Go Ahead and Jump: Early and Often

Team 9

SMT Data Challenge 2025

Abstract

This project analyzes the relationship between two outfield defensive metrics: Outs Above Average (OAA) and Statcast's Jump. OAA evaluates past performance by comparing a fielder's ability to convert balls into outs against league averages, while Jump breaks down the distance covered during the first three seconds of a play into Reaction, Burst, and Route phases. By combining these metrics, the analysis provides a more complete picture of individual performance and potential. A logistic regression model was used to predict whether a ball hit into the outfield would be converted into an out, using hangtime, distance to result, and the direct distance angle as input features. The model achieved a mean accuracy of 92.6% and an F1 score of 95.1%, with time and distance identified as the most impactful features. Further analysis shows that covering more ground for longer durations correlates with higher success rates on difficult plays. Player-level Jump metrics reveal different movement strategies, with early and sustained movement contributing to better outcomes. Instead of result based critiques based on OAA alone, the data supports a more actionable critique: move early and often. This paper is in a similar vein as OAA, and is a result of a much greater underlying process that takes place inside of the Jupyter notebooks attached to this project.

Background

Outs Above Average

Outs Above Average (OAA) is a metric that rewards fielders for making outs that the average fielder has a lower probability of making, and punishes fielders for failing to make an out that an average fielder has a high probability of making (Baseball Savant, 2025a). Tom Tango introduces the Distance/Time math model in the *History of Fielding* where the catch's probability is derived from historical data on the percentage of times a ball was converted to an out given the distance the fielder needs to travel to catch the ball and the amount of time they have to travel that distance (Tango, 2019). OAA is a good metric when describing the past performance of a fielder compared to their peers. However, when using OAA to evaluate individual performance, there is only one critique that can be given: catch everything.

Jump

Unlike OAA, Statcast's Jump focuses only on the amount of distance the outfielder is able to cover during their route to the ball. Jump takes into account the amount of distance covered during different phases over the first 3 seconds of the play, starting when the pitcher releases the ball (Baseball Savant, 2025b). The distances involved in the Jump include:

- **Reaction**: Distance the fielder travels in any direction in the first 1.5 seconds from when the pitcher releases the ball.
- **Burst**: Distance the fielder travels in any direction from 1.6 to 3 seconds.
- **Route**: Distance covered in any direction from pitch release to 3 seconds compared to the direct distance from the fielder's starting location to their location at 3 seconds.

These distances are collected for each fielder for every play ranked 2 Star or higher difficulty. The average of each distance is taken as a whole to get the league average, which individual fielder averages are compared against. This gives average feet either gained (more distance traveled) or lost (less distance traveled) during each phase of the route. The gains and losses are then added up to give the outfielder's Jump metric. The greater the Jump, the better! A detailed explanation and example can be found in the *Analysis* Jupyter notebook attached to this project under “The Jump.” The components of Jump as performance metrics are more robust than OAA alone, as it effectively details how early an outfielder starts their route and begins to cover ground. Using Jump components in conjunction with OAA gives insight into the fielder’s potential success in converting difficult plays into outs, altering the critique to: catch everything as fast as you can.

Data

Content and Storage

The data used in this project was provided by SportsMedia Technology (SMT) and includes anonymized ball and player tracking data from three Minor League venues captured at 20 and 33 frames per second. The data also includes events that occur during the game such as when the ball is pitched, hit into play, hits the ground, acquired, or thrown, along with timestamps for each event and the position value for the player that was involved in the event. To find the specific player attributed to the event, there is an additional table with unique player IDs for the players on the field. Due to each observation in the data being attributed to a game ID and play ID, this data was a perfect candidate for a relational database. Using the DuckDB library in Python, all of the data was stored into 4 tables: two for each type of position tracking, one for game events, and one for information pertaining to the players on the field.

