MLB Offensive Projections: Facilitating Competitive Advantage Using Public Resources

Jenny Goldsher
Georgia Tech Applied Practicum
Employee Sponsor: Philadelphia Phillies
November 2024

The present project aims to thoroughly evaluate publicly available MLB offensive projection systems and use insights from this analysis to direct a novel implementation using public resources. The understanding of the value of public resources, and where they may systematically fall short in encapsulating player value, is critical in many areas of Baseball Operations. Establishing a framework to analyze public resources and establish internal benchmarks for proprietary model performance have an array of applications within the field. Being able to effectively utilize public information while understanding its potential shortcomings can allow teams to leverage information and efficiently manage internal resources.

Background

Over the last few decades, the magnitude of both publicly available baseball data and proprietary team data has grown exponentially. This has presented teams with the novel and dynamic challenge of utilizing information in a competitive way. All 30 MLB organizations have employees dedicated to analyzing information and deriving insights to support their leadership as they make impactful organizational decisions. As time passes, the margin for competitive advantage continues to narrow as teams develop their own novel implementations of forecasting tools.

While it is critical that teams have a high baseline of access to reliable projections, in various domains teams are approaching the ceiling of predictive accuracy after years of research. Furthermore, while it is important to have quality information, there is an upper threshold of quality of predictive tools. Once a reasonable threshold of reliability has been hit, another way to achieve competitive advantage is to maintain predictive validity with less resources. This has highlighted the importance of efficiently managing employee resources and leveraging available information to diminish necessary internal resources. As it relates to the present project, utilizing public resources rather than developing a model from scratch is a way to lessen the internal resources needed to produce quality information.

Public Resources

Various public resources provide open-source baseball statistics and player projections. A popular resource, Baseball Reference, provides a vast set of historical baseball data ranging from individual player statistics to team data. This data is all in the realm of observed performance data, and while relatively exhaustive in nature, does not aim to provide any type of forecasting data. Many of the individual statistics housed by Baseball References are sub-components needed to compute more advanced baseball statistics (Baseball Reference).

In a similar vein, Baseball Savant is another popular resource that houses data and provides insights based on observed performance data. Like Baseball Reference, the majority of the content on Baseball Savant is driven by observed baseball data. In contrast, Baseball Savant focuses heavily on providing visualizations and leaderboards to provide a basis for inference based on these statistics. In many cases, this involves standardizing statistics relative to MLB average and presenting the information in ways that articulate how a player's performance was relative to the league. Still, while the information of Baseball Savant is presented in a way that is intended to be inferential in nature, they do not overtly provide future projections (Baseball Savant).

Fangraphs, a resource that has been available and evolving for the past two decades, is a well-known and utilized public resource that does provide multiple future projection systems. Similar to Baseball Reference, Fangraphs provides observed player and team data, though their content is more focused on advanced statistics. For example, they tend to prominently and by default display advanced statistics rather than traditional observed data. In terms of their future projections systems, they provide implementations in many realms including but not limited to offensive player projections, individual player WAR (Wins Above Replacement) projections, team-level Division winning and World Series odds, and team-level win/loss projections. In contrast to the other resources, the information on their site is intended to be both inferential and predictive in nature (Fangraphs Baseball).

Specific to their MLB offensive projections, Fangraphs provides 7 distinct projection sets (Fangraphs Baseball). Each set of projections provides predictions for the coming season for a predetermined set of MLB players who are expected to hit a minimum threshold of MLB plate appearances. All seven resources provide predictions for the statistics we are interested in investigating: wRC+, wOBA, and OPS. As a part of this investigation, I will detail the possible statistics of interest and articulate why we landed on looking at these 3 specifically.

Shortcomings of Historical Baseball Statistics

Traditional baseball statistics have been a large part of the game since its onset. These traditional stats were originally rooted in game box scores and statistics tables on the back of baseball cards. In most cases, the goal of this information was to provide a way to quantify a player's performance in a simple and digestible way. Key bulk statistics include home runs (HR), doubles, triples, singles, and walks, which track a player's ability to get on base and produce extra-base hits. "Counting statistics" on their own hardly articulate offensive performance though, as they are directly related to the number of opportunities a player has.

The most fundamental attempts to account for opportunity and transform these counting statistics into rate-based statistics are batting average (AVG), on-base percentage (OBP), and slugging percentage (SLG). Batting average is the proportion of a player's at-bats that result in a hit. OBP goes one step further by including walks and hit-by-pitch, better encapsulating a player's ability to consistently reach base. Slugging percentage aims to quantify the value of extra-base hits, dividing the total number of bases obtained by the number of plate appearances.

