The Man in the Middle

An In-Depth Analysis of Cutoff Decision-Making in Baseball

Pranav Rajaram, Jack Kalsched, Alex Frederick, Andrew Zaletski

GitHub Repo: https://github.com/pranavrajaram/smt-2025-submission
Results Dashboard: https://maninthemiddle.shinyapps.io/smtcutoffdashboard/

Abstract

Cutoff plays are a critical but under-analyzed component of public baseball analysis. In this project, we develop a framework to assess cutoff play decision-making using anonymized MiLB player and ball tracking data. First, we construct a logistic regression model to estimate baserunner safe probabilities for extra base advancement paths. Next, we combine these probabilities with RE24 run expectancies to compute expected run values for each possible cutoff action, identifying the optimal cutoff choice for every play. Using these labels, we train a Random Forest classifier with features such as runner and fielder distances, arm strength, and sprint speed to predict optimal actions. Our findings reveal that cutoff men overwhelmingly favor aggressive throws when our model recommends holding on to the ball, leading to notable run expectancy losses. Finally, we present an interactive dashboard that provides team- and player-level cutoff decision-making insights that can supplement scouting and coaching. This project offers a scalable approach to cutoff play evaluation that can be used by teams to improve their defensive strategy in crucial situations.

1. Introduction

The bases are loaded with two outs in the bottom of the 8th inning of a two-run ballgame. The batter slices a base hit, bouncing once to the left fielder. The runner on third scores easily and the ambitious runner from second rounds third and darts for home. The throw home is a bullet, low and on target. The crowd anticipates a close play at the plate…but the pitcher cuts the ball off. The run scores, and a sigh of relief is shared among the UC San Diego fans. As the dust settled, we all wondered, "Should the pitcher really have done that?"

Click on the image to watch the play. UC San Diego's Anthony Potestio hits a 2-run single to left field against UC Davis. The pitcher curiously cuts the throw off from left field, allowing the second runner, Emiliano Gonzalez, to score. [Link to play](#).

Public statistical analyses of cutoff plays are few and far between. Infielders are taught how to line up and where to throw, but do they make the right cutoff decision when it matters? Is it better to cut the outfielder's "hero throw" or let it fly?

In this project, we quantified the abilities of the cutoff man in order to evaluate cutoff decision-making. We discuss our methodology, the challenges we faced, and how we overcame them. We then showcase an application that stores our insights in an interactive dashboard and conclude with big-picture takeaways.

2. Data

The data was provided to us by SMT and can essentially be thought of as an extremely detailed baseball scorebook. Key tables included game info (which team was hitting, which runners were on base), game events (each play broken into events like pitch, ball contact, catch, etc.), ball positioning (where the ball was over the course of a play in 3D), and player positioning (where each player was located over the course of a play in 2D). Importantly, the data only captured the physical events that happened on the field, like pitches, throws, and ball trajectories. It did not include play outcomes like outs, runs scored, errors, and other official scoring results.

2.1 Filtering

We confined our analysis to base hits to the outfield, with runners on first and/or second base. These play types opened the door for potential cutoffs, where the cutoff man would have to choose between continuing the assist throw, cutting and relaying to another base, or holding the ball. Using this criteria, we ended up with around 1,500 unique baserunners over 800 cutoff plays.

2.2 Data Constraints

The anonymized data omitted two necessary components: whether runners reached their target base and the number of outs. We created several conditionals and boundaries for the runner, fielder, and ball coordinates in order to label the baserunner safe or out (see *Appendix 1*). For instance, if the runner was much closer to the base than the infielder when the infielder caught the ball, then we labeled them safe. Additionally, we calculated the number of outs using base occupancy information (see *Appendix 2*). Although these definitions are not perfect, we tested and manually checked enough plays to be confident in our results.

3. Methodology

Our overall methodology consisted of three main steps: modeling baserunner safe probabilities, calculating optimal cutoff actions, and modeling optimal cutoff decisions. To model the optimal cutoff decision for a given play, we first needed to establish what the optimal cutoff actions were. To uncover these, we combined baserunning safe probabilities with RE24[1] run expectancies to determine expected run values.

3.1 Safe Probability Model

To find the safe probabilities, we built a model to predict the probability of a given base runner advancing at least one extra base (1st to 3rd, 2nd to Home, etc.). In the UC San Diego example, our model would produce the probability of Gonzalez reaching home from second at the moment the left fielder fielded the ball.

[1]RE24 - Data representing the expected run value of every base-out permutation. Using historical MLB data, RE24 estimates how many runs a team is expected to score given the number of runners on base and the number of outs. For example, with 1 out and a runner on first, a team is expected to score 0.404 runs in that inning.

We used a set of five logistic regression classification models, considering the outfielder's and baserunner's respective distances from the target base at the moment the outfielder acquired the ball (*Appendix 4*). We trained five individual models, each considering a different baserunner advancement scenario: 1B ⟶ 3B, 1B ⟶ home, 2B ⟶ home, home ⟶ 2B, and home ⟶ 3B.

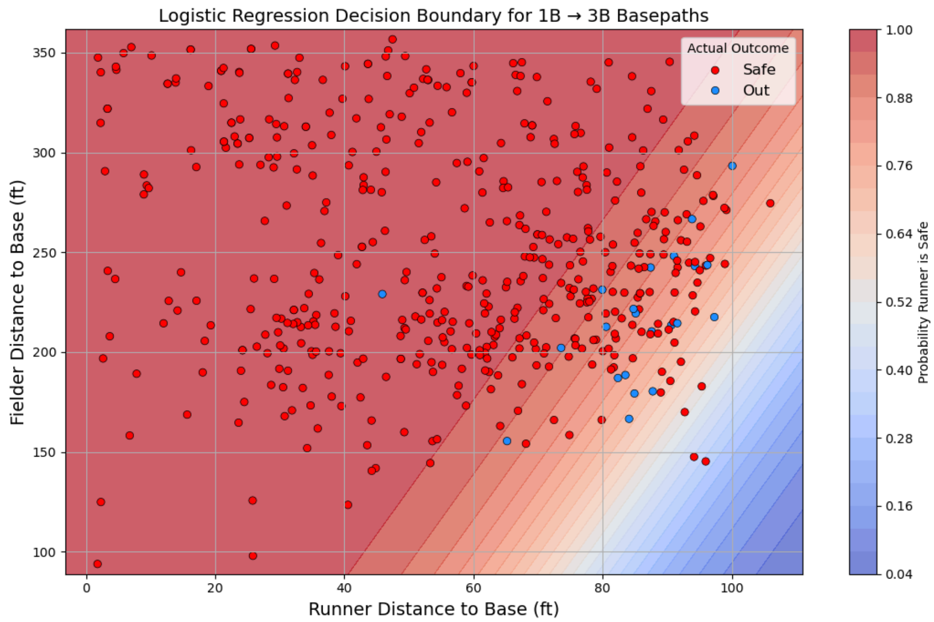


Figure 1: How the probability of the runner being safe changes with the fielder's and runner's distances to the targeted base – the two features in our safe probability model. Shown here for the 1B ⟶ 3B basepath.

All five models performed well, with brier scores below 0.1.x

Safe Probability Model Brier Score by Basepath

BASEPATH	BRIER SCORE
Home → 2B	0.017
1B → 3B	0.028
Home → 3B	0.036
2B → Home	0.054
1B → Home	0.091

Figure 2: Brier scores for each basepath. The Brier score is a metric that evaluates the accuracy of probabilistic predictions. 0 means the probabilities predict the outcome perfectly, and 1 means they never predict the outcome correctly. The lower the Brier score, the better the model.

