Beyond the Stats: Analyzing the Effect of Fatigue on Pitchers in Extra Innings

Heiby Lau
University of Alberta
Undergraduate Division

Abstract

Traditional baseball scouting often relies on game statistics, which fail to capture nuanced factors influencing player performance, especially fatigue and endurance. Noticeably, there is a lack of fatigue analysis of players in the farm systems. This can lead to suboptimal training, management, and rehab strategies for developing pitchers. Therefore, we propose that conducting such an analysis will provide a clearer understanding of the factors distinguishing each class, particularly between High-A and Double-A. Building on existing studies of pitcher fatigue in MLB, this paper uses cluster and regression analysis to explore the relationship between fatigue and performance in extra innings. Insights from this analysis, as well as a UI tool we developed in this project, augment the current approach to player assessment and assist coaches in making more informed decisions regarding pitcher usage and development, particularly during the crucial Double-A jump.

1 Introduction

“Almost all relief pitchers are failed starting pitchers.”[1] This was the stereotype of the relief pitchers when the spotlight for the best pitchers in the league was rarely focused on the bullpen. But in 2018, the Tampa Bay Rays created a new strategy called “the opener”, where they utilized relievers to take the place of starting pitchers to pitch the first one to three innings. This sparked a new discussion where teams began to consider the traditional roles of starters and relievers. In the same year during the World Series, the Boston Red Sox put Nate Eovaldi and Chris Sale, a pair of starters, in the bullpen warming up for the ninth inning[2]. The seemingly radical strategies for both teams turned out to be beneficial, where the Rays’ openers combined an ERA that was better than the league average, and for the Red Sox, winning the World Series.

When we discuss the reason behind such moves in the pitching rotation and bullpen management, we tend to believe pitcher roles are determined by statistics without context. While performance is certainly a factor, the shift in pitching roles also reflects a recognition of the negative impacts on starters who were traditionally trained to pitch as many innings as possible, leading to increased fatigue and risk of injury. Looking back at the Rays, one of their starting pitchers, Blake Snell, went on the disabled list with shoulder fatigue[3]. While there are fatigue analyses conducted on pitchers in the MLB, such as the ones done by Dr. Michael Sonne and Dr. Jim Potvin[4], fatigue analyses are not extensively applied in the MiLB.

In the world of scouting, the Double-A jump[5] has been seen as the most crucial part of a prospect’s future. In addition to the gap in game performance and quality of pitch, we hypothesize that fatigue significantly affects pitcher’s performance, measured by WHIP and K/IP, in extra innings. We also suspect that as pitchers in Double-A exhibit a resilience to fatigue and better muscle recovery, resulting in performance metrics that are less adversely affected compared to those in High-A. By incorporating fatigue, endurance, and recovery metrics into performance analysis, we aim to gain a clearer understanding of the factors distinguishing High-A and Double-A pitchers. This insight could assist in developing targeted training and management strategies specifically tailored to help pitchers successfully navigate the challenges of the Double-A jump, ultimately enhancing player development and success throughout the farm system.

2 Data

For this project, we used data from the 246 games played in season #1884 across the four home teams under a single farm system. We corresponded the levels in the data to the classes in a typical MLB farm system as follows:

Level	Class
1A	Single-A
2A	High-A
3A	Double-A
4A	Triple-A

Table 1: While our focus was on High-A and Double-A, incorporating data from all classes could reduce potential bias, such as model overfitting, and provide the flexibility to extend the analysis to the entire farm system in the future.

We decided not to use data from season #1883 since there was noticeable amount of missing or duplicated data being found, as well as the fact that the data only accounted for half of the season.

We noticed and attempted to correct some of the errors found in the season #1884 data. One example was that a .csv file had "\N" in the cells where the values were supposed to be NA. We manually replaced all such instances with the NA value so that the file could be read properly. We excluded data where pitchers from away team(player_id > 1000) were identified as the pitcher during a top inning. We also wanted to address that some games had an inning of missing values in the game_info files. We ignored those innings in our analysis as attempting to reconstruct the actual play was not time efficient and could potentially lead to more errors.

