITATIONS

The primary limitation of this work is the imputation of outs. This process reflects a best guess at the outcome of a play in a setting where outs information was intentionally removed to preserve anonymity. In any non-anonymous setting, information about outs would be available and trivially correct.

As such, this work is best thought of as descriptive and as a demonstration of a viable framework for the evaluation of first base receives, rather than as a definitive report of the top performers of this skill.

5 FINDINGS

A groundout to first is the culmination of several different sets of interactions. In this analysis, I find differences in thrower, batter, and first baseman behavior between out and not out. Because these samples are small, I largely treat these players in aggregate. With more data, we could better identify player-level differences in receiving skill[1].

5.1 What makes an out at first?

5.1.1 The Thrower and Batter. There are more throws on the 1B-side of the infield, more not outs on the 3B-side of the infield, and a wider spread of not outs (Figure 5(a-b)). Throws for outs seem to travel shorter distances and are released from more conventional infield positions. Similarly, in plays that are not outs, the batter is closer to first base at the point when a ball is released than in plays that are outs.

5.1.2 The Throw. Plays that are not outs are significantly more likely to have a throw that bounced, was thrown harder, and was thrown more offline than plays that are outs (all significant at $p = 0.05$). Though sample is small, some first baseman seem better

[1] And with more confidence in our outs data, we could get around the small sample with an Empirical Bayes approach like the one described in [8].

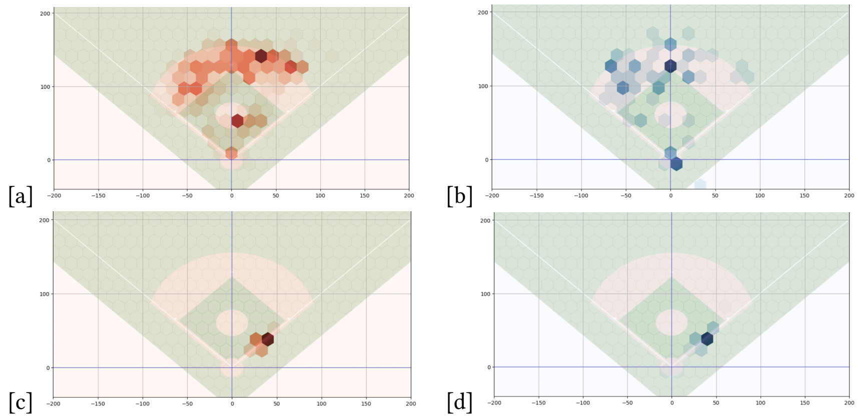

Figure 5: (a) Heat maps of thrower position on outs at first. More outs occur closer to first and in conventional positions. (b) Heat maps of thrower position on not outs at first. There is a much wider spread of postions than in (a). (c) Heat maps of batter position on outs, and not outs [see (d)], at first. As expected, batters further from first at the time of throw are more likely to be out.

at converting bounced and offline throws into outs. While bounces and accuracy results are straightforward to interpret, throw velocity is not. Throwers may sense more urgency in plays that are likely to be closer. Whatever the explanation, this suggests that interactions between several factors determine the likelihood of an out.

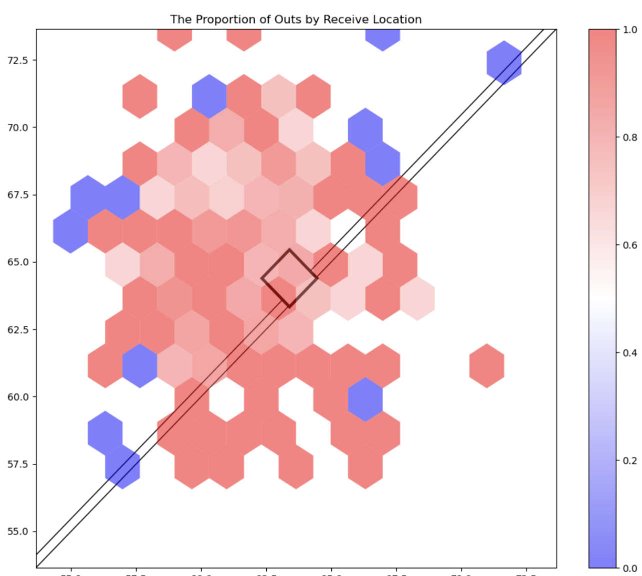

Figure 6: A hex plot showing the proportion of balls received in a particular zone that ended as an out at first. As expected, there appear to be a number of throws that pull the first baseman off the bag.