The safe probability distribution is left-skewed, as shown in *Figure 3*. This skew results from selection bias, since we only considered plays where the runner attempted to take an extra base. Because runners are unlikely to attempt very low probability advancements, the dataset underrepresents those scenarios, concentrating the distribution toward higher probabilities. The example play in *Figure 4* illustrates the capability of this initial safe probability model.

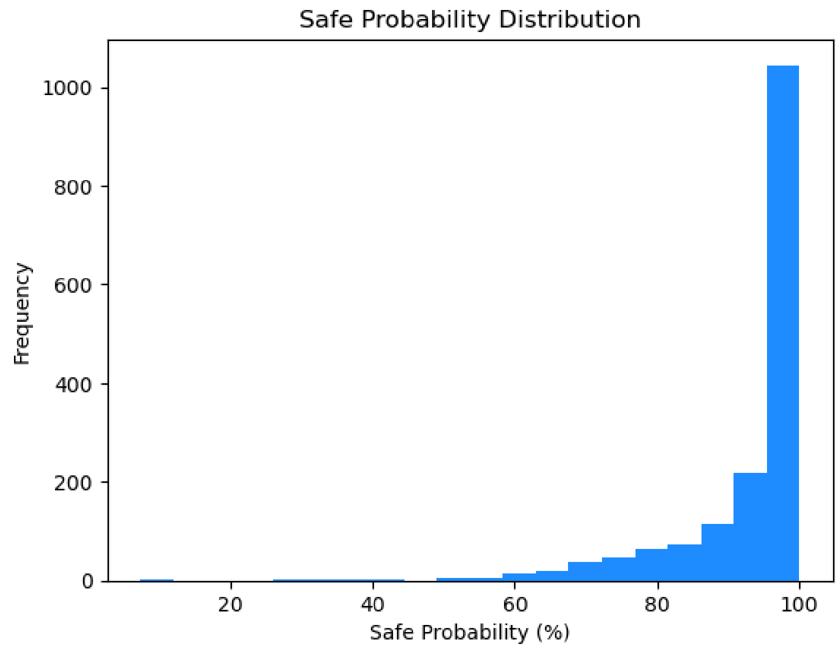


Figure 3: The model's safe probability distribution.

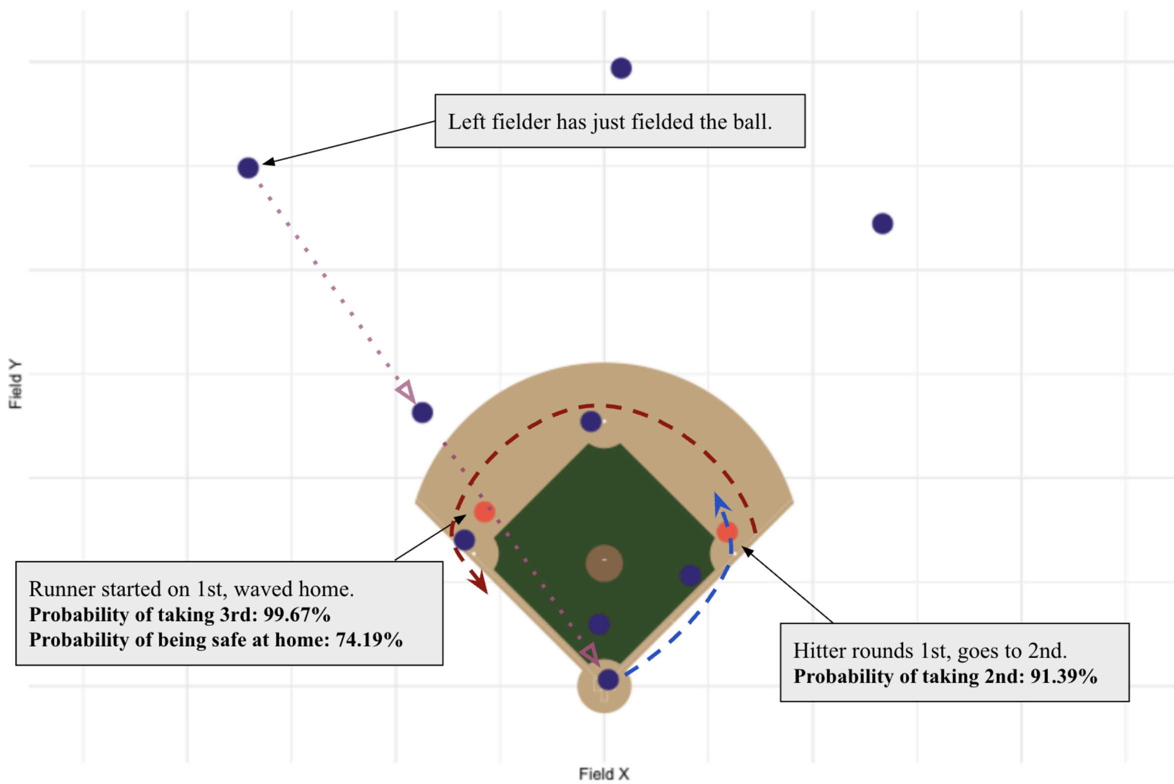


Figure 4: Showcased are the probabilities our model calculates for the baserunner and hitter to reach their respective target base(s). The blue dots represent the fielders, and the red dots represent the hitter and the baserunner. The red and blue dashed lines indicate the approximate basepath route of the baserunner and hitter, respectively, and the purple dotted line represents the approximate ball trajectory. Link to play.

3.2 Determining Optimal Cutoff Actions

Now that we had the safe probabilities for every baserunner in our dataset, we could determine the cutoff man's optimal action. We considered three options:

1. "Continue": The cutoff man throws to the base initially targeted by the outfielder. In the UCSD play, this would be the case that the pitcher throws home.
2. "Cut and relay": The cutoff man throws to a different base. In the UCSD play, this would be the case that the pitcher throws it to third base.
3. "Cut and hold": The cutoff man holds on to the ball. In the UCSD play, this is what the pitcher did.

Note that we assume all throws from the outfield are able to be cut off by an infielder. This may not always be the case in reality, as audacious outfielders may attempt to bypass their cutoff man completely and throw directly to their target base.

Using a combination of the RE24 data and our safe probabilities, we calculated the expected run values (xRVs) for the three potential cutoff actions for each play (*Appendix 3*). *Figure 5* explains our process in further detail with an example. After comparing each, the cutoff action with the lowest xRV was deemed *optimal*.

