Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 2, 2024
A Regional Conference of the Society for Industrial and Systems Engineering

Using Data to Increase Air Force Baseball's Win Percentage for the 2024 Season

Mason Bugg, Jackson Cook, Andrew Schaeffer, Anthony Tonthat, Hayden Archer, Gerry Gonzalez, and Brian Lemay

Operations Research Capstone,
United States Air Force Academy,
Colorado Springs, CO 80841

Corresponding author's Email: andrewschaeffer2@gmail.com

Author Note: The cadet authors are all first-class cadets at the US Air Force Academy, partnering with Air Force Baseball. We thank our faculty advisors for their expertise in completing this capstone research analysis. The views expressed herein are those of the authors and do not reflect the position of the US Air Force Academy, the Department of the Air Force, or the Department of Defense.

Abstract: The goal of the United States Air Force Academy's Baseball team for the 2024 season is to win the Mountain West Conference. To help them achieve this goal, we employ analytics to more effectively use the team's resources. In this paper, we describe a variety of analytical methods used to improve the team's performance. Specifically, we explain our novel altitude conversion model for simulating lower altitude conditions during batting practice, player-specific training plans and a catcher selection model based on game situation and player statistics. The results of our work have enabled data-driven training and decisions for the coaches and players and are being utilized to improve performance in the 2024 season. The conclusions of our work show how Air Force Baseball can overcome limited resources to maximize their performance and compete for a Mountain West Conference Championship.

Keywords: baseball, analytics, sabermetrics, Trackman, optimization, operations research

1. Introduction

Over the past 20 years, analytics and operations research have transformed baseball. Coaches have increasingly relied on quantitative measures to inform coaching decisions and improve performance to ultimately increase their chances of winning. The objective of Air Force Baseball is to win baseball games at the NCAA Division 1 level and ultimately win the Mountain West Championship. To do this, they hope to leverage analytics to improve their performance subject to limited resources, thus increasing their ability to accomplish their goals. Therefore, the aim of this project is to provide analytical support to help increase Air Force's wins. There are three main analytical tools the project focuses on.

The first tool is used to mimic an opposing pitcher's ball movement and velocity at the lower altitudes of away games while training at the high altitude of the Air Force Academy. At higher elevations, the air is thinner. Because of this, pitches with identical spin rates break less at higher elevations than lower elevations. The coaching staff highlighted to us that the team has historically averaged more strikeouts and less hits in away games than when playing in their home stadium, as they are unaccustomed to pitch movement at sea level. This tool allows hitters to better prepare for away games while practicing at high altitude.

The second tool finds players' hitting archetypes and determines their strengths using individual player statistics. The purpose of this aspect of the project was to aid the coaching staff as they create offseason coaching plans. For example, if a player excels in hitting fastballs, then the coaching staff would have that player focus on hitting other types of pitches in the offseason. This area of the project will also prove useful for the coaching staff in coaching situations where they want a clear output from a given player. For example, if a certain game situation requires a player who makes contact at a high rate, these archetypes will inform which players have the best chance of success.

The third tool we developed is a performance metric for the catchers on staff. This tool provides a more informed decision for the coaching staff in selecting the catcher that improves the team's performance subject to various game situations. For example, if the coaches plan to have pitchers throw a high number of curveballs on the bottom corner of the strike zone, this metric will show which catcher on the team 'frames' these pitches the best (i.e., makes pitches that slightly miss the strike zone 'look' like strikes to the umpire).

In this paper, Section 2 describes the methodology of our research. This includes collecting data for mimicking opposing pitchers, developing a metric to evaluate catchers, categorizing players' archetypes (both batters and pitchers), and providing scouting reports for opponents. This section explains where we gathered data and how we manipulated it to solve the problem. Section 3 describes the results and analysis of our research. This section analyzes the collected data using the methodology to create usable tools, scouting reports, and detailed analyses for the client. Additionally, the results section highlights key findings and visuals to help explain the significance of our results. Section 4 contains our conclusion, recommendations, and ideas for future research. This includes the significance of the research, our practical recommendations for the client, and suggested future research on other areas to improve for the client. Section 5 provides a list of references mentioned or utilized throughout the report.

1.1 Problem Statement

Entering the 2024 season, Air Force Baseball's goal is to win the Mountain West Conference Championship. Due to high elevation (players practice with decreased pitch drop at home compared to on the road), current inability to assess catcher performance, and no objective method to quantify and compare hitter characteristics, this research will help the team overcome these challenges and achieve their goal.