Data Engineering

This data was used to recreate Statcast's catch probability model to calculate OAA and Jump Metrics for balls hit into the outfield. To achieve this, key metrics such as hangtime of the ball, the distance the outfielder needed to travel to make the catch, and the direction of the direct route to the ball were extracted from the data. Precise measurements of these metrics relied on timestamps that were located in different tables and needed to be synced to the release of the ball from the pitcher's hand. Additionally, plays needed to be attributed to the responsible outfielders and determinations on whether it was plausible that a fielder could catch a ball were made. However, not all plays have the player information attributed to them and without game state information, not all plays were able to be attributed to a specific player. To navigate this, placeholders were created to track metrics for each venue and outfielder position combination. This allowed for plays to still qualify for model training but were excluded during analysis. After trimming the data, there were 4,029 individual plays of balls hit into the outfield being attributed to 375 unique fielders and 9 placeholders. The full details on this process can be found in the *Data Engineering* Jupyter Notebook attached to this project under the "Data Extraction" and "Data Exploration" sections.

Feature Engineering

The sprint speed of each player was used to create a route intensity metric. Following Statcast's lead of tracking player sprint speeds as the "feet per second in a player's fastest one-second window on individual plays," a rolling average of the player's speed was taken during the route and the max value was attributed to the play (Baseball Savant, 2025c). However, not all plays require fielders to sprint. To remedy this, the sprint speed that was attributed to the fielder was the player's potential sprint speed, which was calculated by taking the average of the

Figure 1

Star Levels per Probability Range

Catch Probability Buckets	
Star Level	Catch Probability
☆ ☆ ☆ ☆ ☆	>95%
★ ☆ ☆ ☆ ☆	91-95%
★ ★ ☆ ☆ ☆	76-90%
★ ★ ★ ☆ ☆	51-75%
★ ★ ★ ★ ☆	26-50%
★ ★ ★ ★ ★	0-25%

Note. (Baseball Savant, 2025a)

fielder's top third max route speeds. Along with the total distance the fielder traveled during the route, the distance the fielder traveled at or above their average sprint speed was tracked. This type of distance will be referred to as their intensity distance and was used to create the route intensity metric which is the percentage of total distance that was traveled at or above the fielder's average sprint speed. After the catch probability model was completed, the plays were bucketed based on Statcast's Star difficulty system which is shown in **Figure 1** (Baseball Savant, 2025a).

Modeling

Logistic Regression

Logistic regression was selected as the classification model for this project due to the interpretability of the results. The target feature for the model will be whether the fielder was able to convert a ball hit into the outfield into an out. The input features to predict this target include:

- **Hangtime:** Number of seconds from the ball being put into play to result.
- **Distance to Result:** Number of feet the fielder needs to cover to make the catch.
- **Backwards:** Binary feature with 1 indicating the route required the fielder to move backwards.

Model Evaluation and Feature Weights

The model demonstrated consistently high performance across 20 stratified k-folds, achieving a mean accuracy of 92.6% with a standard error of 0.4%, and an F1 score of 95.1%

Figure 2

Logistic Regression Feature Weights

Feature Weights	
Features	Weight
Hangtime	4.762
Backwards	-0.474
Distance to Result	-3.833

