California Institute of Technology

When to Pull a Pitcher? Predicting Drops in Pitcher Effectiveness in Major League Baseball

Edward Speer, BEM/Ec 120: Sports Science, Spring '23

Dr. R. Micheal Alvarez and Dr. Oliver Eslinger

Abstract

The decision of when to substitute one pitcher for another in a major league baseball game is one of the most important strategic choices in the game. This paper employs clustering and explanatory analysis of time series data of pitches to determine important markers in statcast data which can forecast a pitcher giving up an offensive advantage to the opposing team, finding that the decline in spin rate of a pitcher's primary pitch is the most important factor, followed by the decline in spin rate of their secondary pitch. Interestingly pitch velocity decline was found to play little to no role. This work can inform baseball strategic decision making regarding pitcher usage based on in-game statcast data.

Table of Contents

I. Introduction

In baseball, one of the most critical and notoriously difficult strategic decisions to make is when to remove one pitcher from the game and put another in. This decision is based on a wide range of game features, including hitter-pitcher matchups, game importance, bullpen rest, and many other factors. However, the most crtical factor, and yet the most difficult to judge, is pitcher fatigue. As a pitcher becomes either physically or mentally tired, they naturally become less effective in preventing the other team from scoring. However, pitcher fatigue can manifest itself in ways that vary between pitchers, and often may only clearly show itself after the opposing team has already gained an offensive advantage. The goal of this paper is to propose a sabermetrics approach which uses sliding in game pitch metrics to determine when a pitcher is growing fatigued, before enabling the other team to take advantage.

In baseball, throughout most games the personnel in the game remain largely te same. Starters typically play a whole game, without the frequent substitutions expected in other sports. The major exception to this is the pitcher, who is expected to be substituted 3 or 4 times throughout the game. The pitcher also has the largest effect on a team's defense by virtue of their role. This makes the substitution of the pitcher a very weighty decision in the game with large ramifications, and consequently one of the decisions of a baseball manager underrth largest scrutiny. The game importance and visibility of these decisions make it important to develop metric based methods which optimize their result.

Baseball pitchers are highly varied athletes. The MLB recognizes 13 different types of pitches, with each pitcher in the league throwing a different combination of pitches (called their pitch arsenal), each of which has unique characteristics when thrown by one arm over another. This incredible variety in pitchers makes it difficult to generalize any conclusions from sabermetrics across pitchers. This paper's approach is to use data clustering techniques to sort pitchers into distinct archetypes based on their pitch arsenals, and then look at specific indicators from game data to uncover indicators of pitcher fatigue common to each pitcher cluster. This allows for an additional measure of specificity allowing the result to apply generally enough for ease of use across MLB pitchers, but enough specificity to be insightful.

Recognizing the characteristic fatigue signals of different archetypes of pitchers in the MLB is widely applicable beyond the limits of MLB pitchers. In youth and amateur sports, many pitchers pattern themselves after professional pitchers, throwing similar arsenals, or arsenals which can be sorted into the same archetypes as MLB pitchers. Recognizing the pitcher archetype in combination with fatigue markers for their pitchers can allow managers at all levels of baseball to make informed decisions about when to change pitchers, even with decreased access to advanced game data.

II. Review of Literature

Research into pitcher removal from a game has been somewhat sparse, and has focused mainly on how managers currently make decisions as opposed to how optimal decision making could be performed. These studies have attempted to determine or measure managers act as they do, and if their decision making is optimal, and have revealed the need for this study by showing sub-optimal behavior in this decision making. Phillips' 2017 study showed that managers emphasize pitch count disproportionately when removing pitchers, showing a strong tendency to remove pitchers on pitch counts which are multiples of 10 regardless of the game state, with no clear statistical reason for this behavior [1]. Clearly this shows managerial behavior which is subjective and not driven by sabermetrics.

Grantham et al. looked at college pitchers, and showed that pitcher fatigue significantly impacted pitching mechanics over the course of a game, with pitchers showing clear biophysical indicators of fatigue with increased number of pitches or innings thrown[2]. For the purposes of this paper, we hypothesize that these biophysical differences in pitching mechanics will result in observable changes in the metrics surrounding ball flight. For example, Grantham found that external rotation was strongly impacted by pitcher fatigue. A decrease in external rotation likely results in decreased pitch velocity, which is observable from the sabermetrics used in this paper.

Whiteside et al. 2016 found that the variables which changed the most over the course of a pitchers performance as the pitcher fatigued were pitch velocity, vertical movement, and proportion of offspeed pitches thrown by the pitcher. However, these variables were not strongly correlated with effectiveness [3]. This paper seeks to find markers hidden in these and other pitching metrics which both change as a pitcher fatigues and have a stronger correlation with effectiveness than those found by Whiteside.

Woodham et al. 2019 used a machine learning model to identify the most significant features in when a starting pitcher would be removed from a game in an effort to predict when a starting pitcher should be removed, and found that the strike count and number of batters a pitcher has faced in the game, in conjunction with the number of outs, inning number, and number of homeruns allowed are the most significant factors in removing a pitcher [4]. However, these conclusions yield sparse information in how to make decisions about removing a pitcher, as they are very standard metrics of performance, and if a pitcher is performing poorly (throwing balls and giving up homeruns), then it is evident they must be removed from the game. An improved predictive model would anticipate from markers in the sabermetric data when the pitcher was going to give up homeruns or walks, and remove the pitcher prior to allowing that advantage to the offensive team.

