Changes In Minor League Umpire Tendencies With The Challenge and Automatic Ball-Strikes Systems

by Jonah Soos, Dr. Jeremy Losak, & Dr. Jason Maddox
General Track Presentation Proposal
Syracuse University

Introduction:

Subjective, and sometimes problematic and inconsistent, umpire pitch calling has sparked intense debate prompting a reevaluation of the need for changes behind the plate. Baseball purists argue that umpires add personality, a unique touch that makes each and every individual game different. They argue that umpires continue to improve over time, but also have negligible net impact in the grand scheme of a 162-game season. A more modernist approach, sparked by the Sabermetrics movement and stoked by the increased availability of pitch tracking and analysis, sides more strongly with the need for consistency, mitigating potential error in a game of inches and milliseconds. In an attempt to modernize the game, Major League Baseball has trialed multiple unique systems in the minor leagues that automate pitch calling and take away various amounts of control and influence from an umpire. This study focuses on the impact of these systems, exploring both their accuracy and evolving umpire tendencies under new technology, increased scrutiny, and tighter observation.

The first system trialed in the minor leagues is known colloquially as the "robot umpire"; a complete automation of strike calling. The system, more formally known as the automatic ball-strike (ABS) system utilizes Hawk-Eye's 3-dimensional pitch tracking software to determine whether any part of the ball touches the strike zone at any point (Whitrock, 2021). The result is then instantaneously transmitted to umpires through an earpiece allowing for the flow of the game to not be significantly altered (Whitrock, 2021). Since its debut during the Single-A Atlantic League All Star Game in 2019, the ABS system has consistently been criticized by players and managers alike for not portraying a realistic strike zone, either too heavily favoring batters or pitchers (Castrovince, 2023). Various tweaks and adjustments have made significant improvements, however, players and coaches have not fully embraced the technological renaissance, instead calling for a sort of happy medium to not remove umpire tendencies and uniqueness entirely but allow for accountability in poor pitch calling.

In response to feedback from players and managers, Major League Baseball has also more recently begun testing on the "challenge system" to allow umpires to control strike calling while giving players the opportunity to challenge calls they disagree with. When initiated, a challenge sparks a replay review, where the Hawk-Eye pitch tracking technology is used to determine if the pitch was a ball or strike – exactly like the ABS system (Castrovince, 2023). Only the pitcher, batter, and catcher are eligible to issue challenges. The challenge system follows a three-strike rule: teams can challenge an unlimited number of times until they fail to do so successfully three times (Castrovince, 2023). The challenge system was extremely popular during Single-A's Florida State League and at Triple-A Charlotte in 2022, with players and managers calling for a further utilization of the challenge system.

In the 2023 season, Major League Baseball decided to implement both the challenge and ABS systems in all 30 Triple-A ballparks for the majority of the season, aiming to gather feedback. During each six-game series (also newly introduced), ABS was featured for the first three games, and challenges for the final three. In response to feedback throughout the season, both systems featured a slightly raised bottom of the strike zone for the final month of the season to avoid calls on the low and outside corner batters thought were too pitcher friendly. Utilizing the 2023 Triple-A season as the focus of our study, our paper analyzes umpire performance and tendencies under the following four treatment groups: ABS before the strike zone change, ABS after the strike zone change, challenge before the strike zone change, challenge after the strike zone change (in this early version of the study, we do not actually analyze the third and fourth treatment groups due to limited sample sizes; we will add this to our analysis in a later iteration). We attempted to answer the question if umpires change their strike calling tendencies with the ABS and challenge systems. Early results indicate that umpire tendencies did change after the implementation of the challenge and ABS systems. Results also indicate that umpire accuracy improved under increased supervision and feedback (challenge system).

Literature Review:

Increased availability of publicly available data has led to amplified scrutiny of umpires and analysis of umpire tendencies. A large publicly available source that analyzes umpire performance is the "Umpire Scorecards" – a website and Twitter page that posts daily updates on the performance of Major League Baseball umpires. Founded by Ethan Singer and Ethan Schwartz, the website uses publicly available pitch-by-pitch data to rate and rank umpire performance. Cumulative run expectancy values on incorrect calls are used to calculate the total run impact of an umpire each game, identifying which team gained an absolute advantage, or was "favored." They further their evaluation by estimating expected umpire call accuracy, visualizing estimated umpire strike zones, and calculating umpire consistency to call strikes within their personalized strike zone. These archived reports are available in a publicly available depository, allowing for significant strides in the accessibility of umpire feedback and evaluation.

Increased data has also allowed for more thorough review of umpire behaviors. Recent literature has found statistical evidence that umpire pitch calling has positively progressed as strike-zone technology improves (Mills, 2016). The improved technology has allowed for umpire feedback to be more precise and consistent, expediting skill development (Mills, 2016). This development is seen most prominently in younger umpires (Mills, 2016). It is expected that umpires will continue to slowly progress over time, becoming more and more accurate, as training and feedback continue to develop (Mills 2016). Understanding how technology has improved umpire pitch calling, our study directly assesses the impact of technological advances and increased pitch-level oversight of Triple-A umpire decisions, and how umpire strike calling reacts to the incorporation of technology. It also documents changing umpire behaviors that may be prevalent if either system reaches the Major League.