1.2 Related Work

Several resources informed our work and provided a precedent to our methodology. Existing research aided our understanding of pitch movement, optimal roster building, and advanced batting statistics.

First, Nagami (2013) showed that pitch motion is affected by three main components: gravity, air resistance, and the magnus effect (a lift force on a spinning object through the air). This information informed our pitching machine conversion model, as it allowed us to determine which pitch movement characteristics were most important to adjust for during batting practice based on which pitch characteristics are most relevant to a batter's perception of the pitch. Therefore, we can aim to replicate these in our pitching machine conversion to increase batter familiarity.

To inform our catcher selection model, we referenced Muniz's (2023) work on balanced roster-building decisions. To find the optimal selection of a catcher for a given game, sensitivity analysis must be performed on their offensive and defensive contributions. However, these contributions vary based on a variety of game-level factors, such as pitcher handedness, score, lineup position, etc. For example, a catcher that excels with left-handed pitchers should be chosen when a left-hander is substituted into the game. Therefore, this study is applicable to determine which Air Force catcher should be playing at any given time based on their overall contribution to win probability.

To better quantify player archetypes and categorize their strengths and weaknesses, we employed the work from Baumer's (2014) study on baseball analytics. This research recommends analyzing advanced game-level performance data to determine which aspects of offense a player should work on the most. For example, a player may be considered 'good' if he can make good contact with pitches at a high rate when batting. However, he could further improve performance by increasing his power output. Therefore, this research recommends that he spends most of his practice time increasing power. This article showed us which in-game statistics to evaluate to gain these insights.

2. Methodology

2.1 Replication of Opponents' Pitches with Pitching Machine

The first step in solving this problem was to find the average pitch characteristics of opposing pitchers at lower elevations than the team's training facilities so that Air Force players could practice effectively against similar-looking pitches. The averages were calculated by gathering data from the opposing team's previous season. Trackman, a data collection device placed above home plate, provides access to data from all teams' previous seasons, allowing us to get accurate data for each pitcher. Then, we combined the data, took the averages of each opposing pitcher by pitch type, and formulated a database with all the calculated values.

We selected four pitchers to analyze based on two criteria: first, they must have pitched over 20 innings last season. This means they played regularly last season, thus increasing the likelihood they would receive playing time against Air Force during the 2024 season. Second, they must be planning to return to play at a school that Air Force will play at an away game during the 2024 season. For each of these pitchers, we used the Trackman data collection device to find average statistics for each pitch type they threw in a 2023 game at sea level, thus giving an overview of what their pitches look like.

Based on correspondence with the Air Force coaching staff and related work from the literature review, several pitch characteristics are important when attempting to replicate a pitch (Nagami, 2013). These characteristics include velocity, horizontal break, vertical break, and induced vertical break. Velocity is measured in miles per hour (mph), and horizontal break, vertical break, and induced vertical break are measured in inches. If our team could replicate each of these pitch characteristics, it would provide Air Force hitters with an accurate simulation of what an opposing pitcher's pitches will look like regardless of elevation, thus providing them an advantage when preparing for away games at lower elevations.

We aimed to replicate the pitch characteristics collected from opposing pitchers using the pitching machine at the altitude of the Air Force Academy. To do this, we created a database of pitch characteristics based on different pitching machine settings. The pitching machine used by Air Force Baseball has three wheels that spin to pitch baseballs: a left wheel, right wheel, and bottom wheel.

To create this database, we created a spreadsheet that included an entry for each combination of wheel settings we planned to test. After narrowing down the sample to relevant combinations (based on realistic velocity and break), this gave us 265 data points to collect. We threw two pitches out of the machine at each setting to take an average of each statistic. Using two pitches allowed us to control for variance between pitches of the same setting and finish the data collection for the tool while constrained for time given the team's schedule. The statistical output of the machine-pitched balls will not be exact on every pitch, but taking the average of two pitches approximates what a batter can expect to see at the plate. Therefore, we threw 530 pitches from the pitching machine. The Trackman device automatically collected the relevant statistics for each pitch.