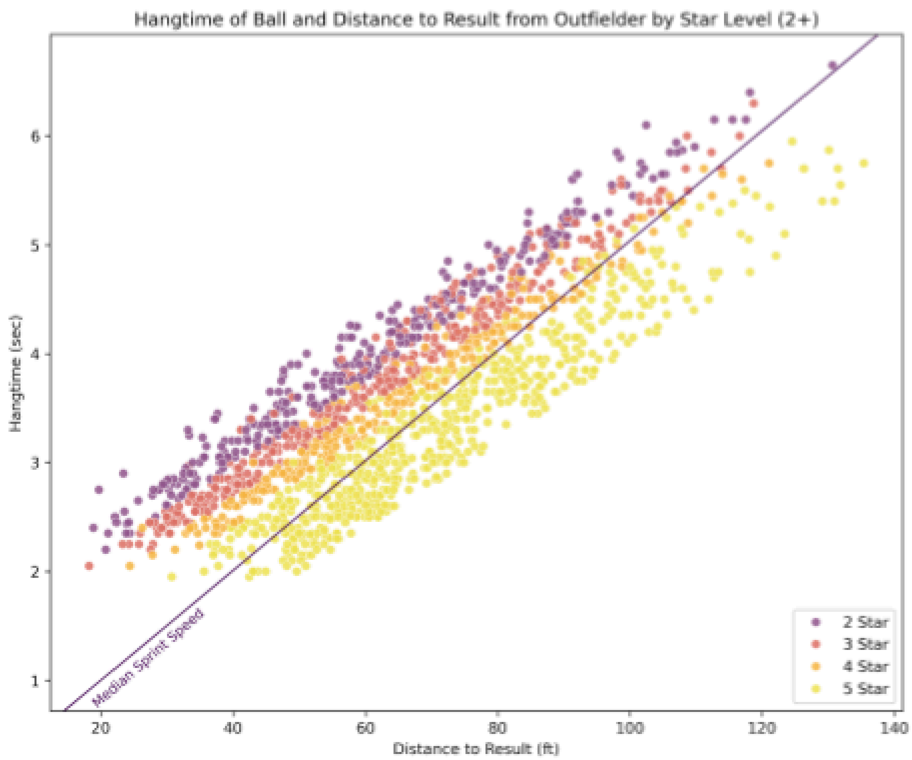
with a standard error of 0.3%. The accuracy of the model is 17.2% greater than the base classification of 75.4%. The features with the greatest impact on converting a ball into an out is the amount of time the fielder has to catch the ball and the distance the fielder needs to travel to catch the ball. For every second the ball is in the air, the probability increases by 4.7% and for every additional 23 feet the fielder needs to travel, the probability decreases by 3.8%.

Analysis

Magnitudes of Difficulty

Figure 3

Relationship between Hangtime, Distance, and Speed



Note. Scatterplot illustrates the relationship between the speed and distance needed to cover to convert a ball to an out at various levels of difficulty. The median sprint speed for this population of outfielders is 19.8 ft/sec. Inspired by the Opportunity Space of Outfielders explored in *The History of Fielding* (Tango, 2019).

Figure 3 illustrates the relationship between the speed needed to catch a ball and the difficulty converting the ball into an out at different play difficulties. All of the 2-3 Star plays require less than the population's median sprint speed and the majority of the 5 Star plays require greater than the population's median sprint speed. The 5 Star plays that require less than the median sprint speed occur within the first 3 seconds of the route. To convert those balls into outs, the fielders would need to be moving at a constant speed of 19.8 ft/sec for the entirety of the route. If this were the case, the fielder would exhibit double digit React distances, and would need to be traveling at high intensity for the entirety of the route. **Figure 4** shows the relationship between play difficulty and route intensity. As the play difficulty increases, the percentage of the route traveled at intensity also increases. The largest increase occurs between 2 and 3 Star plays, which requires the fielder to have a route intensity of 43%. Additionally, the average OAA added between 2 and 3 Star plays increases by 0.2 OAA; the OAA added continues to increase by approximately that amount at each increase of play difficulty. Routes that have high route intensity (>40% intensity) provide more value as they improve the likelihood of success on higher difficulty plays. **Figure 5** shows that routes run at high intensity convert more balls into

Figure 4

Average Metrics of Out Producing Plays at Various Levels of Difficulty

Average Metrics of Out Producing Plays per Star Level

Star Level	React Distance (ft)	Max Sprint Speed (ft/sec)	Intensity %	OAA Added
☆ ☆ ☆ ☆ ☆	4.43	16.49	8%	0.01
★ ☆ ☆ ☆ ☆	5.73	20.42	19%	0.07
★ ★ ☆ ☆ ☆	6.13	21.75	27%	0.15
★ ★ ★ ☆ ☆	6.46	23.54	43%	0.35
★ ★ ★ ★ ☆	7.29	25.23	53%	0.59
★ ★ ★ ★ ★	8.93	26.96	65%	0.79