Each of these previous studies show that pitcher fatigue is observable in in-game metrics seen throughout the course of the game. Combining this observation with a study of the in-game pitching metrics available from the modern MLB statcast system should allow the development of a predictive model which correlates markers of pitcher fatigue with oncoming increases in offensive output by an opposing team.

III. Theory and Hypothesis

Due to the wide variability in pitcher mechanics and style across pitchers in the MLB, it is clear that different pitch metrics will impact the effectiveness of pitchers in different ways. For example, a pitcher who throws a changeup 10% of the time will likely be far less effected in their ability to get outs by a reduction in their changeup metrics than a pitcher who throws a changeup 50% of the time. As a result, the first step to developing a predictive model of which variables impact pitching performance is to sort pitchers into categories, with pitchers in each category of pitcher being likely to to have their effectiveness impacted by the same metrics. This way the predictions of the model can still be generalized across a larger group of pitchers while still having a measure of precision. This means there will be some inherent trade-off between the granularity of the pitcher grouping and the accuracy of the predictions which will need to be fine-tuned for the desired results.

A pitcher's arsenal is the set of pitches that a pitcher uses combined with the frequency and characteristics with which they throw each pitch. The characteristics of a pitch are metrics like the speed, spin rate, and spin axis that characterize the flight and shape of the pitch. A pitcher's arsenal combined with their longevity, handedness, and the opponent completely determine the strategy of how a pitcher will be used.

The best way to cluster pitchers will be based on their pitch arsenal. Pitching style is captured the best by how hard pitchers throw, which different pitches they throw, and how often they throw different pitches. Other factors which might tend to be included in a clustering are pitcher handedness and pitcher arm angle, however these factors are unlikely to be impacted by fatigue and therefore may be excluded from a clustering intended to reveal fatigue characteristics.

With this clustering of pitchers grouping similar arsenals together, an analysis of common fatigue characteristics may then be carried out. Grantham et al. showed that fatigue changes the mechanics of a pitcher's delivery over the course of a game, and with changes in mechanics come changes in the characteristics of the flight of the ball out of the pitcher's hand. Any subtle change in the pitcher's arm angle will cause a change in the spin axis of the ball. Muscle fatigue will show a change in pitch velocity, and forearm musce fatigue will show a reduction in spin rate. Any of these factors can be observed from in-game statcast data.

Notably absent from this discussion of when to remove a pitcher is the current opponent and opposing batter. These factors clearly will influence the optimal time to remove a pitcher, since pitchers match up more or less favorably against certain opponents. However, the goal of this study is to determine how markers of pitcher fatigue can be observed to understand the current effectiveness of the pitcher, and a manager could consider these markers in conjunction with the opponent to inform pitching decisions. For instance if a pitcher belonged to a cluster with curve ball spin rate identified as a predictor of effectiveness, and the manager observed a sliding reduction in the pitcher's spin rate and knew that the upcoming hitter has historically performed will in this matchup, then the manager could use the spin rate in order to inform his decision to remove the pitcher. Performing the analysis in this opponent-agnostic way allows the development of simpler predictions which may be applied more widely. In the future similar analysis could be carried out to analyze certain matchups, however this study is more concerned with identifying more general pitcher fatigue tendencies that impact their performance.

It is hypothesized that for each cluster identified in the paper, the combination of variables which is most important for each pitcher to remain effective throughout a game are the speed of the primary pitch, combined with either the spin axis or spin rate of the secondary pitch. This group of variables indicates a pitcher's ability to throw a primary pitch at a speed which is competitive, and a secondary pitch with good movement, a combination of features which is traditionally sought after in baseball. However, it is suspected that some clusters will deviate from this behavior, as some clusters will likely not throw any pitches at a high speed, in which case primary pitch speed shouldn't provide as much of an indicator of pitcher effectiveness. However, it is likely that for these clusters as well the best indicators of fatigue contributing to a breakdown in effectiveness will be some combination of features of primary and secondary pitches, as it is typically the combination of multiple pitches which make an arsenal effective.

IV. Data and Methods

Data Source

In order to select data to use for this study, it was important to consider several confounding factors which could alter the consistency of pitching data. Firstly, pitcher tendencies change from year to year, with pitchers adding or removing pitches from their arsenals and changing the characteristics of the pitches in their arsenals in the offseason. Sometimes, this change could even be drastic enough to change the cluster which a pitcher could belong to in a clustering analysis. Therefore, clearly, data must be selected from a single season. The most recent available full season is 2022 data, however halfway through the 2022 season, the MLB implemented new substance control policies relating to pitchers which caused sharp changes in the spin rates of different pitchers across the league, making the data too inconsistent for this analysis. Therefore the 2021 season was selected.

Baseball Savant is a website owned and operated by the MLB which makes publically available advanced metrics taken from MLB games Their data set of pitching metrics from

2017 is extensive and contains all the data used for any of the analysis which follows.