While these traditional rate statistics do a better job of capturing a player's overall offensive performance compared to raw counting stats, they still have significant limitations. Most prominently, none of these rate-based statistics accurately quantify the run value of a given event. They don't answer the question: *How much more or less valuable is a particular plate appearance outcome relative to others?* SLG is the only statistic that attempts to assign different weights to events. However, SLG is interval in nature, meaning it treats a double as exactly twice as valuable as a single/walk, a triple three times as valuable, and a home run four times as valuable. Historical game data clearly articulates that these neatly defined weights are not accurate depictions of the game, which was a fundamental thesis in the original push for "Sabermetrics" which I will outline in a future section.

Additionally, traditional rate-based statistics do not attempt to account for park factors. Park factors aim to encapsulate how each ballpark's dimensions and conditions influence players' offensive production. Another common way to reference this concept is in relation to the park itself, where experts may cite that a certain park has a different "run environment". For example, Coors Field (Denver, CO) has been shown to have a "strong" run-environment, meaning there are typically more runs scored in games played there, likely due to the altitude.

The failure to consider park factors in traditional rate-based statistics make comparisons between players less reliable as they may play in vastly different run environments over the course of a season.

Finally, these traditional statistics do not have adjustments for league tendencies and are not standardized in any way. Throughout baseball history, offensive performance has fluctuated in response to various environmental factors. Two notable examples are the "Dead Ball Era" (1900-1919), when scoring was significantly lower, and the massive spike in offensive production during the late 1990s and early 2000s, which was closely tied to widespread steroid use. These league-wide trends are important to consider when trying to understand a player's performance in context. Since the traditional rate-based statistics do not have a way of factoring in league tendencies, it is difficult to compare player’s performances across different seasons.

These limitations have led to the development of more advanced metrics that aim to better quantify player value. The proper quantification of the run value of different events, adjustment for park effects, and normalization for league offensive environments, have led to advanced statistics that allow for more meaningful comparisons across players, eras, and contexts.

Advanced Baseball Statistics and the Sabermetrics Movement

Sabermetrics refers to the strategic use of statistical analysis to better understand the game and encapsulate player performance. The Society for American Baseball Research (SABR) was founded in the early 1970s and has been a major force in driving innovation in the field of baseball analysis. The Sabermetrics movement sparked the creation of many novel statistics that addressed the limitations of traditional counting and rate-based statistics (Society for American Baseball Research).

Namely, two of the main statistics we will look at in this investigation are a result of the Sabermetrics movement: wOBA and wRC+. Weighted On-Base Average (wOBA) aims to provide a more accurate measure of a player’s offensive value by assigning specific weights to each possible plate-appearance outcome. The weights used in wOBA are derived from historical data and represent the true value of each outcome in terms of generating runs. The scale the number is presented on is similar to that of OPB or AVG, where the numbers are a fraction of 1 rather than on a 100-scale. Presenting wOBA in this form provides familiarity to those who have had exposure to AVG and OBP historically, making it valuable within the field. Below is a table of the relative weights of plate appearance outcomes derived from observed game data (Fangraphs Baseball). These numbers may vary slightly year-to-year based on the league run-environment, but they tend to be within .01-.02 of these figures:

Plate Appearance Outcome	Weight
Walk	.69
Hit By Pitch	.72
Single	.89
Double	1.27
Triple	1.62
Homerun	2.10

Weighted runs-created plus (wRC +) is an advanced statistics that weighs plate-appearance outcomes in terms of relative run-value, scales to league average, and adjusts for park factors. The foundation of the wRC+ formula utilizes the same plate-appearance weights as wOBA but goes a few steps further to overtly include park factors and standardization to MLB average. To scale to league average, the outputs are in integer form where 100 is MLB average and each point +/- is 1% better/worse than MLB average. The statistic has become incredibly popular within baseball analytics, and recently has been included more prominently in fan-facing settings such as broadcasts and stadium videoboards. As the exposure to the statistic increases, users have become more accustomed to understanding the 100-scale. Additionally, the inherent calibration to MLB average is incredibly valuable in accounting for changes in the run-environment across seasons.

So, why investigate OPS?

OPS (On-Base Plus Slugging) is a combined metric that combines two traditional rate metrics—on-base percentage (OBP) and slugging percentage (SLG)—by simply adding them together. While it is methodologically flawed to sum these two statistics (one point of OBP is not directly equal in value to one point in SLG), OPS nonetheless provides a useful overview of offensive player performance. By combining a player's ability to get on base (OBP) with their power (SLG), OPS captures multiple important aspects of offensive production in a somewhat simple way. In addition, OPS is highly correlated to wOBA (*corr=.989*) validating its relevance to our other target variables.