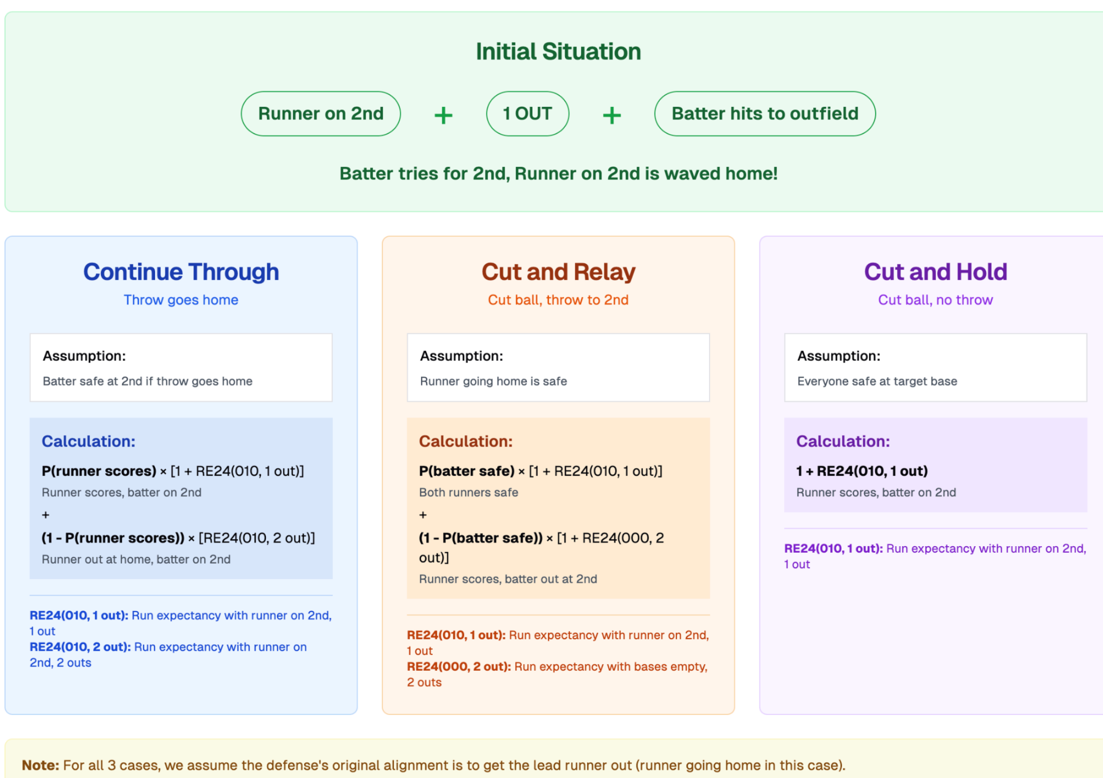
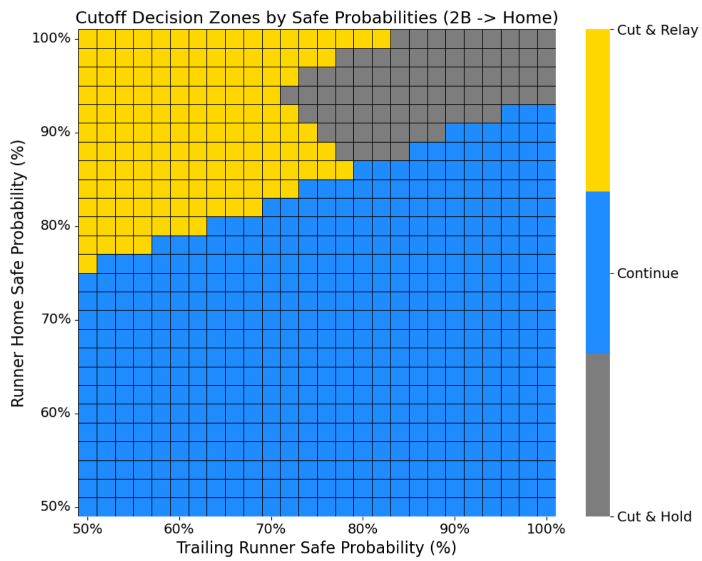


Figure 5: Example situation is runner on second and 1 out. The graphic outlines expected run value calculations using our derived safe probabilities from the initial model, for each possible cutoff action.

Many plays showed small differences in the expected run values between cutoff actions, so to account for "cut and hold" never having the lowest xRV (as we assume all runners to be safe if the cutoff man holds), we set a run value threshold of 0.08. If continuing and cutting-and-relaying were within 0.08 expected runs of holding the ball, we assumed the fielder should cut and hold to minimize the risk of an error (*Appendix 3*). See *Figure 6* for how the optimal cutoff actions are distributed through cutoff decision zones, and *Figures 7* and *8* for plays where "cut and hold" and "continue" are the optimal actions, respectively.

Figure 6: For a basepath such as 2B → home, the probabilities of the runner being safe at home and the trailing runner (batter, in this case) being safe at second. Throwing the lead runner out at home saves far more expected run value than getting the trail runner, so it makes logical sense that it's only optimal to "cut and relay" when it's far more likely to get the trail runner than the lead runner at home. This can explain the sea of blue in this graphic. Also, note that the vast majority of our probabilities fall into the top right of this graph, as was pictured in *Figure 3*.



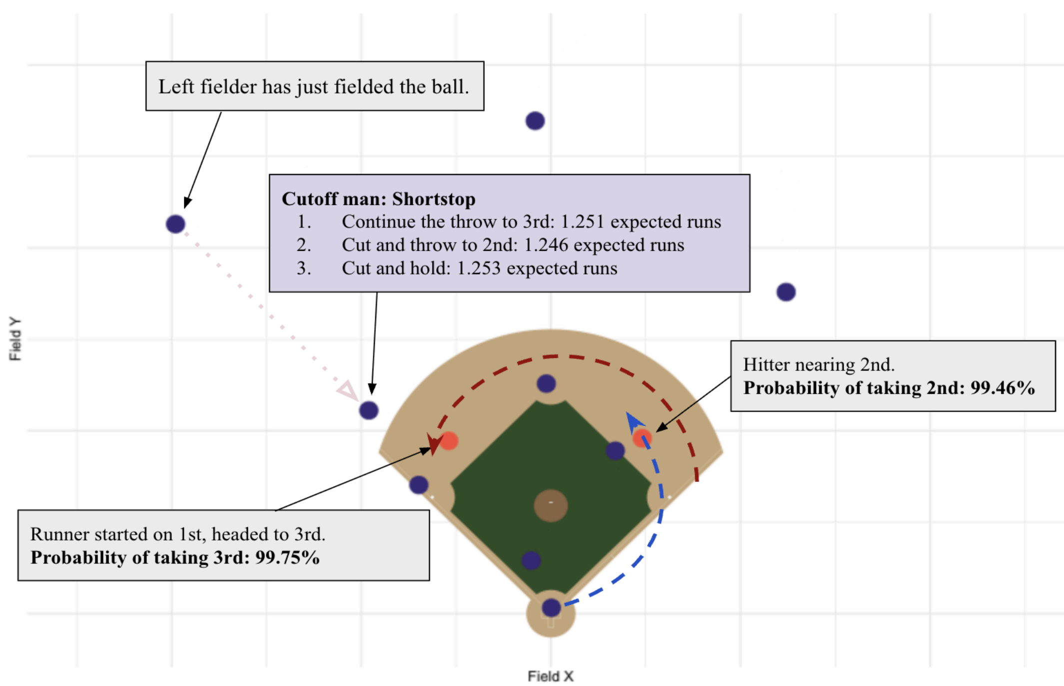


Figure 7: **Click** on the image to see an animation of the play. This play's optimal cutoff action is to "cut and hold," as the expected run values are all within 0.08 of each other. Both runners are very close to reaching their target base, so the cutoff man holding the ball would prevent any errors and inadvertent extra bases. [Link](#) to play.