In order to cluster MLB pitchers into distinct archetypes, the following statistics are scraped for all pitchers who threw a minimum of 250 pitches in 2021:

Metric	Description
Average Speed	The speed of a pitch in miles per hour, averaged over all the pitches of that type thrown by the pitcher.
Average Spin	The rate at which a pitch spins, measured in rotations per minute, averaged over all the pitches of that type thrown by the pitcher.
Use Percentage	The percent of pitches thrown by the pitcher which are this type of pitch.

Table 1: Pitching Arsenal Metrics

Each of these metrics are scraped for each type of pitch thrown by each pitcher, providing a characterization for each pitcher of what pitches they throw, with what speed and movement (for full data see appendix I). The pitches thrown in 2021 include the following pitch types:

Pitch	Data Label	Description
4-Seamer	ff	A fast, straight fastball with close to vertical backspin
Sinker	si	A fastball that drops vertically over the course of its flight
Cutter	fc	A fastball which curves horizontally toward the glove side of the pitcher
Slider	sl	A ”breaking ball” (pitch with an arcing path) which moves mostly horizontally toward the pitcher’s glove side
Changeup	ch	A slower pitch whose flight is straight with some dropping vertical Movement.
Curve	cu	A breaking ball whose movement is primarily vertical, dropping sharply over the course of its flight
Splitter	fs	A slower pitch with lower spin rate that drops as it approaches home plate.
Sweeper	st	A breaking ball which breaks exactly horizontal in its flight path
Slurve	sv	A breaking ball which breaks at a 45 degree angle between horizontal and vertical movement

Table 2: Pitch Types

Data Clustering

K-means clustering is used to identify clusters of pitchers representing distinct pitching styles based on the arsenals thrown by each pitcher. The input vectors are the sets of arsenal data for each pitcher, pre-processed by applying mean normalization of each feature f_m across the entire population for each arsenal metric m_a:

$$f_m = (\mu_{m_a} - m_a)/\sigma_{m_a} \quad (1)$$

For those features which are not populated (i.e a pitcher who does not throw a certain pitch), the absence of the feature is actually a key feature of the dataset, as they pitcher not throwing that pitch certainly says something about their strategy. Therefore in order to capture this information in the data, we assume that features are approximately normally distributed across pitchers, and define empty features as a significant deviation from the mean, initializing all values of the mean normalized empty feature to 4 standard deviations from the mean:

$$f_m = \begin{cases} (\mu_{m_a} - m_a)/\sigma_{m_a} & \text{pitch in arsenal} \\ (\mu_{m_a} - (4 * \sigma_{m_a} + \mu_{m_a}))/\sigma_{m_a} & \text{pitch not in arsenal} \end{cases} \tag{2}$$

In order to identify the number of clusters which should be distinguished by k-means, the elbow method is used in which clusterings of different sizes are applied until an approximate convergence of clustering inertia, defined as the reduction of mean distance from the centroid of each cluster, is reached, shown for the pitching arsenal data in the following plot:

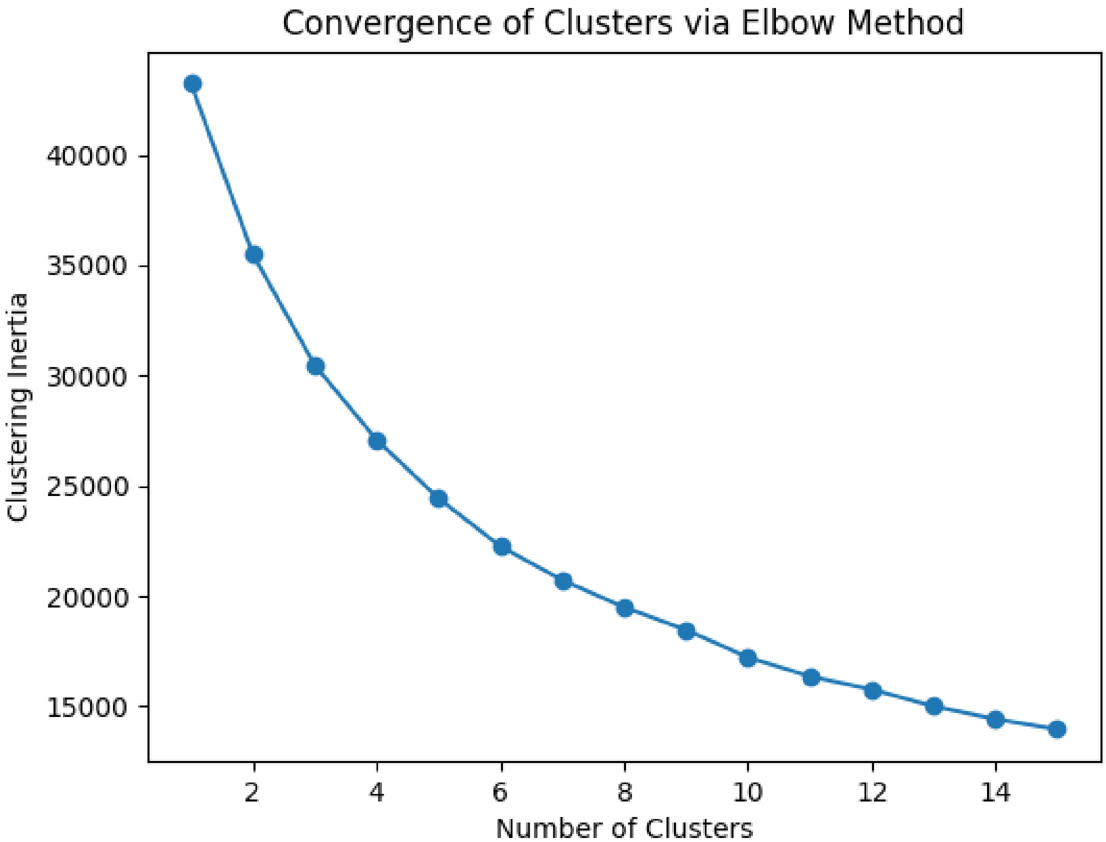
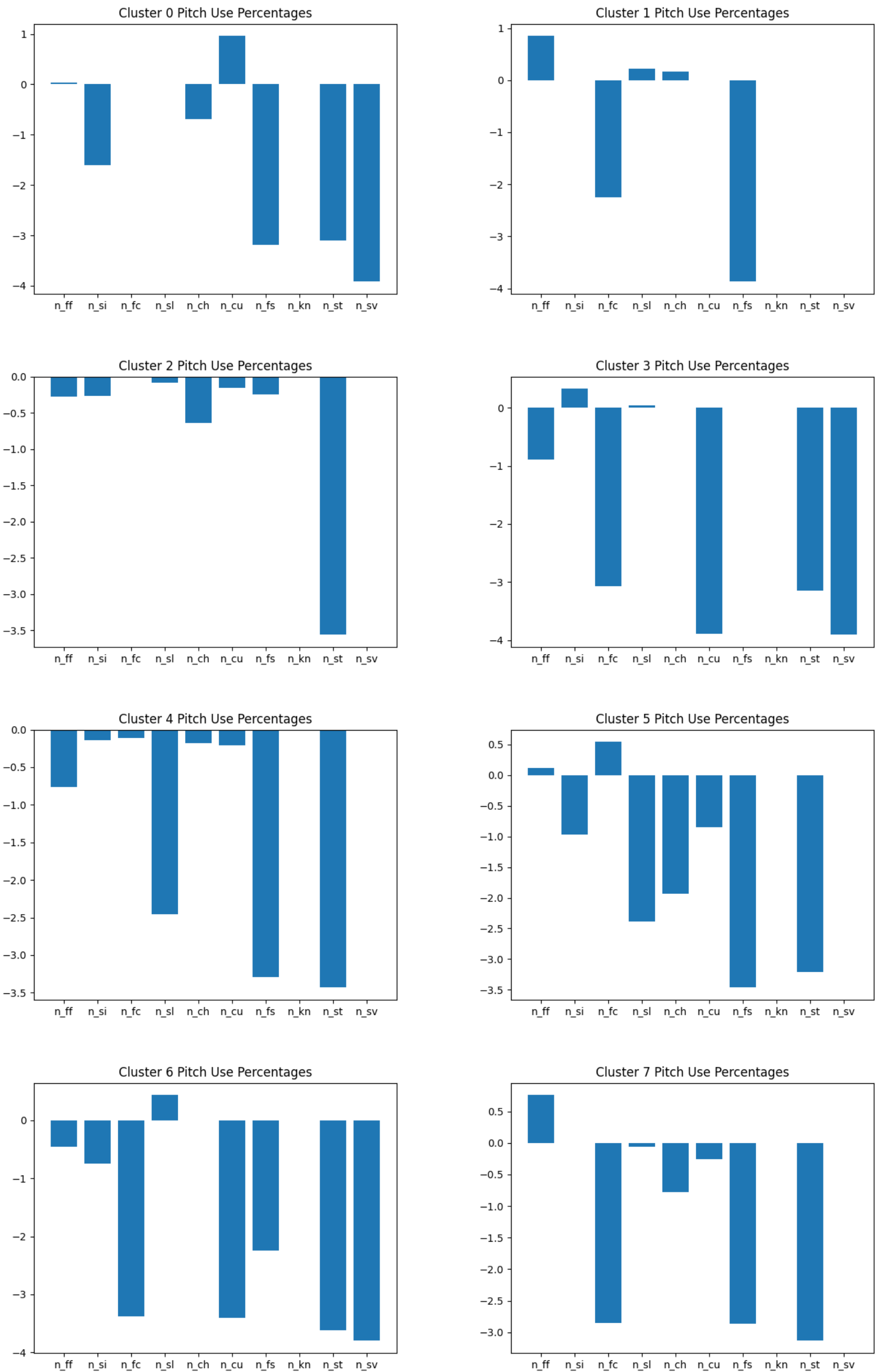


Figure 1: K-means Clustering Elbow Method

In order to minimize clustering inertia while maximizing the amount of information conveyed through clustering, a clustering size of 8 is selected and pitchers are sorted accoring to this clustering. The results of the clustering are 8 groups of pitchers with distinct styles of pitching, characterized best by the frequency with which they throw different pitches. These frequencies are shown in the plots below in standard deviations from the average pitch usage, with any blank bars indicating that the cluster of pitchers does not throw that pitch.