3 Methodology

The following table summarized our definition of fatigue, endurance, and recovery:

Metric	Definition
Fatigue	Muscle fatigue[1], Workload fatigue(Fatigue unit)[2]
Endurance	Total innings pitched, Total games played
Recovery	Average rest days between games

Table 2: Summary of definition of each metric

1, 2: See Appendix for calculations

We used a regression analysis approach for this project. We defined the features as the pitcher's level, as well as the definitions listed in Table 1. We tested three regression models: ridge regression, random forest, and XGBoost, using a 60/40 train/test split.

Model	R^2(WHIP)	R^2(K/IP)
Ridge regression	0.42	0.67
Random forest	0.20	-0.24
XGBoost	0.02	-0.39

Table 3: Performance of each model when predicting WHIP and K/IP in extra innings

The inconsistent performance among models suggested that a pitcher's playing level, fatigue in regular innings and recovery time had **no significant** effect on their performance in extra innings. As a result, another approach needed to be taken. We identified that the regression analysis approach failed to take into the account of pitcher roles. For example, closers specialize in getting strikeouts in high-leverage situations while relievers focus on bridging the gap between starters and closers by maintaining stability. Recognizing these role-specific dynamics, we decided to explore a **cluster analysis** approach. We used the same features as the regression analysis to group pitchers into **four** distinct clusters.

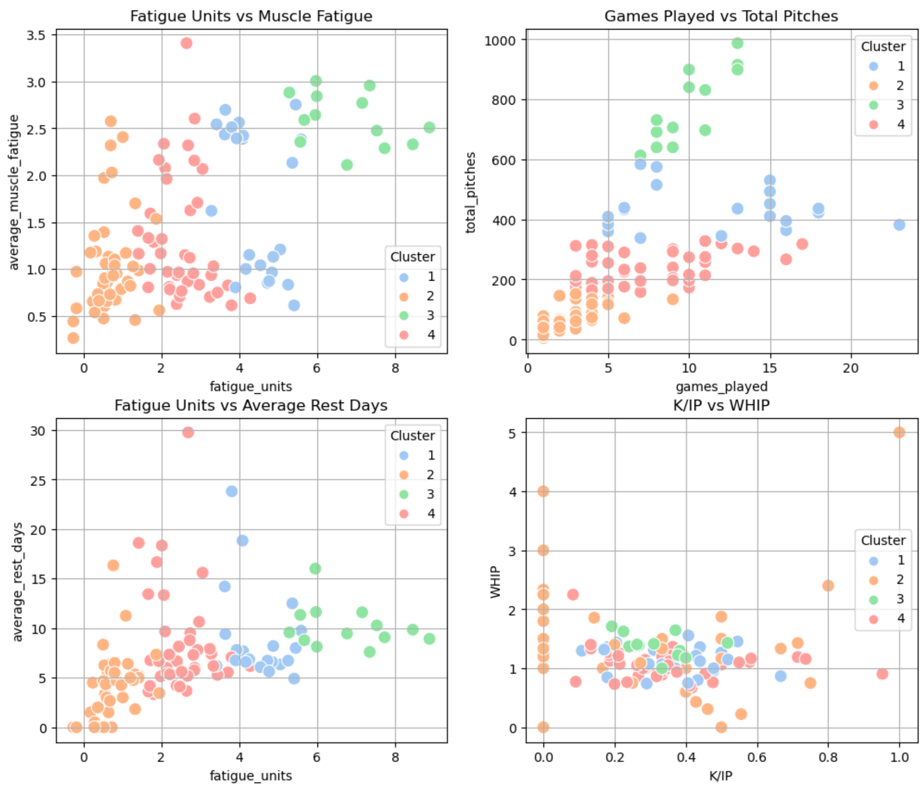

Fig. 1: Distinct patterns can be observed when comparing the fatigue, endurance, and recovery metrics among clusters. However, performance metrics (WHIP and K/IP) alone cannot distinguish one cluster from another.

This time we used the fatigue, endurance, and recovery metrics to predict which cluster a pitcher belonged to. We tested two classification models: XGBoost and random forest, using a 60/40 train/test split. The random forest classifier yielded a better accuracy score of 0.90.