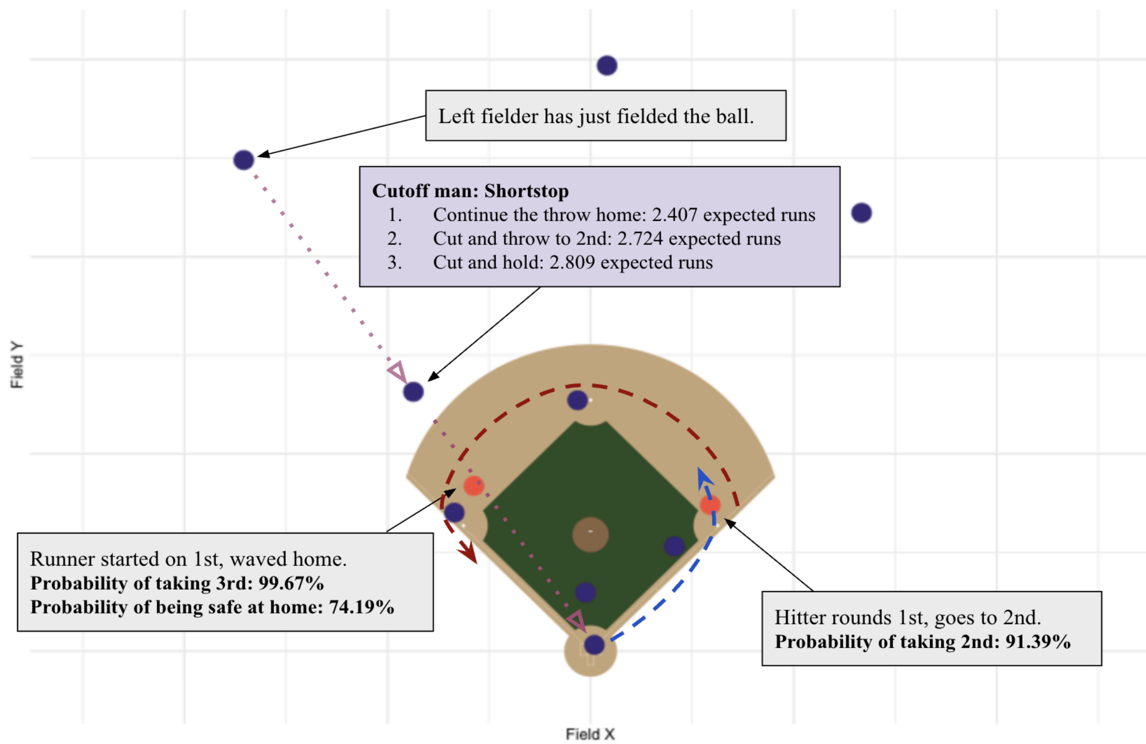
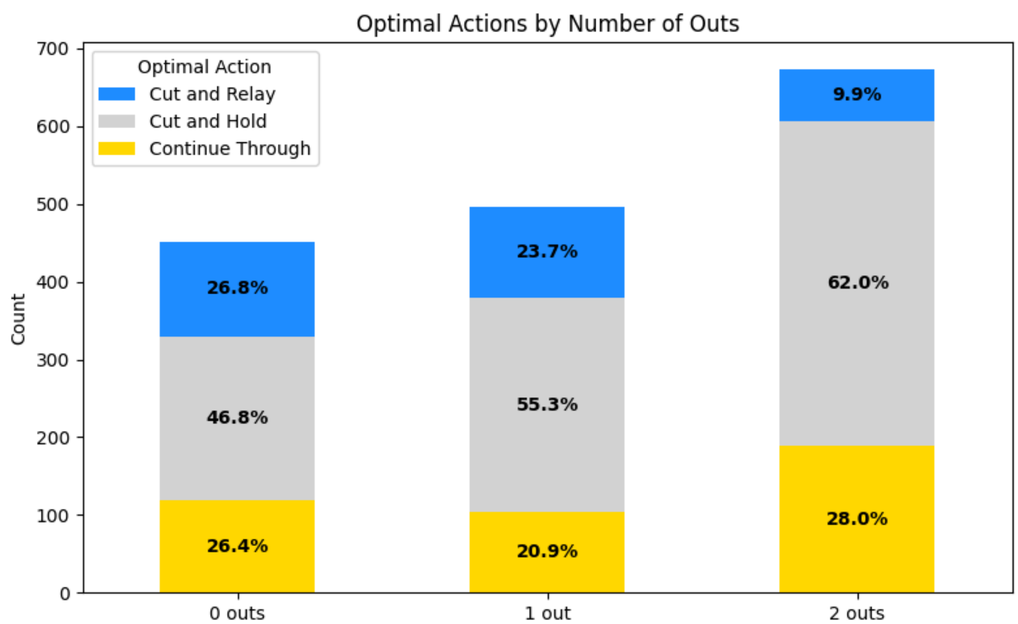


Figure 8: Same play as *Figure 4.* The optimal action for the cutoff man is to continue the throw home, as it results in 0.317 fewer expected runs than relaying it to 2nd. In reality, the cutoff man made the right choice and threw the runner out at the plate!

Additionally, deriving the number of outs on any given play was a key aspect to improving this logic (*Appendix 2)*. You can see the optimal cutoff action distribution by out in *Figure 9*.

Figure 9: Optimal cutoff action distribution by number of outs. Understandably, there are more plays with 2 outs since more runners are on base with 2 outs, leading to more possible cutoff plays as a whole. We can see that a lower percentage of plays are recommended to “cut and relay” with 2 outs, since there is more of an emphasis on getting out of the inning and preventing the lead runner from scoring.



3.3 Optimal Action Model

The optimal actions we determined in section 3.2 effectively communicate the correct decision-making to the infielders in *hindsight*. But, we were missing the "why." We wanted to build a model that could answer the question: "The outfielder has the ball and is about to throw it my way. What should I do?"

We made sure only to include information that the infielder can pick up in real time (like approximate distances) and metrics that could be listed in a pre-game scouting report (like fielder arm strengths and runner speeds). This helped prevent data leakage, as we avoided incorporating any outcome dependent variables that the infielder would not know during the play. Using the optimal cutoff actions as a target variable, we trained a random forest classifier model with the following features:

a. fielder distance to targeted base;
b. baserunner distance to targeted base;
c. cutoff man distance to targeted base;
d. outfielder arm strength;
e. baserunner sprint speed;
f. number of outs;
g. baserunning path (1st to 3rd, 2nd to home, etc.)

Figure 10 showcases the importance of the features in the model.

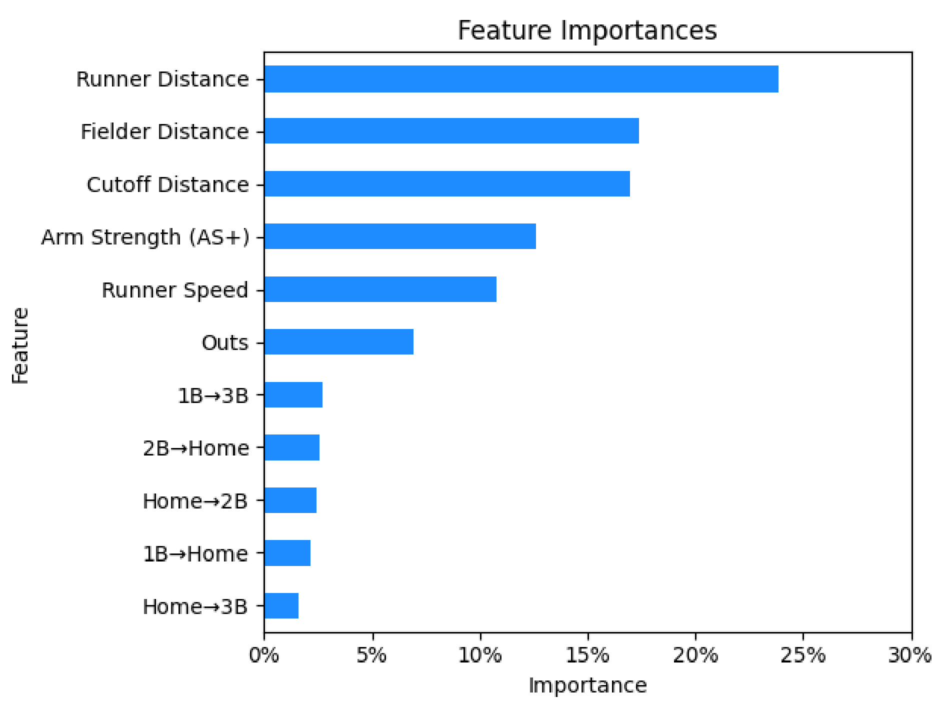


Figure 10: All features used in the cutoff model and their relative weights in the random forest classifier. As you can see, the runner's distance to the base is the most critical factor in the model's decision, followed closely by the distances of the fielder and cutoff man to the base. Arm strength and runner speed also play important roles. We intuitively agree!