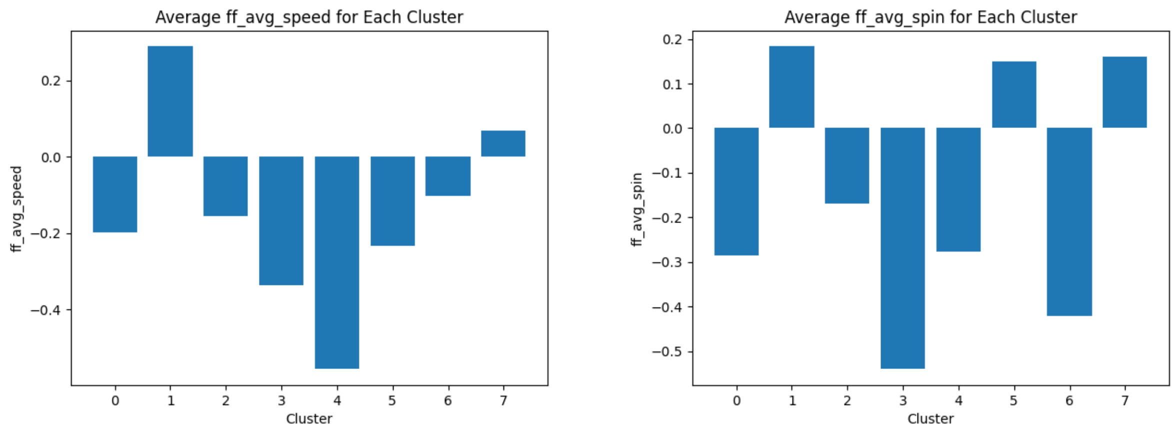
Figures 2-9: Cluster Pitch Usages

A clear characterization of each cluster can be observed from these plots of pitch usage:

Cluster	Description
0	Changeup heavy pitchers. These pitchers throw a combination of fastball and changeup, heavily favoring the changeup as their best pitch, and using their fastball as a change in speed to increase the effectiveness of their changeup.
1	Power fastball pitchers. These pitchers throw far more fastballs than the rest of the league, using their fastball in combination with occassional breaking balls to get outs.
2	2-seam pitchers. These pitchers put a lot of movement on their fastballs, both cutting them and sinking them, without and blending them with the occassional straight fastball and changeup, but without relying too heavily on any one pitch.
3	Sinker ball pitchers. Rely heavily on the sinker to get outs, and mix it with the occassional slider.
4	Nothing straight. These pitchers throw far less straight fastballs than other clusters of pitchers, relying on moving fastballs mixed with changeups.
5	Cutter ball pitchers. These pitchers throw very high amounts of cutters, mixing them with a power four seam to blow the ball past guys when they need a strikeout.
6	Slider heavy pitchers. These pitchers throw a power slider more frequently than any other group of pitchers, mixing it with occassional fastball when speed is needed.
7	Only fastballs. These pitchers throw mostly fastballs with the occassional slider, mostly looking to blow the ball past the hitter with velocity.

Table 3: Cluster Characteristics

The characteristic arsenals as well as the reasons for why each cluster throws the way it does can also be seen from the plots of average metrics across each cluster. For example, it is clear why clusters 2 and 7 throws the the most fastballs, with cluster 2 throwing 2 seams and custer 7 throwing 4 seams, as these clusters very clearly throw the fastest and highest spin rate fastballs across all clusters of pitchers:



Figs 10, 11: Cluster Fastball Characteristics

Note that cluster 5 also throws very high spin fastballs, which is to be expected since a cutter is a fastball with a high horizontal spin rate. Clearly this clustering makes a lot of sense; it has sorted pitchers by the metrics of their pitches, and they throw those pitches with the best metrics the most frequently. Now, this clustering with these pitcher archetypes can next be used to determine which fatigue characteristics impact pitcher effectiveness for each classification.

Fatigue Analysis

Now, for each cluster identified by the clustering analysis, an explanatory analysis is to be carried out to understand how fatigue experience by a pitcher throughout a game can maifest itself in observable changes within the data. In order to begin this analysis, we need to first identify and collect data for instances of "fatigue events" in which a pitcher allowed the offensive team to gain an advantage. In order to do so, a fatigue event is defined as an outing in which after a pitcher threw at least as many as half of their average number of pitches thrown per outing, the offensive team gained a large swing in expected winning percentage from a discrete event. A large swing in expected winning percentage is defined as a delta of more than .1 over the course of the window.

For each cluster, the events matching this definition are queried using Pybaseball and the event identification stored via python's inbuilt data pickling module (See full event listings by unlpickling the events file from the link provided in Appendix I). Events are scraped from querying all pitches thrown by each pitcher in a cluster, sorting them by game, then running checks on the events of the game to determine if the definition above is satisfied. In order to prevent any biasing in the analysis based on the clustering, a random sample of 50 such events is taken from the set of clusters identified for each cluster.

