Automated Force-Velocity Profiling of National Football League Athletes

by

Mark Joseph Wright

B.S. Computer Science and Engineering, Massachusetts Institute of Technology, 2021

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Engineering in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2022

Author . . . .
Department of Electrical Engineering and Computer Science
May 6, 2022

Certified by . . . .
Anette (Peko) Hosoi
Neil and Jane Pappalardo Professor of Mechanical Engineering
Thesis Supervisor

Accepted by . . . .
Katrina LaCurts
Chair, Master of Engineering Thesis Committee

Automated Force-Velocity Profiling of National Football League Athletes

by

Mark Joseph Wright

Submitted to the Department of Electrical Engineering and Computer Science
on May 6, 2022, in partial fulfillment of the
requirements for the degree of
Master of Engineering in Electrical Engineering and Computer Science

Abstract

Force-velocity profiles are a well-established approach to generating key parameters of an athlete's overall fitness profile. They are currently utilized by NFL teams for their players. However, athletes run the risk of injury while testing to create these profiles since they must sprint with a weight attached to them at max speed. As such, teams are not utilizing these profiles as well as they could as they prefer not to jeopardize their athletes.

In this paper, we present a novel approach to generating force-velocity profiling inspired by former work in the MIT Sports Lab to create these profiles directly from tracking data generated by wearable technology sensors. The techniques presented in this paper allow NFL teams to create force-velocity profiles over any time frame of tracking data they have available and allow them to better assess, train, and rehabilitate their players.

Thesis Supervisor: Anette (Peko) Hosoi
Title: Neil and Jane Pappalardo Professor of Mechanical Engineering

Acknowledgments

There are so many people I need to thank for getting me to the point in my life where I am able to submit this thesis. First and foremost, I would like to thank Peko and the entire MIT Sports Lab for the support I've received during this process. I wouldn't have been able to complete this work without your insights into where to go next in the thesis. Peko, you have a way of coming up with ideas that are simple and yet make so much sense as a method of going forward with an idea. I would also especially like to thank Kevin Lyons who led the charge on generating force-velocity profiles directly from tracking data in his thesis last year. Not only that but also thank you for being a brother I could always ask questions when I got lost on what to do next.

Without my family, I would have never gotten into MIT, let alone considered graduate school or writing a thesis. Mom, Dad, Jeff, John, Taylor, Arthur, Ale, thank you for being the most important thing in my life and for helping shape me into the man I am today. There's no way to put into words how important each of you is to me, but I hope you all know how much I appreciate and love each of you.

I also want to thank all of my fraternity brothers at DKE and members of the MIT football team for being a great group of friends and helping push me through MIT. I will always cherish my time at MIT thanks to y'all. I want to especially acknowledge DKE 21' for living together during the pandemic year and permanently bonding us together during those times. Those were some of the best days of my life. Dan West, you have been a rock during all of MIT and you deserve a special shoutout for keeping up with us despite being a boomer.

I finally want to thank all my other friends that I haven't mentioned yet since they all provide so much to my life and experience. Better than Goodell deserves a special shoutout for getting me interested in football way back in the day and for still being so tight-knit even after high school, especially you Dusty for having the most quotable name of all time.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Professional sports has always been a field where small advantages can mean the difference between winning and losing. The legendary Vince Lombardi once said that "Football is a game of inches and inches make the champion", and nowhere is that more true than in the modern data-driven era of professional sports. Teams have whole data analytics teams developing new techniques in an attempt to get a leg up on their opponents.

Techniques to get the better of your opponent tend to fall into two main categories: analyzing the other team's tendencies/players and finding strengths/weaknesses in your own team. One technique that has come into vogue in recent years is the concept of a force-velocity (F-v) profile. These curves present teams with three key metrics that can help a trainer or a coach understand how an athlete may perform on the field: maximum theoretical force output, maximum theoretical velocity, and maximum theoretical power output. They are plotted using curves similar to figure 1-1. It has been empirically established that maximum force output is a decreasing linear function of velocity [2]. Techniques to get these curves are well established for sprinting and weightlifting applications[5, 18].

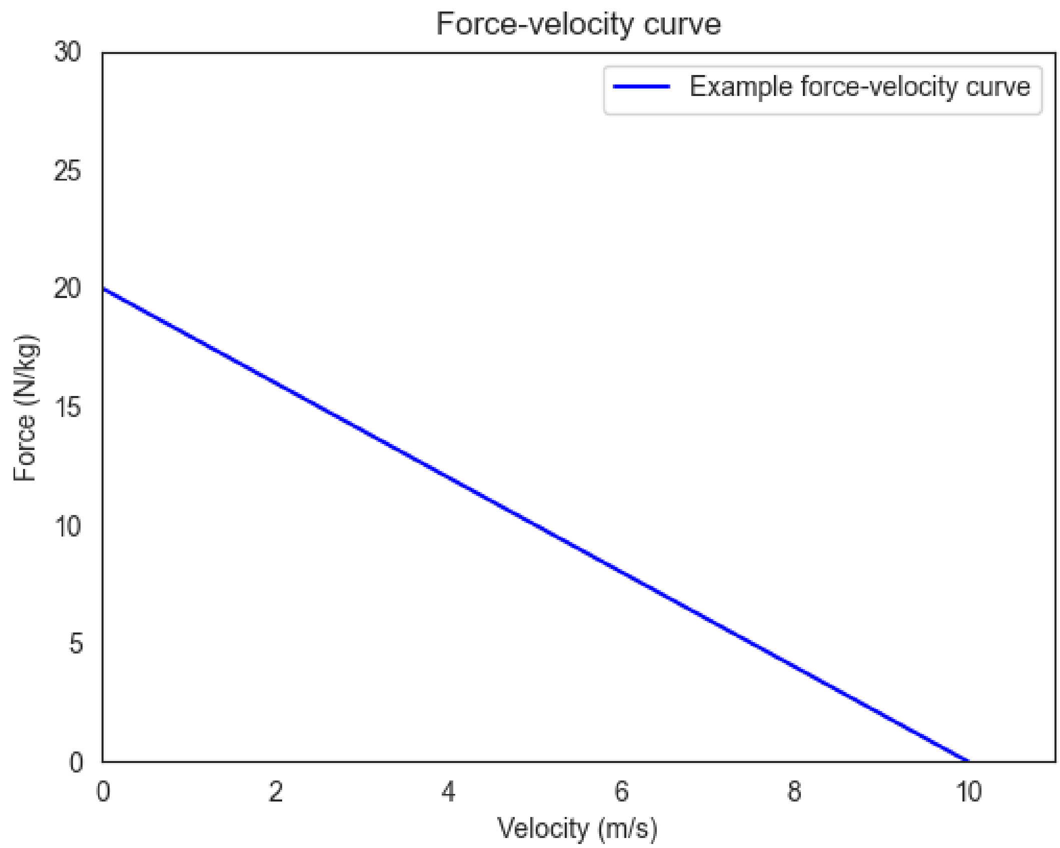


Figure 1-1: Example of a model force-velocity profile, where maximum theoretical force is the y-intercept and maximum velocity is the x-intercept.

1.2 Why Football?

Techniques to generate these force-velocity profiles are easy to implement for sprinters and weightlifters due to the maximum exertion requirements in both of those sports. When sprinting, you are continually accelerating until you reach your maximum velocity so it is simple to calculate this profile by tracking an athlete's speed at various points in the run and fitting a line.

In football, it becomes more complicated. During a football play, the different responsibilities of different position groups and the contact nature of the sport mean that a player may not reach their peak output in every play. Additionally, they may not accelerate in the same way that a sprinter might at the beginning of a play. This makes it more difficult to generate force-velocity profiles from natural football actions.

Due to this, most NFL teams generate these profiles by having their athletes run a sprinting test to generate a force-velocity profile. These are difficult to get done consistently and add an increased risk of injury to athletes.

1.3 Contributions

Due to the risks discussed above, National Football League (NFL) teams do not test for force-velocity profiles typically during the season unless a player is injured. Force-velocity profiles are still useful to have, however, so this paper presents a novel technique to generate force-velocity profiles directly from tracking data. The benefits of this technique are evident as it utilizes data that teams already track for their players in both practice and games. We leverage this data in several ways to come up with key insights about each player that can be used by NFL and college teams to further assess their players for potential.

1.4 Outline

Chapter 2 discusses the related work that contributed to the inspiration and development of this paper and presents the current state of generating force-velocity profiles in the NFL. Chapter 3 describes the techniques used to generate force and velocity data directly from NFL tracking data as well as a brief background on football and the dataset we are using. Chapter 4 details the novel approach taken to generate force-velocity profiles directly from tracking data as well as a kernel density estimation approach to creating a load analysis estimate for athletes. Chapter 5 names several applications to professional sports and some of the benefits of our approach before chapter 6 evaluates the accuracy of our model in the context of different position groups. Chapter 7 concludes with the key insights of our model as well as how future work could expand upon the work presented within this paper.

Chapter 2

Related Work

This chapter breaks down the prior work in the field of utilizing force-velocity profiles for improving athlete performance. There has been extensive work in this field on the best way to derive force-velocity profiles as they can inform the training and rehabilitation of athletes to produce better results.

2.1 Applications to Sprinters

Work on computing the relationship between muscle shortening and lengthening and a force-velocity profile began in 1938 in Hill's seminal paper [7]. He theorized that there was an inverse hyperbolic relationship between force and velocity, which was later directly measured to be true but only for one muscle. Future work has shown that the relationship for multi-joint activities is quasi-linear and can thus be reasonably well fit with a simple linear model [19].

The application of force-velocity profiles to sports began with sprinters since there is an obvious benefit to improving horizontal force, velocity, and power output: winning more races [13]. The work done by Keller to compute the optimal velocity for a race laid the groundwork to understand how to develop a race strategy to achieve this optimal result [9]. Figure 2-1 shows the result he derived for a typical sprinting race with a clear asymptote for the maximum speed.

There have been a multitude of different ways to calculate the force-velocity profile

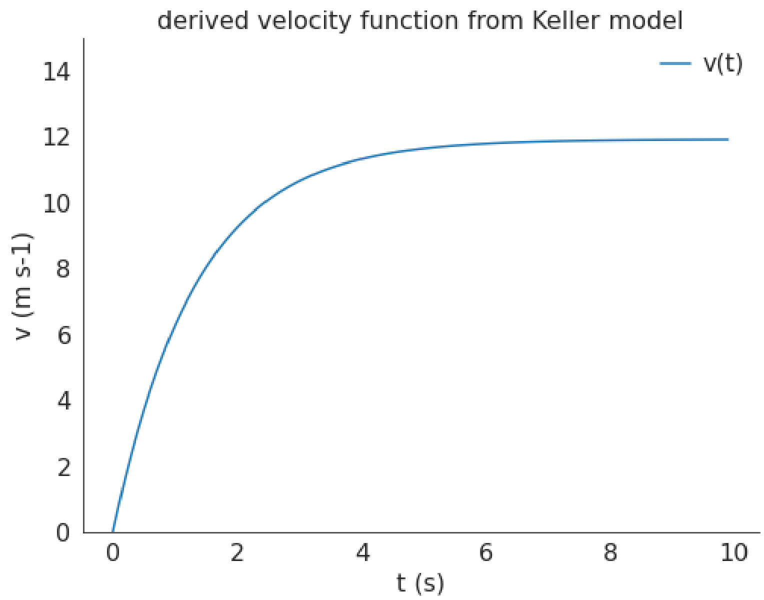


Figure 2-1: Velocity function from the Keller model of sprinting.

of athletes [13, 17], but all methods share one commonality: they require a specific test and setup to calculate their results. These take the form of various different movements: jumping as high as possible, bench pressing for speed, or sprinting through a set of velocity sensors [17]. The most common test that involves sprinting involves setting up an athlete with a known weight attached to them (ideally a small weight). The athlete then sprints for a set distance and their speed and time between different points over the run are recorded. Those can be fit using the model from Keller [9] and the maximum force and velocity can be derived from the test.

This test informs how to train a sprinter and see what sort of strength imbalance they hold. The optimal profile for the distance that the athlete runs can be computed and then used to train the athlete. Figure 2-2 shows how a strength imbalance can look once it has been identified. This athlete is more speed-biased, suggesting that they could use training in force to reach their optimal profile. Charts like these help coaches to target athlete training to get better at their event, or sports medicine staff to help get an injured athlete back to their previous baseline.

