DTA: Directional Throw Angle

https://github.com/LukePsyh/SMT_DTA

Abstract

In this paper we outline our reasoning and methodology for creating an ideal directional throw angle model (X, Y plane) for any given throw to first base and we prove that the direction of the throw has a significant impact on the result of the play. Our aim was to push the boundaries of defensive analytics and enter a domain that has been left unexplored publicly in regards to differentiating between the quality of the throw by an infielder and the quality of the "scoop" by the first baseman. While the attempt by a first baseman to catch a throw by the infielder is a substantial piece of the full picture of any infield play, public research has not covered this important part of player defense. A lot of factors matter when it comes to the success of a bang bang play at first base, but the angle and direction of the throw towards the first baseman is certainly one of the most crucial.

[Optimal Throw Animation](#) [Nonoptimal Throw Animation](#)

Figure 1A & 1B: Screenshots of animation demonstrating throw with an optimal DTA

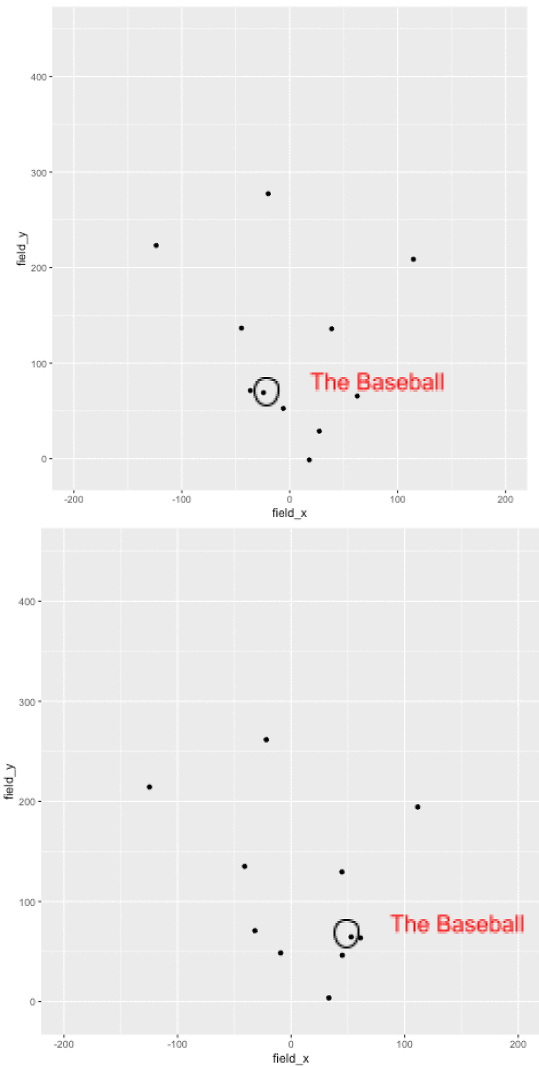
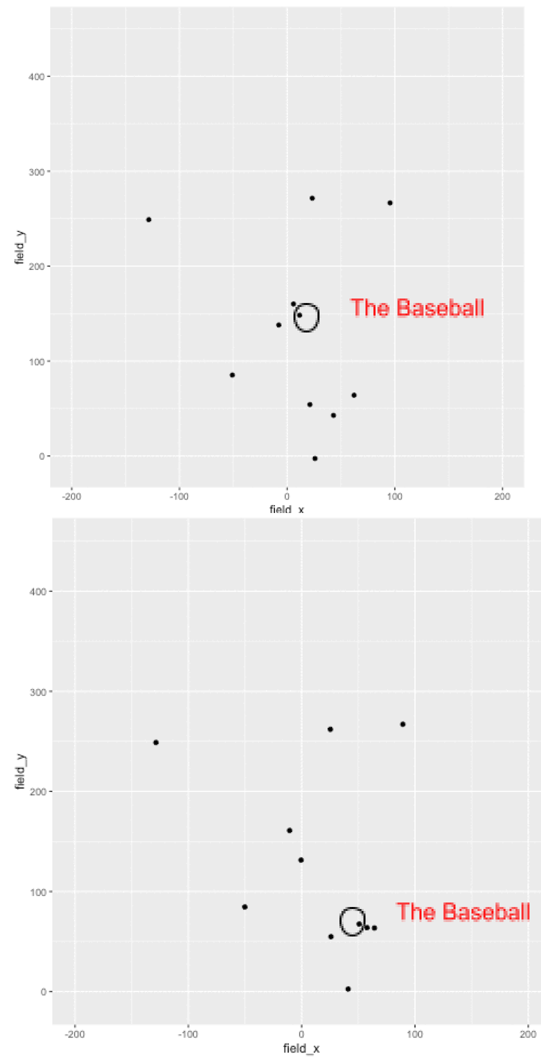