Note. Table contains average metrics of plays that produced outs at various Star Levels. The average distance traveled in the first 1.5 seconds and the fielder's sprint speed increase as the difficulty of the play increases. The percentage of the route that is traveled at intensity has the largest increase between 2 and 3 Star Levels.

Figure 5

Outfield Fielding Percentage at High Intensity and Various Difficulty Levels

Outfield Fielding Percentage per Star Level and High Intensity

	All Plays			>40% Intensity			Percent Outs at >40% Intensity
Star Level	Opps	Outs	%	Opps	Outs	%	
☆☆☆☆☆	2,196	2,167	98.7%	135	128	94.8%	5.9%
★☆☆☆☆	193	183	94.8%	42	41	97.6%	22.4%
★★☆☆☆	321	288	89.7%	100	90	90.0%	31.3%
★★★☆☆	358	264	73.7%	186	155	83.3%	58.7%
★★★★☆	296	118	39.9%	172	88	51.2%	74.6%
★★★★★	665	18	2.7%	312	17	5.4%	94.4%
Total	4,029	3,038	75.4%	947	519	54.8%	17.1%

Note. Table contains the fielding percentages for the different Star Level plays, as well as fielding percentages for routes that were ran at high intensity. The table also gives the percentage of total outs that were recorded when the fielder was traveling at high intensity.

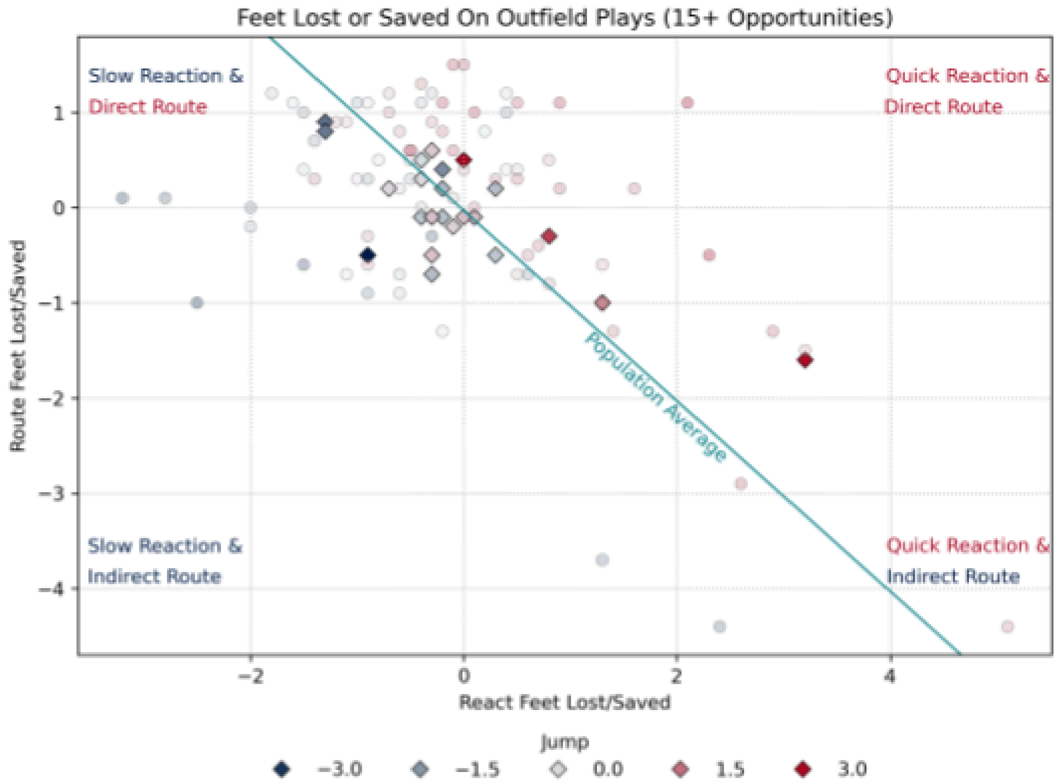
outs at higher difficulty ratings, with the fielding percentage increasing by 10% for 3 Star plays and 12% for 4 Star plays. Outs recorded on routes with high intensity make up 65% of all 3+ Star difficulty outs recorded in this sample. Clearly, covering ground quickly and for as long as possible has a positive effect on recording outs, but when should outfielders be covering the most ground?