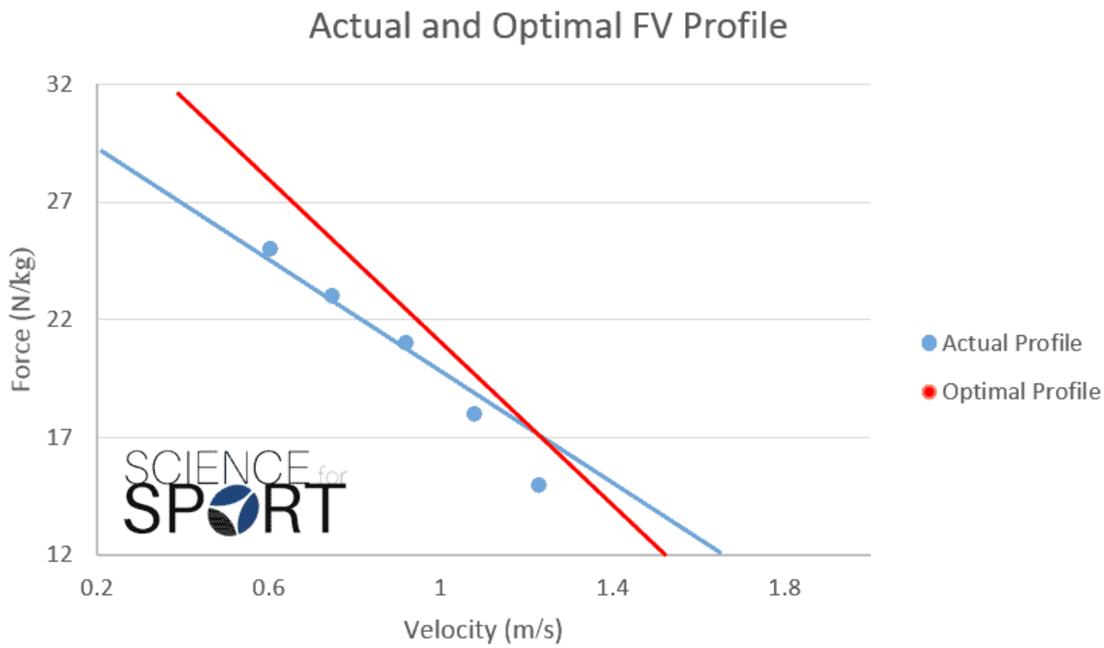


Figure 2-2: Example of how a strength imbalance can be identified using a force-velocity profile [19].

2.2 Current Use in the NFL

Force-velocity profiles are currently used by many NFL teams to help train and rehab their athletes. However, there are a few issues with football specifically that make it more difficult to generate these profiles than for sprinters. NFL athletes need a different optimal profile than sprinters since they are not trying to win a race. Additionally, each position played is significantly different from one another with no clear indication of what an optimal profile would even look like. As such, force-velocity profiles are typically used in the NFL for comparison purposes and to rehabilitate injured athletes. For example, one team I spoke with said they track a player's maximum velocity in each practice relative to their theoretical maximum and if it is a certain percent lower multiple days in a row they know that player needs more rest.

Teams tend to find the profiles useful, but they have difficulties in getting up-to-date force-velocity profiles for every player. The difficulty here falls on the specific sprinting test that is typically performed to get the profiles. NFL teams try to limit the amount of sprinting their players do outside of game time in order to minimize injuries and overuse. Adding another test into the already packed week of an NFL athlete is something most teams will not risk [14]. As such, teams tend to get a baseline for all players at the beginning of the season that they use as a "gold standard" for the rest of the season to compare to. The only time they are typically tested again during the season is if an athlete is injured and the sports medicine team wants to test how their recovery is going and if they are ready to return to play. Being able to get more up-to-date force-velocity profiles would benefit NFL teams as they could see how their players are doing during the season and potentially identify weak spots to work on to improve the team.

2.3 Prior Work in MIT Sports Lab

The most significant previous work to this paper is the wonderful thesis by Kevin Lyons, who worked in the MIT Sports Lab the year prior to me [10]. His thesis was also focused on finding an automated way to generate force-velocity profiles of NFL athletes and made several interesting novel advancements in the field. His work involved taking the tracking data for a player on each play and extracting increasing and decreasing segments of acceleration for an athlete on a given play and finding segments over a second in length. With that data he could use the Keller model to track their velocity and derive the acceleration and force [9].

After extracting the forces and velocities directly from the tracking data, he also took a percentile-based approach to calculate the upper envelope of a force-velocity profile. His approach was to go directly from the data to see what the player did instead of taking the direct linear approach and trying to estimate the theoretical max force and max velocity of a player. He was able to derive several upper envelope methods that are improved on throughout this paper. Figure 2-4 shows the results of his work to produce an upper envelope for a player. Kevin's work is the direct inspiration from my thesis, and I was able to heavily iterate and improve upon his initial findings in the coming chapters.

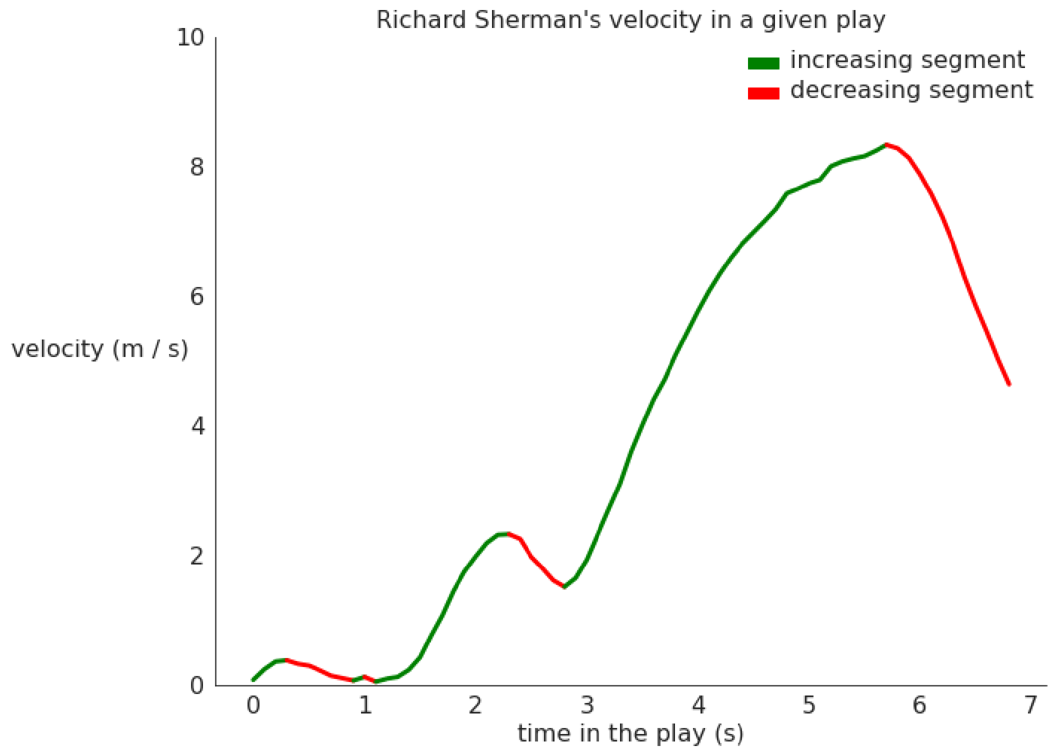


Figure 2-3: Extraction of sufficient segments from Kevin's thesis [10].

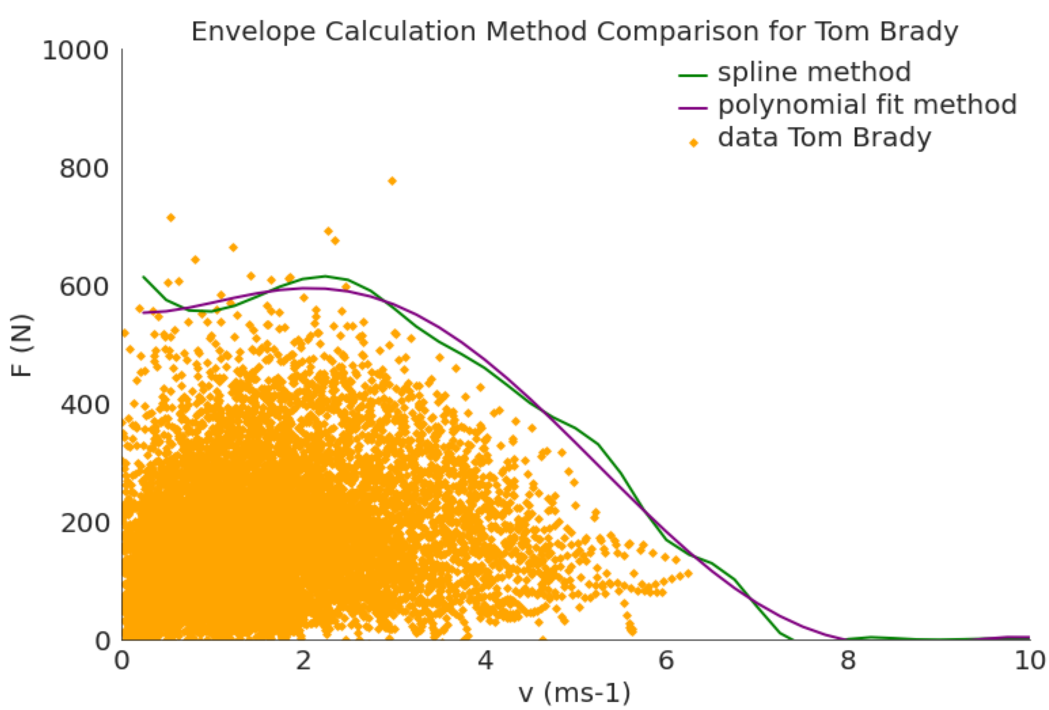


Figure 2-4: Final upper envelope for Tom Brady from Kevin's thesis [10].

Chapter 3

Calculating Force-Velocity from Player Tracking Data

Given the merits of using force-velocity plots to rehabilitate and train athletes, professional sports teams use force-velocity (F-v) curves regularly in their athletic training and sports medicine departments. They help the staff to make informed decisions about when athletes are ready to return to full play and monitor their ongoing progress.

3.1 Motivation to Use Player Tracking Data

The biggest issues teams encounter in utilizing these curves is that the only time they can update them is when an athlete runs a specific test. The standard test is to have an athlete attached to a known weight (say 2.5kg) and run 40m as fast as possible. This technique generates accurate force-velocity curves but has several major drawbacks. The most significant is the risk of injury during the test. NFL teams try to minimize the time athletes are spending at maximum exertion as injuries there occur more frequently and are typically more severe [11]. This is especially concerning in rehabbing athletes, as they need the test to see how their recovery is progressing, but doing the test too early risks reaggravation of the injury. Another issue is finding time for athletes to perform the test. NFL athletes have grueling schedules that have them watching film, lifting, practicing, and receiving treatment [14]. Adding time in the week for another specific test is tough for one player and infeasible at the scale of all 52 players on an NFL roster. Given these injury risks and scheduling concerns, I propose a method of calculating a force-velocity profile directly from the tracking data that is already recorded for every player in practice and games.

3.2 Football Definitions

It is necessary to briefly lay out the different positions in football and describe some of the difficulties in calculating force-velocity plots for certain positions. Unlike runners, whose movements are mostly unimpeded except by natural terrain, football players don't spend a lot of time at a maximum speed sprint. Figure 3-1 shows how a football play may begin, with different groups moving around based on the play call after the snap.

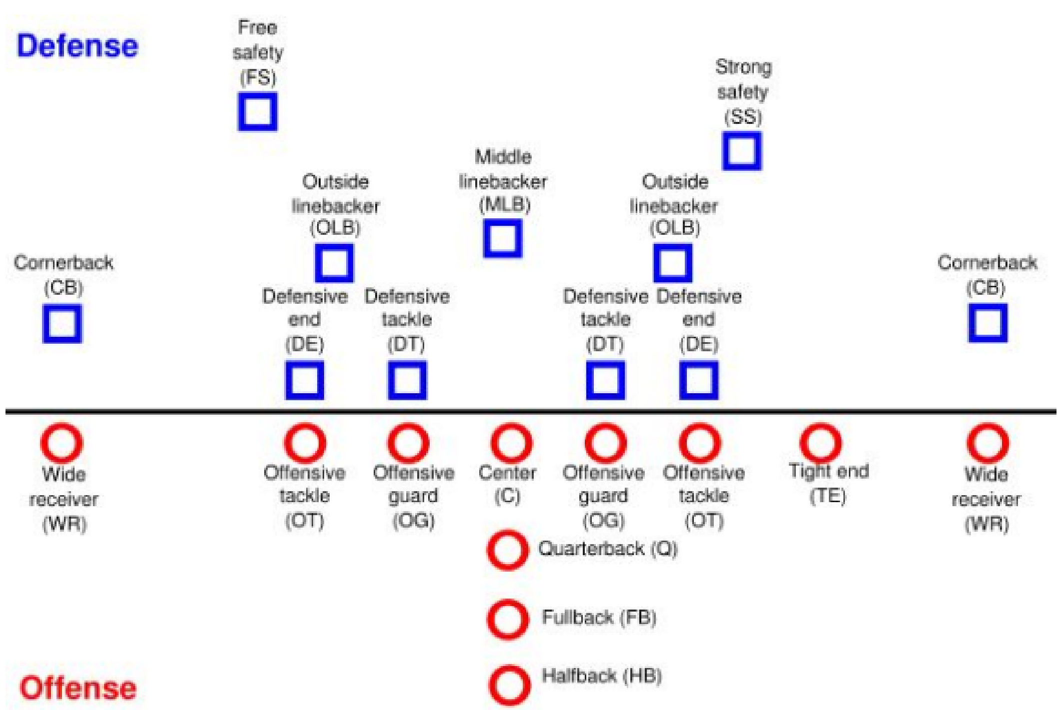


Figure 3-1: Aerial view of a football field, showing where each position group roughly starts a play [10].