But still, why include the analysis of a statistic we know has limitations? The answer is rooted in what I believe to be the most important part of baseball analytics: effective communication of information. While OPS has methodological limitations that certain advanced statistics do not, it has the key advantage of being far more familiar to most stakeholders and people within the field. It is important when considering this tradeoff to present the caveats along with the information itself, but I believe the communication upside outweighs the potential danger. Additionally, since OPS will be presented alongside the advanced metrics, the three statistics can be interpreted in-tandem to holistically represent player production.

Application of Offensive Projections in Baseball Operations

Offensive player projections have an array of applications throughout Baseball Operations. The most prominent applications are within the sphere of player evaluation, where offensive projections can influence which free agent players to target, which players to target in trades, and which players within an organization to promote. Having reliable offensive projections is crucial in addressing the value of a player and providing information necessary to make important decisions.

Within an organization, having reliable offensive projections can help player development staff determine whether a player may be ready to be promoted to a higher level. While this decision is broader than just the projection itself, having reliable projections can aid staff in their forecasting of players' future performance. Additionally, within a roster, offensive projections can serve as a tool to the manager to construct their lineup. Managers aim to create a batting order that will generate the most runs per game, and leaning on offensive projections can help them envision how to accomplish this in the context of a 162-game season.

Throughout all the possible applications, it is critical that the offensive projections provided have a meaningful degree of reliability and are presented in a way that highlights their strengths and weaknesses. Providing decision makers with both quality information and necessary context empowers them to make the best decisions possible in critical situations. Being cognizant of audience, and which statistics to present and how to present them is imperative in making meaning of analysis.

Methodology

Population of interest

When making comparisons between model projections and actual player performance, the present project focuses on players who had at least 400 plate appearances in the 2024 MLB season. While this is not quite the number to have a "qualified" MLB season (3.1 PA per game), 400 plate appearances is enough to establish a reasonable sample to reflect a player's true talent. In 2024, 207 players had at least 400 MLB plate appearances.

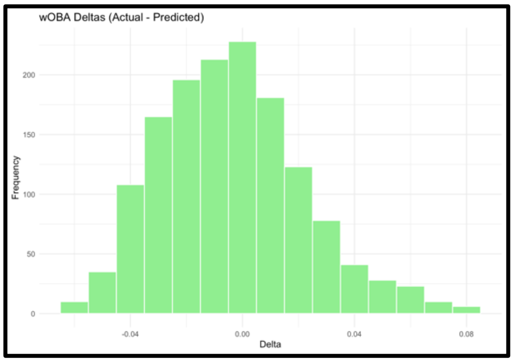
Anonymization

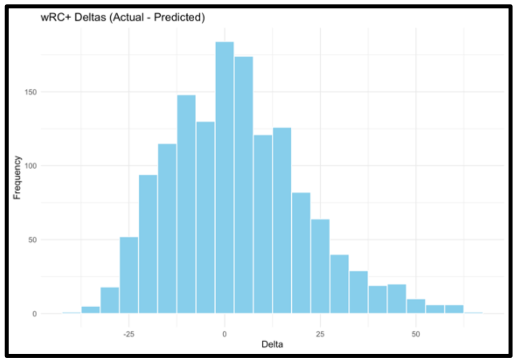
To maintain necessary discretion for proprietary analysis and intellectual property, the projection systems have been randomly assigned "System ID's" rather than their names. In addition, the visuals and aggregation of data for this report has been conducted with this goal in mind.

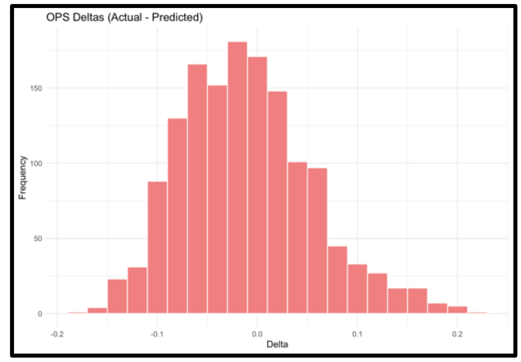
Analysis of Public Projection Systems

To gain insights necessary to conduct a novel projection framework, I conducted an in-depth analysis of the 7 existing projection systems. Firstly, across all projection systems I looked at the delta between all predictions and the actual values for the 3 variables of interest. I wanted to verify that the distribution of these predictions were balanced, and that there was no concern that the projection systems were consistently over/under predicting any of the target variables.