With the availability of pitch-by-pitch data becoming more accessible, the evaluation of umpire pitch calling tendencies has been explored in multiple settings with various intentions. Numerous studies implement generalized additive models to express the odds of a called strike for any given location around a strike zone (Mills, 2014; Deshpande & Wyner, 2017). These models have been used in the context of catcher framing methodology to account for umpire fixed effects (Mills, 2014; Deshpande & Wyner, 2017). The umpire fixed effects are then applied in different settings, most notably to account for umpire strike calling differences when incorporating pitcher and batter handedness (Deshpande & Wyner, 2017). Deshpande and Wyner (2017) apply these to a hierarchy model using a Bayesian logistic methodology. However, they designate an "average" strike zone using mean player heights for the top and bottoms. This creates major potential inaccuracies, especially when determining incorrect ball or strike calls. This is less of a concern in our paper, as recent developments in the data allow for individualized strike zone dimensions (Spencer, 2017). We do not include individual umpire effects due to limitations in available data.

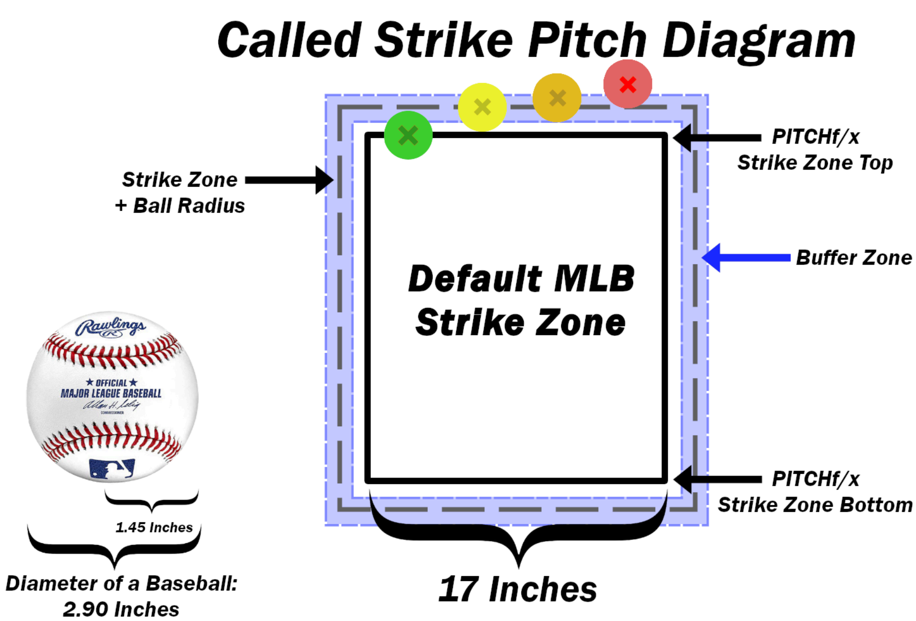
Data:

To conduct our analysis, pitch-level data for each 2023 Triple-A regular season game was obtained from the MLB API which includes pitch tracking data provided by Sportvisions PITCHf/x. Since the mid 2000's, PITCHf/x has provided fans the ability to track Major League pitch trajectories, velocities, and location through publicly available data and graphics. For the first time in 2023, PITCHf/x was fully adapted by Triple-A, a key factor as it contains individualized estimates for the top and bottom of a batter's strike zone, utilized to create a more accurate depiction of the strike zone at an at bat level. The recent integration of PITCHf/x brings with it limitations however, as we do not have reliable or comparable pitch-level data for previous Minor League seasons nor other Minor League levels in 2023 besides Triple-A. With use of the PITCHf/x data, we conducted a multi-treatment analysis of Triple-A umpires to analyze umpire strike calling accuracy and changes with the implementation of the ABS and challenge systems.

To conduct our multi-treatment analysis, we chose four distinct treatment groups to best fit each system and adjustments made throughout the season. Our control group consists of games in which both the challenge and ABS systems were not in use. MiLB implemented the challenge and ABS systems for the entirety of the Pacific Coast League (PCL) season, while the International League (IL) did not begin using them until April 25th. This gave us roughly a month of pitches (control, n = 60,583) without any newly introduced systems to use as a control for our data. We considered using MLB umpire performance as a control group, but 2023 comparable analysis provides statistically significant evidence that there are substantial accuracy differences between MLB and Triple-A umpires. With the remaining data, we created and categorized our four treatment groups: pitches from challenge system games (treatment 1, n = 269,463), pitches from ABS games (treatment 2, n = 250,231), pitches from challenge system with adjusted strike zone games (treatment 3, n = 42,474), and pitches from ABS with adjusted strike zone games (treatment 4, n = 41,539). With limited data for Treatment 3 and 4, we only consider the first two treatments in subsequent analysis.

In our models, we utilized a binary variable that indicated whether the call on the field was incorrect. To do so, our first step was to create an accurate definition and representation of the strike zone. Our strike zone slightly differs from the rulebook Major League Baseball strike zone to account for how pitch data are recorded. In *Diagram 1*, the rulebook MLB strike zone is defined as the black box, where the width is the set width of home plate (17 inches) and the top and bottom of the strike zone are fluidly defined as PITCHf/x's defined Strike Zone Top and Strike Zone Bottom variables, respectively. It is important to note that both of these variables vary from batter to batter depending on height and stance to match the official rulebook definition of the strike zone. Critically, we have to adjust the strike zone because PITCHf/x records each pitch's coordinates at the center of the ball. By definition, the ball needs to only touch the plate at some point while crossing, meaning the ball can graze the corner, and the center can be off the plate, but it is still by definition a strike. In *Diagram 1*, this is represented by the Pitch 1 (Green) and Pitch 2 (Yellow); both by definition are strikes, however, the data only records Pitch 1 (Green) as being in the strike zone. To account for this, the strike zone was adjusted by adding a ball radius in each direction so that Pitch 2 (Yellow) is correctly identified as a strike. That is represented by the gray dotted line in *Diagram 1* and known as the "adjusted strike zone." Also seen in *Diagram 1* is the measurements of a standard Major League Baseball. Currently, the average official game used baseballs have an average radius of 1.45 inches and a diameter of 2.90 inches. Adding two radii, one to each side of the plate, the width at which the center of the ball must be within to be considered a strike is 19.90 inches, or within 9.95 inches of the center of the plate in either direction.