Table 3.1 breaks down the general position groups you will see in football, where the differences between how these groups play would have a noticeable effect on which regions of their force-velocity curves are regularly accessed. Wide receivers, for example, spend a lot more time sprinting and changing direction quickly, so may have higher maximum velocity output on tracking data. Linemen are the most difficult to calculate since they tend to spend time blocking or trying to get through the blocks of other linemen. Tracking data provides us with velocity, but true force (e.g. the resistance from other players) is not measured. The differences in force-velocity between "skill" positions (QB, RB, WR, TE, DB, LB) is not as significant as the difference between linemen and skill positions. For example, a linebacker spends a majority of the beginning of the play reading what the offense is doing, but will still have to accelerate quickly in an attempt to tackle the ballcarrier or to cover a wide receiver, so they should still hit their maximum power output over the course of a season.

Position	Abbreviation	Side of Ball
Quarter Back	QB	Offense
Running Back	RB/HB	Offense
Wide Receiver	WR	Offense
Offensive Lineman	OL	Offense
Tight End	TE	Offense
Defensive Back	DB	Defense
Linebacker	LB	Defense
Defensive Lineman	QB	Defense

Table 3.1: Table of Football Positions.

3.3 Dataset and Limitations

The work in this thesis is derived from a publicly available dataset hosted on Kaggle, which has positional tracking data for every passing play in the 2018 NFL season [1]. Although the dataset does not contain any running plays, I am confident that the passing play data contains enough information and time spent at maximum power output to derive the force-velocity relationship of each athlete. The addition of run plays would potentially change the profiles of some players (in particular running backs), but the general methodology developed can be adapted to account for this new data once added into the dataset. This dataset also does not include linemen, but as discussed above linemen are the most difficult to assess with tracking data as most of their force comes from colliding with others.

Name	Count
Games	253
Players	$1,303$
Plays	$19,329$
Unique Entries	$18,309,388$

Table 3.2: Key metrics from the Kaggle dataset [1].

Table 3.2 shows the size and type of data in this dataset. The dataset was collected using RFID sensors placed into the should pads of each player during regular-season games, which were then used to track the positional data of each player. The data was collected at a rate of 10Hz, giving a significant amount of unique entries for each player. The dataset contained metrics about each player, giving their name, height, weight, position, date of birth, college, and a unique identifier. The identifier was used for the data from each play to record the x, y position of each player, as well as their speed, acceleration, orientation, and direction. Figure 3-2 shows the orientation of the axes used to standardize the measurements. One key thing to point out is that the data is not in SI units but instead in yards, so some preprocessing was required.

3.4 Deriving Acceleration and Velocity from Positional Data

After running several tests on the acceleration and speed values from the Kaggle dataset, I concluded that they were not consistent enough to run an analysis on as they seemed to be overly smoothed. I decided to take the positional data (which is what was tracked originally in the dataset), and derive the velocity and acceleration using derivatives, where the smoothing is known. To smooth the data I used a

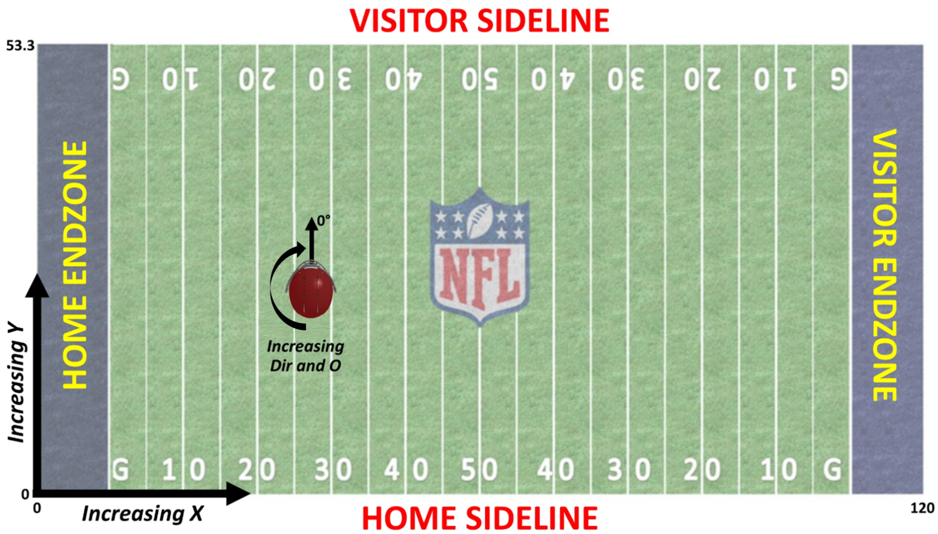


Figure 3-2: Axes used to describe a player's location and direction in the Kaggle dataset [1].

Savitzky-Golay filter [6] with order 2 and window size 5, which were selected through experimentation. A Savitzky-Golay filter works by taking a window size w and calculating a polynomial fit of order o on the $\lfloor \frac{w}{2} \rfloor$ points to the left and right of a given point. With the polynomial fit, it smooths the given point by replacing it with the value of the polynomial at that point then moving to the next point and shifting the window. This method tends to produce nicely smoothed data where we can adjust the window size and order as needed.

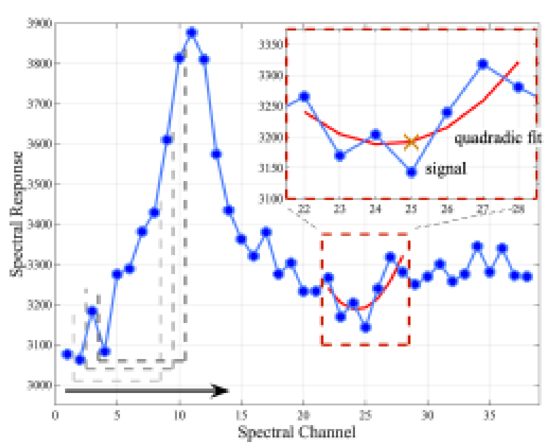


Figure 3-3: An example of how the Savitzky-Golay filter is calculated with window size 7 and order 2 [6].

With our smoothing function decided, the next question was how to calculate the derivative of this discrete dataset. This was done by fitting cubic splines to the raw data and taking the derivative in both x and y separately. This allowed me to find the velocity and acceleration in x and y independently from each other, giving v_x, v_y, a_x, and a_y. We could have alternatively used the Savitsky-Golay polynomials directly, which would have given similar results. The algorithm to derive these follows from the steps of smoothing, fitting a spline, and taking the derivative to produce the final algorithm:

1. Smooth raw position data using the Savitzky-Golay filter
2. Fit a spline to the smoothed position data
3. Differentiate the spline to find velocities
4. Smooth velocities using the Savitzky-Golay Filter
5. Fit a spline to the smoothed velocities
6. Differentiate the spline to find accelerations
7. Smooth accelerations using the Savitzky-Golay filter.

After applying this process to both the x and y data, we now have v_x, v_y, a_x, and a_y, which can be combined to give us our final outputs of velocity and acceleration for an athlete using:

$$v = \sqrt{v_x^2 + v_y^2}$$

$$a = \sqrt{a_x^2 + a_y^2}$$

Now that the velocities and accelerations were calculated, all that remained was to calculate the force the athlete was exerting to generate the force-velocity profile.

3.5 Calculating Force exerted by an Athlete

To estimate a lower bound on the force an athlete is exerting, the first step is to generate a free body diagram of the forces on the athlete. Figure 3-4 shows this free body diagram and the forces acting on the runner. To compute the force from this, we must find the net force acting on the athlete.

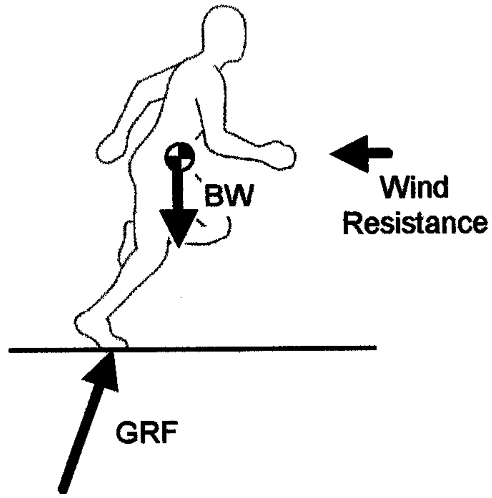


Figure 3-4: Free body diagram showing the three external forces that determine the acceleration a runner: ground reaction force (GRF), gravitational force (equivalent to body weight, BW), and wind resistance [8].

Some key assumptions made were that the force of gravity of the runner would be cancelled out by the ground reaction force's y-component, and so the acceleration calculated in the prior section would be associated with the net force in the x-direction caused by the ground reaction force from the athlete. In total, the F we were looking for was governed by the equation:

$$\sum \mathbf{F} = (GRF_x - F_{drag})\hat{\mathbf{x}} + (GRF_y - mg)\hat{\mathbf{y}}$$

where F_{drag} is the wind resistance on the athlete. Since we assume that the y-components cancel out, we are now looking to find GRF_x (which we will refer to as $F(t_i)$ from here onwards) since we strictly focus on the x-components and simplify the equation to:

$$F(t_i) = ma(t_i) + \frac{1}{2}\rho C_d A_c v(t_i)^2$$

This includes the formula for calculating drag as well as the acceleration of the athlete, ma. C_d is the drag coefficient of a running person, which is on the order of 1 [16]. ρ is the density of air, which is 1 kg/m^3. $v(t_i)$ and $a(t_i)$ were calculated in the prior section, so the only unknown left to estimate is A_c, the cross-sectional area of the athlete. We estimated the shape of a person to be approximately a cylinder to calculate A_c.

We started by calculating the weight of a cylinder, which is

$$w = \frac{1}{4}\rho h \pi D^2$$

In this case, ρ is the density of water (10^3kg$/m^3$), h is the height of the cylinder/person, and D is the diameter of the cylinder. D is what we're interested in to find the cross-sectional area of a person, so rearranging for D we find:

$$D = \sqrt{\frac{4w}{\rho h \pi}}$$

The cross-sectional area of a cylinder is $A_c = hD$, so plugging in the values from before we get:

$$A_c = hD = \sqrt{\frac{4wh}{\rho\pi}} \approx \sqrt{\frac{wh}{\rho}}$$

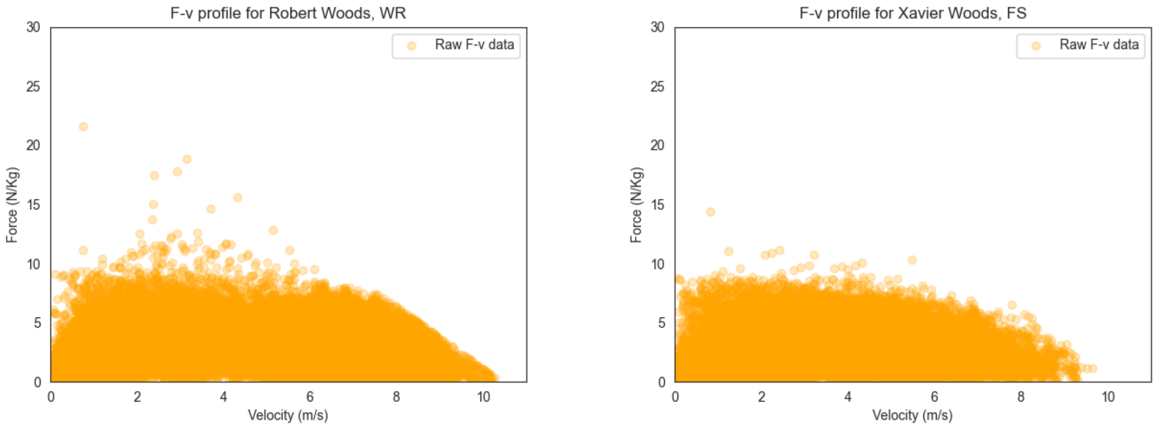
Plugging all these constants back into the force equation, we get the final equation to estimate the force:

$$F(t_i) = ma(t_i) + \frac{1}{2}\sqrt{\frac{wh}{\rho}}|v(t_i)|^2$$

Now that we can calculate both force and velocity, I calculated those for each of the players and was able to produce force-velocity plots for each athlete in the dataset. Figure 3-5 shows two examples of these total computed point clouds for two different players. you can already begin to see some of the differences between even these two players. For example, you can see that Robert Woods has a higher and more consistent time spent in the higher velocity section of the graph.