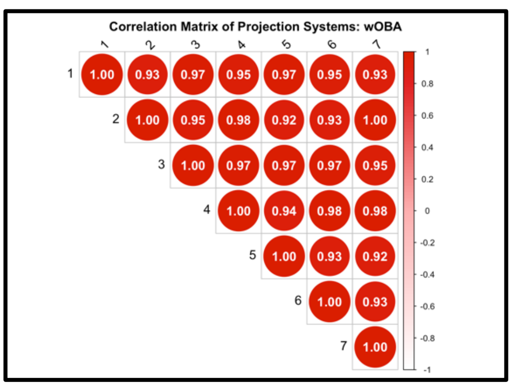
All three histograms verify that the aggregate of the projection systems yield a consistent delta, as all are normally distributed and roughly centered around 0.

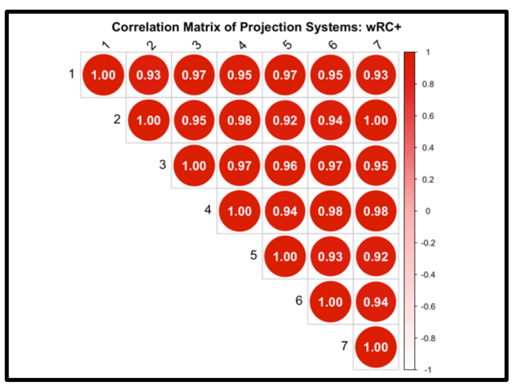


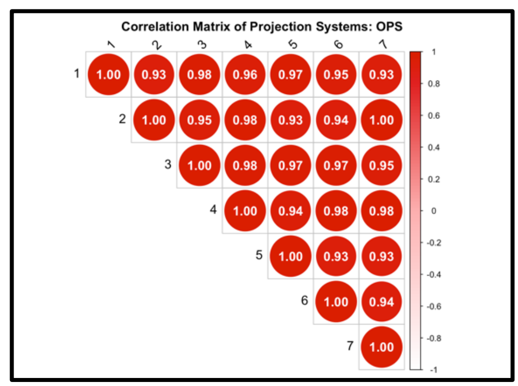




In addition, I investigated the relationships amongst the project systems by looking at their correlations (See correlation matrices below). Because these projection systems are working off similar resources, mainly differing in their proprietary implementation, we would expect them to be highly correlated but not identical. As expected, the projection systems are all highly correlated with each other across our target variables. Further, the projection systems tend to evaluate players similarly, and the relative projections of players tends to be consistent amongst projection systems.



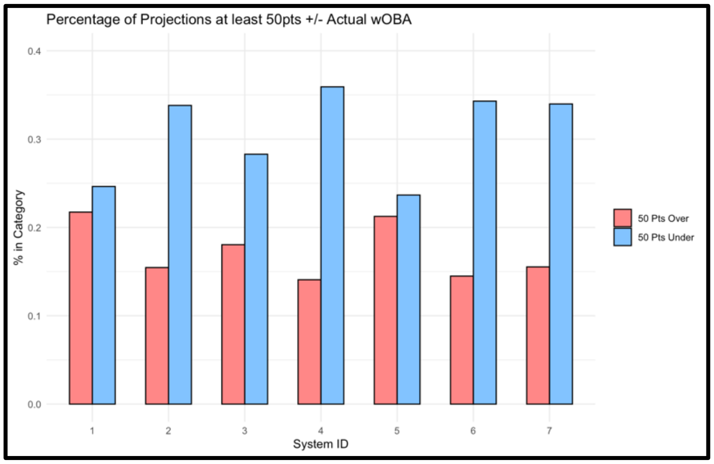




Next, to look more specifically at the relative quality of the 7 projection systems, I looked at the error terms for each projection system across the three target variables. In addition, for each target variable I looked at the percentage of predictions that were notably divergent from the actual player performance in each direction. A table of the error terms of interest for each projection system are included for each metric (ordered by largest to smallest magnitude of error). The sections below are breakdowns of this information grouped by target variable.

wOBA

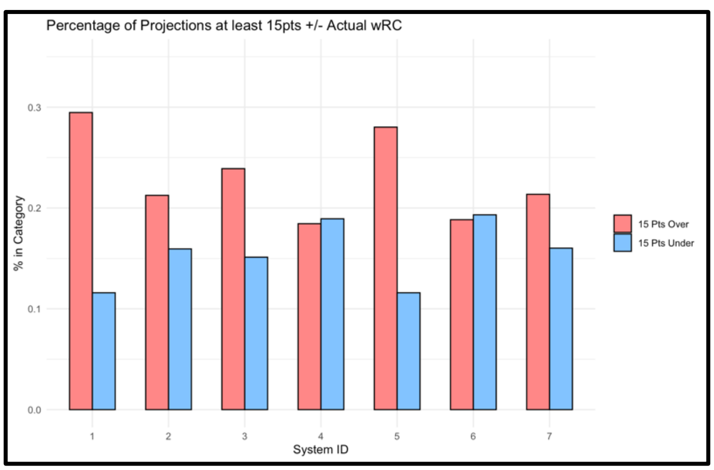
Below is a chart depicting the percentage of wOBA predictions 50pts (.050 on wOBA scale) over or under actual 2024 player performance. The chart indicates that each projection system had more instances of projecting at least 50pts under the actual player value than over. Projection System #5 was the most balanced in large over and under predictions, had the lowest error terms, and the highest Spearman's correlation.