Overall, this model performed fairly well in predicting the optimal cutoff action based on the features, achieving about a 77% accuracy rate across all three cases. The full accuracy report is shown in *Figure 11*, and the decision matrix is presented in *Figure 12*. Considering that choosing the optimal cutoff action at random would yield 33% accuracy, we feel confident in our model's results.

Model Performance by Action

	PRECISION	RECALL	F1 SCORE
Cut and Hold	0.80	0.86	0.83
Cut and Relay	0.69	0.75	0.72
Continue	0.77	0.61	0.68
Overall Accuracy			**0.77**

Figure 11: Accuracy report for random forest classifier model. The model can predict the optimal cutoff action 77% of the time using the distance, arm strength, and runner speed features.[2]

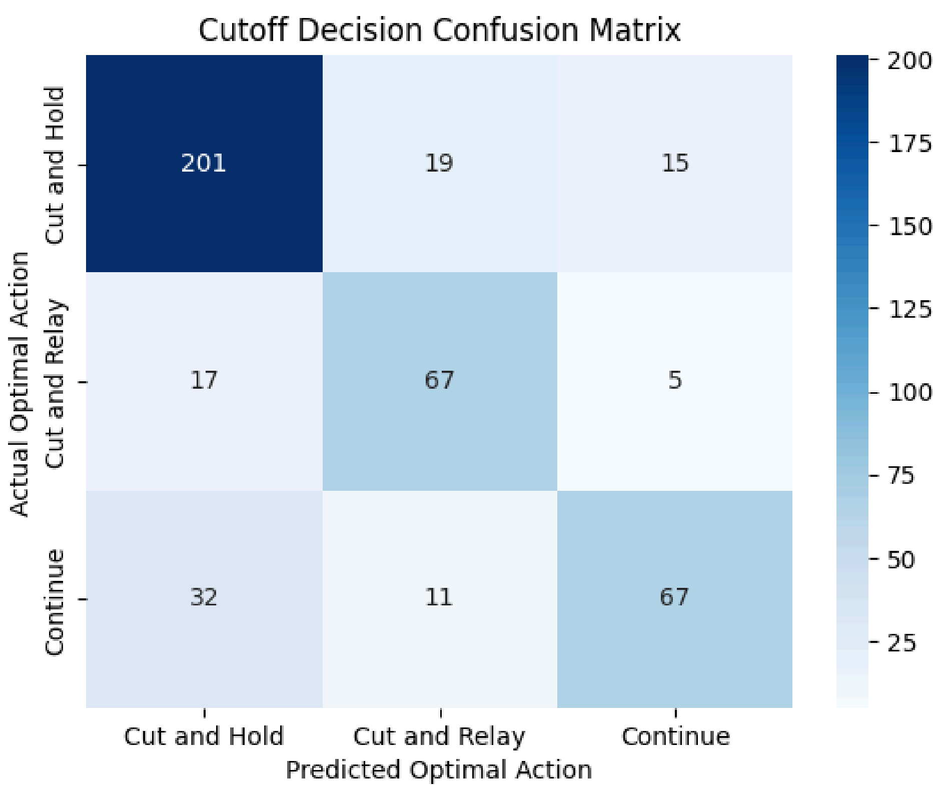


Figure 12: Confusion matrix for model predicted optimal cutoff actions vs. actual optimal actions. For instance, out of the 235 plays where the optimal action was to "cut and hold", our model predicted to "cut and hold" in 201 of them.

[2] Precision measures the proportion of predicted positives that are truly positive, while recall measures the proportion of actual positives that were correctly identified. The F1 score is the harmonic mean of precision and recall, providing a single metric that balances both.

4. Takeaways & Analysis

Our analysis compares our model-predicted optimal cutoff actions with the actual decisions made by fielders, providing both big-picture strategy insights and team- and player-specific scouting takeaways. The takeaways are supplemented by an interactive dashboard designed to communicate our findings on cutoff decision-making.

4.1 Model vs. Observed: Are Cutoff Men Making the Right Decisions?

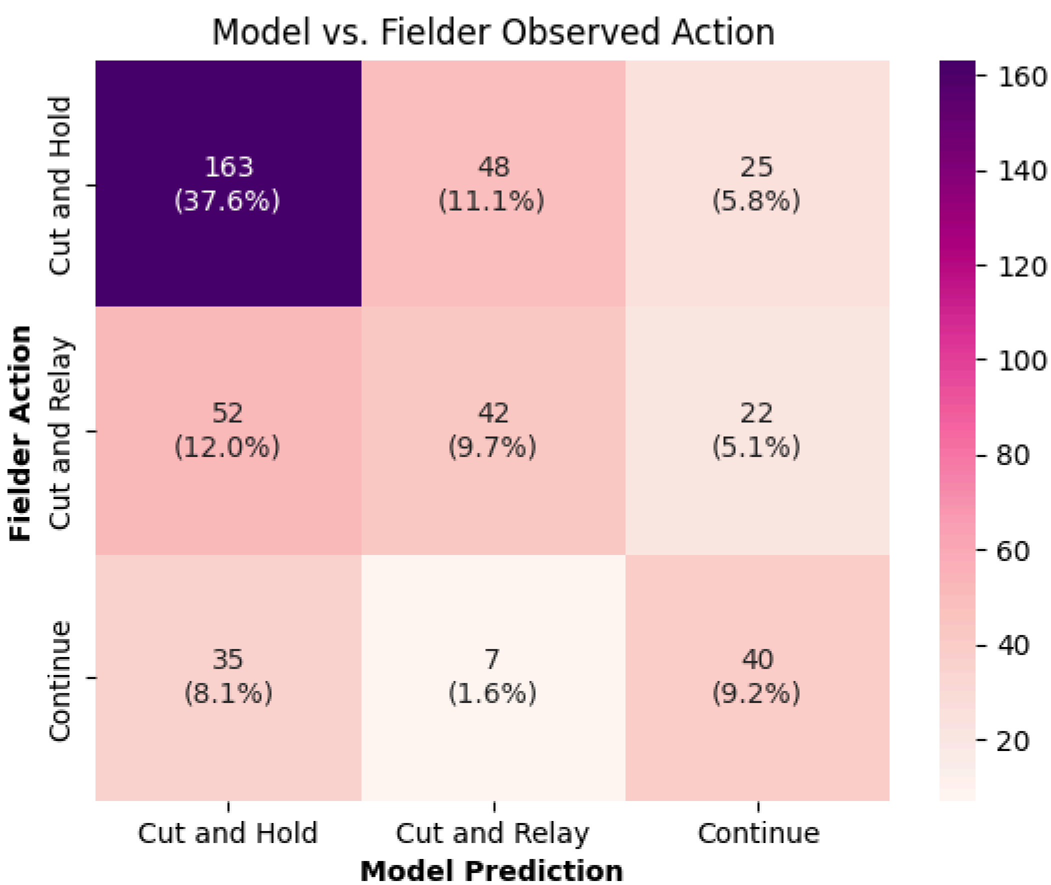


Figure 14: Confusion matrix comparing our model predicted optimal cutoff action vs. the observed fielder action.