3.6 Cleaning the Data

Since we are taking the second derivative of the tracking data, there will still be some outliers despite our smoothing functions. We have derived a simple way to remove



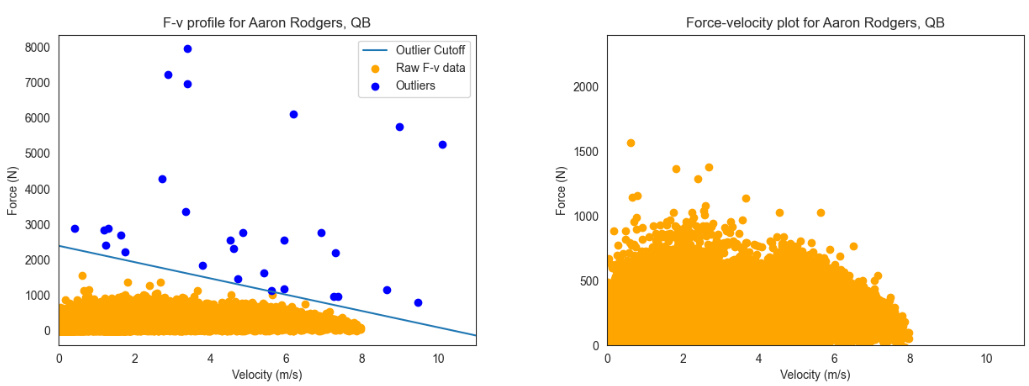
(a) Robert Woods F-v plot

(b) Xavier Woods F-v plot

Figure 3-5: Two Examples of computed force-velocity plots.

outliers from the dataset that is consistent for most of the data.

The general flow of the plan is to bound the max values of the data based on certain constants. We cut off the velocity max at the Olympic world record for the 100m, which comes out to 10.44 m/s, and we cut the force off at 2400N, which is around a 600-pound slow squat, which we estimate is around the average an NFL player could perform. This works well since the players are not performing this type of movement at low velocities in games so there will only be outliers at that force. We connect these two maxes to form a line and designate any point above the line as an outlier. Figure 3-6 shows the results of this outlier designation process.



(a) Aaron Rodgers Outliers

(b) Aaron Rodgers Cleaned F-v plot

Figure 3-6: Example of computing outliers.

We then divide by the player's mass to get a measurement of athlete strength per kg. This sets up the basic framework of computing force-velocity curves for different players and enables us to use these curves to analyze players.

Chapter 4

Force-Velocity Profiling and Kernel Density Estimation

With the ability to compute clean force-velocity data directly from tracking data, we can analyze different players using force-velocity profiles (FVPs) and a player's distribution in different output ranges using kernel density estimation.

4.1 Force-Velocity Profiling

Force-velocity profiling is an effective way to keep track of different aspects of an athlete's training and provides an overview of how well an athlete will perform in different scenarios. It additionally provides a convenient way to track how well an injured athlete is progressing in returning to full strength and might be able to track what muscles an athlete specifically needs to work on to completely rehabilitate. Figure 4-1 shows how an FVP may change after training specific areas.

This example figure has a hyperbolic relationship between force and velocity, but the data and more recent literature show a different story. The original paper by Hill in 1938 suggested this hyperbolic nature due to the heat of shortening and lengthening muscles, but this was shown to be true only for isolated muscles [7]. More recent studies have shown that the F-v relationship of multi-joint performance tasks is quasi-linear and can thus be modeled with a simple linear model given good data [15].

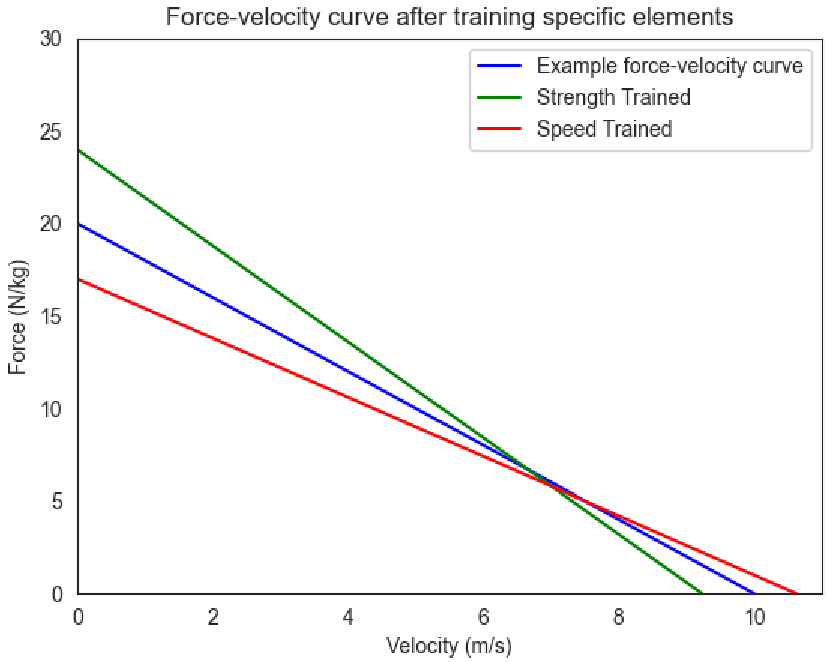
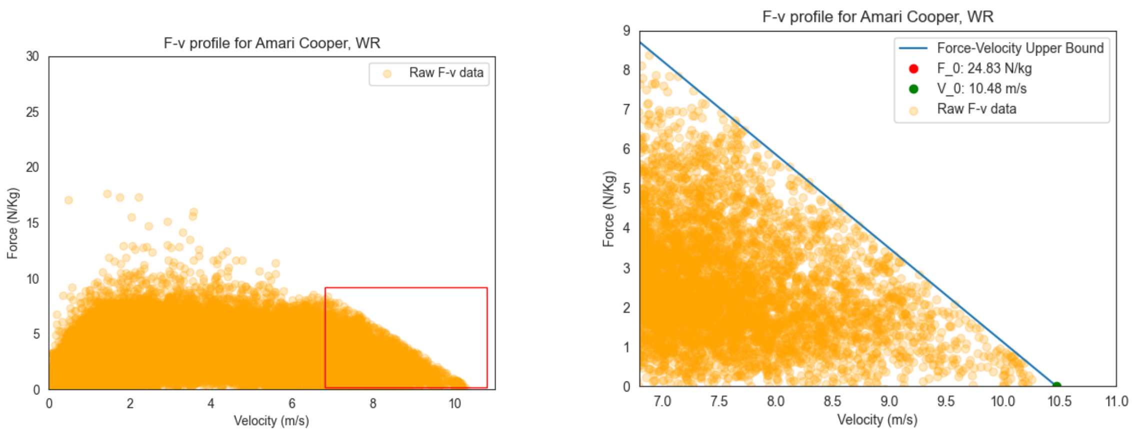


Figure 4-1: Effects of performing different types of training on a Force-velocity profile.

4.2 Generating Force-Velocity Profiles

With an understanding of the necessity for generating FVPs, I was able to derive a consistent way to find the maximal FVP for an athlete. The general approach is to find an area where the athlete is almost at maximum exertion. From examining the plots, I saw that the top range of an athlete's velocity seemed to show the linear trend we are searching for. This made sense because of the unique aspects of football at low velocities. As opposed to sprinting, where athletes tend to use a lot of force at low velocities to accelerate faster, football has a lot of technical work done at low velocities due to the contact nature of the sport. A defensive player may need to read what the offense is doing before accelerating to attempt to cover a receiver or tackle the ball carrier, and wide receivers have to make lots of quick cuts while trying to avoid defenders. All of this leads to a lack of data in the lower velocities where we expect some higher forces.

At high velocities, we see a different story. At this point, athletes are typically in the open field not making contact with other players, and are running as fast as they can to get to the end zone. At this level, we see a clear inverse linear relationship between force and velocity. An example of this lies in figure 4-2, which shows the area where the linear relationship lies. We can use this to extrapolate the FVP even without the low-velocity high force points.



(a) F-v plot showing linear section.

(b) Zoom in of plot to the left, with a best fit line plotted.

Figure 4-2: Linear section of F-v plots, zoomed in to show how well a line fits.

The technique to find the best fit line then relies on finding the area where the data is roughly linear. The best fit I found was to take the data points that were between 80% and 99% of the max velocity. This was found by testing different values and calculating the average r^2 values of the resulting lines. Then, the space is split up evenly to segment the area into 15 different equally spaced sections. A new data point is then calculated at the 99th percentile for force of all data in that section and used as a new data point. These points are then used in a linear regression to create a line that results in the FVP for that player. Figures 4-3 and 4-4 show the process of this fitting and the final results for a player. Note that F_0 and V_0 are the intercepts and key metrics in analyzing the FVP of a player. They indicate the upper bounds of an athlete's estimated abilities in both force output and velocity. They can be used to see how an athlete is progressing as well as how close an injured athlete is to getting back to full performance.

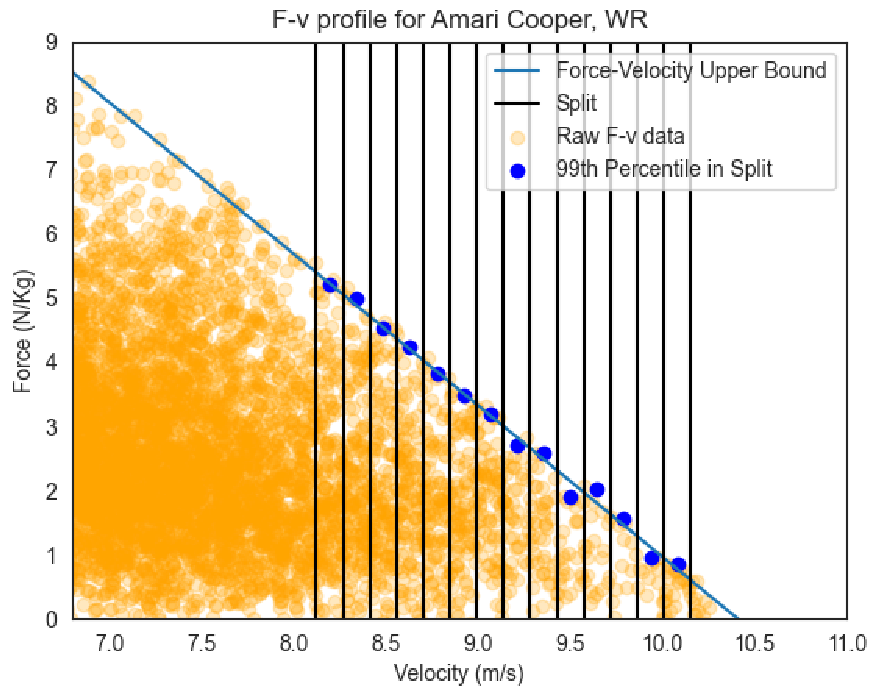


Figure 4-3: Example splits and fitting to Amari Cooper's F-v data.

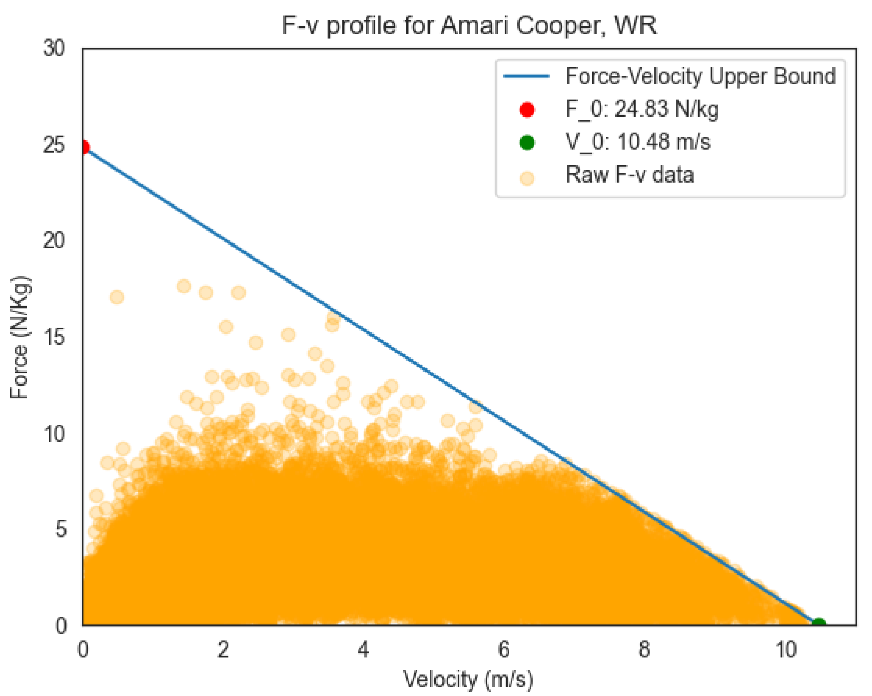


Figure 4-4: Fit line displayed on all of Amari's data.

