Using Tracking and Charting Data to Better Evaluate NFL Players: A Review

Eric Eager, Ben Brown, George Chahrouri, Timo Riske, Brad Spielberger, Lau Sze Yui, Zach Drapkin, Tej Seth

The game of football is undergoing a significant shift towards the quantitative. Much of the progress made in the analytics space can be attributed to play-by-play data and charting data. However, recent years have given rise to tracking data, which has opened the door for innovation that was not possible before. In this paper we combine charting and tracking data to build metrics for pass rushers, linebackers and receivers and show that combining charting and tracking data can help us evaluate players better than either data source by itself.

1. Introduction

The game of American football is undergoing a significant shift towards the quantitative. While the most explicit biproduct of this shift can be found in the decisions that coaches are making during games [5], [1], [2], [23], the way in which teams scout [24], evaluate [29] and value players [36], [17] has also changed substantially during the last decade. The genesis of these changes has been the ubiquity of available data.

Until recently, football data was largely available only through NFL-sanctioned box scores and play-by-play data, with companies like Football Outsiders [21] and people like Brian Burke [4], [5] moving the conversation forward with the invention of defensive value over average (DVOA) and expected points and expected points added (EPA), respectively. Ben Baldwin and Sebastian Carl recently made nflfastR [3] publicly available, building on the work of Ron Yurko, Sam Ventura, and Max Horowitz [36] in getting play by play data and EPA into the hands of aspiring analysts and fans. A great deal of the breadth in public football analysis and consumption can be traced back to these advancements.

During the middle of the last decade, Neil Hornsby started Pro Football Focus [30], which has assigned a numerical grade between -2 and 2, in increments of 0.5 to every player on every play in each National Football League (NFL) game from 2006 to the present, based on a set of criteria aimed at emulating traditional scouting methods. PFF added NCAA football to its data collection in 2014. In addition to these grades, over 200 data points are collected for each player on each play of each game, and this data is sold to each NFL team and over 100 NCAA teams in raw form, along with other analytics tools. The grades have since been translated onto a 0-100 scale for consumer customers, who have used the data and tools for things like covering the sport to fantasy football and sports betting. Prior to PFF, evaluation of players at the line of scrimmage (e.g., offensive linemen) was limited, whereas now many postseason accolades and contract values are derived using PFF's data and analysis. Metrics like PFF WAR [17], which assigns a win share to each player in the NFL and NCAA, or more finely grained metrics like quarterback accuracy or running back rushing yards above expected [33], [34] have moved the football conversation forward drastically, to the point where almost no serious football analysis does not include some PFF data in her or his work. The vast majority PFF's data is manually obtained, data that we will refer to as "charting" data moving forward.

Finally, since 2017 the NFL has placed radio-frequency identification (RFID) chips in the shoulder pads of all of its players, as well as the football, in an effort to obtain (x,y) coordinate data, along with things like orientation, direction, speed and acceleration. This data, known as "tracking data," is provided each tenth of a second, and has been made available to all NFL teams and select media entities. Additionally, thanks to the work of Michael Lopez and his staff, the NFL created the Big Data Bowl each year since 2019, which has posed a set of problems to prospective data scientists and engineers using this data. Example outputs for this competition can be found in [10], [12], [37] and [27].

Despite the promise of tracking data, and the great work of people like Burke [6], [7], [8] and NFL's Next Gen Stats [27], the promise of tracking data is still in the early stages of being fully realized, both within NFL teams - but especially in the public sphere. Burke's work in evaluating play in the trenches has caught on well [7], [8], and the 2020 Big Data Bowl winning solution – a model for rushing yards above expected (RYOE, [27]) has made its way into television broadcasts of games, but for the most part the outputs of these models are trying to replicate or improve on the work of charting companies like PFF, rather than using it to enhance our ability to evaluate and predict the performance of players.