Diagram 1: Representation of a Called Strike



The next challenge to consider is that Hawk-Eye has 3-Dimensional capabilities while the data PITCHf/x provides only allows for a two dimensional strike zone. To combat this, we include a "buffer zone" for umpires, that provides leniency against incorrect calls within half a ball radius (a fourth of the diameter of the baseball). In *Diagram 1*, that is depicted by the blue buffer zone surrounding the adjusted strike zone. What this buffer zone depicts is that if the center of the ball falls within the blue shaded area, it will not be considered incorrect, whether the call is a ball or strike. This can be seen in Pitch 3 (Orange) and Pitch 4 (Red). For Pitch 3 (Orange), the center of the baseball is beyond the adjusted strike zone and not touching the rulebook strike zone, therefore should be considered a ball. Because this is a 2-Dimensional zone, there is a great likelihood that despite not touching the default strike zone in the graphic, in a 3-Dimensional zone, it will hit the strike zone at some point. As a result, Pitch 3 is not considered an incorrect call, and since Pitch 3 is neither correct nor incorrect, it is excluded from further analysis. The center of Pitch 4 (Red), though, exceeds that buffer zone, and will be recorded as an incorrect call (if originally called a strike). This would leave our four pitches called strikes being recorded as the following: Pitch 1 (Green) is a correctly called strike, Pitch 2 (Yellow) is a correctly called strike with the adjusted strike zone, Pitch 3 (Orange) is not labeled correct or incorrect as the ball's center is in the buffer zone. Pitch 4 (Red) is an incorrect strike.

Utilizing the adjusted strike zone and buffer zone, we are able to determine whether a pitch is correctly or incorrectly called a strike on any given pitch. Based on this definition, out of 360,199 calls made by an umpire (either called balls or strikes) there were 11,290 incorrect calls over the course of the Triple-A season. This equates to a 3.134% incorrect call rate, with a vast majority of these being pitches far from the edge of the plate. Due to the buffer zone exclusion, 29,074 observations were excluded from the data set, accounting for roughly 8% of the data. Despite this being a loss of a sizable percentage of data, it significantly improved the accuracy of our model when predicting the probability of an incorrect call when using the ABS system (which theoretically should be zero or close to it) without sacrificing too large of a portion of our data. This was why we chose the buffer zone to span one half radius in each direction rather than a full radius in each direction.

Modeling Approach:

In modeling incorrect call probability, we include numerous control variables in addition to our treatments to account for extraneous variables that may affect umpire decision making. First, pitch movement was accounted for using PITCHf/x's horizontal and vertical movement parameters, as well as a binary indicator on whether the movement was armside or gloveside for a pitcher. Along with armside and gloveside movement, binary variables for pitcher and batter handedness cover the difference between pitch calling tendencies with righties and lefties on the mound and in the batter's box. Other factors such as umpire fatigue and weather were accounted for using an index for the at bat number in the game and a binary variable for night games.

Next, we incorporate variables that explicitly allow us to consider umpire behaviors. First, to account for asymmetry in strike-calling versus ball-calling behaviors, a binary variable was used to indicate the original call on the field (reverting challenged pitches) to indicate whether a pitch was called a strike or a ball. Second, we consider umpire behaviors at various bat defining counts, specifically 3-0, 0-2, and 3-2 counts, compared to all other counts. Each of these counts represent various extremes, best illustrating umpire behavioral differences. Third, we include the pitch's distance from both the horizontal and vertical edges of the strike zone. Intuitively, we would expect incorrect call rates to increase the closer the pitch is to a strike zone edge. In total, we consider the effect each of these factors has within each of our treatment groups.

Many of our predictors could have large interactive effects on the response with other predictors. For example, how far a pitch is from the zone very well may have a larger effect on the probability of it being called a strike given a 3-0 count than an 0-2 count. For this reason we modeled the probability of a pitch being called correctly with an xgboost model, due to the model's ability to detect multi-level interactions more efficiently than other parametric models (logit, for example) and semi or non parametric models (GAM). Furthermore, we evaluate the difference in umpire performance based on their accuracy of calling balls and strikes. Several studies, such as Dwidarma et al. (2021) have found xgboost models to outperform logistic regression models when comparing predictive accuracy.

To evaluate the performance of the umpires given the traditional umpiring, challenge, and ABS systems we fit an xgboost model and evaluate predictive performance of a ball or strike using misclassification rates. The data were randomly split into a 60% training group and 40% test group to fit the model. The initial fit of the model was run on 10 trees with a down weight of .3 for each tree in the xgboost model. This process was repeated for 5000 bootstrap samples for building and testing the models. The average, along with a 90% bootstrap confidence interval of the misclassification rate, are reported below.

Discussion:

To present and discuss our results, we calculate predicted incorrect call rates given various pitch locations, count, and treatments. *Figure 1* and *Figure 2* display predicted incorrect call rate of Triple-A umpires in each of the ABS, Challenge, and control (Ump) groups based on vertical and horizontal distance from the zone edge, respectively. *Figure 3* and *Figure 4*, analyze differences in projected percentage of incorrect calls for each system by both count and distance. *Figure 5* and *Figure 6* do the same, but present the predicted values in a different order. As our conclusions rely on observations from the different figures simultaneously, we organize our thoughts and repeatedly reference the various figures.

First, we note that in all figures the predicted incorrect call rate under the ABS system is nearly zero, which is what should be expected. Of course, exactly zero should not be expected, as even with our buffer zone we are still classifying based on a two-dimensional strike zone. The most notable inaccuracy rates occur within a half-ball of the plate.