4.3 Kernel Density Estimation

Now that we can calculate the FVP for any given player, there is one more tool for analysis that is quite useful in determining how a player plays: kernel density estimation. Kernel density estimation (KDE) is a technique to estimate the underlying probability density function (PDF) of a dataset [4]. with the estimate of the PDF, you can display confidence bands on a dataset at various levels to see where the data is concentrated.

Let $X_1, ..., X_n$ be an independent, identically distributed random sample from an unknown distribution P with a distribution p. The KDE can be expressed formally as:

$$\hat{p}_n(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

Where K is a smooth function called the kernel function and h is the smoothing bandwidth that controls the amount of smoothing. In our KDE we use a gaussian kernel, represented by:

$$K(x) = \frac{exp(-||x||^2/2)}{v_{1,d}}, \quad v_{1,d} = \int exp(-||x||^2/2)dx$$

With this same smoothing in each direction applied by the Gaussian, each data point is smoothed and accumulated to get the final density estimate $\hat{p}$ [4], which we can then use to draw levels on a graph.

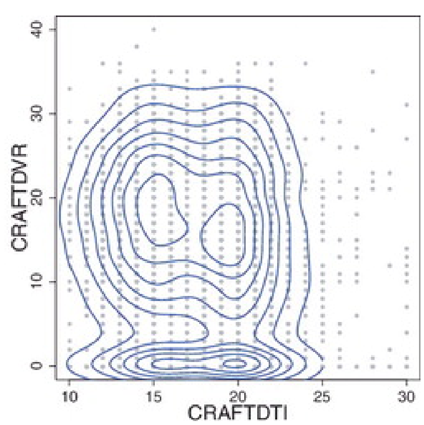
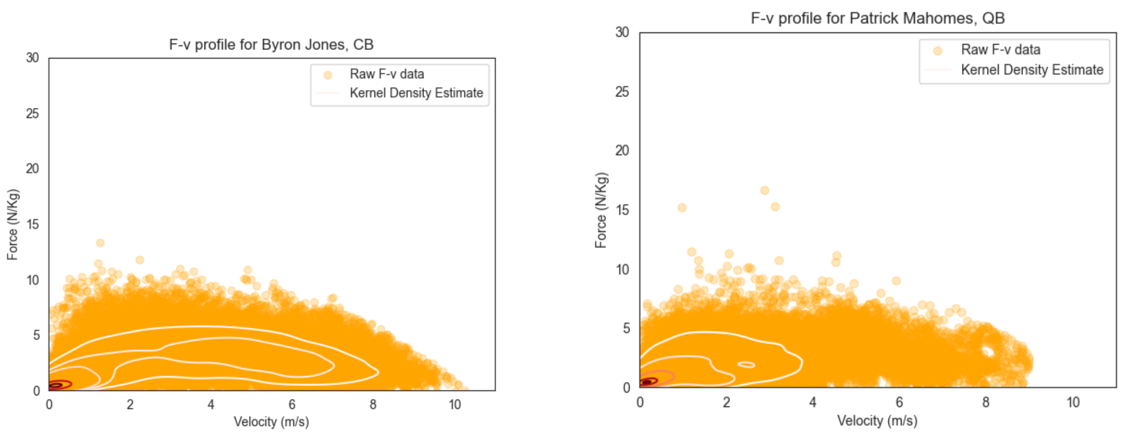


Figure 4-5: Example of a KDE with contour lines drawn over a dataset [4].

4.4 Kernel Density Estimation for NFL Athletes

In the case of NFL athletes, we must use bivariate kernel density estimation since we are trying to see force-velocity bands that an athlete spends the most time in. I applied KDE estimation to the data for each athlete, which helps show the total load an athlete has accumulated during the season. This can be utilized to perform load analysis on players throughout a season to prevent overuse.



(a) KDE of Byron Jones (b) KDE of Patrick Mahomes

Figure 4-6: KDE's of two different NFL players.

Figure 4-6 shows an example of KDE applied to two athletes, with levels 20%, 50%, 80%, 95%, and 99% drawn. We immediately notice a difference between the two KDEs that is explainable by the difference in position between the two players. Cornerbacks are typically required to cover fast wide receivers and have to spend a lot of time at high velocities, so you see the KDE stretch to the right near the upper bound of their velocity. Patrick Mahomes on the other hand is a quarterback who spends the start of most passing plays reading the defense and finding a receiver to pass to. As a result, he spends a lot of time at low velocities during plays, which is indicated in his KDE.

With both force-velocity profiles and kernel density estimates able to be calculated directly from tracking data, we can get a complete profile of each athlete in the dataset. An example final evaluation plot is shown in figure 4-7

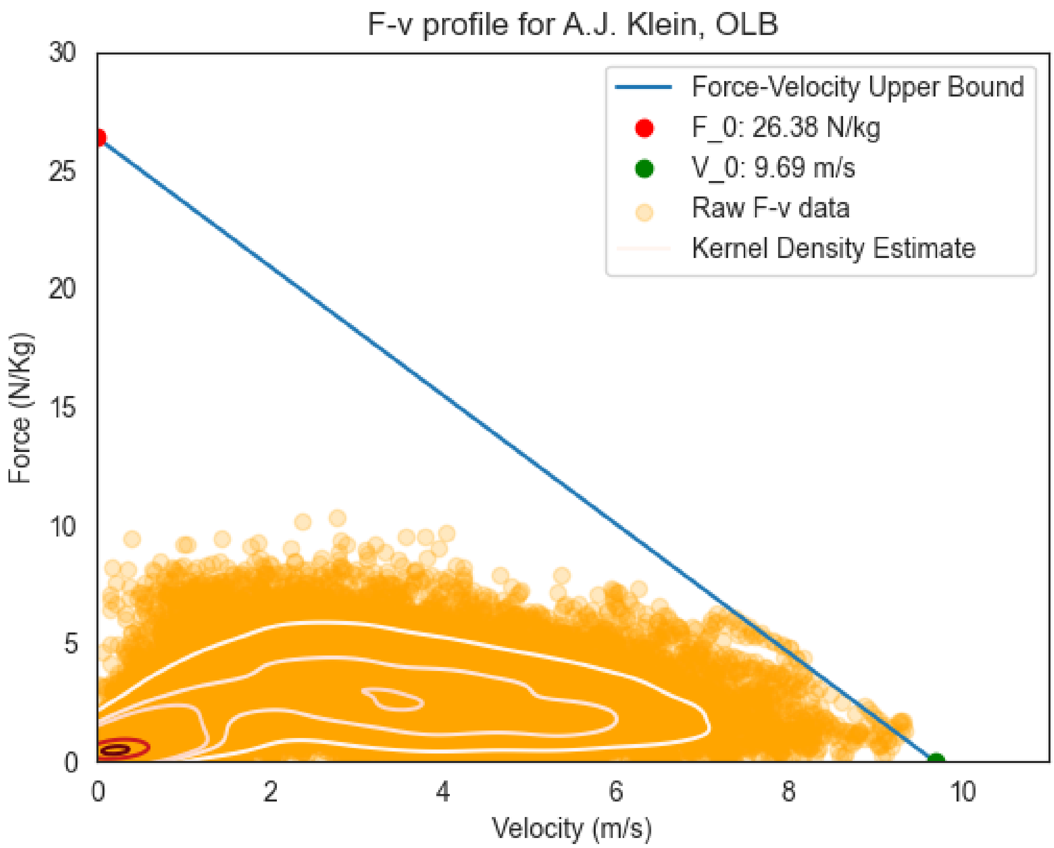


Figure 4-7: Complete composite image of an NFL athlete with intercepts, KDE, and FVP plotted.

Chapter 5

Applications to Professional Sports

This chapter presents two distinct applications where the work in this paper could be used to benefit the training and recovery of NFL athletes. NFL teams already use force-velocity profiles to inform these two areas, but as discussed earlier require specific tests that are tough to coordinate and risk injury to the athlete [11]. Given these concerns, NFL teams could utilize the techniques presented in this paper to track force-velocity profiles and load analysis directly from the tracking data they already record on the athletes.

5.1 Sports Medicine

Sports medicine staffs have in recent years have started to use FVPs to more accurately assess whether athletes are fit to return to play. It enables them to better understand what has changed in an athlete and how they may be compensating and risking further injury in other muscle groups. The most prevalent example of reinjury risk is hamstring injuries. Athletes will change their running style without realizing it or affecting their maximum velocity, and overcompensate to different muscles [11].

Figure 5-1 shows a real example of a soccer player who tested sprinting for an F-v profile before and after a hamstring strain [12]. As you can see, there is no effect on the athlete's maximum velocity, which is something you can see by eye during on-field play and practice reps, so to the naked eye the athlete may appear to be fully

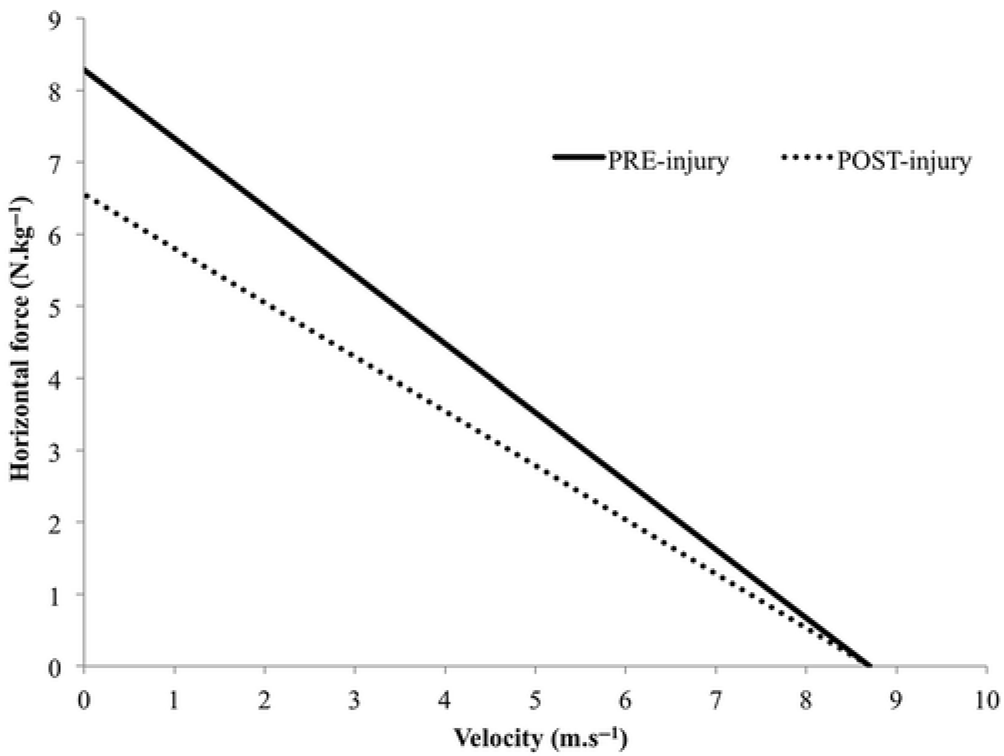


Figure 5-1: F-v profile pre and post hamstring injury for an athlete [12].

healed. However, by plotting the force-velocity profile, you can see that the athlete is most likely compensating using different muscles and is not producing the same level of force that he was before. This is a good indicator to sports medicine staff that an athlete either still needs to rehabilitate their injury or needs to be held out of team activities for longer than they might otherwise.

Utilizing force-velocity profiles could help reduce reinjury rates among athletes, especially in hamstring injuries which are notoriously high risk for reinjury and are tough to tell when they have fully healed as the first time an athlete tries to produce more force than their hamstring can handle they can reinjure their hamstring. They provide another metric to sports medicine staff of when an athlete is fit to return to play. Being able to estimate the FVP of a player directly from tracking data would help this process even more as athletes would not have to run sprints where they risk reinjury.

5.2 Load Analysis and Training

Another benefit comes from the kernel density estimation aspect of this player analysis. With the FVP we know a player's approximate theoretical max velocity and force, so teams could keep track of how long an athlete is spending doing high-intensity activity. NFL teams already track how many "high output yards" a player runs, as well as how many reps in-game and practice they get. With a kernel density estimate, they could classify what percentile of an athlete's output counts as "high output" and see how much time they spend in that on different practices/games. This would enable insights on when an athlete needs to rest to recover the ability to spend more time at high output.