Figure 2A & 2B: Screenshots of animation demonstrating throw with a nonoptimal DTA



ii Introduction:

Consider these two separate throws from Fernando Tatis Jr. to Eric Hosmer

Play 1

Figure 3A: Fernando Tatis Jr. bobbles the ground ball

Figure 3B: Tatis' throw lands in Hosmer's glove

Figure 3C: The ball gets past Hosmer's glove and Tatis is charged with an error

Play 2

Figure 4A: Fernando Tatis Jr. fires the ball across the diamond

Figure 4B: Tatis' throw hits the infield grass, leaving Hosmer no chance

Figure 4C: The ball gets past Hosmer and tatis is charged with an error

Despite both of these plays being scored as errors by Fernando Tatis Jr., and with more modern defensive numbers, such as Statcast's Outs Above Average, charging him with a similar malfeasance, the level of responsibility on his end is unquestionably unequal. In play 1, we see Tatis charged with a fielding error because of the initial bobble which forces Tatis to make a tough throw. The throw, while not perfect, hits Hosmer right in the glove well before the runner reaches first. Hosmer could have made this play, and in that world, Tatis gets a different level of credit for the same quality of play. Therefore, Hosmer carries a proportion of the responsibility for this play that is currently unrecognized in the realm of defensive analytics. In play 2, however, Tatis makes a throw that bounces off the infield grass well to the right of first base, leaving Hosmer with no chance to make the play. Tatis should get the lion's share of the blame for said throw.

Other examples of assigning credit have been executed in sabermetric research. Pitch framing measures have effectively given credit to the catcher for the pitcher's performance (although one could argue that the pitcher has a lot of impact on his own framing as well). Similarly, an infielder is not the only player on the defense responsible for the result of their throw. In these two plays from Fernando Tatis Jr., he gets credited with an error, and his Statcast Outs Above Average (OAA) number decreases because he could not get outs on what seemed to be plays with relatively high out probabilities.

Statcast OAA is the current gold standard for public defensive measures, as it utilizes the most in depth level of data currently available. The OAA model for infielders currently factors in three parts of the play: how the fielder arrives at the ball, how he fields it, and finally, the frantic race to first between the ball and the batter. The probabilities of the fielder arriving at the ball, retrieving it cleanly, and beating the runner to first with the throw are then used to form an overall probability x that the out will be made at first. If the play is successful, the infielder is credited with 1-x OAA, and if it is not, then the infielder gets -x on his resume. The final piece, the race, is where we believe the most value is left unevaluated in the OAA model. This estimate is made using the average speed for someone running down the first base line, and the average time it takes for a ball to get to first at the infielder's coordinates. Because of this structure within the model, strong high velocity throws to first are properly rewarded. However, the model cannot properly determine whether or not the throw was accurate because credit is not differentiated between the first baseman and infielder. OAA assigns value in terms of outs added. If OAA deems there's a 60% chance a roller to the third baseman gets converted for an out at first, the third baseman will be rewarded +0.40 outs for getting the runner at first, and -0.60 outs if the play is not made. The issue with OAA lies within the assignment of an out value on each individual play, specifically when it comes to the race. While OAA is a good tool for describing the range of a particular player on the field, we believe it leaves value on the table by failing to assign credit to both the thrower and receiver. This does not just go for OAA, but also any defensive statistic attempting to be descriptive. In order for the public to ever get an accurate assessment of how good an infielder is as a defender, there needs to be a tool to accurately assess how skilled an infielder is at consistently delivering on target throws to first and making life easy on his first baseman. On the flip side, it would be prudent to evaluate first baseman for the quality of their picks. In those two throws from Tatis, he lost 0.x and 0.y outs respectively. However, these estimates are incomplete because one of these throws is relatively accurate and the other is way off the mark. While the value lost on play 1 should be split between Tatis and

Hosmer, Tatis was graded negatively for both plays, while Hosmer's defensive value remained unchanged. In order to accurately understand the value of a first baseman or infielder, we believe that defensive models must figure out an evaluation of the first baseman's ability to field competitive throws to first.