Projection System ID	RMSE	MAE	Spearman's
6	0.0267	0.0216	.506
1	0.0263	0.0210	.504
4	0.0261	0.0209	.514
3	0.0258	0.0208	.502
7	0.0261	0.0208	.507
2	0.0261	0.0207	.508
5	0.0251	0.0200	.521

wRC+

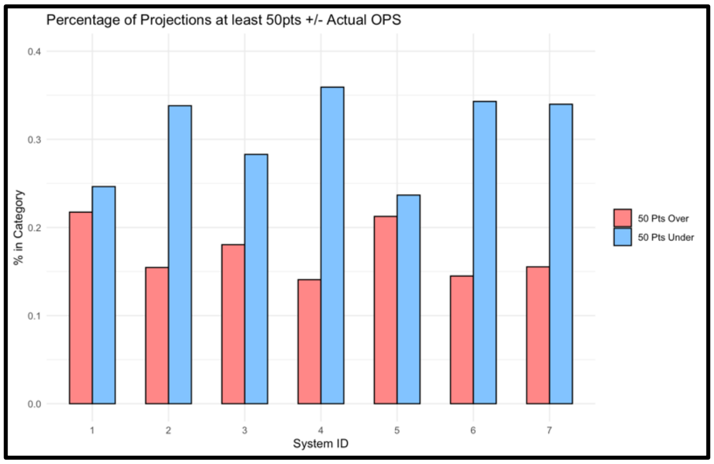
The chart below shows the percentage of wRC+ projections at least 15pts over or under actual 2024 player performance values. Notably, Projection Systems #1 and #5 had far more instances of predicting at least 15pts over a player's actual production. This is further reflected in the error metrics, where Projection Systems #1 and #5 had the highest RMSE and MAE figures. Projection System #4 had the lowest error metrics as well as the highest Spearman's correlation.



Projection System ID	RMSE	MAE	Spearman's
1	18.92	14.67	.528
5	18.12	14.32	.539
3	17.88	13.92	.532
7	17.66	13.83	.536
2	17.62	13.78	.537
6	17.49	13.69	.539
4	17.34	13.62	.548

OPS

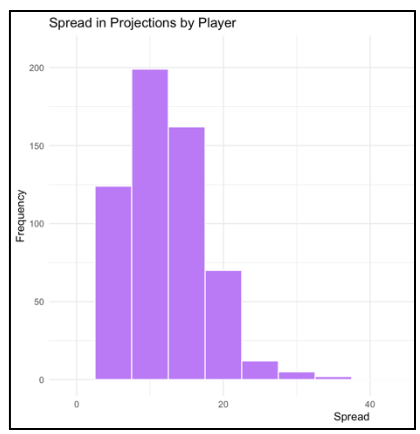
The chart on the left shows the percentage of OPS projections that were at least 50pts (.050 on OPS scale) above or below the actual player performance OPS. Across all systems, there were more instances of predicting 50pts under the actual value than over. Projection System #5 was the most balanced in over/under prediction and also recorded the lowest RMSE and MAE. Interestingly, while Projection System #4 had the highest Spearman's correlation, it recorded higher error metrics, indicating that it may encapsulate relative player performance well but may have a scaling flaw.



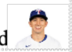
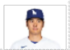
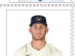
Projection System ID	RMSE	MAE	Spearman's
6	.0685	.0556	.511
7	.0670	.0541	.507
4	.0669	.0541	**.520**
1	.0673	.0539	.502
3	.0662	.0539	.504
2	.0668	.0538	.507
5	.0642	.0510	.515

Variance in Predictions by Player

Finally, I looked at the variance in predictions from a player-to-player perspective, looking specifically at wRC+ predictions. I decided to look at wRC+ as it is the most holistic player production statistic of the three target variables and is standardized to MLB average. The histogram on the left shows the density of the spread of wRC+ predictions per player. Most players had a spread in predictions between 2-15pts, with players with a spread 15pts or above "high-spread" players.



The table below lists the players with the highest variance in wRC+ predictions, ordered with highest variance players at the top descending lower in variance with each row. To maintain necessary anonymity, the columns are not labeled by their projection system. Rather, the columns left to right are the highest to lowest projections for a given player. Each cell value is the difference from the highest prediction to the corresponding prediction.