Player Jump Metrics

The components of Jump give insight into where the outfielder travels the most. The focus of this analysis is on outfielders that have at least 15 opportunities to convert 2+ Star plays into an out. This filters the 375 unique outfielders down to 23 that are represented by diamonds in **Figure 6,** which is a plot that can be used to profile different types of outfielders. Using this figure in conjunction with **Figure 7**, the top and bottom outfielders by Jump can be analyzed.

Figure 6

Combination React and Route Jump Components Compared vs. Population Average



Note. Plot compares the React and Route Jump components to the population average. Fielders with quick reactions and direct routes are up and to the right of the population average line, whereas fielders that have slow reactions and indirect routes are located on the bottom left of the population line. Fielders with a high Jump metric have a red point and fielders with a low Jump metric have a blue point. Diamonds indicate fielders with at least 15 2+ Star Level opportunities. Inspired by Statcast's Jump Leaderboard (Baseball Savant, 2025c)

Figure 7

Top and Bottom 5 Outfielder by Jump

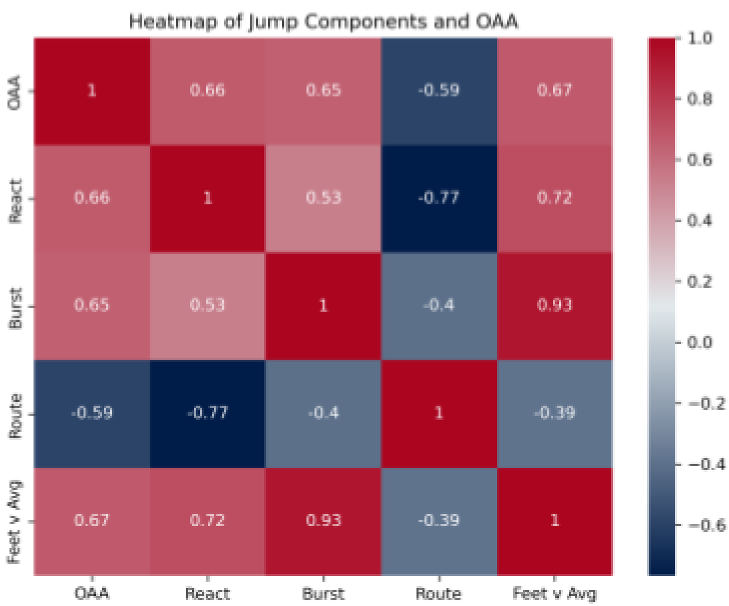
Top and Bottom 5 Outfielders by Jump

Top 5							Bottom 5						
Rank	Fielder	OAA	React	Burst	Route	Feet v Avg	Rank	Fielder	OAA	React	Burst	Route	Feet v Avg
1	QEA-0252	2.1	0.0	2.5	0.5	3.0	19	YJD-0185	0.0	−0.2	−0.9	0.2	−0.9
2	RZQ-0346	4.3	3.2	1.3	−1.6	2.9	20	RZQ-0014	0.1	−0.2	−1.8	0.4	−1.6
3	YJD-0210	3.1	0.8	1.7	−0.3	2.2	21	RZQ-0333	-2.9	−1.3	−1.4	0.8	−1.9
4	YJD-0019	3.6	1.3	1.1	−1.0	1.4	22	QEA-0249	-0.9	−1.3	−1.7	0.9	−2.1
5	YJD-0345	1.0	0.0	0.4	−0.1	0.3	23	RZQ-0275	-3.2	−0.9	−2.0	−0.5	−3.4