With events identified for each cluster, the time series data of pitch metrics over the course of each pitcher outing approaching the fatigue event must be collected. The metrics chosen to analyze per event are given in the following table:

Metric	Description
Fastball Velocity	The velocity of any pitch which qualifies as a fastball, including cutters, sinkers, four-seam fastballs, and two-seam fastballs (mph)
Breaking Ball Velocity	The velocity of any pitch which qualifies as a breaking ball, including curves, sliders, and sweepers (mph)
Changeup Velocity	The velocity of a changeup (mph)
Fastball Spin Rate	The spin rate on any pitch qualifying as a fastball (rpm)
Breaking Ball Spin Rate	the spin rate on any pitch qualifying as a breaking ball (rpm)
Changeup Spin Rate	the spin rate on a changeup (rpm)
Curve Ball Vertical Break	The total vertical translation of a curveball over the course of its flight in feet
Slider Horizontal Break	The total horizontal translation of a slider over the course of its flight in feet
Release Height	The total height off the ground at which a pitch is released

Table 4: Time Series Metric Definitions

These metrics were chosen from the pool of avaliable statcast metrics for their likelihood to influenced by fatigue for the following reasons:

- Pitch Velocities: Traditionally thought to decrease withn fatigue, though in the MLB today pitchers tend to pace their own usage to maintain velocity
- Spin Rate: Is highly effected by grip, meaning that forearm fatigue over the course of the game could reduce spin rate, which is one of the most important features that make a pitch effective.
- Break: Is also a proxy for the combination of velocity and spin rate and is therefore impacted by fatigue for the same reasons as those listed above, and is what makes

breaking pitches effective.

- Release height: Whiteside et. al showed that this is one of the most impacted biometrics correlated with pitcher fatigue

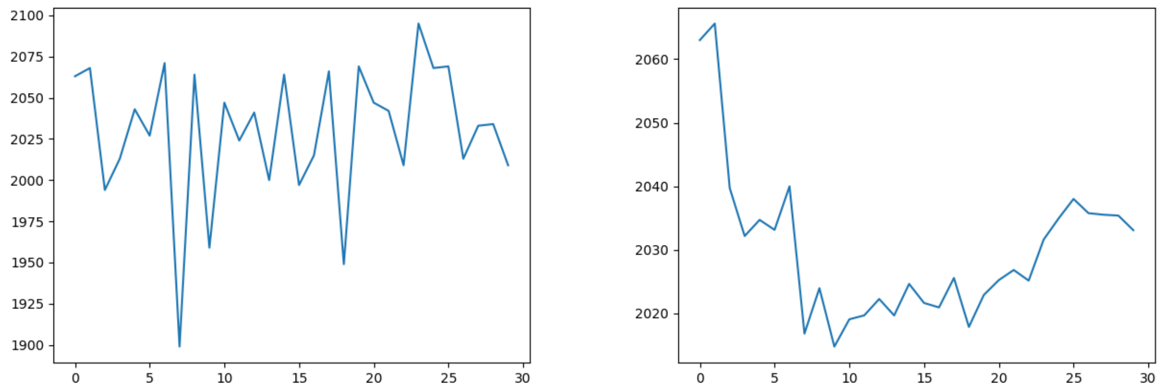
These metrics are scraped from Pybaseball so that for each fatigue event the time series of each metric from the beginning of the pitchers' outing up until the fatigue event is stored.

Next, a smoothing algorithm is applied to the time series of metrics in order to obtain a sliding average of pitch metrics over the course of the game. This is necessary as pitching metrics show a lot of noise over the course of a game, with very high variability from pitch to pitch. The following exponential smoothing algorithm is applied such that the time series of a certain pitcher's metrics at any given pitch number in an outing represents their mean performance with that pitch and statistic over a sliding window of all previous pitches:

$$F_{t+1} = \alpha(A_t) + (1 - \alpha)F_t \quad (3)$$

where A_{t+1} is the $t+1$th observation in the series, and α is a decay factor determined by the desired half life of any particular pitch in the time series. The decay factor is set such that the half life $T_{1/2}$ is 8 pitches, so that the cumulative averaging occurs over a sliding window of length equal to an average inning. This way the metric data recorded in the time series represents the pitchers average performance over the course of the previous inning.

Exponential smoothing is chosen so that more recent observations are weighted more heavily in the course of averaging, in order to attempt to capture the cumulative impact of fatigue. This captures the way that the previous pitch thrown by a pitcher better demonstrates their current level of fatigue than a pitch which was thrown longer ago, so the markers present in that pitches metrics should be weighted more heavily. The results of exponential smoothing are vizualized for a time series below:



Figures 12, 13: Unsmoothed Vs. Smoothed Fastball Spin Rates for Marco Gonzales, 7/3/21

Explanatory analysis may now be applied to these time series to understand the relationship between these fatigue proxies and pitching performance, by analyzing relative changes between the time series to understand which metric is most strongly correlated with the ensuing decrease in performance. This makes the reasonable assumption that larger decreases in the metrics selected as fatigue proxies cause larger decreases in pitch efficacy. In order to determine which time series show the largest changes over the course of an a game, first, the delta of each metric is taken over the smoothed time series, such that $\Delta F = F_{event} - F_0$, giving the total change in average metrics between the beginning of a pitchers performance and their performance at the time of a fatigue event. Define this delta as the drift of a metric m in an event e, $D_{m,e} = \Delta F$ Since this ompares with the pitcher's baseline performance on the day of the game instead of the pitcher's average performance, confounding variables such as the stadium, opponent, play condition, etc. are screened out as much as possible.