In this paper we use charting and tracking data together to show the promise of using both data sets to evaluate and predict the performance of players in several settings. In Section 2 we describe the methods we used to blend charting and tracking data to measure pass-rusher get-off rates, linebacker bite distances on play action, and wide receiver speed on deep routes. In Section 3 we dive into model diagnostics and results as they pertain to stability and predictive power from one year to the next. In Section 4 we summarize the scope of the results in this paper and foreshadow things to come in the football analytics space using tracking data.

2. Methods

In each section we use PFF play-by-play data (charting data) and the NFL's NGS data (tracking data), alongside data sets like those gathered at the NFL Scouting Combine. There are so many things one can do when armed with these data, and in this paper, we will choose a subset of three player evaluation tools. For a good guide on the differences between the two types of data sets, there is a summary in a StatsBomb research paper [9]. The metrics and their approaches are outlined here:

2.1. Pressure Rate, Win Rate and Sack Rate
Getting pressure on the quarterback is one of the more important things a defense does. Pressure is the process by which sacks, tackling the quarterback behind the line of scrimmage before he can throw the ball, occur. PFF's pressure rates have been shown to be one of the more effective ways of predicting sacks and sack rate by individuals and teams on a year-to-year basis [14]. PFF pass-rushing player grades can also add context to pressure and sack rates, as they can help distinguish the pass-rush plays in which a player produced a pressure that was not his doing (e.g., a "clean-up" pressure) or a player did not produce a pressure but would have in league-average circumstances. A positive PFF grade is considered a "win" in pass rushing.

However, even though PFF pressure rates have been predictive, there is evidence that athleticism matters more for pass rushers than other positions, especially edge players, than any other facet or position [14]. Thus, in Section 3.1 we use the tracking data to measure a player's "get off" on pass-

rushing snaps, by measuring their total Euclidean and vertical (y) distance from their position at the snap during the first one second of the play (see [23] for an overview on how ESPN approached this problem). Given how much context can influence a player's get off, we translated these get-off distances into a probability of obtaining a pressure and the probability of gaining a pass-rush win on a given play, using an XGBoost modeling framework [35]. The following features were used in our XGBoost model: season, whether the offense is the home team, the home team stadium (to account for stadium effects in the tracking data), down, distance, quarter, time left in quarter, whether the game was played in a dome, whether the game was played on turf, how much rest the player had prior to that game, the football position of the pass rusher (using charting data), as well as x position of the pass-rusher at the snap, y position of the pass-rusher at the snap, speed, and vertical speed. The caret package in R was used [30], with five-fold cross validation.

In Section 3.1 expected pressure and expected win rates were then calculated by aggregating these play-by-play probabilities and were then investigated for their relationship to traditional measures of athleticism for edge and interior pass rushers, as well as future pressure, win and sack rates in conjunction with raw PFF charting data.

2.2. Linebacker Coverage Grades and Play-Action Awareness
Along with pass rushing, coverage is the second part of stopping a passing game. We showed in [16] that coverage performance is both more highly correlated with team defensive success in a season and predictive of it in subsequent seasons. However, coverage grades and statistics are nowhere near as stable year-to-year on the individual or team level as compared with pass-rushing metrics. This makes sense to a degree, as coverage activity is largely in response to what an offense does, and a defense has relatively little control over what an offense does. Contrast this with pass rushing, which is by-and-large an attacking action by a defense.

Speaking to this, play action is one of the ways in which an offense can get an opposing defense in a precarious position, and has thus been studied extensively as a means by which offenses have been able to gain a sustainable edge [2], [3], [22]. In this paper we measure how much off-ball linebackers are displaced by play action, above or below what would be expected based on play- and opponent-level features. We use the change in the player's y coordinate (as measured with tracking data) as his measure of displacement. As with pass rushing, we use an XGBoost model to calculate this, using the features: season, down, distance, whether there was a blitz on the play (players who are charted as blitzers are not considered for this analysis, but an indicator variable for the existence of a blitz is necessary), position (based on charting data), x and y position at the snap of the ball, the percentage of time the opposing offense runs power, counter and man run concepts, the percentage of time the opposing offense runs outside and inside zone, the number of players in the box (using charting data), dropback type and depth for the quarterback (determined by charting data), whether the offense was home team or away team, whether the game was played in a dome, on turf, along with the stadium in which the game is played.