While all of this seems intuitive, we knew that in order to make any sort of tangible ground in the sabermetric community, we needed to quantify this effect. The emphasis of our work was in the field of throw angle with respect to the X and Y coordinates, essentially being the basis for the direction that the ball goes. Other factors, such as throw height, velocity, and batter speed, were considered and are definitely crucial variables in this context, but the focus was made on throw direction/angle in the X and Y planes as these other variables are at least somewhat incorporated into the current standard for defensive measurement.

iii. Data

The dataset contained intricate tracking data, with an emphasis of both player and ball positioning on each play, for 97 anonymous minor league games. Within the dataset, there were a few variables we focused on. Field_x and field_y in the ball_pos and player_pos dataframes, measures of where any player or ball were at any point, were instrumental in determining certain calculations involving the throw. The state of the game was important when it came to determining whether or not the batter reached base, and therefore whether or not the throw was successful. The timestamp variable for each measured frame was utilized in order to properly determine the sequence of events.

iv. Methodology

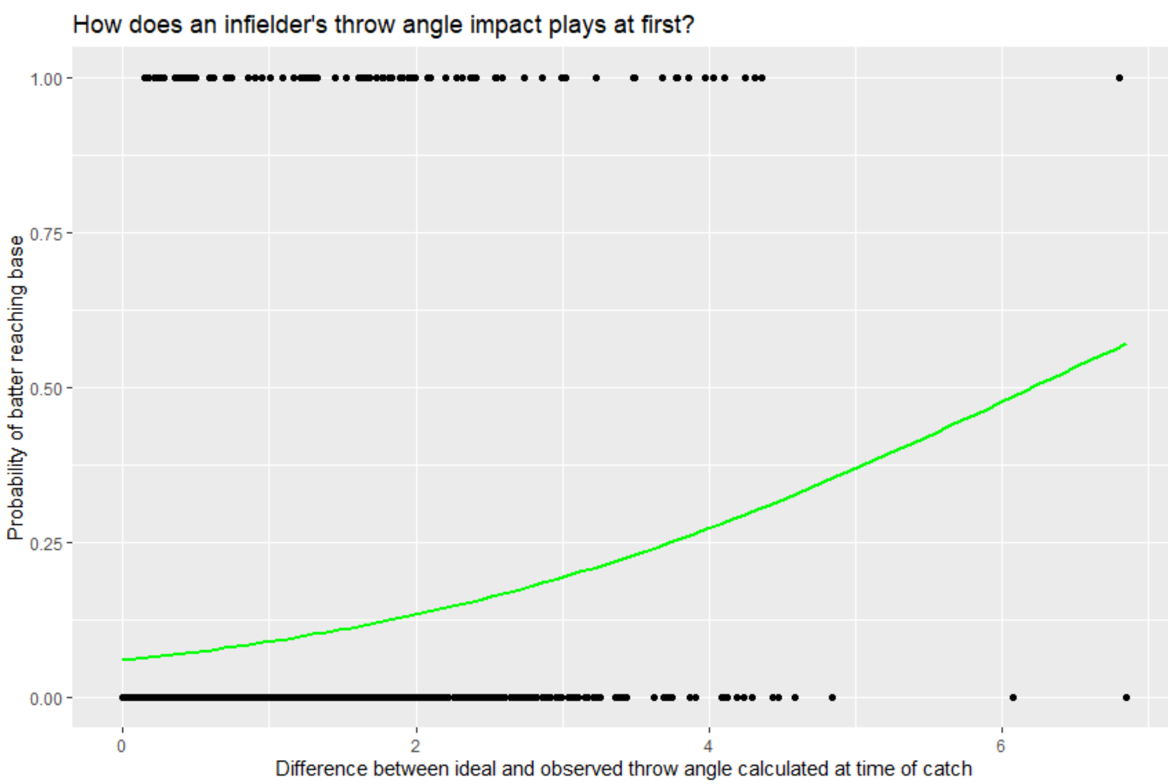
The approach to solving this problem took a few steps. The first and biggest hurdle was to identify plays that actually applied to the context that we were evaluating: competitive throws to first base. Figuring out how to filter plays that involved a batter trying to beat out a throw to first was easier said than done. In order to solve this, we had to filter such that we had plays that were labeled as throws, calculate the difference in distance on the x and y plane of the ball between the time of the throw and 15 measured frames after the throw in order to determine throw angle, and then use said throw angle in some nifty trigonometry to interpolate the point at which the ball would hypothetically cross the first base line as if it continued straight along the path it was going forever. If this result was between 80 and 100 feet along the first base line, then it would have a chance to be included in the sample. We then had to only include plays in which the batter was running towards first, make sure that the throw went towards the right side of the field, and only include throws from infielders. This gave us all plays in which an infielder was attempting to gun a runner out at first.