In order to compare the deltas of each time series with each other, for a cluster c, an average delta of each metric is taken across the set of all events identified for a the cluster E:

$$\bar{D}_{m,c} = (\sum_{e \in E} D_{m,e})/|E| \quad (4)$$

In order to compare metrics to each other, measurement standardization must be applied so that the average drifts are comparable to each other. In order to do so, apply mean standardization so that drifts are defined in terms of standard deviations of each metric m, so that each drift $d_{m,c}$ is converted to the average number of standard deviations by which the exponentially smoothed metrics change over the course of a time series approaching a fatigue event:

$$d_{m,c} = (\mu_m - \bar{D}_{m,c})/\sigma_m \quad (5)$$

Now, time series are comparable in an explanatory sense so that the drifts in each metric relative to each other demonstrate the importance of each fatigue-proxy metric correlated to the occurrence of a fatigue event.

V. Results

The following table presents the drift $d_{m,c}$ for each metric m measured across each cluster c for all events of that cluster. Measurements are standardized in number of standard deviations.

Cluster	FB Velocity	BB Velocity	CH Velocity	FB Spin	BB Spin	CH Spin	CU Vert. Break	SL Hor. Break	Release Height
0	-0.143	0.132	0.028	-0.84	-1.43	-.25	-0.126	-0.386	0.134
1	-0.068	-0.229	-0.031	-1.25	-0.019	-0.83	0.05	-0.162	0.017
2	-0.213	-0.23	-0.16	-0.71	-0.67	-0.65	-0.016	-0.269	-0.45
3	-0.008	-0.136	-0.22	-1.31	-0.97	-0.07	-0.449	-0.283	-0.304
4	-.208	-0.095	-0.039	-0.32	-1.57	-0.89	-0.512	0.173	-0.41
5	0.079	-0.256	-0.305	-1.83	-0.45	-0.33	-0.368	-0.216	-0.038
6	-0.151	-0.269	0.16	-0.56	-0.27	-1.34	0.154	-0.481	-0.043
7	0.153	-0.392	-0.378	-1.19	-0.28	-0.37	-0.321	0.127	0.144

Table 5: Time Series Metric Drifts per Cluster

It is clear from the above computations that the drift in pitch velocities throughout a game do not present a significant correlation with the occurrence of fatigue events. With velocities presenting drifts of magnitude less than half of a standard deviation from the mean without a clear pattern in relation to the pitch usages of each cluster, there is no evidence here that pitch velocities are a marker of fatigue or that changes in pitch velocity are a typical cause of offensive advantage for the opponent.

Similarly, there is no evidence shown here that a pitcher's fatigue presents itself in the form of breaking ball break distances or release heights. Break distances present slightly higher magnitude drifts for those pitcher clusters who utilize their breaking balls a higher percentage of the time (clusters 4 and 6), however the correlation is weak and still the drift is around only a half of a standard deviation from the mean.

From the computed drits, it is clear that of the metrics studied in this account, spin rates show the clearest correlation to the occurrence of fatigue events, and present a pattern across the clusters. For each cluster, the spin rate of the primary pitch thrown by the cluster (the pitch type thrown with the highest frequency on average across the cluster) presented a drift of -1.33 standard deviations on average, and secondary pitches showed a deviation of .7 deviations on average, both showing significant drift over the course of a pitcher's performance for each cluster.