Outfielder QEA-0252 has the highest Jump metric, traveling the most distance during their Burst phase and running a slightly direct Route. QEA-0252 starts moving later in their route and uses their speed in the Burst phase to get to the ball. The second highest Jump metric is with RZQ-0346, with the majority of their distance occurring during the React phase and running an indirect Route. RZQ-0346 starts moving earlier in their route and utilizes their speed in the Burst phase to make up the initial distance traveled in the wrong direction, running an indirect route. Based on the three components of the Jump, there is already a clearer picture of how these outfielders make their outs compared to just OAA. Looking at the bottom 5 outfielders by Jump, they are all below average on distance traveled over all phases but run rather direct routes. The only outfielder in the top 5 by Jump to run a direct route is QEA-0252, but they also have the second lowest OAA of the top 5 outfielders.

Figure 8

Heatmap of the Correlation between Jump components and OAA



Note. The heatmap shows the correlations between the components of Jump and OAA. Values close to 1 or -1 have the strongest correlations and values close to 0 have the weakest correlations.

Figure 8 shows that Route has a strong negative correlation to OAA, meaning that the closer the outfielder's total distance traveled is to the direct distance needed, the worse their OAA will be. Additionally, React has a negative correlation with Route. This makes intuitive sense that the distance the outfielder is traveling at the beginning of the route is when they have the least amount of information about where the ball will end up, leading to indirect routes.

Conclusion

Converting balls into outs is about more than running fast. Outfielders need to start moving as soon as they possibly can. By analyzing the three phases of a fielder's Jump in conjunction with their OOA, more specific feedback can be given to that fielder to convert balls into outs. Outs recorded with high route intensity account for 65% of all outs on the most difficult plays, and what is route intensity if not a measurement of the fielder covering more distance for longer periods of time. What is important is that the fielder covers distance, regardless of the direction. This is seen with Route's strong negative correlation to OAA and positive correlation with Jump components that measure the amount of distance in any direction. So instead of just telling players to catch more balls, we can tell them to move early and often.

Acknowledgements

The work in this project was done by a single individual, however, it was not done alone. A special thanks to those who supported me “on and off the field.” Thanks to my family for supporting me through this project and my wife for telling me numerous times that my explanations could be better. Thanks to Billy Fryer and Meredith Wills for coordinating the challenge and the continued support throughout the project by helping me see the forest through the trees and encouraging me to be confident in my abilities. Thanks to Cameron Adams for giving me the inspiration to bring the audience along with me in my analysis, and to remind them of the things that I just told them. Thanks to Gabe Riedel for providing the starter code for DuckDB, as this project would not have been able to get off the ground without that. Finally, thanks to SMT for allowing me the opportunity to work with this type of data as this is not a public resource and is usually reserved for teams and organizations.

References

Baseball Savant. (2025a). *Statcast Catch Probability*. Major League Baseball. https://baseballsavant.mlb.com/leaderboard/catch_probability

Baseball Savant. (2025b). *Statcast Outfielder Jump Leaderboard*. Major League Baseball. https://baseballsavant.mlb.com/leaderboard/outfield_jump

Baseball Savant. (2025c). *Statcast Sprint Speed Leaderboard.* Major League Baseball. https://baseballsavant.mlb.com/leaderboard/sprint_speed

Tango, T. (2019). *History of the fielding*. http://tangotiger.com/images/uploads/History_of_the_Fielding.pdf