In 3.2 we analyze this new metric, which we call Bite Distance Under Expected (BDUE), for stability as well as its ability to track with team defensive performance. Additionally, we use BDUE to enhance predictions of PFF coverage grades for linebackers from one year to the next.

2.3. Wide Receiver and Tight End Deep Speed
Since tracking data has been available, both at the NFL and collegiate level [11], player speed has been the lowest of low-hanging fruit for analysis. However, without context, simple miles per hour measurements are rarely useful for anything other than possibly detecting injuries in players

within a game or a season or finding players that play faster than their traditional times during offseason and pre-draft drills.

However, if armed with charting data, data that contains information as to how a player is using his raw speed, these readings can be put in the context of the game of football. To do this, we used PFF charting data to determine a player's performance on deep routes as a function of his speed. The model uses a player's PFF receiving grade – a number between -2 and 2 – as the response variable, with the following as features: season, whether the offense is the home team, the home team stadium, down, distance, quarter, time left in quarter, whether the game was played in a dome, whether the game was played on turf, how much rest the player had prior to that game, the football position of the receiver (using charting data), whether the receiver was lined up on the line of scrimmage, whether the receiver was pressed whether there was a blitz and/or a pressure on the quarterback, whether the play involved a run-pass option or a screen, whether a pass was actually attempted, the time to throw (or sack/scramble if a throw was not attempted), whether there was play action, and finally the player's max speed, average speed or median speed of the receiver during the play using tracking data.

The reason for using the PFF grade for a player on a pass play was due to the fact that, for the most part, players only earn positive PFF grades on deep routes when they earn a target and given how stable target rate is for a player, there's a lot of evidence that earning a target is a skill a wide receiver possesses and hence should be selected for and rewarded by analysis [24]. Additionally, a negative grade earned on a deep pass is likely the product of not being able to use one's speed effectively to track the ball, gain separation from a defender, etc., when the ball is thrown his way.

In Section 3.3 we then compute a player's tracking-data-adjusted 40-yard dash time by fitting a simple linear model with each player's per-route tracking-data grade as the feature and his actual 40-yard dash time prior to entering the NFL as the response. Only plays where the receiver ran an "8" route (a post) or a "9" route (e.g., a go route) from a wide receiver position (either out wide or in the slot) were used in this analysis, and the minimum threshold to make the above training set was 150 total deep routes total from 2017-2020. These adjusted 40-yard dash times were then analyzed for trends at the individual player level.

3. Results

3.1. Pressure Rate, Win Rate and Sack Rate

3.1.1 Predictive Power of Charting and Tracking Data
Models for pressure probability and win probability at the play and player level were compared with naïve models that include no features at all, and those that include all of the features stated in Section 2.1, except for (x,y) position at the snap and both speed variables (using tracking data). The values are in Table 3.1.1

Model	ROC (Naïve)	ROC (Naïve to Tracking)	ROC (Full Model)
Pressure probability	0.500	0.611	0.620
Win probability	0.500	0.584	0.604

Table 3.1.1: ROC values for naïve model, a model naïve to tracking data and a full model for pressure probability and probability of a pass-rush win.