5.1.3 The First Baseman. First basemen set up differently depending on the throw position and the receiving positions seem to be related to out likelihood. Interestingly, there appears to be a region near the bag with a relative cold zone (near (65, 60) in Figure 6). It could be that inaccurate throws from the 3B-side of the infield pull lefty first basemen off the bag. With handedness data, this could be examined.

5.1.4 A Caveat and Summary. The analysis only considers throws that are actually made, and presumably the thrower in each case believes that a throw could become an out. Missing from this analysis are plays where a would-be thrower eats the ball instead, either because they don't think they could make the throw, or they don't believe the first baseman could field a less-than-perfect throw. This "trust" might be a confounding factor in this analysis.

5.2 Example Scenarios

I modeled out likelihood at the time a throw to first is released (See Appendix B). While not well-calibrated, this model is useful to identify the types of plays that have either high or low probabilities of being an out (see Figure 7).

5.2.1 Limitation of this Approach – Errors. A good outcome on terrible throws is the first baseman knocking the ball down (See Figure 8 (a-b)). These plays both save probable advancement of all baserunners. Consideration of these second-order impacts are left as future work.

6 DISCUSSION AND FUTURE WORK

I have examined a ubiquitous, but overlooked aspect of defense – the first baseman receive. In this work, I have shown that first baseman have a significant role in what makes an out. It is possible that we undervalue first base defense in the same way that catcher framing and blocking was historically undervalued [2].

While Run Expectancy [11] would provide additional perspective on how to value this specialized skill in relation to batting, or baserunning, the available data did not support this. As such, I have chosen to leave that, and the examination of other second-order impacts, as future work.

I argue that first base receiving is worth additional consideration, and have made inroads to explain this phenomenon. Even so, I am

(Click to play. Open in Adobe Acrobat or use the link in the caption.)

[a] [b]

[c] [d]

Figure 7: (a) A likely out that was converted for out. These plays are considered routine. (Play ID 150, from Game 1902_26_TeamMH_TeamA3) (b) An unlikely out that was not converted for an out. These plays are long shots. (Play ID 132, from Game 1903_32_TeamNB_TeamA1) (c) An unlikely out that was converted for an out. This is a great play by the first baseman. (Play ID 106, from Game 1903_30_TeamNF_TeamA2) (d) A likely out that was not converted. (Play ID 193, from Game 1903_05_TeamND_TeamA2) These plays are hard to interpret without additional context. It could be a blunder on the first basemen, or a throw bad enough that preventing the throw from getting away is more important than the out. See Figure 8 for additional examples of this play type. Click here to view these plays online.

limited by the fidelity of the data that reports the center of mass as the player positions. Pose tracking [9] is a promising extension that could be used to even more precisely examine first baseman reach and flexibility.

Additionally, this framework begins to give a quantitative value to a tool – "hands" – that is notoriously hard to measure. As a corollary, minor adjustments to this framework could be used to explore fielder "arm utility" through an examination of the quality of off-platform throws [7].

7 CONCLUSION

In this analysis, I have provided additional evidence that first baseman receiving could be a differentiated skill worthy of additional attention, though I leave the player-level assessment of this skill as a future exercise. By doing so, we might more appropriately direct our cheers (or jeers) to the most deserving player, and give weight to a phenomenon that today largely exists only in reputation.

8 ACKNOWLEDGEMENTS

Thank you to Dr. Meredith Wills for supporting this competition and for being so generous with her time. Thanks to <REMOVED FOR REVIEW> for their comments on an earlier version of this draft.

REFERENCES

[1] [n. d.]. Glossary. https://www.mlb.com/glossary

[a]  (Click to play. Open in Adobe Acrobat or use the link in the caption.) [b]

Figure 8: (a) Mets first baseman Pete Alonso saves an error on a throw well off the bag. (b) A similarly bad throw pulls the first baseman off the bag in Play ID 25 of Game 1902_05_TeamML_TeamB. While there is value in stopping a bad play from getting worse, the data anonymization hampered my ability to explore that in this analysis.