Figure 14 represents a confusion matrix between the *model-predicted* and *observed* cutoff actions. It is important to note that this is different from the confusion matrix in section 3.3 (*figure 12)*, which compared *model-predicted* cutoff actions with *optimal* cutoff actions. In short, the matrix in this section lets us analyze fielder decision-making, while the previous one helped us analyze the model's accuracy. Comparing model-predicted actions with observed actions allows us to fairly evaluate whether fielders made the correct decision based on the information available to them during the play.

One main trend emerges from the matrix: fielders are not cutting and holding the ball nearly as often as they should. Of the 250 plays where the model predicted that the fielder should cut and hold the ball, the fielder cut and held it in only 163 of them. This indicates a widespread room for improvement for cutoff men, specifically related to over-aggressiveness. Fielders appear to be opting for riskier decisions to try and get runners out, costing their team significant run value in the long term.

4.2 Dashboard

Explore the dashboard here: https://maninthemiddle.shinyapps.io/smtcutoffdashboard/

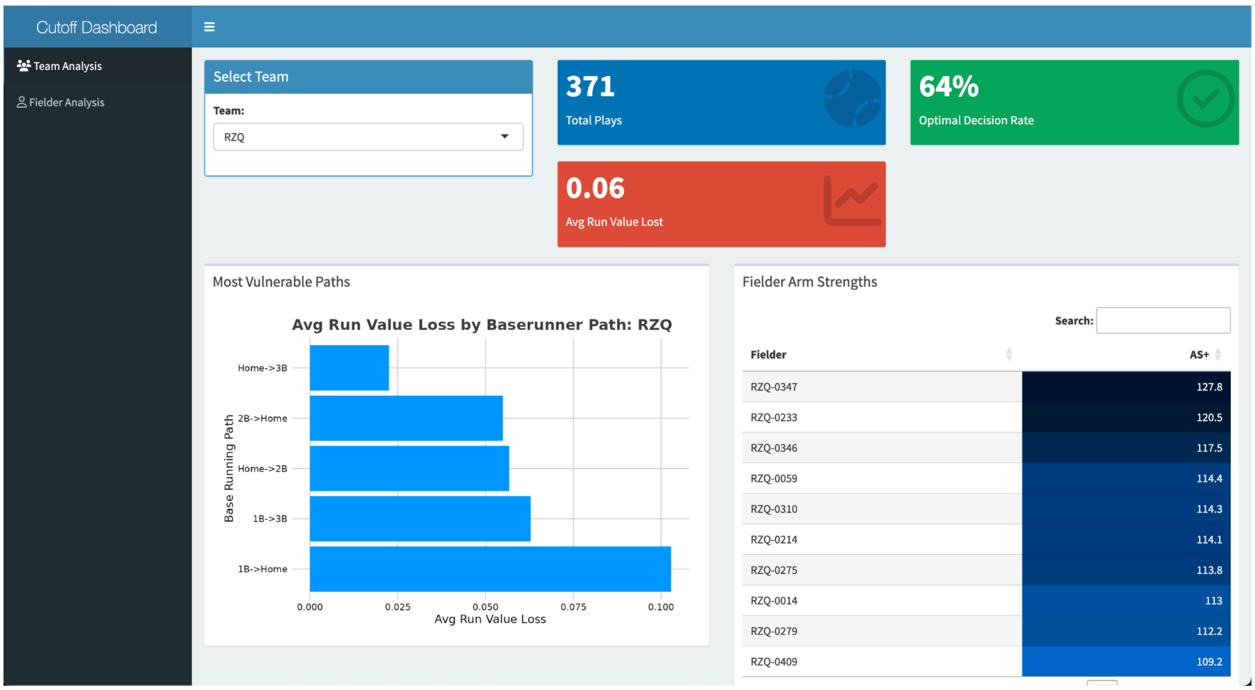


Figure 16a: An example of what the team tab of the dashboard looks like. You can select a team, and get a breakdown of their performance on cutoff plays, as well as their most vulnerable baserunning paths and their outfielders' arm strengths.

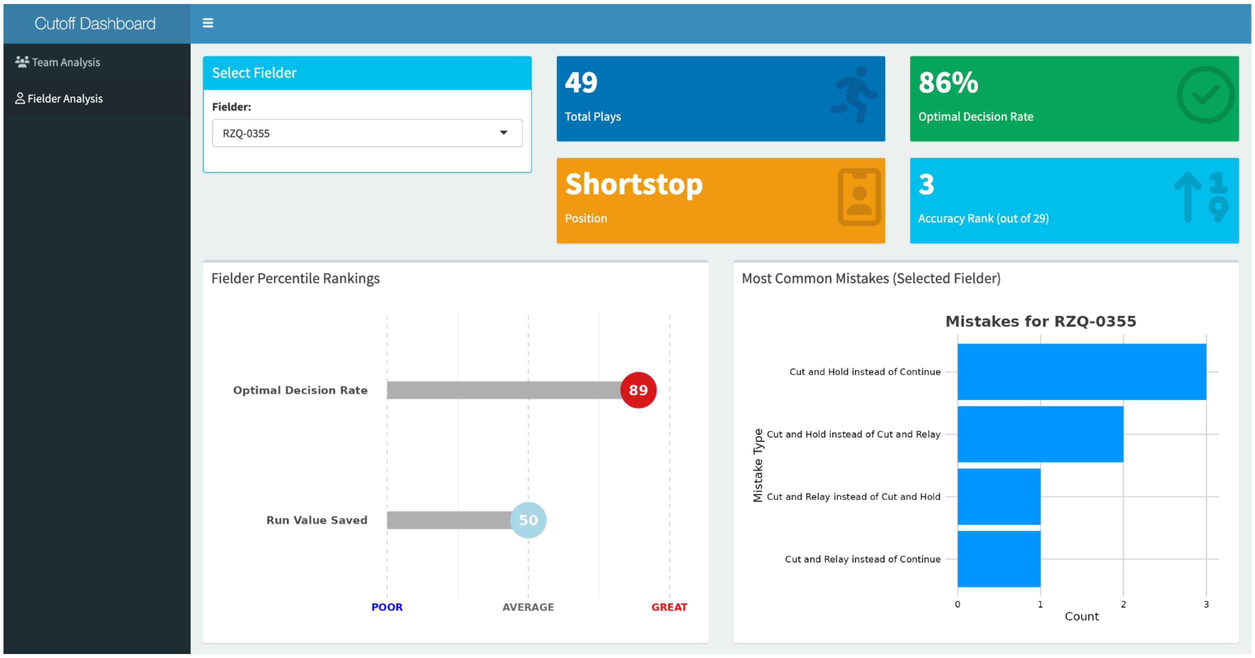


Figure 16b: Dashboard fielder tab view. You can select a player, and get a breakdown of their performance on cutoff plays, including their accuracy and most common mistakes.

In order to make these insights more digestible, we created an interactive dashboard (*Figure 16a and 16b)*. With this dashboard, you can toggle between teams and fielders to understand their decision-making skills, as well as identify particular areas of weakness. With more data, we foresee a tool like this being incorporated into a player or team's internal scouting platform to round out their fielding evaluations.

The dashboard communicates several insights that reinforce our belief that teams should be cutting and holding the ball more. Across the board, fielding teams were most vulnerable on plays with a runner going from 1st base to home, losing an average of 0.1 expected runs on those plays. This is likely explained by over-aggressive cutoff men trying to get a runner out at the plate when they are almost certainly safe.

The dashboard also serves as a valuable scouting tool. Coaches can gameplan their baserunning strategy around the decision-making habits of opposing cutoff men. For example, if the shortstop throws the ball home a lot more than he should, then teams should encourage their trailing baserunners to try and advance more (i.e. from 2nd to 3rd).