Second, we analyze general umpire trends under the various systems. While not statistically significant, there is weak evidence to suggest that umpires are more accurate when the challenge system is used compared to when no system is used. In *Figure 1* and *Figure 2*, predicted inaccuracy rates are greater under the treatment group. Sample size issues for the control group make it difficult to draw definitive conclusions, however. Statistical significance can be seen in *Figure 5* and *Figure 6,* which expands on the previous two figures breaking down each distance by the four count parameters. In *Figure 5*, the trend continues at a count level where in all but one category, it is predicted to produce less inaccurate calls than the control group. Statistical significance can be seen in the "other" (far right) count treatment, consistenting of all pitches in non 3-0, 0-2, and 3-2 counts. With the increased sample size, smaller error bars allow for us to see statistical significance at six of seven distances that umpires are more accurate when using the challenge system than using no system at all. *Figure 6* produces the same results, where the challenge system consistently produces more accurate predictions than the umpire control group, and in the other count treatment group, four of seven contain statistical significance that umpires are predicted to make less incorrect calls.

Next, when viewing the accuracy improvement in the challenge system, evidence exists to suggest that it is as a result of umpires shrinking their strike zones to match that of the Hawk-Eye system. This can be seen most prominently in *Figure 1*, where half and full balls length out of the strike zone have the largest discrepancy in where mean predicted values for the Challenge treatment group are below the control group by roughly 3%. This difference is considerably larger than that of balls in, where there is near parity when the ball is vertically within or at half a ball's length from the strike zone. This can again be seen in *Figure 2,* where the discrepancy in half ball out is larger than that of very close and half ball in. This observation is supported within the distance versus treatment graphs of *Figure 5* and *Figure 6*. In both figures, the largest difference in predicted accuracy is in either the half-ball or ball's length outside of the strike zone, most prevalent in the other count treatment. In the half-ball out treatment, there is statistical significance of a difference amounting to a roughly 4% accuracy difference between the challenge and control groups. This is considerably larger than any other differences seen in *Figure 5* or *Figure 6*, as well as other counts, although, critically, this result is not statistically significant. While somewhat weak given the small treatment group sample sizes, these results support that initial conclusion that umpires have shrunk their strike zones, leading to more accuracy a half and full balls length outside of the zone, while marginal gains have been seen in other areas.

Another aspect when looking at a shift in umpire tendencies comes from analyzing differences in strike calling in different counts when using the different systems. Through our analysis, we found statistically significant evidence that umpires call pitches differently depending on the count. In both *Figure 3* and *Figure 4*, estimated incorrect call accuracy in 3-0 counts is statistically greater than that of 3-2, 0-2, and other counts when both very close and a half ball out of the strike zone, and at a full balls length out in *Figure 4*. The trend continues throughout the other distances, albeit with limited statistical significance. In 0-2 counts, significance can be seen but differently horizontally and vertically from the strike zone. In *Figure 3*, statistically significant evidence exists that umpires are more accurate horizontally in 0-2 counts at five of seven distances from all other counts. This differs slightly from *Figure 4*, where statistical significance exists that umpires are more accurate vertically in 0-2 counts at three of seven distances from other and 3-0 counts, however, it is not different from 3-2 counts. For 3-2 counts, umpires tend to be more accurate than both 3-0 counts and other counts with statistical significance, though, only seen in the very close and half-ball distances in *Figure 4*. In summary, these results suggest that umpires are statistically most inaccurate in 3-0 counts, followed by other (non 3-2 and 0-2 counts), with 3-2 counts trending to be slightly more accurate, and 0-2 statistically being the most accurate.

Using these conclusions an interesting correlation can be found between challenge frequency and predicted incorrect call percentage. *Table 1*, contains challenged pitches by count and are ordered by challenge frequency. The most challenged pitch, intuitively, should be the 3-2 pitch, as it defines an at bat no matter what if called by an umpire. This intuition is supported by the data, with the highest proportions of pitches being challenged in the 3-2 count (just over 3%). with a challenge percentage of just over 3% of the sample. The second and third most frequently challenged counts are 3-1 and 3-0 pitches. Despite the high challenge rate, umpires are relatively accurate for both batters (challenged strikes) and pitchers (challenged balls), with challenge success rates well below 50%, and challenged 3-2 balls being the least overturned challenged out of all counts and challenges. 3-0 pitches, however, show an opposite trend. Batters have seemingly picked up on the expanded and inaccurate zone in three ball counts, and have challenged at one of the highest frequencies, and at the highest overturn percentage of nearly 75% accuracy. This gives light into how batters and pitchers are seeing increased inaccuracies, especially in three ball counts, and are utilizing the challenge system to correctly adjust for umpire tendencies. This, of course, provides real-time direct feedback back to umpires, as if they see these calls being overturned at a higher percentage, it can allow them to adjust zones and become more accurate. On the opposite end of the table, 0-2 pitches are the least challenged pitches. As established earlier, umpires tend to call them most accurately, likely due to a significant amount of waste pitches. An interesting observation, however, is that challenged balls (by pitchers), are the second most overturned call. This, combined with a high accuracy of challenged strikes in 3-0 counts, point towards a trend of risk aversion from umpires, as in 3-0 counts we saw consistent inaccuracies outside the zone, and in 0-2 counts, we see a high probability of overturns on called balls. This claim is one we intend on diving into further in later iterations of this study.