[2] Chuck Brownson. 2009. The most underrated catchers in the game. https://tht.fangraphs.com/tht-live/the-most-underrated-catchers-in-the-game/

[3] John Dewan. [n. d.]. Frequently Asked Questions about Plus/Minus and Runs Saved. https://fieldingbible.com/Fielding-Bible-FAQ.asp

[4] Daniel R. Epstein. 2021. Best of BP: 1,000 Grounders: What We Still Don't Know About First Base Defense - Baseball ProspectusBaseball Prospectus. https://www.baseballprospectus.com/news/article/71706/best-of-bp-1000-grounders-what-we-still-dont-know-about-first-base-defense/

[5] Mitchel Lichtman. 2009. First Basemen "Scoops" – The Value of Handling Errant Throws at First Base. https://blogs.fangraphs.com/first-basemen-scoops/

[6] Mitchel Lichtman. 2010. The FanGraphs UZR Primer. https://blogs.fangraphs.com/the-fangraphs-uzr-primer/

[7] Eric Longenhagen and Kiley McDaniel. 2020. *Future Value: The Battle for Baseball's Soul and How Teams Will Find the Next Superstar.* Triumph Books.

[8] David Robinson. 2017. *Introduction to Empirical Bayes: Examples from Baseball Statistics.*

[9] SABRvideos. 2021. 2021 SABR Analytics: MLB Statcast Player Pose Tracking and Visualization. https://www.youtube.com/watch?v=uLbzUkNi7oU

[10] Tom Tango. [n. d.]. History of the Fielding. http://tangotiger.com/images/uploads/History_of_the_Fielding.pdf

[11] Tom Tango, Mitchel Lichtman, and Andrew Dolphin. 2007. *The Book: Playing the Percentages in Baseball.* Potomac Books.

[12] Larry Wasserman. 2013. *All of Statistics: A Concise Course in Statistical Inference.* Springer Science & Business Media. Google-Books-ID: qrcuBAAAQBAJ.

A IMPUTING OUTS

This section provides more details into my "Sudoku solver" approach to imputing outs. This method relies on two main functions, Algorithm 1 and a function to check whether a given proposed assignment of outs to a play is valid, described in the next section.

A.1 Checking whether a proposed outs assignment is valid

I consider the following scenarios to be valid:

- if this play is a home run, there can be no outs on this play.
- if this play was hit, caught, and then the play ends, this must be exactly 1 out, Note that this will fail if the outfielders get into a run down and tag someone out for a double play, but that is rare.

Algorithm 1 Sudoku Solver Outs

```
function SOLVE_OUTS_SEQUENCE(df, sequence)
    if no more cells to fill in sequence then
        return True
    end if
    for this play outs in {0, 1, 2, 3} do
        if valid outs assignment then
            Fill in this play outs into the sequence
        end if
        if solve_outs_sequence(df, sequence) then
            return True                  ▷ This sequence works so far
        end if
                          ▷ If you make it here, this sequence is wrong
        remove those things you tried to add in sequence
    end for
    return False
end function
```

- if the batter is not on the bases when there is a new batter and the play wasn't a HR then the batter got out.
- There must be an out if there is a baserunner on this play that isn't on the bases on the next play, and there are were runners ahead of them on this play.
- It must be 1 out if there baserunners stay the same, and there is a new batter – this is a strikeout. This fails if there is a mid at bat pinch hitter.
- Outs on this play must be 0 if there is the same batter and baserunners on the next play.
- Outs on this play can't be more than the number of baserunners + 1.
- if we are on the last play of a half inning the outs on this play and the previous plays in the half inning must sum to

three. Failures on this rule indicate missing plays from the "game_info" sequence.

I also explored an alternative imputation process that treated each play in isolation, relaxing the requirement that all plays in a half inning must be "trusted." This did not yield significantly more plays to first, so I selected the more conservative approach of the two. It is also worth reiterating that in a setting with full access to game state data, this imputation process would not be necessary, and the out information would be trivially correct.