The main limitation of the dashboard right now is the lack of a strong sample size. With no individual cutoff man logging more than 50 plays, it is difficult to generalize these results to a higher level takeaway about that player's actions. However, with more high-quality data, we believe this tool has the potential to reshape how teams evaluate and approach cutoff play decision-making, and allow them to exploit tendencies in high-pressure moments.

5. Discussion

Our big picture takeaway is that cutoff men are too eager to target the leading base runner on relay plays, and should be holding onto the ball more. But, that's just the bird's eye view. Our model and analysis is most valuable to coaches and players on a team and fielder level. It provides them with accurate insights into an opponent's cutoff decision-making habits, coupled with the opportunity to smoothly test different scenarios and draw conclusions with our interactive dashboard.

However, several limitations still remain. The most pressing challenge was data quality. The lack of play-by-play and game state information in the data forced us to define our own logic to determine if players were safe or out, which could cause inaccuracy in the initial safe probability model. The limited number of recorded cutoff plays also restricted the sample size for player-level conclusions. In addition, we assumed that all throws could be cut off and that players had complete awareness of the play, which may not always reflect reality. Finally, while our method of aligning the optimal and observed distributions may work in the short-term, future work could incorporate historical error likelihoods or game-theory based models to replace this methodology with a more dynamic and context-aware adjustment.

Looking ahead, there are several ways to expand our project to make it more impactful. Scaling up our methodology with quality MLB data would improve the accuracy and flexibility of our results. Future versions of our dashboard could integrate qualitative scouting, such as player tendencies or coaching tips, and compare specific runner-fielder situations over a larger sample size. Ultimately, our hope is that this work leads to further exploration into cutoff play decision-making. Maybe the smartest throw in baseball is the one that's never made.

6. References

We drew inspiration from Billy Fryer's [2022 SMT Data Challenge Project] to structure our safe probability model and use the RE24 run expectancy data.

7. Acknowledgements

A huge thank you to SMT's Dr. Meredith Wills and Billy Fryer for organizing this data challenge and supporting us throughout the whole process. Also, thank you to Adam Esquer, VP of Baseball Research and Development for the San Diego Padres, and Ajay Patel, Junior Data Scientist for the Pittsburgh Pirates, for offering us advice on our optimal action determination logic and giving us feedback on the paper. We couldn't have completed this project without their support!

8. Appendices

Appendix 1 - Baserunner Safe/Out Criteria

As previously mentioned in sections 2.1 and 5, plays didn't come with "safe" or "out" labels. This was a major problem: the safe probability model needed some indication of whether the runner was safe or out.

Before we began the process of labeling runners safe or out, we needed to filter for instances where the runner attempted to take an extra base. We established a threshold about halfway down each baseline where if the runner was to cross this threshold, we deemed that they were "attempting" to reach the next base and were a "competitive" baserunner. This would exclude instances where the batter takes a big turn around first on a base hit and scurries back to the bag, or where an overly aggressive runner on 2nd base rounds 3rd on a base hit, but is held up by the third base coach.

The tracking data had the back of home plate at (0,0) on the x-y coordinate plane. If a baserunner passed (-35, 35), meaning if their x-coordinate was ever greater than -35 and their y-coordinate less than 35, they were labeled as attempting to reach home plate. For third base, we used (-35, 90) where the x- and y-coordinate had to be less than -35 and 90, respectively. For second base, the coordinates were less than 40 on the x-axis, but greater than 90 on the y-axis.

We began the process of finding safe/out instances by taking a deeper look into who was on base the following play. This was an easy way to tell if the runner was safe at their targeted base, barring extreme or rare scenarios. In our data, we created new binary columns to indicate whether or not the baserunner appeared at their targeted base during the next play. This was the most concrete data that we had for labeling a runner as safe. However, we couldn't rely on this methodology 100%. The dataset was messy with various null instances through the dataset. Additionally, this was only applicable to plays at second or third, failing to account for runners at home plate.

Additionally, we created several boundaries and rules using baserunner and fielder tracking data to confidently label a runner safe or out. This process was backed by manually checking dozens of animated plays. At the moment the fielder fielding the targeted base (e.g. third baseman if third base was targeted) acquired the ball, we calculated the distance of the fielder and the baserunner to the base. The following requirements were to be met to deem a runner safe at the moment the fielder acquired the ball:

a. the runner was fewer than five feet away from the base and
   the runner was at least one foot closer to the base than the fielder

or…

b. the fielder was more than nine feet away from the base.

In combination with this methodology, for plays at second or third, if the runner appeared at the targeted base in the following play's data, we would force the runner safe despite any contradicting estimate using our set of arbitrary rules.

Using our safe/out methodology, the brier scores are generally better in the models where runners are targeting second or third (showcased in *Figure 2*). This is because in the data, we can't see if a runner reached home by checking the base state of the following play, so basepaths 2B → Home and 1B → Home relied on our arbitrary set of rules.

It was important to devise a strong procedure to identify whether a runner was safe or out at their targeted base. Through the two steps outlined in this appendix, we feel confident in the accuracy of our labeling. Of course, if the dataset included those labels to begin with, we would be more confident with the data's accuracy, and therefore our initial safe probability model's accuracy, and would've been able to focus our attention elsewhere. Yet, we are happy and satisfied with our methodology given the brier scores for each of our five safe probability models.

Appendix 2 - Determining the Number of Outs

As previously mentioned, the dataset did not include any direct information about the number of outs. However, it did include two extremely helpful data points that helped us infer outs. First, the data showed whether each base was occupied or not at every frame, along with unique player codes that allowed us to track players around the bases. Second, each play contained the following contextual details: a unique game code, its relative place in the game sequence, and whether it occurred in the top or bottom of the inning.

We began the process of calculating the number of outs by sorting all plays by game and their sequence within each game to see the whole season's data in chronological order. With the plays properly ordered, we then calculated the inning number and the batter's position within each half-inning.

The batter number within each half-inning was essential for the rest of our calculations. By combining it with the unique codes assigned to all batters, we were able to track current baserunners, along with how many had been on base at any point in the inning as a whole. These unique codes made it possible to follow a batter's progress, for instance, identifying when a player reached first base during their own at-bat and then advanced on subsequent plays. Due to data limitations, we assumed that if a current baserunner no longer appeared on base later in the inning, he had scored. The dataset included play-type labels, so when a batter hit a home run and subsequently did not appear on base (because he rounded all at once!), we counted it as a run scored for the offense.

By combining information about the batter sequence within the inning, the number of runs scored, and the number of players who had reached base, we were able to estimate the number of outs using the following formula:

Number of outs = batter number of the inning - runs scored - current baserunners - 1

For example, if the current batter was the fifth of the inning, one run had already scored, and two men were on base, we inferred that one out had occurred.

Accurately estimating the number of outs was important for our optimal action model. While this approach is not perfect, since it can miss outs on the bases, it's a reasonable determination given our data's limitations, and we are satisfied with this calculation and its inclusion in our model.

Appendix 3 - Optimal Cutoff Action Determination

The exact mathematical formula we used to determine the optimal cutoff actions is shown below:

Let $A = \{a_1, a_2, \ldots, a_n\}$ be the set of attempted base advances on a play. Each $a_i \in A$ has a probability $p_i = \Pr(\text{Success of } a_i)$. We compute the **expected run value** (xRV) by summing over all 2^n possible combinations of outcomes, where n is the number of runners attempting an extra base. Each runner attempt can independently result in either a success (safe) or a failure (out), which is what leads to 2^n total binary outcome vectors.

$$\text{xRV} = \sum_{\mathbf{o} \in \{0,1\}^n} \left[\left(\prod_{i=1}^{n} p_i^{o_i} (1 - p_i)^{1-o_i} \right) \cdot (\text{RE}_{\mathbf{o}} + \text{RS}_{\mathbf{o}}) \right]$$

where:

- $\mathbf{o} = (o_1, o_2, \ldots, o_n)$ is a binary outcome vector (1 = safe, 0 = out)
- $\text{RS}_{\mathbf{o}}$ = number of runners that successfully scored (i.e., advanced to Home)
- $\text{RE}_{\mathbf{o}}$ = run expectancy value from the RE24 table for the resulting base state

After coming up with these values, we set the action with the smallest xRV to be the optimal action. A key part of this process was determining a threshold value to default the optimal cutoff action to "cut and hold." This was required because, based on our mathematical process using probabilities, the expected run value of continuing or relaying would always be less than cutting and holding, and therefore more optimal.