Conclusion:

Our paper adds to the evolving landscape of umpire evaluation, with the focus on tendency changes due to technological advances. We found trends that indicated umpires were more accurate when calling games with the challenge system. Evidence also existed to show umpire strike calling tendencies have changed, with umpires shrinking their strike zones when utilizing the challenge system, thus increasing their accuracy. This may have been caused by direct umpire feedback from challenges, as well as more accountability for risk aversion tactics, however these claims must be further explored. Statistical significance was also found to show that umpire strike calling varies based on count, umpires being least accurate in 3-0 counts as compared to 0-2 counts.

As mentioned previously, a major limitation in this study is the relatively limited sample size for the control group. As MiLB pitch data were limited to 2023 Triple-A only, it was necessary to rely on a short window of no systems to build our control group. Prior season Triple-A data or other Minor League levels may serve as stronger control groups, conditional on data being available. Another limitation is the lack of umpire identifying information in the data, preventing us from controlling for umpire specific effects. Finally, the lack of a three-dimensional strike zone introduces random error to our analysis that makes it more difficult to identify statistically significant differences.

The results of our study are especially prevalent for MLB as it considers potential application to the majors. While not perfect under the challenge system, increased oversight led to Triple-A umpires improving their accuracy, especially in calling a tighter strike zone. In counts most likely to be scrutinized, 3-2 counts, umpires see the greatest increase in accuracy under the challenge system. In many respects, the challenge system may reflect a happy medium for fans, matching enthusiasm from coaches and players. We saw improved umpire performance, while still allowing for umpire-specific effects to impact the game, but creating the possibility of correction at the most critical junctures. This is in addition to the fun game-theoretical elements introduced by requiring one of the batter, pitcher, or catcher to call the challenge instead of the manager. More testing is needed, but early results seem especially promising.

Figure 1: Predicted Incorrect Call Percentage of Umpires by Treatment Vertically

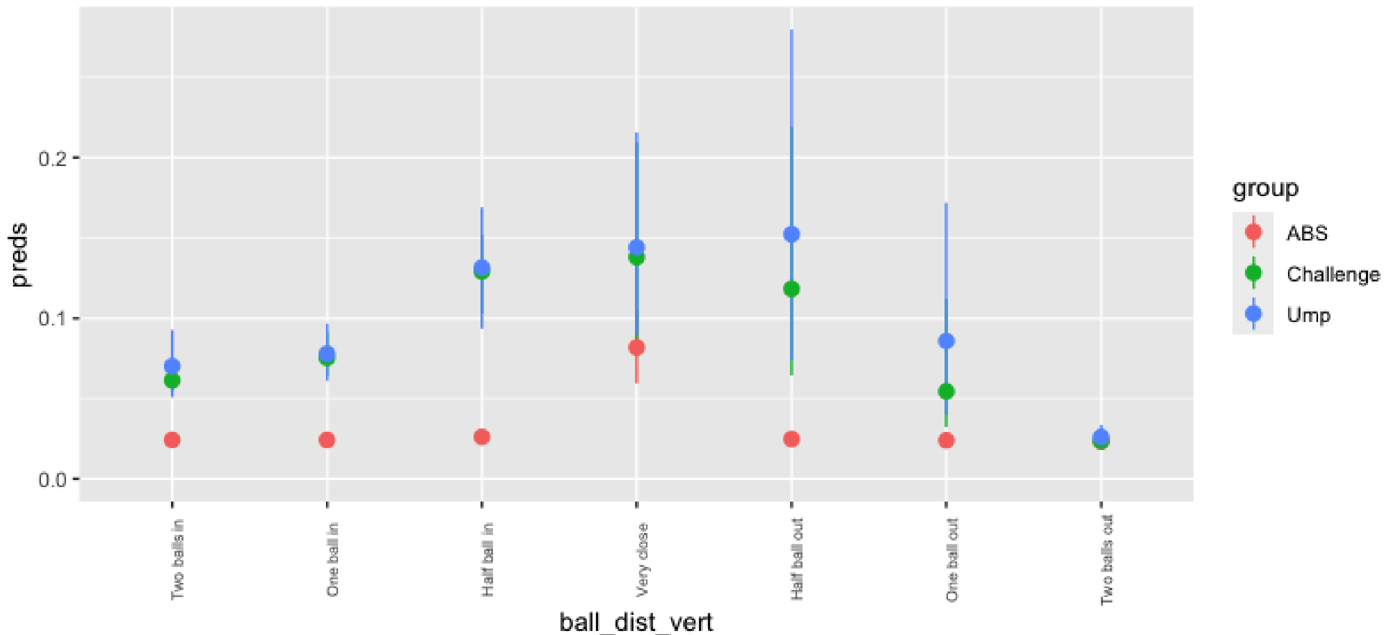


Figure 2: Predicted Incorrect Call Percentage of Umpires by Treatment Horizontally