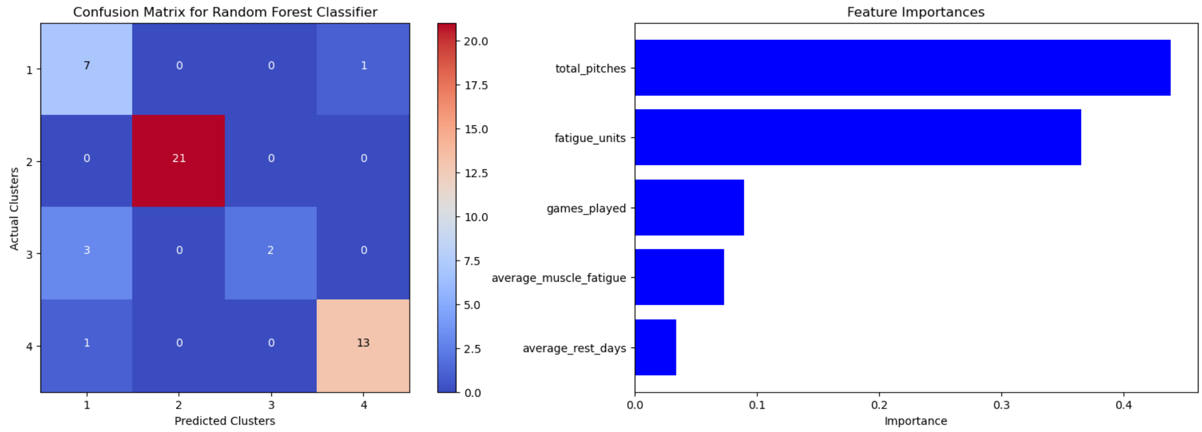

Fig. 2: Confusion matrix and the feature importance of the random forest classifier model.