Before the model could be built, a difference measure was designed. We created a difference measure to be applied to each wheel setting using the aggregate difference of the relevant statistics compared to the pitch type in question. By minimizing the difference measure, we were able to find the most similar pitching machine setting to a given pitcher's pitch. The absolute value of each statistic is used to account for horizontal break being measured as left or right (negative values for left), and vertical breaks being measured as up or down (negative values for down). To account for different scales, each difference is normalized by dividing by the analyzed pitcher's respective value. These normalized values are multiplied by weight of each statistic measured, giving the coaching staff more control over which aspects of the pitch they want to most accurately replicate. With a weighting system, higher weights will more strongly penalize differences in each statistic. This equation is shown below in Equation 1:

$$\begin{aligned}&\left(Vertical\ Break\ Weight\ x\ \left|\frac{Vertical\ Break\ Difference}{Pitcher\ Vertical\ Break}\right|\right)+\\&\left(Induced\ Vertical\ Break\ Weight\ x\ \left|\frac{Induced\ Vertical\ Break\ Difference}{Pitcher\ Induced\ Vertical\ Break}\right|\right)+\\&\left(Horizontal\ Break\ Weight\ x\ \left|\frac{Horizontal\ Break\ Difference}{Pitcher\ Horizontal\ Break}\right|\right)+\ \left(Velocity\ Weight\ x\ \left|\frac{Velocity\ Difference}{Pitcher\ Velocity}\right|\right)\end{aligned} \tag{1}$$

After populating our database with the relevant characteristics for all 265 pitch combinations, we used the data and our likeness measure to build a tool that finds which wheel setting combination most closely replicated each pitch type for a given pitcher. The first step was to input the pitcher's statistics to be replicated into the tool's input sheet. The tool has separate sheets with the same database and likeness measure, but each sheet calculates the likeness measure for a given pitch type. Using each separate sheet the tool denotes where the likeness measure is minimized for each pitch type. Because unique combinations of wheel settings produce different pitch statistics, it was not feasible to test every setting the pitching machine could achieve. To remedy this, we focused on the whole number settings between three and nine, and interpolated metrics for intermediate wheel settings using a linear average. For example, if we had settings 3, 3, 5 and 3, 3, 6, we averaged the statistics measured for these two settings to find 3, 3, 5.5. This increased the number of pitches in our dataset to 1765 pitch settings. We tested these intermediate values before presenting the pitching machine tool to the coaching staff, and the differences between interpolated intermediate values and those same tested values from the pitching machine were negligible.

Minimizing Equation 1 for each pitch type allows for a translation of a desired pitcher's pitch characteristics to the wheel settings on the pitching machine that will accomplish the most similar characteristics. Returning to the input sheet, the tool compiles each wheel setting most like each pitch type of the replicated pitcher.

2.2 Categorizing Hitters

To categorize the players' hitting ability and identify their strengths and weaknesses, we focused on three main characteristics hitting power, contact ability, and plate discipline. We assigned the players scores for these three main categories by taking weighted averages of a variety of hitting statistics from the previous spring and fall seasons that relate to each category. If it is better to have a low value for a certain statistic, then we made the value from that statistic negative when calculating the weighted average for the category score. Descriptions for all of these hitting statistics are shown below in Table 1. The power score was composed 20% from the player's barrel percentage, 30% from their slugging percentage, and 50% from their exit velocity. The contact score was composed from 30% of their swinging strikeout rate, 30% of their contact rate, and 40% of their barrel percentage. The plate discipline was composed 40% from the player's chase rate, 40% from their strikeout looking rate, and 20% from their percentage of time in a hitter's count. These percentages were determined based on correspondence with the coaching staff.

Table 1. Description of Hitting Statistics used for Hitter Categorization

Statistic	**Description**
Barrel Percentage	Percentage of batted balls with an exit velocity of 95mph or faster and a launch angle between 10 and 35 degrees.
Exit Velocity	The speed of a batted ball (mph).
Slugging Percentage	Average bases per at-bat.
Chase Rate	The percentage of pitches out of the strike zone a batter swings at.
Looking Strikeout Rate	Percentage of time striking out on a called strike (i.e. pitch not swung at).
Time in Hitter's Count	The percentage of time batter has a 2-0, 3-0, 2-1, 3-1, 1-0 count (advantage for batter).
Contact Rate	Total pitches where contact was made divided by the total swings.
Swinging Strikeout Rate	Percentage of time striking out on a swinging strike.