Player	1st	2nd	3rd	4th	5th	6th	7th
Wyatt Langford 	-	-3.1	-6.3	-6.3	-11.5	-32.5	-32.5
Jon Singleton 	-	-0.7	-1.3	-1.3	-19.0	-24.7	-26.6
Michael Busch 	-	0.0	-7.4	-14.9	-15.3	-25.7	-25.8
Mark Vientos 	-	-3.9	-7.8	-7.8	-15.7	-19.6	-22.5
Shohei Ohtani 	-	-2.9	-6.0	-16.9	-19.7	-22.4	-22.4
Alec Burleson 	-	-3.5	-7.0	-7.0	-15.1	-18.0	-21.8
Ernie Clement 	-	0.0	-0.6	-1.2	-9.0	-16.2	-21.2
Joey Ortiz 	-	-7.5	-11.3	-14.9	-14.9	-20.1	-20.2
Tyler O'Neill 	-	0.0	-7.3	-12.3	-12.9	-14.6	-19.5
Isaac Paredes 	-	-2.1	-2.8	-4.1	-4.1	-14.9	-19.0

Taking a deeper look at the highest variance players, 5 of the top-10 players were rostered on a team different than their 2023 team (Busch, Ohtani, Ortiz, O'Neill, Parades). This may suggest that across projection systems switching teams impacts projections differently. The impact of switching teams could be universal in nature (same penalty/boost regardless of teams), or implementations could consider the potential overlap in projected pitchers to be faced. For example, a player switching to a team within the same division would be expected to see more similar competition to previous seasons than a player switching to a new division. Additionally, Langford (Rookie), Clement, and Ortiz had limited major league plate appearances coming into the 2024 season. The projection systems may rely on different factors when there is limited MLB data to drive predictions. Some systems may look at trends in minor league performance, and there may be variance in how strongly each system regresses future performance of rookies to league central tendencies.

Novel Projection Frameworks

After analyzing the projection data from the 7 systems, I implemented 3 aggregation schemes in an attempt to out-perform the best system per metric. The goal of experimenting with these frameworks was to identify if we could create more reliable predictions just using public information, and to establish predictive accuracy thresholds for additional internal resources. The three frameworks implemented were a simple weighted aggregation, a linear weighting system, and a gaussian weighting system. Each of these implementations considered each prediction made by all 7 systems, with different underlying logic.

The simple weighted aggregation gave equal weight to each of the 7 projection systems. Thus, the final prediction was the average of the 7 predictions for each player. In the linear weighting system, each prediction was assigned a weight based on its deviance from the mean of all predictions. That is, predictions that were closer to the mean of all predictions were given more weight than those further from the mean. The gaussian weighting system operated similarly to the linear weighting scheme, but the assignment of weights was computed using a decay function. Congruently, values closer to the predictive mean were more impactful than those further from the mean. The formulas for the Linear and Gaussian weighting frameworks are as follows:

Linear Formula:

prediction_weight = 1 / (abs(prediction - mean_preds))

Gaussian Formula:

prediction_weight = exp(-.5 * ((prediction – mean_preds)/sd_preds)^2

(where *mean_preds* = average value of all predictions, *sd_preds* = standard deviation of all predictions)

Below are the error terms for each novel framework across the three target variables:

Simple Weighted

Target Variable	RMSE	MAE	Spearman's
wOBA	.0254	.0203	.521
wRC+	17.4	13.6	.548
OPS	.0650	.0525	.523

Linear Weighted

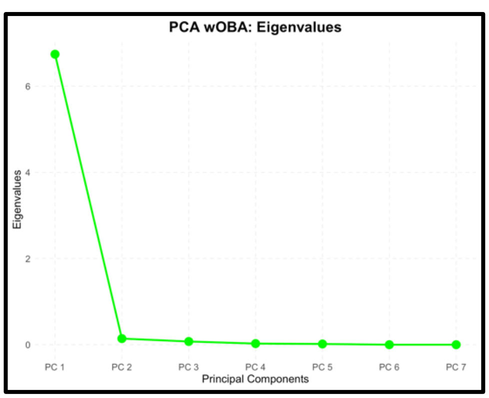
Target Variable	RMSE	MAE	Spearman's
wOBA	0.0251	0.0201	.523
wRC+	17.2	13.4	.548
OPS	0.0646	0.0520	.521