B A PROBABILITY MODEL FOR OUTS

At an earlier stage of this analysis, I created a probability model for outs at the time of throw. Using a number of the features described in Figure 4 with the intention of using this model the compute a player-level metric for "receiving outs above expectation," I used a similar methodology to OAA. The remainder of this section describes this process, and why I ultimately chose to exclude that from the analysis.

B.1 Making the Model

Following the prepocessing described in Section 3.2, I computed a number of additional features and give attention to some here:

B.1.1 All about vectors. For every throw in "game_events," I computed a vector in the x-y plane from the x-y point of ball release. The magnitude of this vector gives the distance to first. I also computed a velocity vector of a throw over the first 0.25 seconds after release. Using dot products, these vectors can also be used to compute the x-y angle of the throw as follows with algebra, where $\vec{f}$ is the vector to first and $\vec{t}$ is the throw vector:

$$\vec{f} \cdot \vec{t} = \|\vec{f}\|\|\vec{t}\| \cos\theta$$

In additional to this being a useful bit of information to have on the throw, this allowed me to select only throws in the direction of first. This is necessary because many plays have more than one throw and the first baseman often serves as the relay man on an extra base hit with no one on base. I only wanted to consider throws to first where the intention was to get an out. In practice, I found that filtering throws of 7 degrees or less was a balance that allowed for the throws where the intent is not to throw the ball directly towards the bag (e.g., as in the case on a throw along the first baseline by the catcher), while also removing throws to other bases that are along a similar line (e.g. a throw from left to second).

A related idea was to compute the distance that a first baseman stretches and the distance that they shorten the throw by stretching. To compute this, I defined two vectors, an x-y-z vector from the ball position when it was received and the x-y-z position of the bag (call this $\vec{a}$), and a second x-y-z vector from the ball position at release to the bag (call this $\vec{b}$). I then found the component of the first in the direction of the second. Formally:

$$\text{Shortened Distance} = \frac{\vec{a} \cdot \vec{b}}{\|\vec{b}\|}$$

This information can provide some useful context to a given play, without access to additional information about the play and players beyond the position of the first baseman's center of mass, it's utility is limited. Without video or handedness (for example), it is difficult to tell both *why* a particular player stretched and to evaluate *how* a particular player positioned their body. Pose detection like [9] might be able to address this in the future.

B.2 Evaluating the Model and Why I Didn't Use it

Computational methods like machine learning often rely on large data sets, and while these data are not "small," at 562 observations in total, these are certainly not "big" data. Given this, the possible set of appropriate models are limited and the risk of overfitting is high. Logistic Regression is a good candidate in this case, since it is simple and directly interpretable as a probability.

I used a 50-50 train-split because of the small amount of total data, and performed F-testing [12] for the feature selection, and ultimately trained a Logistic Regression model with the following features:

- x-y throw angle in degrees
- the elevation angle of the throw in degrees,
- the velocity of the throw in mph
- an indicator for whether the ball bounced
- batter distance to first at time of throw

This model performed well on F1-score and poorly on ROC AUC, meaning the actual predictions of out or not are largely sound, but the predicted probabilities are not. See the calibration plot in Figure 9. As such, I elected not to exclude the final stage computing "receiving outs above average" from this analysis, since it would rely heavily on the model predicted probabilities.

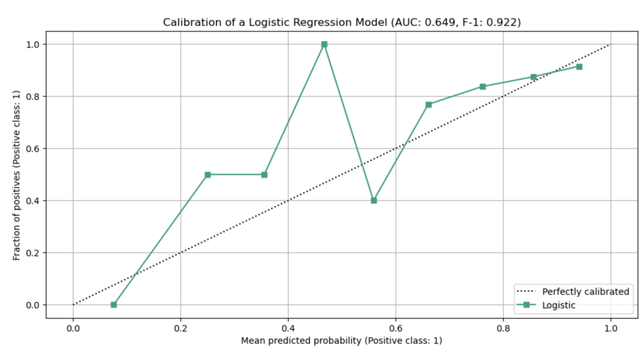

Figure 9: A not particularly well-calibrated Logistic Regression model on the "test" dataset. It is especially bad for the mid to lower range of predicted probabilities, likely due to limited amount of "not out" data. This model performed well on F1-score and low on ROC AUC, meaning the actual predictions of out or not are largely sound, but the predicted probabilities are not.