The next goal was to calculate the ideal place for the first baseman to catch the ball. This was done by simply taking the average catch location of all plays in which the runner did not reach base. Using this figure, we determined the optimal throw angle given the fielder and ball's position. Given the difference between the optimal throw angle and the actual throw angle, we were able to determine that the accuracy of the throw is a significant variable in determining whether or not the runner would eventually reach base.

v. Conclusions

The variable optimal_residual, the difference between observed and ideal throw angle, had a very low **p value of 0.00000164**, indicating a strong dependence on fielding success given just the direction of the throw.

Figure 5:



As shown in the graph above, the further the observed throw angle deviates from the ideal path of the throw, the more likely it is that the runner will reach first. Due to this result, we can conclude that throw angle is a drastically overlooked factor in the public work done to quantify defensive value. We have found that in order to fully understand a player's impact on any given throw to first base, the angle at which the ball has been thrown must be a factor. This makes sense, intuitively. If you have ever seen a play where the throw might look a tad bit off target, but you still think another first baseman might've made that play, there should be a measure which accurately places blame on each defender. Similarly, if you've ever been left wondering "How did the first baseman stretch his glove so far to make that play?" we believe the first baseman is adding unquantified value that the sabermetrics community's analysis could benefit from.

```
Call:
glm(formula = runner_reached_base ~ optimal_residual, family = binomial,
    data = within_range)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.3010  -0.4944  -0.4165  -0.3698   2.3524

Coefficients:
                 Estimate Std. Error z value             Pr(>|z|)
(Intercept)      -2.76687    0.18715 -14.784 < 0.0000000000000002 ***
optimal_residual  0.44592    0.09302   4.794           0.00000164 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 598.74  on 882  degrees of freedom
Residual deviance: 576.91  on 881  degrees of freedom
AIC: 580.91

Number of Fisher Scoring iterations: 5
```

Figure 6: Summary Statistics of our model

Through our optimal throw angle model, which is able to predict whether a throw will result in an out or not with relative accuracy by simply looking at the directional throw angle. This is **proof that directional throw angle matters in the result of any given throw to first base, and factoring it into player analysis will provide an edge for any team that uses the information effectively.**

vi. Improvements/Future Work

When we set out on this project, we wanted to deliver something to the public sphere that could be studied for years to come. The advances to be made within the sabermetrics community with throw angle analysis is undoubtedly what excites us the most about our research. There is no doubt that some MLB teams are already quantitatively factoring throw angles into their talent evaluation, and that the edges that throw angle analysis provides are likely already being exploited by savvy teams including but not limited to the Rockies, Rays and Dodgers. However, we believe there are a multitude of ways our research can be utilized to help fans get a better understanding of who the most accurate infielders are, and how valuable that truly is.

i. **OAA Adjustments**
ii. **Throw Accuracy and 1B Stretching/Picking leaderboards**
iii. **Biomechanics**

Our model is by no means a perfect attempt at quantifying directional throw angle because we simply did not have the resources necessary. A model that could accurately factor the Z-Coordinates of the throw and a first baseman's movements at the catch point should undoubtedly yield better results than ours. Our model should not be used to make conclusions about any individual player, however, it is proof that directional throw angle analysis yields quantifiable results.

The work left to be done with directional throw angle and the insights it will provide into first base defense is vast. We believe our research is the first step towards a new frontier within the broader sabermetrics research, but we greatly anticipate the work that can be built upon ours.

vii. References

http://tangotiger.com/images/uploads/History_of_the_Fielding.pdf

viii. Appendix

TRIG USED:
Given different positions of the ball relative to home plate, these were the trigonometry equations used.

$\sqrt{(2 \times (\tan(throw\_angle\_radians) \times -field\_x + field\_y)/(1 - \tan(throw\_angle\_radians))^2)}$

$\sqrt{(2 \times (\tan(throw\_angle\_radians) \times field\_x - field\_y)/(1 + \tan(throw\_angle\_radians))^2)}$

$\sqrt{(2 \times (\tan(throw\_angle\_radians) \times -field\_x - field\_y)/(1 + \tan(-throw\_angle\_radians))^2)}$

$\sqrt{(2 \times (\tan(throw\_angle\_radians) \times field\_x + field\_y)/(1 + \tan(-throw\_angle\_radians))^2)}$