Another benefit here comes from the force-velocity profile of a player. NFL teams would be able to update each athlete's profile after every practice/game and be able to see if an athlete is performing up to the standards that they have previously set. There are multiple advantages to this approach. One, the athlete doesn't need to run a specific test so teams can generate an FVP over any period of time they need to see how an athlete is performing. Two, the athlete can be compared to themselves to see if they need rest or if the training regiment the athlete is on is improving the maximum force or velocity that they wish to improve. Three, it provides numbers for maximums that can be given to coaches to see how an athlete's numbers compare to other members of the team when making starting lineups. Having only a few numbers/plots for each athlete that encapsulate the athlete's athletic profile allows the coach to make informed decisions without being overwhelmed by a lot of data.

These are just two potential applications to professional sports but the techniques presented in this paper are adaptable to whatever approach a sports team would like to use in their organization. The hope is for more teams to take this approach as it provides benefits to player safety as well as the team as a whole.

Chapter 6

Evaluation and Discussion

This section holds evaluation and sanity checks on the data to make sure that the force-velocity profiles and kernel density estimates are giving data that is consistent enough to be used at the highest level of the sport.

6.1 Positional Differences

The most logical place to begin a discussion of the accuracy of this dataset is to understand the position differences and lineups of different positions in football. Since the Kaggle dataset [1] has only passing plays in it, there are a few position groups that should be closely correlated to each other; Defensive backs cover wide receivers on most plays, and linebackers typically cover running backs on passing plays. Tight ends and quarterbacks are slightly different from other positions. Tight ends sometimes block during a play and sometimes go out for passes. As such, they have a different expected force-velocity profile since their typical weights and heights are different from other positions. Quarterback is unlike other positions in football since they stand back during the play and see how it develops. Some quarterbacks tend to run with the ball more as well if they don't like what the defense is showing them, while some tend to stay in the pocket to pass the ball more, so one might expect to see a lot of variability in quarterback profiles.

Given the different requirements of each position, the average height and weight

Position	Height (Ft, ins)	Weight (Lbs)
Quarter Back	6'3.4"	225.0
Running Back	5'10.7"	214.5
Wide Receiver	6'0.4"	200.3
Offensive Lineman	6'4.8"	314.2
Tight End	6'4.5"	254.3
Defensive Back	5'11.7"	200.1
Linebacker	6'2"	244.6
Defensive Lineman	6'3.2"	309.0

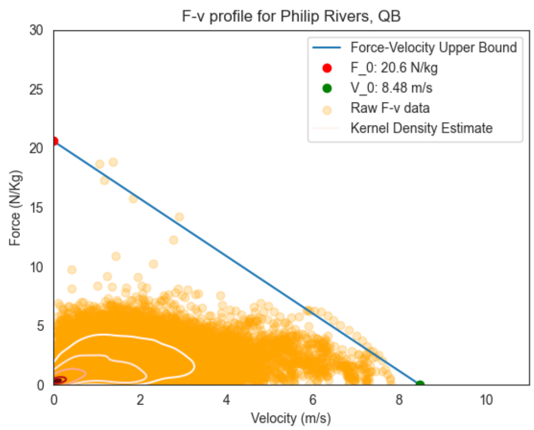
Table 6.1: Table of Football Heights/Weights by Position Group [3].

by position group varies greatly in the NFL, and a different FVP is expected for each group. Table 6.1 shows a breakdown of the different position groups by height and weight [3]. Note that the position groups mentioned have some similarities in their height/weight, such as wide receivers and defensive backs being almost identical in height and weight averages. This understanding of football knowledge and differences by position groups helps to explain the data aggregation in the rest of this chapter.

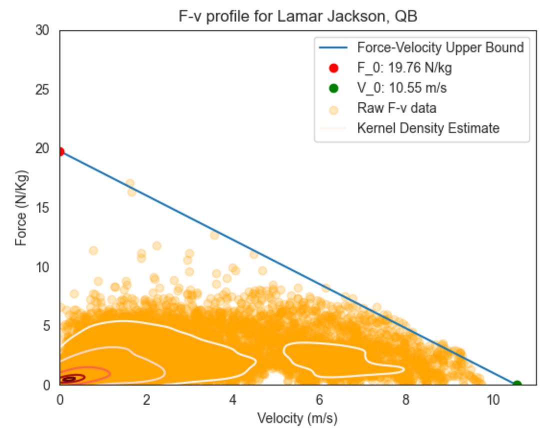
Figure 6-1 shows several different force-velocity profiles split by position group and you can already see some of the differences discussed above in that plot. Philip Rivers is more of a pocket passer and thus his max velocity is lower than Lamar Jackson, a more mobile quarterback. You can also see in the KDE that Lamar Jackson spends a lot more of his time at higher velocities. You can also see that the defensive backs are similar in profile to each other while running backs spend less time at the high velocities that the defensive backs do. This makes sense as a lot of passing plays involving running backs have them either blocking for the quarterback or running shorter routes.

6.2 Distributions of Force and Velocity Maxes

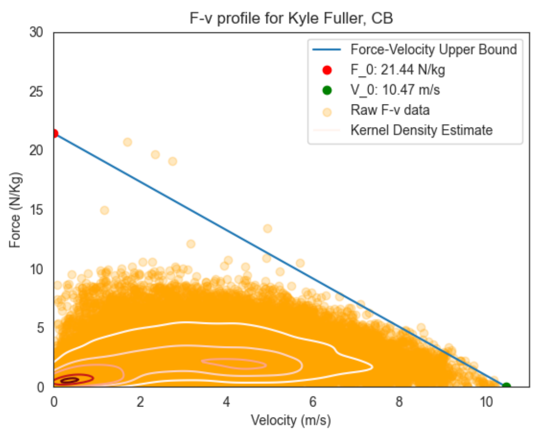
Looking at individual plots makes it difficult to see the distribution of the data, so this section presents several different ways to view the data based on two key data points we can extract from each plot: F_0, the theoretical maximum force of a player, and v_0, the theoretical maximum velocity of a player. These can be used separately



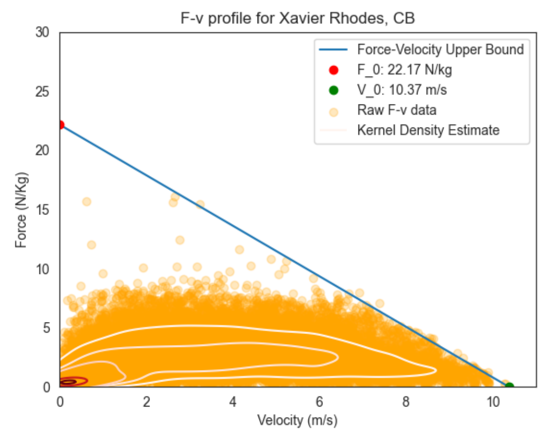
(a) Philip Rivers (QB)



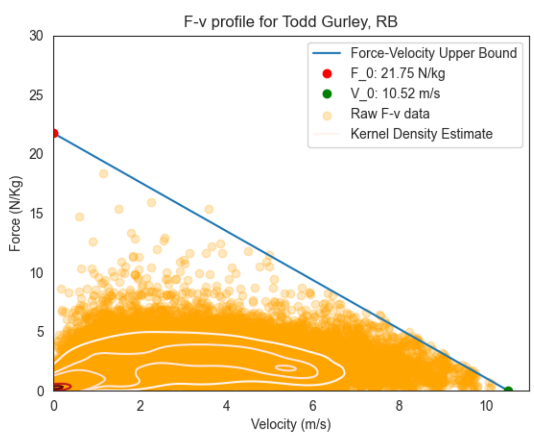
(b) Lamar Jackson (QB)



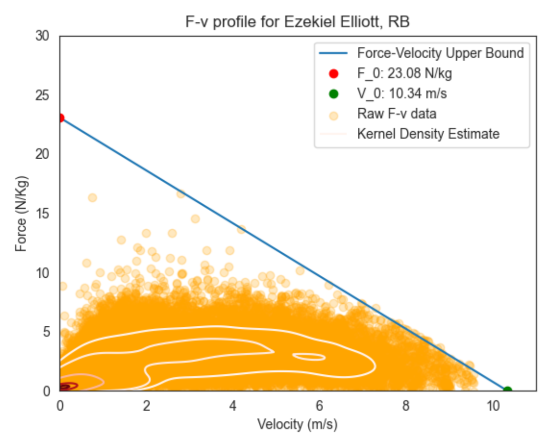
(c) Kyle Fuller (DB)



(d) Xavier Rhodes (DB)



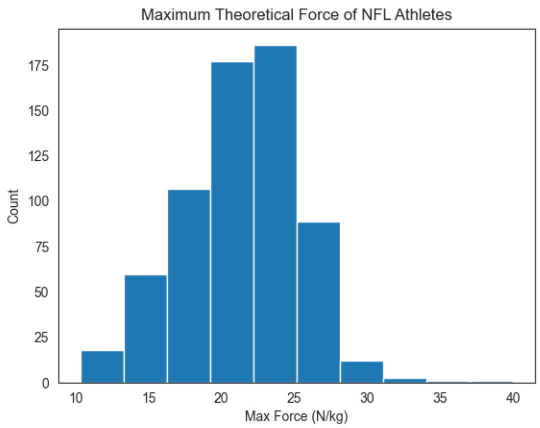
(e) Todd Gurley (RB)



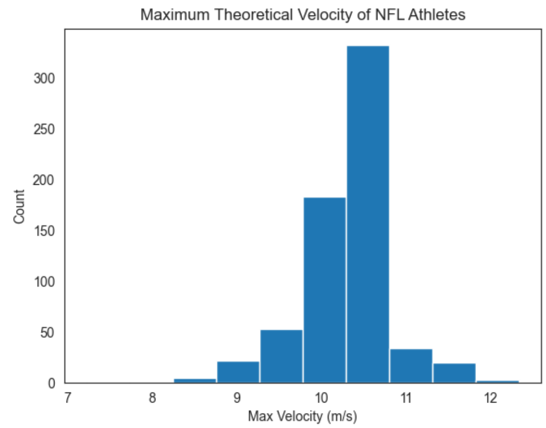
(f) Ezekiel Elliott (RB)

Figure 6-1: Example KDE's of two different players from several position groups. From top to bottom: Quarterbacks (QB's), Defensive Backs (DB's), Running Backs (RB's).

or together to get a feel for how the data distribution of different players and position groups is shaped in this dataset.



(a) Maximum force distribution



(b) Maximum velocity distribution

Figure 6-2: Maximum force and maximum velocity histograms for the Kaggle dataset.

To start, we can create histograms of F_0 and v_0 for the entire dataset. These need to be cleaned up a little as there are a few cases where certain linear estimators are affected by outliers at the top level of a dataset and create lines that are too horizontal. This was fixed by bounding the values to be positive and below the value of the max recorded human speed (12.3 m/s by Usain Bolt). Figure 6-2 shows the histograms for F_0 and v_0. The shapes of each are roughly Gaussian, which is good for consistency as we would expect a distribution of athletes to have a roughly normal distribution. The means and standard deviation of each parameter are listed in Table 6.2.

Parameter	Mean (μ)	Std. Dev. (σ)
F_0	21.31 N/kg	4.01 N/kg
v_0	10.31 m/s	0.534 m/s

Table 6.2: Mean and Standard Deviations for F_0 and v_0.

These values seem to be consistent with the maximum force and velocity of players that we would expect, although the results do tend to be slightly biased towards speed values. This extra speed result seems to come from estimating at the top level of speeds only, but as seen earlier this is the main area where we see the linearity in the dataset, so without better ways to estimate impact forces between players the speed bias is acceptable for the data and mostly consistent with example data we have received from NFL teams.

6.3 Analysis by Position Group

Another area where we can assess whether the data is consistent is by plotting the maximum force and velocity values for each player and grouping them by position. A plot of F_0 vs. v_0 split with a color for each position group is shown in figure 6-3.

