Batter Pitch Mix Breakdown-Coaching Insights

Patrick Mellady

2024-10-19

Objective

We are interested in a solution to the following question:

Given pitch-by-pitch data from the 2021-2023 MLB regular seasons, we would like you to predict the pitch mixes that batters faced in 2024 across three groups: fastballs, breaking balls, and off-speed pitches

In this report, we examine the solution to this question for certain players and how we can use this information for game preparation and in-game decision making.

Background

The model works by using the number of times a hitter is at-bat during certain game situations. To illustrate this, we will examine the data and predictions for Paul Goldschmidt. The data used to make the predictions is shown below.

Paul Goldschmidt Pitch Mix Data (2021-2023)

Year	FB	BB	OS	RHH	RHP	Hitter's Count	RISP	Shift	Home Team Winning
2021	1528	975	248	2751	2143	650	592	530	1105
2022	1498	976	227	2701	2115	576	595	787	892
2023	1670	955	222	2847	2155	601	634	447	883

The first three columns indicate how many of each type of pitch- fastball, breaking ball, and offspeed- were seen by Paul Goldschmidt from 2021 to 2023. The remaining six columns give the total number of times Paul Goldschmidt saw a pitch under certain situations. For example, Goldschmidt saw a pitch in 650 hitter's counts in 2021, he saw a pitch with a runner in scoring position 595 times in 2022, and he saw 1670 fastablls in 2023. For the predictions, we use the game information prior to the pitch being thrown because it is the same information available to the opposing team at the time of the pitch. Using this method, our predictions are effectively modeling the thought process of an opposing pitcher before each pitch is thrown.

The above data translates to the following pitch mixes:

Paul Goldschmidt Historical Pitch Mixes (2021-2023)

Year	FB	BB	OS
2021	0.555	0.354	0.090
2022	0.555	0.361	0.084
2023	0.587	0.335	0.078

So, in 2021, 55.5% of the pitches that Goldschmidt saw were fastballs, 35.4% were breaking balls, and 9% were offspeed pitches.
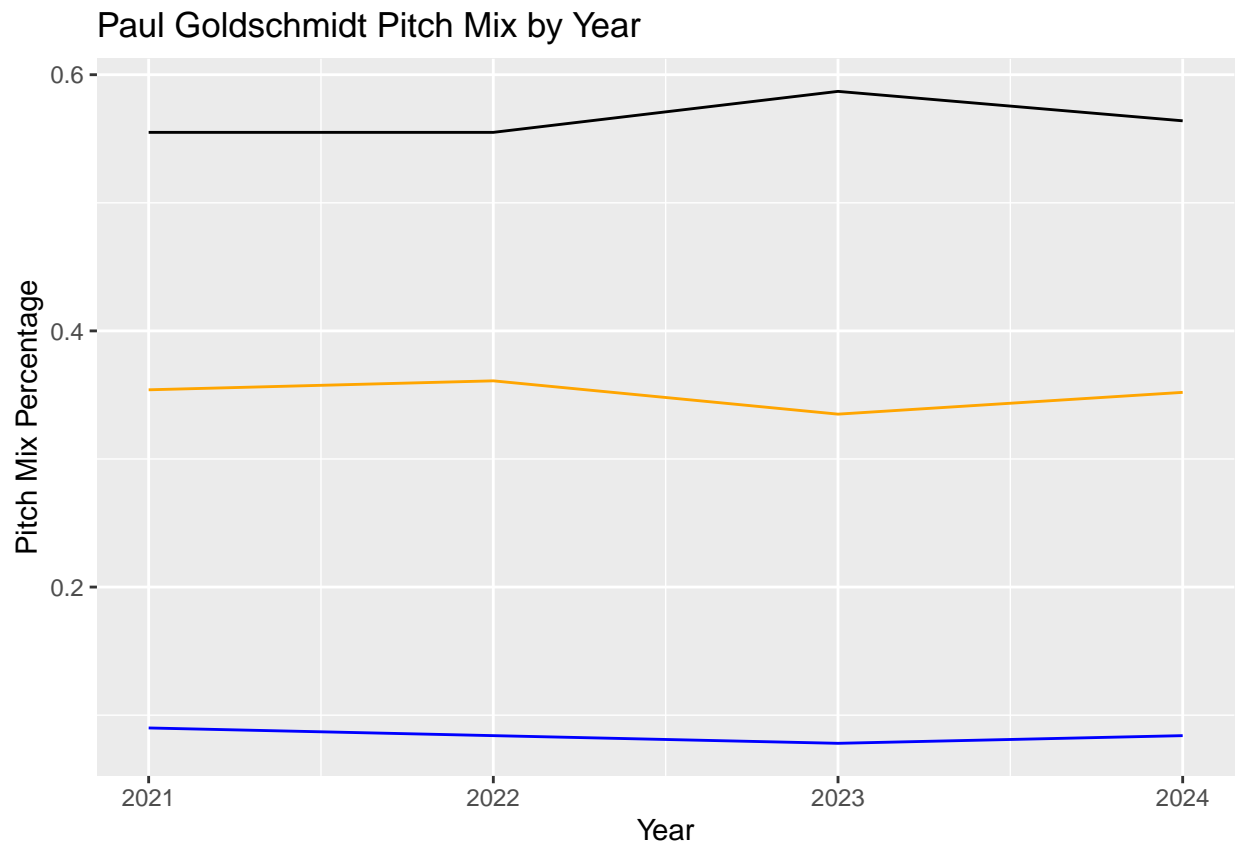
Results