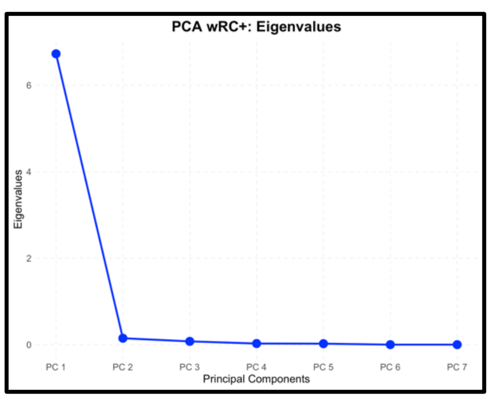
Gaussian Weighted

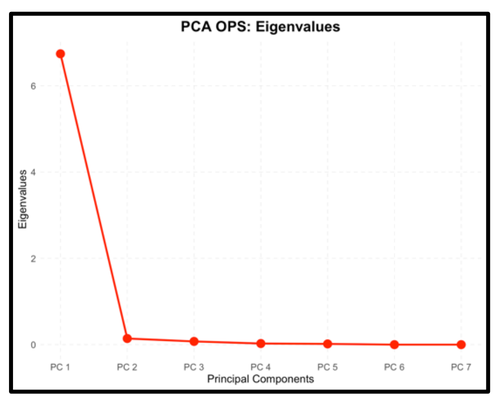
Target Variable	RMSE	MAE	Spearman's
wOBA	0.0253	0.0202	.522
wRC+	17.3	13.5	.550
OPS	0.0650	0.0522	.523

Principal Component Analysis

The nature of the projections dataset provided a use-case for Principal Component Analysis (PCA). Using PCA allows for reduction of dimensionality of the data while preserving variance accounted for across variables. As shown by the initial analysis of the data, the projections systems had linear relationships amongst variables, correlations between systems, and consistent scaling – thus meeting the core assumptions of PCA. Additionally, the sample size was just large enough to implement models using the principal components with cross-validation techniques, allowing us more reliably assess the effectiveness of the analysis. Below are scree plots and tables of variance accounted for across PCA for all three target variables. The scree plots display the eigenvalues of the principal components, which are directly related to the variance accounted for across components. The plots and variance tables illustrate that the first 2 principal components sufficiently account for variance across the dataset.







wOBA

Component	PC1	PC2	PC3	PC4	PC5	PC6
Variance Explained	. 96057	.02196	. 01146	.00351	.00250	< .00000

wRC+

Component	PC1	PC2	PC3	PC4	PC5	PC6
Variance Explained	. 96125	.02117	.01078	.00364	.00316	< .00000

OPS

Component	PC1	PC2	PC3	PC4	PC5	PC6
Variance Explained	.96305	.02027	.01060	.00369	.00240	< .00000

For model implementation and evaluation, a 5-fold cross-validation framework was used to test a linear regression model based solely on the first two principal components. The cross-validation technique allowed for more reliable results given our small dataset. The implementation of the linear model using just the first two components allowed for suitable model performance while minimizing the risk of overfitting. Below are average summary statistics across the 5-folds for all target metrics:

Target Metric	**Avg. R-Squared**	**Avg. RMSE**	**Avg. MAE**
wOBA	.395	.0252	.0198
wRC+	.410	17.2	13.5
OPS	.412	.0651	.0518

Other Model Considerations

During this process, several alternative modeling frameworks and techniques were considered. Most namely the possibility of hyperparameter tuning or testing various regression models. However, the driving factor in selecting applicable frameworks was data availability. Since we only have predictions from the 2024 season (FanGraphs does not provide a historical record of previous years predictions) the sample size was relatively small. This created difficulties in satisfying the assumptions needed for these methods, specifically a concern for insufficient data for appropriate model training.

Basic implementations of Ordinary Least Squares (OLS) regression produced only moderate correlation terms, with concerns of whether the framework was an adequate fit. Since the models were all shown to be highly correlated, this violated the assumptions needed for multiple additional frameworks. Due to the limited data, there was not enough sample for reliable forward or backward selection. As a result, we decided that the final deliverable would focus on comparing different modeling frameworks rather than pursuing additional novel model implementations. If we were to continue our investigation, we could collect predictions for future seasons, continue to test the existing frameworks, and explore using new methods.

Insights

The Novel Projection Frameworks provided sufficient predictive reliability and the ability to output a set of predictions at or above the best individual projection system. All three aggregations schemes had similar Spearman's correlation terms across the target metrics. Both the Linear Weighted and Gaussian Weighted frameworks outperformed the Simple Weighted framework, with the Linear Weighted framework slightly outperforming the Gaussian implementation. The performance benefits of these two systems articulate that calibration to the predictive mean is additive in blending across projection systems.