4 Results

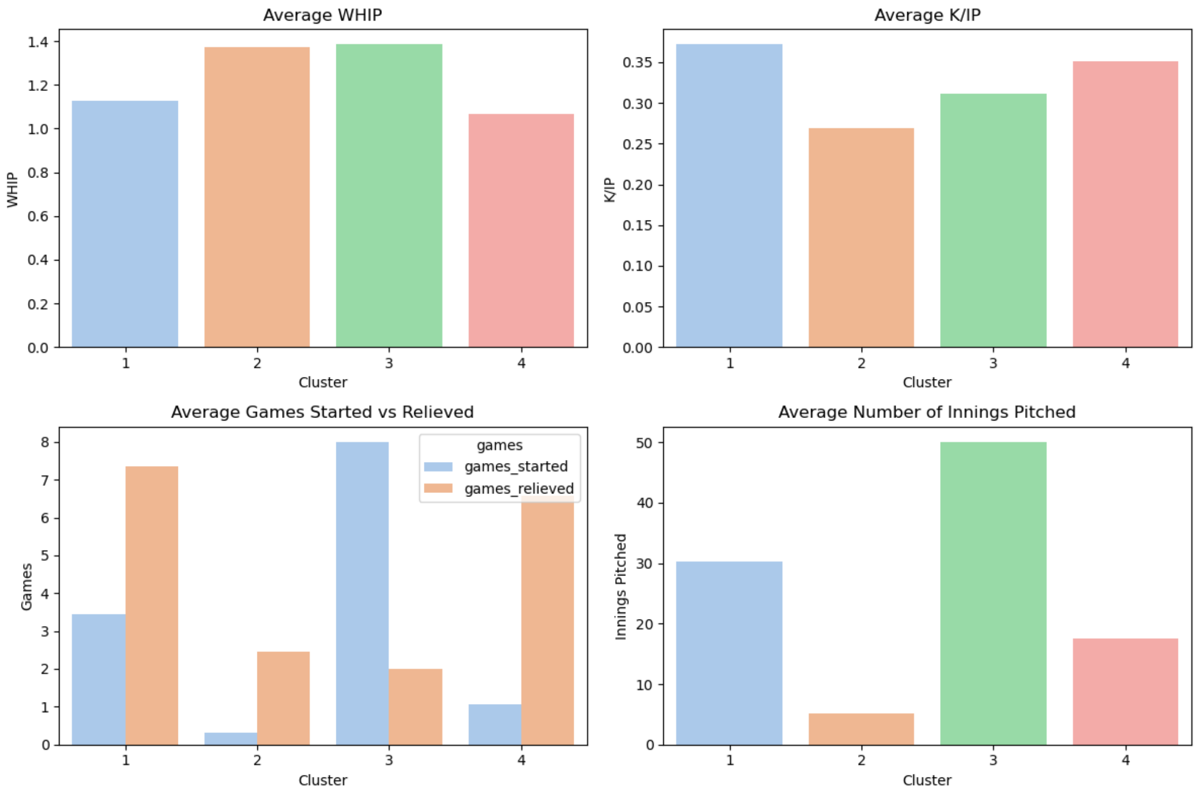

Fig. 3: Key characteristics of each cluster.

From the result of the cluster analysis in Fig. 3, we suggested the following role for each cluster: cluster 1 as **versatile relievers**, cluster 2 as **middle relievers**, cluster 3 as **starting pitchers**, and cluster 4 as **closers**.

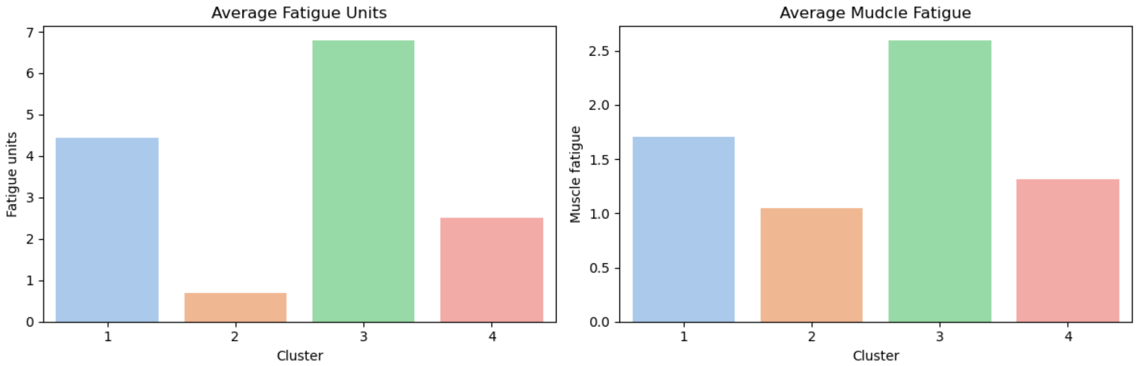

Fig. 4: The high workload and muscle fatigue that starting pitchers experienced in the season can make them be prone to injuries in the future. One way to prevent this is to utilize the versatile relievers as starters more.

The distinct duties of versatile relievers, middle relievers, and closers contributed to their management of fatigue and performance in extra innings. Versatile relievers, tasked with handling various situations throughout a game, prioritized consistency and adaptability. This allowed them to maintain a low WHIP in extra innings, although their focus on inducing contact rather than overpowering hitters resulted in a lower K/IP. Fatigue was not a main factor for extra inning performance by middle relievers, but rather their game experience. As they appear on the field once in a while, they often faced inconsistent situations, leading to fluctuations in their performance under the increased pressure of extra innings. Closers, who routinely handle high-pressure situations at the end of games, were trained to maximize strikeouts and shut down opponents even if they became exhausted. Therefore they were least affected by fatigue and managed to deliver effective performance in extra innings.