Our model predicts the following pitch mixes for Goldschmidt in 2024:

Paul Goldschmidt Projected Pitch Mixes (2024)

FB	BB	OS
0.564	0.352	0.084

Visualizing the historical and predicted pitch mixes over time shows us the following trends:


Most significantly, this chart shows us that 2023 saw a change in the pitch mixes that Goldschmidt received, yielding an increase in fastballs and a decrease in breaking balls. The chart also shows that we predict the pitch mixes for Goldschmidt to move back towards their historical values in 2024. This means we should expect to see fewer fastballs and more breaking balls thrown to Goldschmidt in the 2024 season. With this information, we can utilize practice time to focus on the pitches we most expect Goldschmidt to see, suggesting that Goldschmidt would benefit from an increase in individual work with breaking balls.

We have applied this predictive model to our entire database of hitters to calculate predicted pitch mixes for each one. To get a sense of the model's accuracy, we look at a sample of 4 hitter's projected pitch mixes below:

Projected 2024 Pitch Mixes

Name	FB	BB	OS
Pham, Tommy	0.565	0.327	0.108
Díaz, Elias	0.541	0.335	0.124
Bell, Josh	0.527	0.245	0.228
Turner, Trea	0.528	0.351	0.120

This table gives an idea of the model's accuracy, as the predictions for these players differ from the observed values by no more than 3%. We can also visualize the year-by-year trends of the pitch mixes for each of the above players:
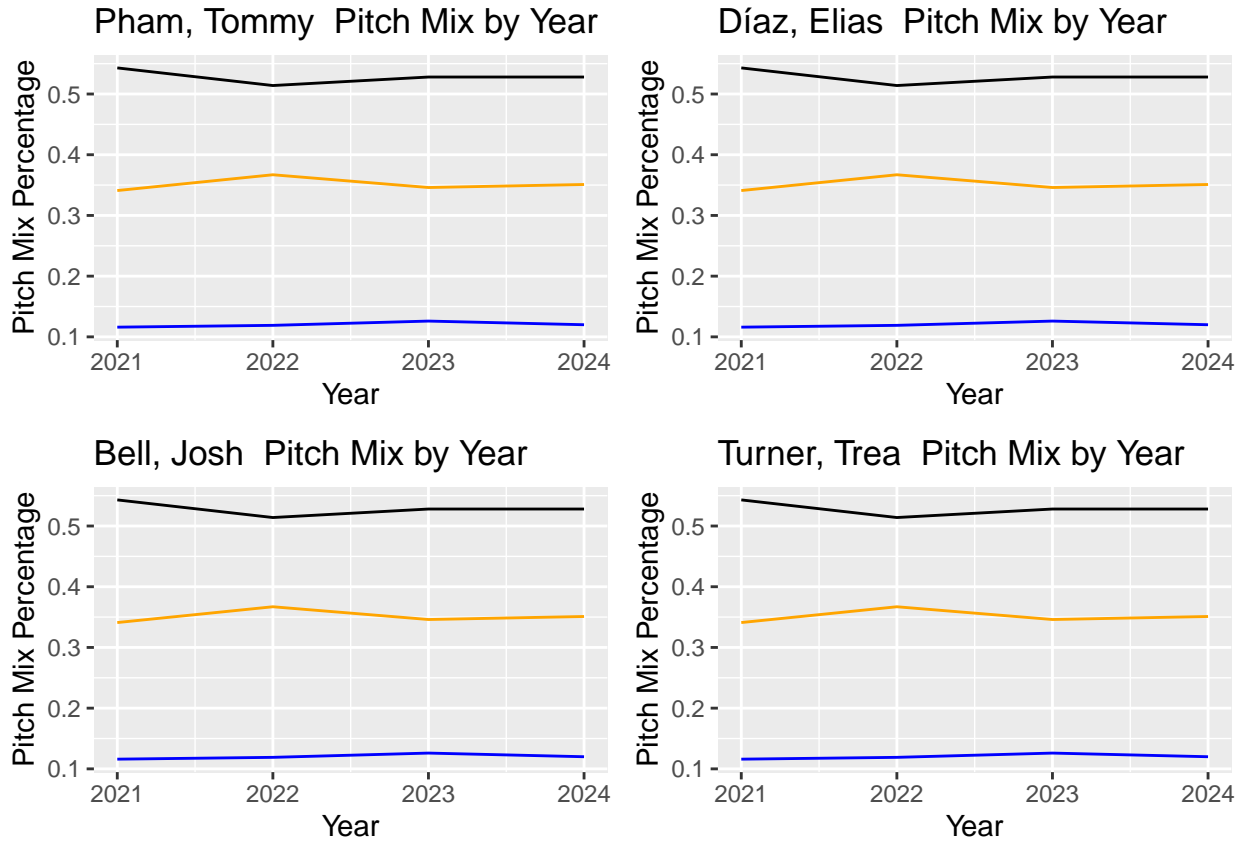
Limitations and Recommendations

While this model proves to be useful for allocating practice time, further understanding how of the model predicts is necessary to maximize the benefit og these predictions. As mentioned above, the model predicts based on the number of times a hitter sees pitches in certain game situations, but we don't know the number of each situation a hitter will see in the coming year. So, in order to use the model, our predictions work based off of the median of the number of times that situation was seen in the past.

As an example, consider the following data for David Peralta:

David Peralta Pitch Mix Data (2021-2023)