In terms of wOBA, the Linear Weighted framework had nearly identical error terms to that of the most effective individual projection system and outperformed the other 6 systems. The Linear Weighted approach had lower error terms than all wRC+ projection systems. As it relates to OPS, the Linear Weighted approach had similar error terms to the most effective projection system and outperformed the other 6 systems. Across all target variables, the Linear approach had a Spearman's correlation at or above the best individual projection system. Overall, the Linear Weighted framework provided predictive value across our target variables.

The linear regression model implemented using Principal Component Analysis (PCA) yielded moderate R-squared values and error metrics comparable to those of the aggregation systems. This approach offered another viable strategy for effectively combining projection systems, providing reliable results while accounting for variability across systems. The dimensionality reduction achieved through PCA offers the long-term benefit of increased robustness for future inputs. This method can be applied not only to incorporate additional external resources but also the inclusion of internal projection systems. The scalability benefit of PCA provides future flexibility during model implementation. Additionally, examining the variance explained by the principal components serves to investigate disagreements across projection systems.

Since we cannot be certain which of the 7 systems will be the most effective prior to the conclusion of a season, the weighting schemes and PCA add value in ensuring the predictions we employ are similar to or slightly better than the unknown, best individual system. The responsible aggregation of public information added certainty to the quality of projections, which can allow stakeholders to utilize the information without having concern that there is a better individual system of projections. Providing the relevant error terms with sufficient guidance to their meaning in-context can aid decision makers in using projections to evaluate players.

By combining the examples of players with the most individual predictive variability with specific baseball insights, there are grounds to identify theoretical constraints of where projection systems may be most unreliable, even without knowing the intricacies of each model.

The small subset of high-variance players indicated that switching teams and the number of major league plate appearances prior to the start of the season may be differentiating factors amongst public resources. Seeing concrete examples of which players have higher variance across projection systems can add value when discussing with stakeholders by increasing interpretability.

Another meaningful interpretability bonus was enlisting novel frameworks with high explainability. All 3 weighting schemes are highly interpretable to a variety of audiences. Since these projections are shared with various stakeholders with differing degrees of technical familiarity, it was imperative that we chose frameworks with high interpretability. The interpretability consideration, along with the understanding of assumptions of combining model outputs, suggested against considering frameworks like Neural Networks and Gradient Boosting frameworks.

Perhaps the most impactful insight and expandable piece of this investigation is the concept of responsibly establishing a minimum benchmark threshold for internal resources. In the present project, we were able to quantify the level of predictive accuracy yielded from the use of public resources in terms of RMSE and MAE. This enabled us to have an objective benchmark for which any internal projection system should add value to justify the use of additional internal resources. Additionally, interpreting the value of these error terms can allow leadership to set target goals for error terms before embarking on creating a novel model implementation. Setting this target before creating a model can alleviate bias in the model evaluation process.

Moreover, the concept of establishing benchmarks using public resources does not apply only to offensive projection systems. It can be applied consistently across other instances of model implementations throughout Baseball Operations. Since there are public projection resources available for many player-level and team-level statistics, there are many opportunities to apply this concept. Evaluating the quality of publicly available resources and the time required to develop proprietary models that outperform them can help leadership plan future research most effectively. Having a robust framework for interpreting and maximizing the quality of public resources has vast future applications.

Reflection

Conducting this research provided the opportunity to apply skills gained throughout OMSA coursework in a professional setting. Specific to OMSA coursework, my analysis of the existing projection systems was driven by concepts learned during the program. Doing an in-depth investigation to verify the validity of the projection systems was key before considering which model frameworks to proceed with. Looking at the relationships amongst the projection systems and articulating why the findings made sense in-context was a direct application of fundamental concepts of the OMSA program. The selection of relevant error terms (MAE, RMSE) and novel frameworks (Linear weighting, Gaussian weighting, PCA/Regression) were driven by experiences conducting previous analysis during the program.

When considering the novel frameworks, I was able to combine the skills gained during my coursework with my domain knowledge. As I initially planned this project, I went through all the possible analysis frameworks and novel implementations that may have been relevant. I considered the options, evaluating them primarily based on their underlying assumptions, predictive power, and interpretability. Combining this knowledge with my familiarity with the industry, I decided that interpretability was the most important concern and let that drive the framework selection process. Ultimately, these considerations led to processes that were effective in evaluating public resources, combining them into a set of predictions with increased utility, and creating a general framework for benchmarking internal resources against public projection systems.

Sources

"A Guide to Sabermetric Research." *Society for American Baseball Research*, https://sabr.org/sabermetrics.

Baseball Reference, https://www.baseball-reference.com/.

Baseball Savant, https://baseballsavant.mlb.com/.

Fangraphs Baseball, https://www.fangraphs.com/.