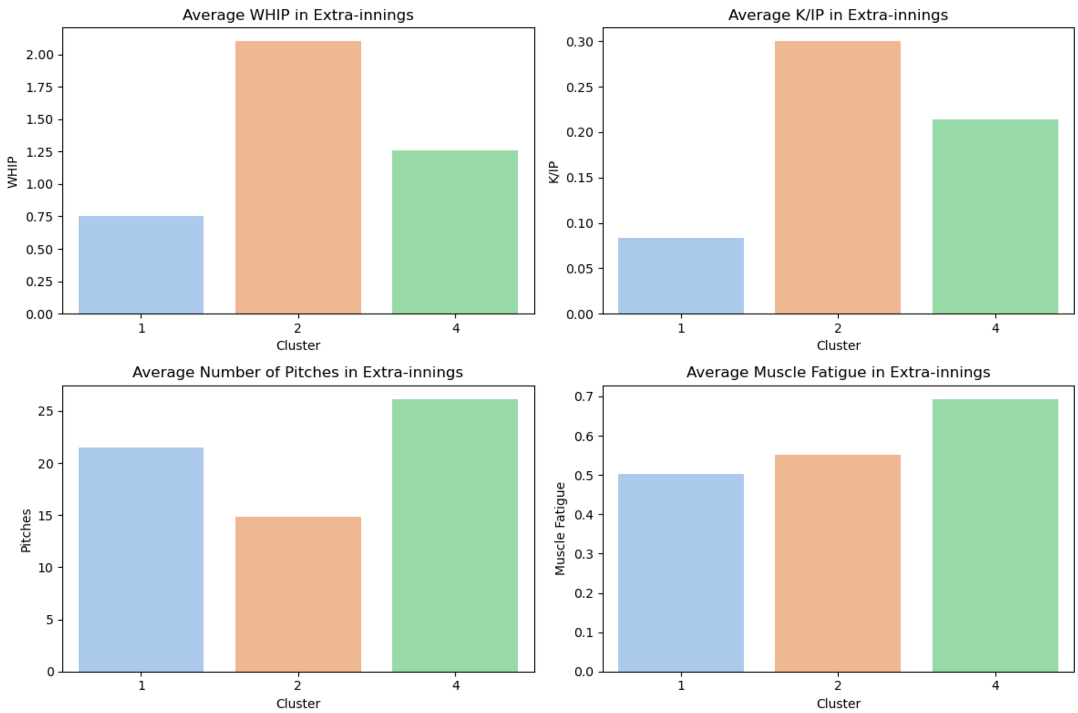

Fig. 5: Performance of each cluster in extra innings. While there is no significant relationship between fatigue in regular innings and performance, we can speculate that fatigue lead to less power in the throw by versatile relievers, leading to a lower strikeout rate.

Looking back at the farm system, there is a higher proportion of closers and versatile relievers in Double-A than in High-A. This shift suggests that in Double-A, teams prioritize pitchers who can handle the complexities and demands of extra innings. The increased presence of roles known for their resilience to fatigue supports the idea that Double-A teams are strategically equipped to handle extra innings. On the other hand, the heavy reliance on middle relievers in High-A can lead to a lack of role specialization. This composition may hinder pitchers' development, as they might not gain the experience needed to excel in high-pressure, fatigue-intensive environments, leading to inconsistent performance in extra innings. As a result, High-A pitchers may face difficulties when transitioning to Double-A, where there are more pitchers who are better adapted to managing fatigue.