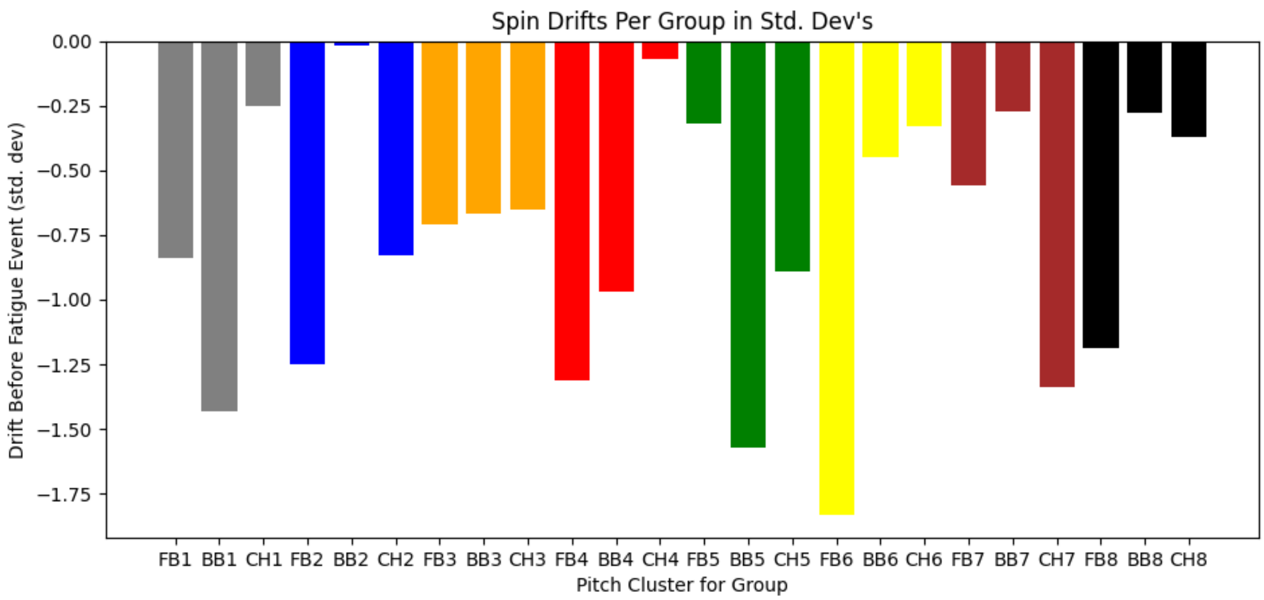


Figure 14: Spin Drifts Per Pitch Across Clusters

The important distinction in which pitch's spin rate shows the highest drift is a clear indication that the clustering method applied is justified. With primary pitches showing the greatest drifts, we have unconvered the relationships between how a pitcher strategically uses their arsenal and the most significant markers of fatigue seen in the performance of a pitcher. Without having applied the clustering, the drift in spin rates would have been the higher static in the group of all pitchers, but the relative amount of drift per pitch would be lower, and the lower drift metrics observed in pitchers who throw certain pitches less frequently would have brought down average spin rate deltas, masking the significance of the spin rate drift in correlation with the occurrence of fatigue events.

VI. Discussion and Conclusion

The prevalence of spin rate as the predominant indicator of pitcher fatigue runs contrary to expectation, in that declining velocity is typically assumed to be symptomatic of fatigue. While overall the results of this study do show that velocity tended to decrease over the course of a pitching performance, it was by a far less significant extent than the spin rates in those instances where it was identified that fatigue played a role in giving the opposing team an offensive advantage. This significance of spin rate as a fatigue marker has a number of implications for pitcher usage, strategy, and training.

For teams who play at levels for which statcast data is available, the prevelance of spin rate as a fatigue marker allows a team to monitor the fatigue of their pitcher closely by tracking the spin rate data aquired by the statcast system as the pitcher pitches. The smoothing algorithm aplplied in this paper can be applied in real time to track filtered time series of spin rate data, with the understanding that the higher the spin drift, the greater the likelihood of a fatigue event which deeply impacts the team's chances of winning occur. A focus of future work could be to identify risk thresholds of spin rate drift such that a team could understand the risk they take when leaving in a pitcher whose spin rate has drifted by certain margins. Dependent on the game situation including the score and the significance of the game, teams can then make more informed decisions about how willing they are to leave a pitcher in who is showing these specific indicators.

In training pitchers to increase their stamina and longevity, it is clear that avoiding negative drift in spin rate should be a major goal. While which biomechanical factors influence spin rate and how to train those is worthy of a future study, it is currently clear that grip strength must play a major role in spin rate, as the force imparted by the fingers on the seams of the baseball is what causes the ball to spin. With this being the case, forearm and grip stamina present themselves as training targets for pitchers seeking to increase their stmaina and resistance to fatigue. In pitching training, for those pitchers who have statcast available in their training facilities, tracking their own filtered time series of spin rate would provide real time feedback during their bullpen sessions as to their current fatigue level.

Similarly, this result can be used for recruiting starting pitchers. Starting pitcher stamina is a very important target to seek when recruiting strarting pitchers, as starting pitchers who are able to throw more innings effectively allow teams to utilize their bullpen pitchers less frequently and therefore have more pitchers available in strategic situations. When evaluating the performance of starting pitchers for their stamina, the results of this paper can be utilized by tracking the spin rate drift of pitchers when statcast data is available, uncovering potential fatigue markers even when the offensive team does not take advantage of them, helping to provide an opponent-blind method of investigating a potential recruit's stamina.

VII. References

[1] David C. Phillips. Stopping on Nine: Evidence of Heuristic Managerial Decision-Making in Major League Baseball Pitcher Substitutions. *Southern Economic Journal* 84(2): 577-599, 2017.

[2] W. Jeffrey Grantham et al. The Impact of Fatigue on the Kinematics of Collegiate Baseball Pitchers. *The Orthopaedic Journal of Sports Medicine* 2(6), 2014.

[3] David Whiteside et al. Variations in Pitching Performance During a Major League Baseball Game: What Can We Learn From Ball Tracking Data? *International Conference of Biomechanics in Sports* 33, 2016

[4] Michael Woodham et al. When to Pull Starting Pitchers in Major League Baseball? A Data Mining Approach. *Applied Data Science*, 2016

Appendix I: Link to Data and Code

Full data and code may be found at: https://github.com/Espeer5/PullingPitchers