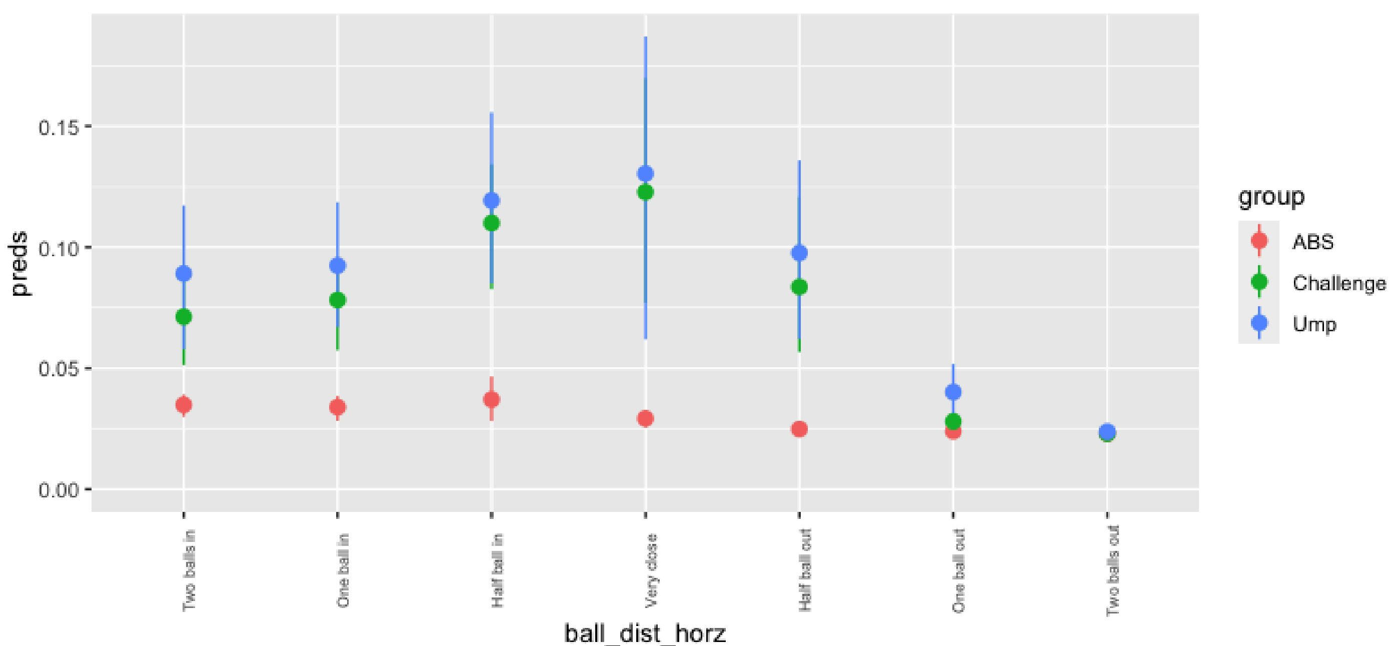


Figure 3: Treatment vs Count analysis of Predicted Incorrect Calls Horizontally

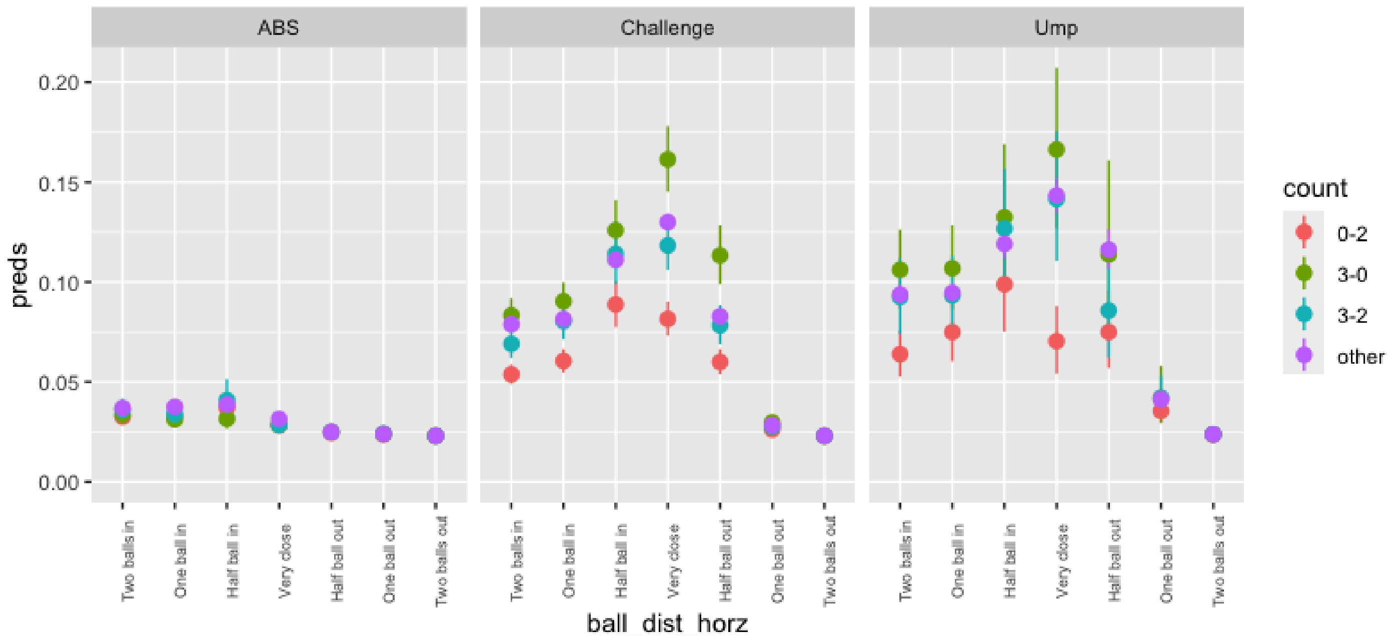


Figure 4: Treatment vs Count analysis of Predicted Incorrect Calls Vertically

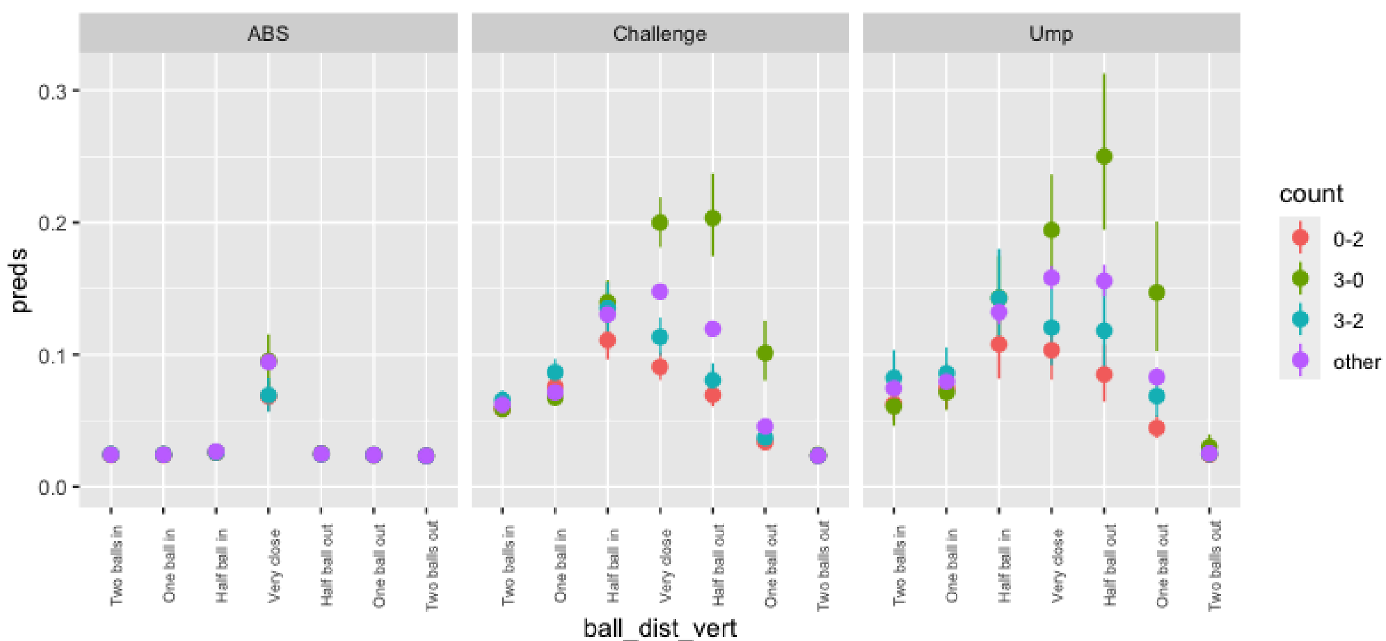


Figure 5: Distance vs Treatment analysis of Predicted Incorrect Calls Horizontally