Both 2023 spring season and 2023 fall season data were used, as this allowed us to calculate scores for incoming freshmen, increasing sample size for returning players, thus making our results more reliable. To account for the number of plate appearances in each season, statistics were multiplied by the amount of plate appearances that made up that statistic (for example, a player with a 100% swinging strikeout rate in the fall isn't penalized significantly if he only had one fall plate appearance). This allowed us to use all available data for each player without penalizing them for a small sample during a specific season. Because pitcher handedness can have a large effect on a batter's performance, we assigned two different scores for each of the three main categories based on facing right-handed pitchers or left-handed pitchers. After assigning each player a power, contact, and plate discipline score for each pitcher handedness, we normalized the scores based on the team's best and worst performer in each category. Next, we created player rankings for each main category split by pitcher handedness using the players' normalized scores. Additionally, we created an overall ranking by averaging the score in each category for a player against both pitcher hands separately. This was done by combining each player's two handedness scores, weighing by how many their percentage of plate appearances against each hand (for example, if a player hit poorly against left-handed pitchers and well against right-handed pitchers, and 95% of their plate appearances were against right-handed pitchers, their left-handed score will not penalize them by much). These results are shown in Figure 1 below for the hitters and coaches to analyze further.

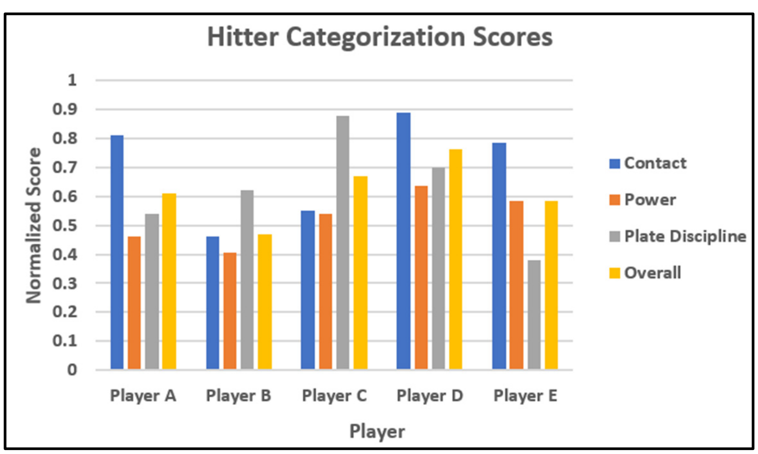

Figure 1: Example Hitter Categorization Scores for Top Five Hitters

ISBN: 97819384962-4-0

2.3 Catcher Framing

Framing refers to a catcher's ability to make pitches that slightly miss the strike zone 'look' like strikes to the umpire. A catcher that is skilled at framing helps their team by effectively increasing the size of the strike zone and helping their pitcher get more called strikes than they otherwise would have. Therefore, framing is an important defensive skill that we sought to quantify. To do this, we found four percentages for each of the catchers on the team for each pitch type: percentage of strikes called in the strike zone, percentage of strikes called on pitches one baseball size out of the strike zone (1BO), percentage of strikes called on pitches two baseball sizes out of the strike zone (2BO), and percentage of strikes called on pitches further than two baseball sizes out of the strike zone. Differentiations are made between each zone because umpires will tend to call strikes in the 1BO and 2BO zones if the catcher frames those pitches effectively.

To accomplish this, we used data on each pitch thrown by Air Force pitchers during the 2023 fall season. We began by sorting to find only the pitches in which the batter did not swing, as a swing would not provide insight on pitch location and/or whether the pitch would have been called a strike. Then, we found coordinates for the horizontal and vertical dimensions of the strike zone, the 1BO zone, and the 2BO zone. Because these dimensions can vary slightly based on batter height, we used average coordinates across all batters.

We then categorized each pitch by which zone it was in by comparing the actual vertical and horizontal pitch coordinates to the dimensions shown above. This was done using the Trackman device, which records the exact true pitch location. After categorizing each pitch into one of the four categories (including outside of 2BO zone), we found if each pitch in the dataset was called a strike by the umpire. Then, we sorted by catcher and pitch type. This showed us the percentage of strikes that each catcher got called by the umpire for each pitch type in each zone. These results are shown in both table and graphical form in Figure 2 below.

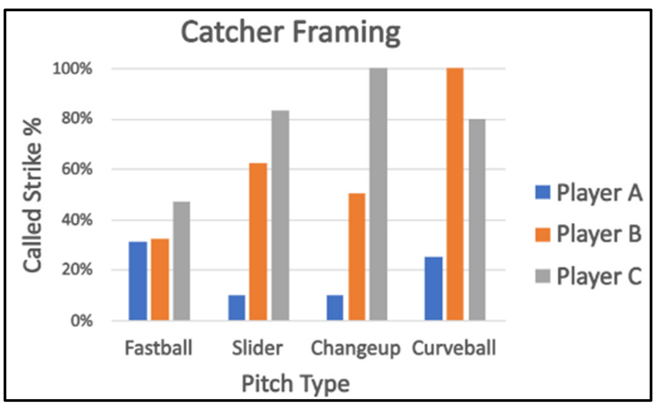

Figure 2: Framing Results by Catcher and Pitch Type

Figure 2 reveals that Player A has the worst framing ability of the three catchers by a large margin because he has the lowest called strike percentage for each of the main pitch types. Player C frames fastballs, sliders, and changeups the best, while player B is the most effective at framing curveballs.