In order to arrive at this threshold number, we first generated the *observed* distribution of cutoff actions. In other words, what cutoff men actually did in our data. We used the following logic to assign each play with an observed label:

a. "Continue": The ball was thrown from the outfield to a cutoff man, and the cutoff man subsequently threw the ball to the player assigned to the target base of the leading runner (e.g. catcher for a runner going home, third base for a runner going to third).
b. "Cut and relay": The ball was thrown from the outfield to a cutoff man, and the cutoff man subsequently threw the ball to the player assigned to the target base of a trailing runner.
c. "Cut and hold": The ball was thrown from the outfield to a cutoff man, who did not make any subsequent throws. Note that if the outfielder threw it directly to the 3rd baseman with the leading runner approaching 3rd, we considered this a "continue" and not a "cut and hold" since the intent was to get the leading runner out.

The observed distribution is shown below:

Action	**Count**
Cut and Hold	530
Continue	177
Cut and Relay	92

After determining the observed distribution, we grid searched through various threshold values, measuring the difference between the resulting optimal cutoff distribution and the observed one. We then picked the threshold value that made the optimal distribution look closest to the observed one. That value was 0.08. With more time and higher quality data, we would definitely try to improve this process to more accurately assess the point at which cutting and holding is optimal, but we thought that this grid searching process was suitable for our purposes, and generally aligned with our intuitive thoughts on the correct decision through manual validation of the plays.

Appendix 4 - Feature Engineering

We outlined the features used in our initial safe probability model in Section 3.1 and described those used in our larger, optimal cutoff action model in Section 3.3.

We want to emphasize that distance-related features were calculated at the moment the outfielder acquired the ball. We felt this moment was a good indicator of when the cutoff man has to start processing all the developing aspects of the play to make his cutoff decision.

The distance-related features are all calculated very similarly – the distances from the fielder, baserunner, and cutoff man distance to the baserunner's targeted base. These were calculated by taking the euclidean distance[3] from the player to the base, on the xy-plane.

The outfielder arm strength and sprint speed features were calculated a little bit differently. These were individual metrics at their core. Each player and runner ID represented a unique MiLB player with a unique arm strength or speed on the basepaths.

To calculate arm strength, we used a similar process to that of calculating velocity out of a pitcher's hand. We identified the two closest timestamps after the outfielder released the ball, calculated the euclidean distance between the ball at those two timestamps, and divided by the

[3] distance = $\sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2)}$

difference in the timestamps. Using this method was superior to calculating velocity by taking (total distance / total time), because the latter leaves room for air resistance, wind, and the possibility of the ball bouncing and slowing down. We felt our calculation was a more accurate representation of true individual arm strength. After that, we converted the metric to miles per hour for ease of comprehension.

We saw many instances where there was no competitive play at a base, meaning the outfielder's throw into the infield was a noncompetitive lob. To avoid taking this into consideration, we used the max arm strength for each fielder ID as their true arm strength. But, there were several fielders who we only had one or two play's worth of data for, so if their throws were uncompetitive, their max arm strength would be noticeably low. To work around this, if their max arm strength in the dataset was less than 50 miles per hour, we imputed their arm strength metric with the mean arm strength for their position. This raised the average arm strength in the dataset from 57.6 to 68.1 mph, a more accurate representation holistically.

Finally, we standardized the arm strength metric to 100, similar to how OPS+ and ERA+ are calculated. We aim for our analysis to be replicated and used as an advanced scouting tool for teams, so using a standardized metric can help with coach or player understanding. The metric is called AS+. If a fielder's arm strength is 20% weaker than the average, their AS+ would be ~80, and vice versa, if their arm strength is 20% stronger than the average, their AS+ would be ~120.

Lastly, we needed to calculate sprint speed for unique baserunner ID's. Our philosophy for finding an accurate sprint speed was slightly different from arm strength. Instead of taking the max sprint speed, we felt that taking each baserunner's 90th percentile sprint speed would more accurately represent their top speed in most plays. To do so, we looped through each baserunner's running paths to find their fastest 1-second interval. Keeping the data clean, we discarded any instances where the runner ID was null. Next, we calculated for the 90th percentile sprint speed, only keeping runner IDs that appeared in five or more plays to ensure a decent sample size.

In *Appendix 2*, we outline our process for identifying the number of outs and how it shaped the model's performance, our findings, and allows for deeper analysis. We used outs as part of our optimal cutoff action logic because we feel this is a notable variable in any high-pressure cutoff play, especially affecting the expected run value equation when there are two outs.

Appendix 5 - Modeling Details

We developed two different models as part of our pipeline: a safe probability model and an optimal action model.

Our safe probability model was a logistic regression model modeled from the point the outfielder received the ball. We used two features, runner distance to target base and outfielder distance to target. We chose to use a logistic regression model because of its simplicity and interpretability for binary classification. We used an 80/20 train/test split on the data, and received strong results without adding unnecessary complexity. To reiterate what was said in Section 3.1, we trained five different models, one for each baserunning scenario, and got brier scores of under 0.1 for all five models.

Our optimal action model was a random forest classification model trained on 11 input features and the optimal cutoff action as our target variable. We chose a random forest model because it is adept at handling nonlinear feature interactions, and is robust to overfitting. To create train/test sets, we used a stratified 75/25 split, ensuring that the distribution of cutoff action classes (cut and hold, cut and relay, continue) was preserved in both sets. This step is important for producing a representative test set, particularly when class imbalance exists. We then performed 5-fold cross-validation on the training set to evaluate performance across multiple partitions and reduce variance. We one-hot encoded basepaths to make them numerical features, telling the model which specific baserunning path the runner was on during a play. For instance, a column labeled "is_1B_to_3B" would have a value of 1 if there was a runner attempting to go from 1st to 3rd on that play.

Since our target variable (the optimal cutoff action) was imbalanced with a higher proportion of "cut and hold" decisions, we used SMOTE (Synthetic Minority Over-sampling Technique) to address this class imbalance in the training set only. SMOTE generates synthetic examples of the minority classes (cut and relay and continue), balancing the class distribution without contaminating the test set. This helps the model better learn decision boundaries for all classes.

Finally, we performed a grid search over various different hyperparameters to find the ideal set. The metric for evaluating for this hyperparameter tuning was cross-validation accuracy, and resulted in the following configuration:

Hyperparameter	**Value**	**Explanation**
n_estimators	100	Number of trees.
max_depth	20	Maximum depth of the trees.
min_samples_split	2	Minimum samples required to split a node.
min_samples_leaf	1	Minimum samples required at a leaf node.

For this final model, we also tried using an XGBoost modeling approach. However, the results for the XGBoost model were fairly similar to the Random Forest model, so we opted to use the Random Forest for the sake of clarity and ease of interpretation.

In the future, we could improve this model by considering features like ball path, throw difficulty, and additional game context (inning, score, etc.). We could also explore strategies such as reinforcement learning to better model the sequential decision-making of cutoff plays.