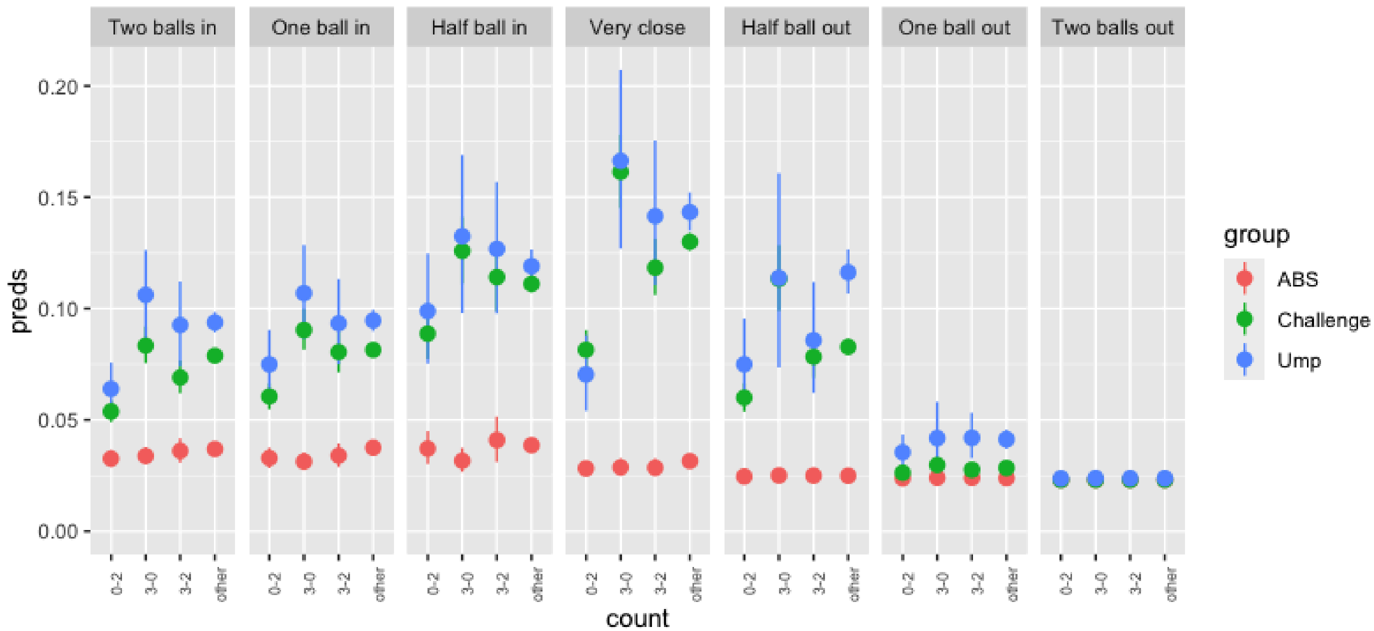


Figure 6: Distance vs Treatment analysis of Predicted Incorrect Calls Vertically

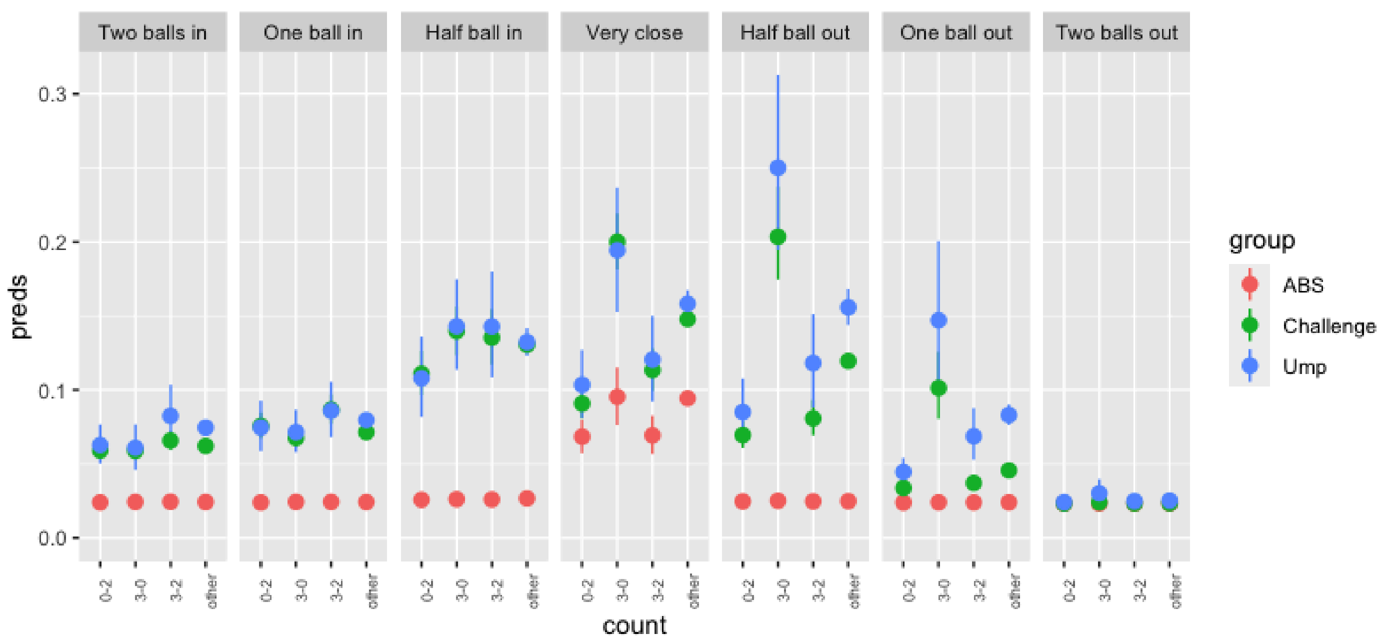


Table 1: Analysis of Pitch Challenges

Count	Challenges	Pitches	Challenge %	Challenged Ball %	Challenged Ball Accuracy	Challenged Strike %	Challenged Strike Accuracy
3-2	494	16122	3.06%	39.47%	31.28%	60.53%	41.81%
3-1	232	8445	2.75%	52.59%	45.90%	47.41%	53.64%
3-0	113	4665	2.42%	61.95%	38.57%	38.05%	74.42%
2-1	429	17745	2.42%	46.62%	43.50%	53.38%	46.29%
2-0	278	12853	2.16%	51.08%	38.03%	48.56%	50.37%
2-2	480	25107	1.91%	44.38%	39.44%	55.63%	38.95%
1-1	577	31508	1.83%	46.10%	44.36%	53.90%	43.09%
1-0	619	34263	1.81%	52.18%	49.85%	47.82%	54.39%
0-0	1416	81164	1.74%	53.81%	51.44%	46.12%	51.76%
0-1	623	37737	1.65%	45.43%	56.54%	54.57%	45.88%
1-2	425	28521	1.49%	45.41%	43.01%	54.59%	34.91%
0-2	280	19624	1.43%	45.36%	58.27%	54.64%	37.91%

References:

Castrovince, Anthony. "Ball-Strike Challenge System Could Someday Be a Game Changer." *MLB.Com*, MLB, 8 Feb. 2023, www.mlb.com/news/explaining-ball-strike-challenge-system.

Deshpande, Sameer K., and Abraham Wyner. "Journal of Quantitative Analysis in Sports: A Hierarchical Bayesian ..." A Hierarchical Bayesian Model of Pitch Framing, Journal of Quantitative Analysis in Sports, 2017, wsb.wharton.upenn.edu/wp-content/uploads/2020/02/Journal-of-Quantitative-Analysis-in-Sports-A-hierarchical-Bayesian-model-of-pitch-framing.pdf.