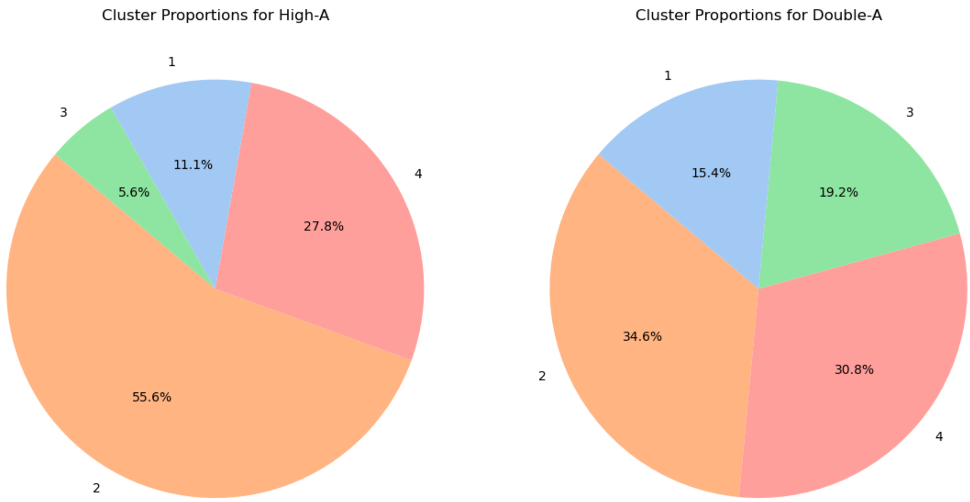

Fig. 6: The increased proportion of more specialized roles(closers, versatile relievers, and starters) may suggest that only those pitchers who are capable of handling fatigue are going to get promoted.

5 Application

In Section 3, we have trained and tested a random forest classifier model that predicts the cluster that a pitcher belongs to based on his fatigue, endurance, and recovery metrics. We have built a website incorporating the model using Plotly.

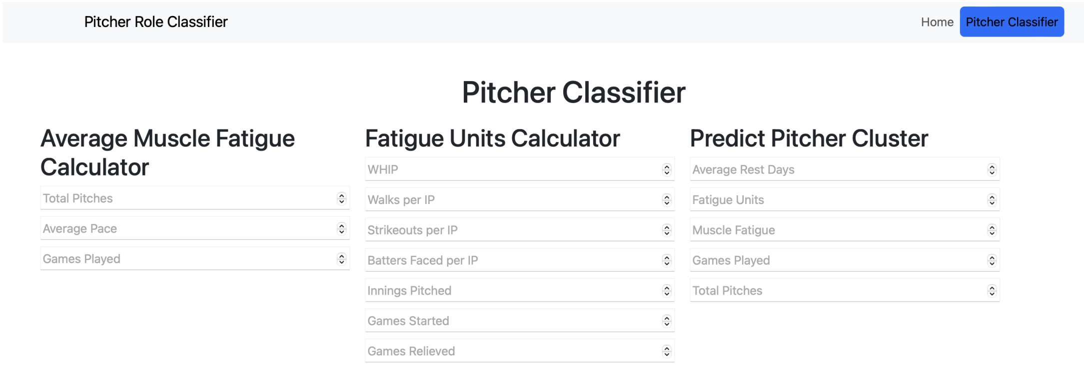

Fig. 7: Users can first use the calculators to calculate muscle fatigue and fatigue units for a pitcher, then using those outputs in the "Predict Pitcher Cluster" section to get the predicted role.

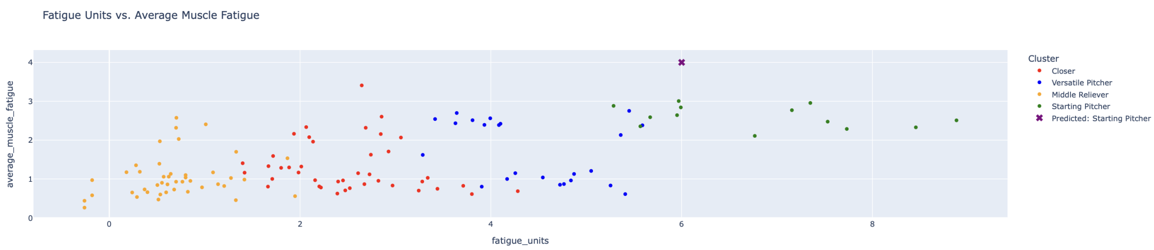

Fig. 8: Users get to see that pitcher's performance relative to the pitchers in season # 1884 similar to Fig 1.

6 Conclusion & Future Improvement

In this paper, we have explored the relationship between pitcher roles, fatigue, endurance, and recovery metrics in extra inning performance within a baseball farm system. Initial regression analysis demonstrated limited results, suggesting that a pitcher's level of playing and their fatigue in regular innings do not have a direct relationship with their extra inning performance. The shift to a cluster analysis approach revealed more meaningful insights by recognizing the role-specific characteristics and duties of pitchers. Cluster analysis identified four distinct roles: versatile relievers, middle relievers, starting pitchers, and closers. Each role exhibited unique performance attributes and challenges in extra innings. Versatile relievers excelled in maintaining low WHIP through consistency and adaptability, while middle relievers faced variability in performance due to their bridging role. Closers were most effective in high-pressure situations, maintaining high strikeout rates even with fatigue. The analysis further reveals strategic differences between High-A and Double-A teams. The Double-A team prioritizes closers and versatile relievers, adapting to the demands of extra innings. In contrast, the High-A team rely heavily on middle relievers, which may impede pitchers' development and transition to Double-A or even higher levels where fatigue management is critical.