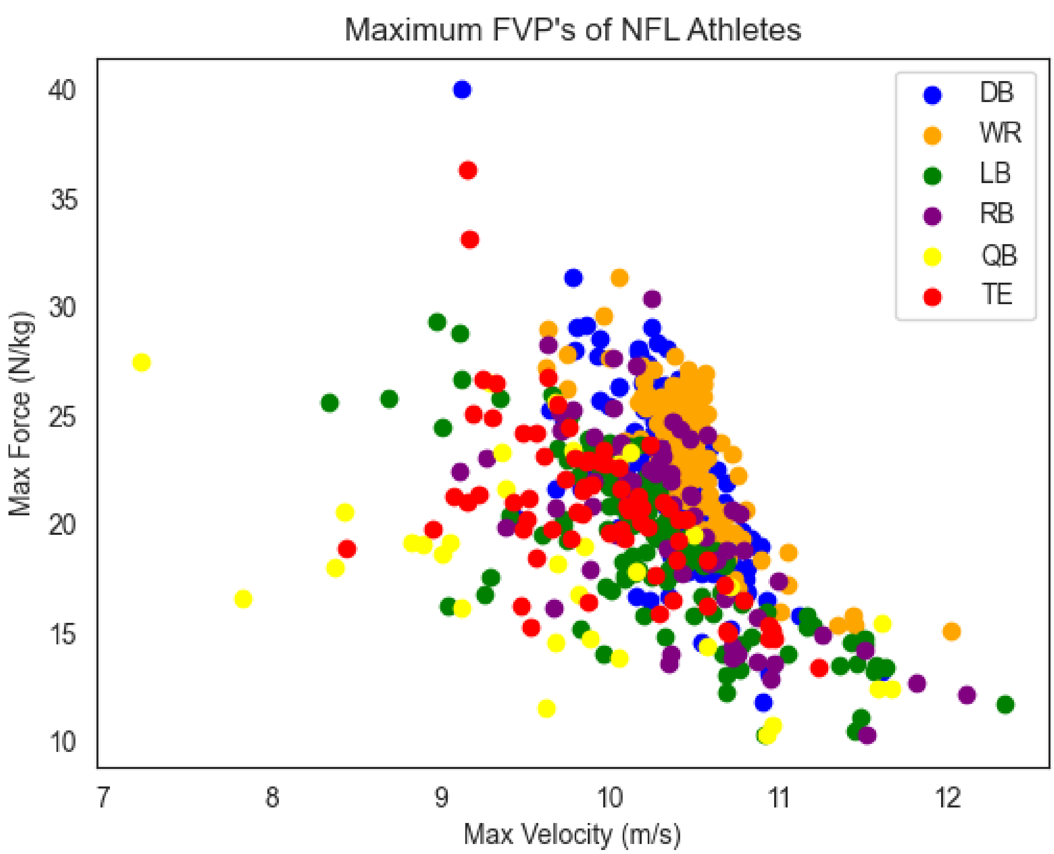


Figure 6-3: Plot of F_0 vs. v_0 for each player in the dataset, split by position group.

This plot is slightly messy since it shows all the players, but you can see trends from different position groups starting to form. We can clean up and analyze this plot using kernel density estimations we have used before. We split up the analysis for the KDEs based on the positional differences discussed earlier in this chapter. Namely, the figures will be split to show 2 position groups if we expect them to be similar and a singular position group if we expect them to be different from other position groups. We ran a kernel density estimate of all the players in that position group and set the level to show to be 50%. The groups we chose are as follows:

- Figure 6-4: WRs vs. DBs
- Figure 6-5: LBs vs. RBs
- Figure 6-6: TEs
- Figure 6-7: QBs

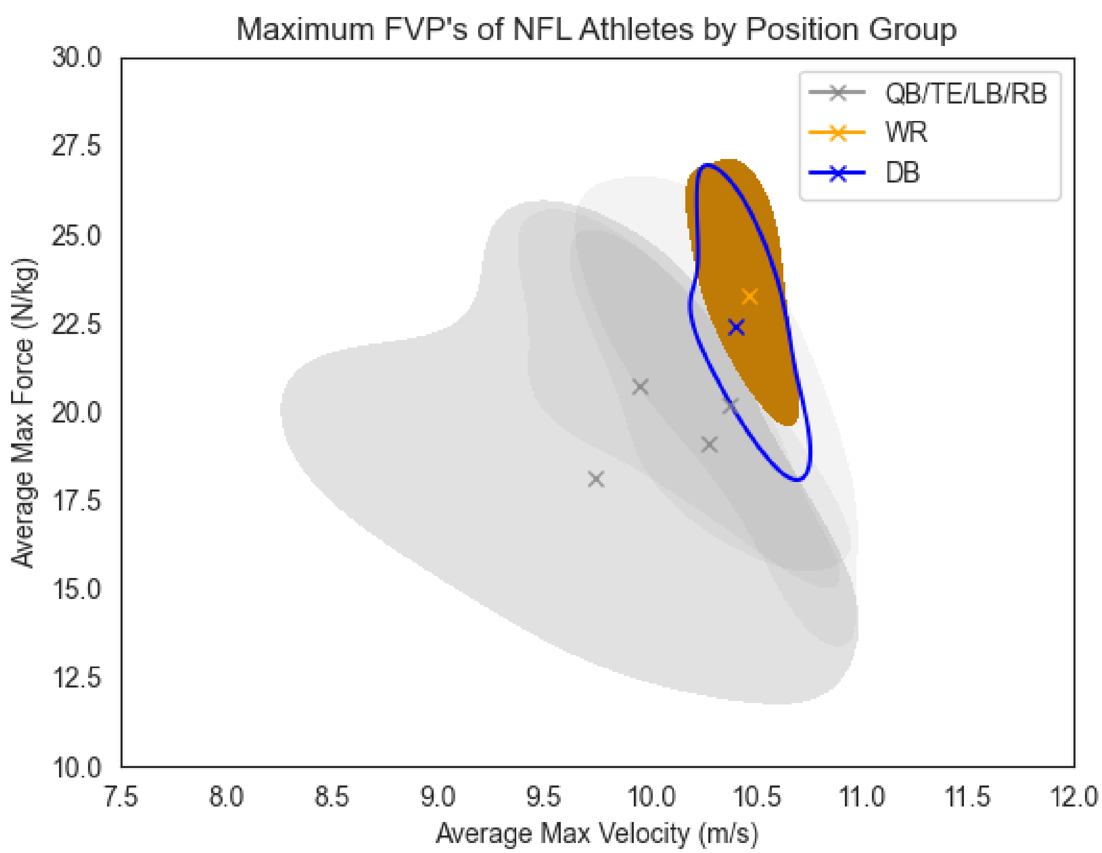


Figure 6-4: WR vs. DB Comparison.

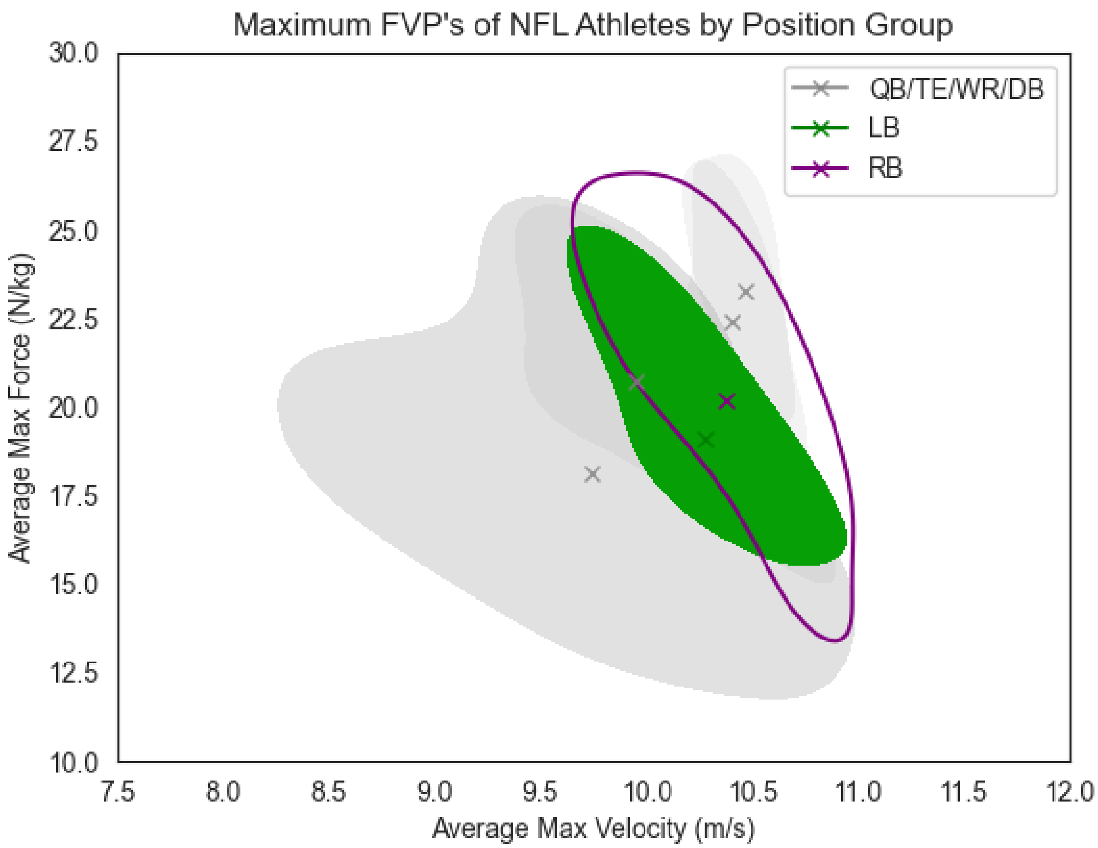


Figure 6-5: LB vs. RB Comparison.

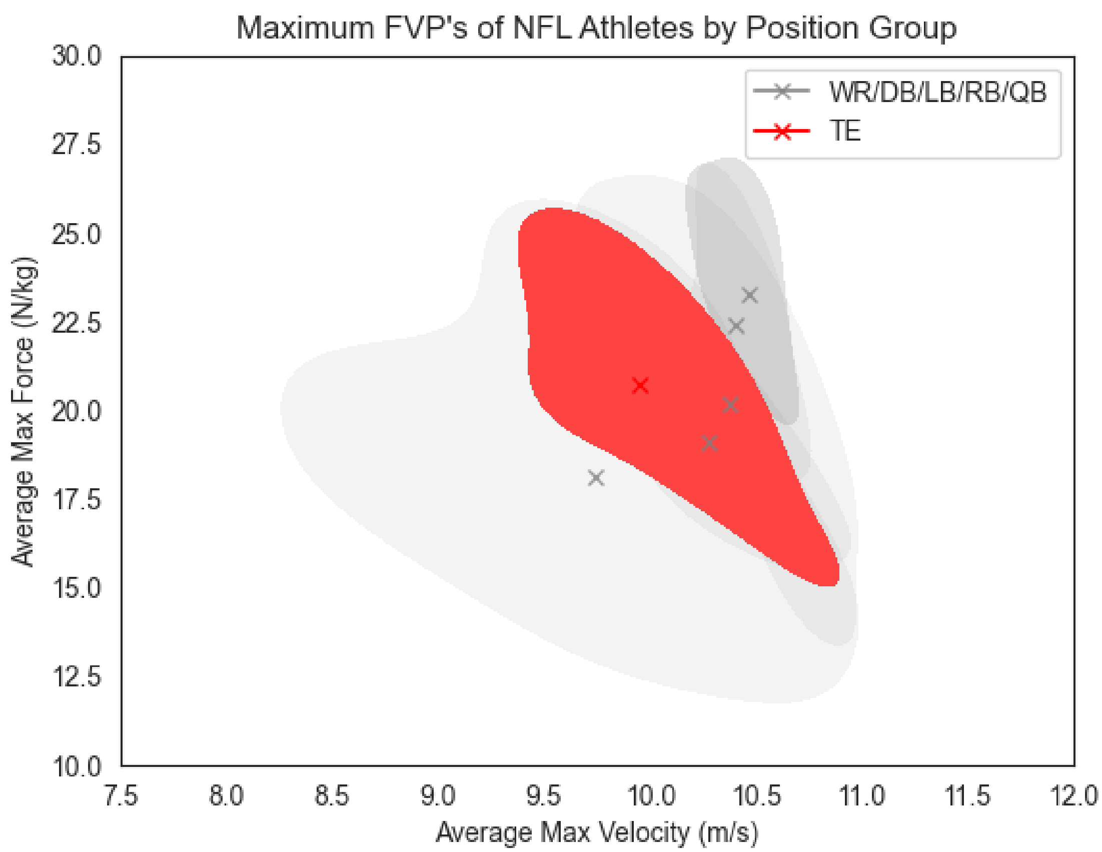


Figure 6-6: TE Comparison.