R. Dwidarma, S. D. Permai and J. Harefa, "Comparison of Logistic Regression and XGBoost for Predicting Potential Debtors," 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS), IPOH, Malaysia, 2021, pp. 1-6, doi: 10.1109/AiDAS53897.2021.9574350.

Mills, Brian . "Technological Innovations in Monitoring and Evaluation: Evidence of Performance Impacts among Major League Baseball Umpires." Labour Economics, North-Holland, 27 Oct. 2016, www.sciencedirect.com/science/article/pii/S0927537116302184?via%3Dihub#s0070.

Mills, Brian M. "Technological Innovations in Monitoring and Evaluation: Evidence of Performance Impacts among Major League Baseball Umpires." Labour Economics, North-Holland, 27 Oct. 2016, www.sciencedirect.com/science/article/pii/S0927537116302184.

Spencer, Scott. "Modeling Umpire Calls." Modeling Umpire Calls :: P(Ssp3nc3r | Columbian), 25 Apr. 2018, ssp3nc3r.github.io/post/modeling-umpire-calls/.

Whitrock, Matthew. "Developing MLB's Automated Ball/Strike System (ABS)." Medium, MLB Technology Blog, 14 May 2021, technology.mlblogs.com/developing-mlbs-automated-ball-strike-system-abs-d4f499deff31.