Year	FB	BB	OS	RHH	RHP	Hitter's Count	RISP	Shift	Home Team Winning
2021	1168	555	339	0	1716	481	553	1246	550
2022	977	512	341	0	1562	399	416	1266	615
2023	831	410	289	0	1409	310	401	965	669

To make our predictions for the 2024 season, we would use the following values:

David Peralta Predicted Situation Totals (2024)

Year	RHH	RHP	Hitter's Count	RISP	Shift	Home Team Winning
2024	0	1562	399	416	1246	615

Understanding how the model is fit is important in knowing the applicability of the results. The model's predictions will be the most useful when a hitter is used in similar ways to the past. For example, if a hitter is historically number four in the lineup, but is moved to number eight, the results won't be as accurate. This is because changing how a hitter is used in games will change the number of game situations that they see. So making a move from four to eight will cause the data that the predictions are based off of to be less applicable yielding less accurate results.

To illustrate the concept above, consider how changing the frequency of certain game situations between Paul Goldschmidt and David Peralta affects their pitch mix breakdown.

The predictions for Peralta using his predicted totals are:

David Peralta Projected Pitch Mixes (2024)

FB	BB	OS
0.543	0.273	0.184

However, when the situations that each player sees are switched, we have the following predicted pitch mixes:

David Peralta Projected Pitch Mixes (2024 with Paul Goldschmidt Predictors)

FB	BB	OS
0.535	0.28	0.185

Paul Goldschmidt Projected Pitch Mixes (2024 with David Peralta Predictors)

FB	BB	OS
0.572	0.343	0.083

This tells us that if we were to use Peralta in situations that Goldschmidt would typically see, Peralta will see fewer fastballs and more breaking balls, while if we use Goldschmidt in situations that Peralta would usually see, he would see more fastballs and fewer breaking balls.

We can apply the insights gained from the model to prepare hitters when we expect some kind of change to the situations that a hitter will see. For example, new additions to the lineup forcing changes in the batting order, a player returning from injury, or even in game decision making.