While the results of the analysis are promising, there are opportunities for further improvement. One area for enhancement is the calculation of muscle fatigue, where incorporating pitch velocity could be beneficial, as the power of a throw significantly contributes to fatigue. Additionally, we assumed that the number of pitches thrown per inning is as important as fastballs thrown per inning when substituting variables. Although this assumption appears reasonable, it might lead to inaccuracies in fatigue estimation. This is because the Sonne & Potvin model[4] did not originally use the number of pitches thrown per inning as a feature, which may partly explain why the initial regression analysis was unsuccessful. Another significant assumption is that the classification of pitcher roles is based solely on data from one season. To make the model more robust, incorporating data from multiple seasons would be advantageous. Finally, in MLB, there are more subcategories of relievers that this paper has not considered, which could further refine the analysis and enhance its applicability.

7 Acknowledgement

We would like to thank Dr. Meredith Wills for her support and guidance in this project, as well as Cameron Adams for the feedback for this paper. We would also like to thank Andrew Steenkamer, Kenny Bores, and fellow participants for their contributions on the Slack channel.

Appendix

The calculation of muscle fatigue is inspired by a regression model developed by Dr. Michael Sonne & Dr. Jim Potvin[4]. Due to lack of classification of pitch types(fastball, curveball, etc.) in the dataset, we cannot directly use fastballs thrown per inning(*#fastballs*) as a variable suggested by the Sonne & Potvin model. To rectify this, we replace the variable by pitches thrown per inning(*#pitches*). The modified equation for average muscle fatigue is as follows:

$$aFatigue = 0.10963 + 0.0032 * \#pitches - 0.0023 * pace(s)^{\mathbf{1}} \tag{1}$$

1. For definition and calculation of pace, see Fangraphs[6]

The calculation of workload fatigue(also called fatigue units) comes from a regression model developed by Dr. Sonne[4]. The equation for predicted workload fatigue units is as follows:

$$pFatigueUnits = 0 + 0.18 * WHIP + 0.14 * BB/IP + 0.34 * K/IP - 0.16 * Batters/IP + 0.08 * IP + 0.3 * Games\ Started + 0.14 * Games\ Relieved \quad (2)$$

References

[1] Rigney, Jed. *Ranting and raving about the All-Star Game ... again.* TTFBaseball, 20 July 2013, https://throughthefencebaseball.com/all-star-game-rants. Accessed 17 July 2024.

[2] Miller, Sam. *Red Sox starters won the World Series – from the bullpen.* ESPN, 29 Oct. 2018, https://www.espn.com/mlb/story/_/id/25024544/boston-red-sox-starters-won-world-series-bullpen. Accessed 17 July 2024.

[3] Taylor, Jon. *How the Tampa Bay Rays Reinvented the Concept of Starting Pitching.* SI, 23 Aug. 2018, https://www.si.com/mlb/2018/08/23/tampa-bay-rays-bullpen. Accessed 17 July 2024.

[4] Sonne, Mike. *An Introduction to Fatigue Units: A New Method for Evaluating Workloads.* Fangraphs, 1 Apr. 2017, https://fantasy.fangraphs.com/an-introduction-to-fatigue-units-a-new-method-for-evaluating-workloads/. Accessed 17 July 2024.

[5] Skillin, Alex. *The Double-A Jump.* Fangraphs, 22 Sep. 2015, https://tht.fangraphs.com/the-double-a-jump-2/#. Accessed 17 July 2024.

[6] Weinberg, Neil. *Pace.* Fangraphs, 25 May 2015, https://library.fangraphs.com/offense/pace. Accessed 17 July 2024.