3. Results and Analysis

The results of our analysis are limited because Air Force has only played 19 games so far in the 2024 season. However, as of 19 March 2024, Air Force is in first place in the Mountain West conference and has improved away-game performance compared to last season, suggesting that our models may be effective.

4. Conclusions, Recommendations, and Future Research

Each tool's methodology, analysis and results give insight to the coaches on their decision-making and offseason training plans. The pitching machine tool replicates pitches at away-game elevations for when the team travels to states such

ISBN: 97819384962-4-0

as Texas, California or North Carolina. Hitter categorization shows the coaches each player's strengths and weaknesses for creating an optimal lineup and how each player can improve in the offseason. Catcher evaluation informs the coaching staff on which catcher has the highest defensive performance, who the starter should be, and who would be the best substitute for games throughout the week. The most important aspect of these tools is that they can be utilized into the foreseeable future.

The pitching machine altitude conversion model provided us with several important insights. It allows the team to effectively practice against sea level pitches while at the altitude of the Air Force Academy to improve hitting statistics on the road. Throughout formulating the pitching machine altitude conversion model, we learned several important lessons. First, collecting data inside allowed us to avoid weather confounding our results. This led to more reliable pitch replications. Second, we realized the need for intermediate values. Intermediate values allowed us to improve our results without unrealistically increasing our data collection requirements. After spot-checking these interpolated values, we found the difference in our predicted values and pitched values to be negligible. Potential future work for this section includes a higher number of collected setting statistics, or a model that accounts for elevation differences and predicts how the pitching machine will perform at any elevation—not just the elevation of the Air Force Academy. Additionally, a model could be formulated to perform the inverse of the current model by predicting how an away pitcher's pitches will look at altitude. This would help the team improve offensive performance at home.

The hitter categorization scores provided us with several valuable insights. First, for each player we can see more specifically what their strengths and weaknesses are. Although the coaches may know a player is a good hitter, we might not understand exactly why they are a good hitter. However, these results can provide the team with such insights. Similarly, we can understand what skills a player should work on to improve. For example, we knew before the project that Player L is one of the best hitters on the team. This analysis shows us that his power trait is what most significantly contributes to this success (his slugging percentage, barrel percentage, and exit velocity are all among the highest on the team). However, he could improve his hitting skills even further by improving in the contact and plate discipline characteristics as he is ranked lower on the team in these characteristics. In other words, by focusing on improving his chase rate, working the pitch count, and making contact more often, he could improve to be an even better hitter. The coaching staff can use insights like these to create training plans for each player on the team to maximize their strengths and mitigate their weaknesses.

The catcher framing scores allowed us to quantify which catchers help pitchers throw more strikes. Based on the composite framing score and the two-baseball-outside performance score, we concluded that Player C is the best defensive catcher on the roster. In this area of the strike zone, this player is able to make more pitches appear to be strikes than the other two catchers, which gives pitchers a measurable advantage over the batter. However, these results also show us zones and pitches in which each catcher can improve. For example, even though Player C frames effectively, he lags behind his teammates when he frames a curveball. This means that he could practice receiving these pitches more often in practice to improve his strike percentage rate even further.

Our work is currently being used by the AF Baseball team in their 2024 season and we plan to compare the team's performance to previous years as the season progresses.

5. References

Baumer, B. and A. Zimbalist. (2014). The Sabermetric Revolution: Assessing the Growth of Analytics in Baseball. Philadelphia, PA: University of Pennsylvania Press. 10.9783/9780812209129

Muniz, M., & Flamand, T. (2023). Sports analytics for balanced team-building decisions. Journal of the Operational Research Society, 74(8), 1892-1909.

Tomoyuki Nagami, Takatoshi Higuchi, Kazuyuki Kanosue, (2013), How Baseball Spin Influences the Performance of a Pitcher, The Journal of Physical Fitness and Sports Medicine, Volume 2, Issue 1, Pages 63-68, Released on J-STAGE April 08, 2013, Online ISSN 2186-8123, Print ISSN 2186-8131, https://doi.org/10.7600/jpfsm.2.63, https://www.jstage.jst.go.jp/article/jpfsm/2/1/2_63/_article/-char/en.