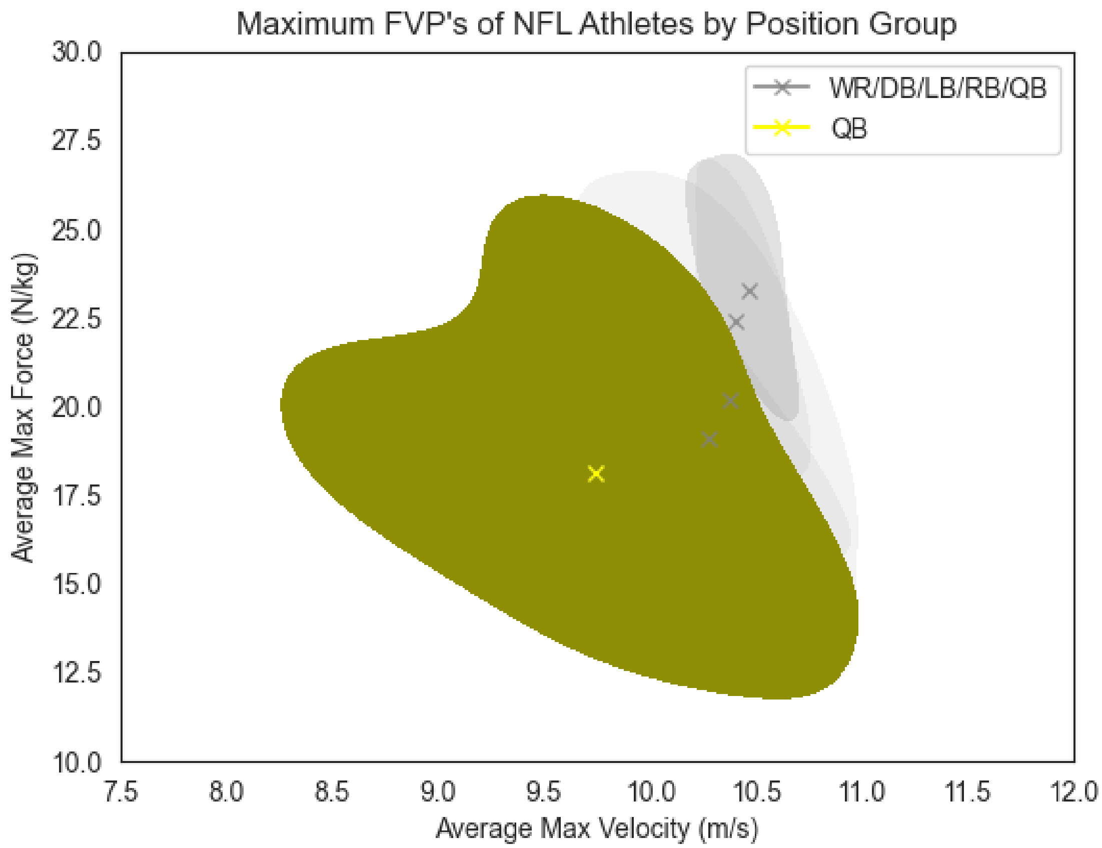


Figure 6-7: QB Comparison.

We can see some clear trends and similarities between the position groups when plotted in this way. Here are some key takeaways from each of the plots.

- Figure 6-4: You can see clearly that the WRs and DBs are closely associated with each other. This is what we expected as in passing plays DBs cover WRs. Additionally, they are at the top end of both velocity and force, which we would expect on passing plays since they are involved in the most sprinting on these plays.

- Figure 6-5: LBs and RBs are also closely associated with each other. They are both lower than the WRs and DBs, which is most likely due to them being involved in more contact on each play as LBs need to read the play to see if it is a pass and then either rush the QB or attempt to cover an RB/WR.

- Figure 6-6: TEs are similar in terms of force output to LBs and RBs, which we would expect since they are also blocking on these plays. They however tend to be slower than both of those other groups, which makes sense since their average weight is significantly higher than both LBs and RBs

- Figure 6-7: QBs as discussed earlier vary significantly from other positions and also vary greatly within the group due to pocket passers versus mobile quarterbacks. Their average is lower in both force and velocity than any other group due to having to read defenses and make passes on passing plays most likely and is stretched in both velocity and force due to the variation within the group, given them the largest area covered for a KDE.

These are the most concrete insights we can take from these plots and the dataset. Our results seem to be consistent based on the differences between position groups discussed earlier, which is a good sanity check to say that our method is correctly seeing the difference between the groups. To further confirm the soundness of the techniques developed in this paper, a dataset containing more than just passing plays or the ability to get more low-velocity force data caused by blocking would be required. Figure 6-8 shows an aggregation of KDE comparisons for all position groups at once.

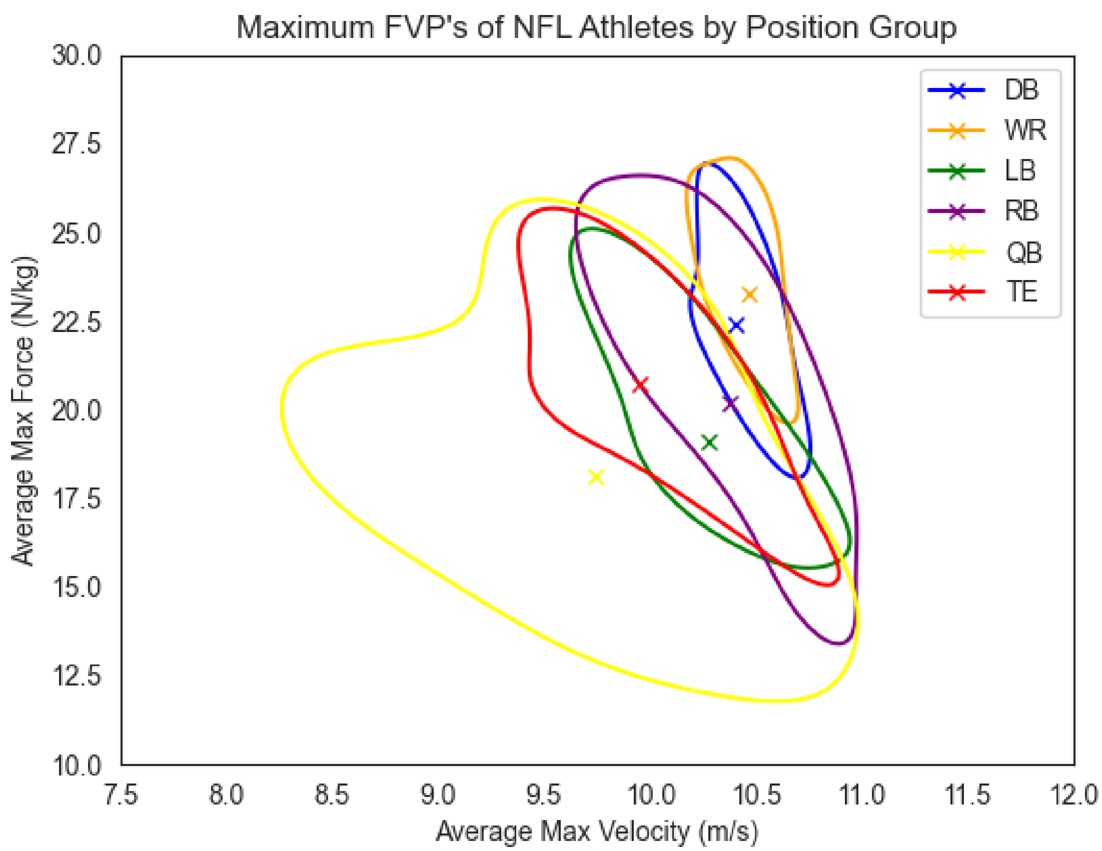


Figure 6-8: Aggregated kernel density estimates for each position group, set at a 50% level for each estimate.

Chapter 7

Conclusion

The methods developed in this paper serve two main purposes for football teams: preventing injuries caused by the current testing process for both rehabilitating and healthy athletes and providing more consistent and up-to-date estimates of force-velocity profiles of NFL athletes. The risk of injury decreases by not requiring athletes to run a test where they sprint with a weight attached to them. This is especially true for athletes returning from injury who are at an increased risk for reinjury when attempting maximum output activity [11]. NFL teams also tend to not generate force-velocity profiles for NFL athletes during the season due to this injury risk, so they instead rely on a baseline preseason profile. Over the course of a season, athletes' bodies can change significantly, and monitoring the force-velocity profile of a player can provide insights into when athletes are overloaded and need rest or if a bench player exceeds their prior performance. The techniques presented in this paper would allow teams to generate Force-velocity profiles using data from any period of data they would like so they could get up-to-date data without burdening their athletes, sports medicine staff, or jeopardizing winning chances with the risk of injury.

As a whole, our methods seem consistent with the Kaggle dataset we are using. For future work to continue to verify consistency, our techniques should be tested against a dataset containing more than just passing plays. It should also be tested over shorter time periods to make sure that the results stay consistent with different periods, such as one practice. This hasn't been tested but should be consistent as the minimum amount of time for a player to need to perform in our dataset is 5000 datapoints at 10Hz, so at minimum, it has been tested with 8 minutes and 20 seconds of data.

A few longer-term techniques that could further enhance this method involve calculating more high force, low-velocity data points. These appear for sprinters as they are attempting to accelerate as fast as possible, but are difficult to get for football players as there are different priorities for them throughout a play. Some sort of force-sensing gloves or method to identify momentum changes when two players come into contact should enable finding linear data at these high force, low-velocity points, which should further improve the model.

Overall, we view these opportunities for future work as improving and further verifying the accuracy of the model. As is, the technique is robust enough to be put into use on NFL or college football teams that record tracking data of their players throughout practice and games, and we believe would greatly benefit a team's ability to unlock these key metrics about their players.

Bibliography

[1] NFL Big Data Bowl 2021.

[2] Julian Alcazar, Robert Csapo, Ignacio Ara, and Luis M. Alegre. On the Shape of the Force-Velocity Relationship in Skeletal Muscles: The Linear, the Hyperbolic, and the Double-Hyperbolic. *Frontiers in Physiology*, 10:769, 2019.

[3] Cole R Blender. Average NFL Height and Weight by Position. page 16.

[4] Yen-Chi Chen. A tutorial on kernel density estimation and recent advances. *Biostatistics & Epidemiology*, 1(1):161–187, January 2017. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/24709360.2017.1396742.

[5] John B. Cronin, Peter J. McNAIR, and Robert N. Marshall. Force-Velocity Analysis of Strength-Training Techniques and Load: Implications for Training Strategy and Research. *The Journal of Strength and Conditioning Research*, 17(1):148, 2003.

[6] Neal B Gallagher. Savitzky-Golay Smoothing and Differentiation Filter. page 2.

[7] Archibald Vivian Hill. The heat of shortening and the dynamic constants of muscle. *Proceedings of the Royal Society of London. Series B - Biological Sciences*, 126(843):136–195, October 1938. Publisher: Royal Society.

[8] Joseph P. Hunter, Robert N. Marshall, and Peter J. McNair. Relationships between ground reaction force impulse and kinematics of sprint-running acceleration. *Journal of Applied Biomechanics*, 21(1):31–43, February 2005.

[9] Joseph B. Keller. Optimal Velocity in a Race. *The American Mathematical Monthly*, 81(5):474–480, May 1974. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/00029890.1974.11993589.

[10] Kevin Lyons. *Automated Force-Velocity Profiling of NFL Athletes via High-Frequency Tracking Data*. Thesis, Massachusetts Institute of Technology, 2021.

[11] Jack Lysholm and Jorgen Wiklander. Injuries in runners. *The American Journal of Sports Medicine*, 15(2):168–171, March 1987. Publisher: SAGE Publications Inc STM.

[12] J. Mendiguchia, P. Edouard, P. Samozino, M. Brughelli, M. Cross, A. Ross, N. Gill, and J. B. Morin. Field monitoring of sprinting power–force–velocity profile before, during and after hamstring injury: two case reports. *Journal of Sports Sciences*, 34(6):535–541, March 2016. Publisher: Routledge _eprint: https://doi.org/10.1080/02640414.2015.1122207.

[13] Jean-Benoît Morin and Pierre Samozino. Interpreting Power-Force-Velocity Profiles for Individualized and Specific Training. *International Journal of Sports Physiology and Performance*, 11(2):267–272, March 2016. Publisher: Human Kinetics, Inc. Section: International Journal of Sports Physiology and Performance.

[14] Irwin Nunez. An Inside Look at the Grueling Schedules of NFL Players After Game Day, March 2020.

[15] Jernej Pleša, Žiga Kozinc, and Nejc Šarabon. The Association Between Force-Velocity Relationship in Countermovement Jump and Sprint With Approach Jump, Linear Acceleration and Change of Direction Ability in Volleyball Players. *Frontiers in Physiology*, 12, 2021.

[16] L. G. C. E. Pugh. The influence of wind resistance in running and walking and the mechanical efficiency of work against horizontal or vertical forces. *The Journal of Physiology*, 213(2):255–276, 1971. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1113/jphysiol.1971.sp009381.

[17] P. Samozino, G. Rabita, S. Dorel, J. Slawinski, N. Peyrot, E. Saez de Villarreal, and J.-B. Morin. A simple method for measuring power, force, velocity properties, and mechanical effectiveness in sprint running. *Scandinavian Journal of Medicine & Science in Sports*, 26(6):648–658, 2016. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sms.12490.

[18] H. Vandewalle, G. Peres, J. Heller, J. Panel, and H. Monod. Force-velocity relationship and maximal power on a cycle ergometer. *European Journal of Applied Physiology and Occupational Physiology*, 56(6):650–656, September 1987.

[19] Owen Walker. Force-Velocity Curve - Science for Sport. Section: All.