Thus, the speed at which a player gets off the ball does affect his probability of obtaining a pressure or a pass-rush win, even after controlling for contextual variables. Figure 3.1.1 has the variable importances for the expected pressure rate, which are qualitatively similar to those for the expected win rate:
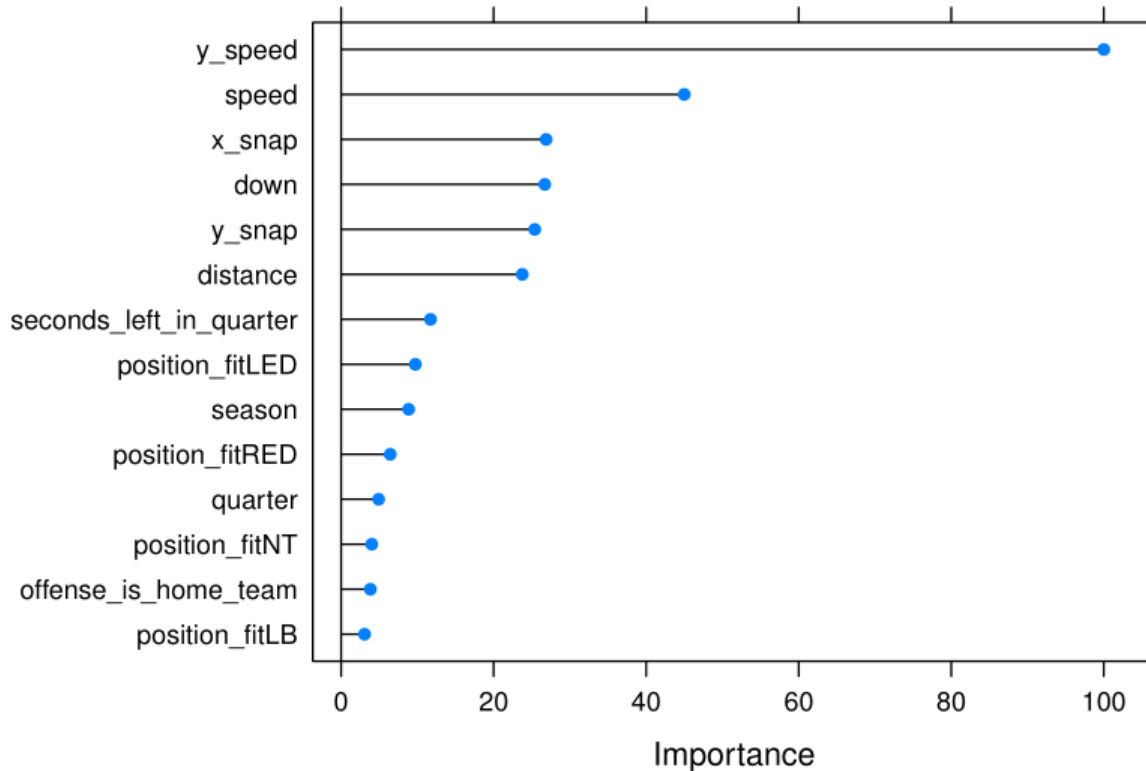

Figure 3.1.1: The highest variable importances for the XGBoost model for pressure probability at the player and play level variable importances are on a scale of 0 (not important given the model construction) to 100 (the most important given the model construction). x_speed and speed are the player's speed in the y direction and overall, x_snap and y_snap are the x and y positions at the snap of the ball, respectively, and position_fit is the charting-data position of the player. down and distance, seconds_left_in_quarter, quarter, offense_is_home_team and season are self-explanatory.

As expected, the tracking-data-related variables are more important to the probability of earning a pressure than the play-level charting data that is used in the naïve models. It also makes sense that vertical speed is more important than overall speed (which comes into play more on exotic pass-rushing plays like stunts and loops and less on more traditional rushes). Positioning is very important, both in a raw tracking sense but also in the traditional charting sense (i.e., if you're an edge player or an interior player). Stadium effects are not the most important, but they have a non-zero variable importance, as do things like playing on turf or on different rest differentials.

When aggregated over an entire season, these probabilities can be averaged into an expected pressure and expected win rate that are analogous to a player's actual pressure and actual win rate. These rates are extremely stable at the player level relative to other football metrics. For players with more than 150 pass-rushing snaps in back-to-back seasons from 2017 to 2020, expected

pressure rate has an r-squared value of 0.858 from year n to year n + 1 and expected win rate 0.861.

As for how these metrics predict other metrics we care about, it's an encouraging sign that not only do expected pressure and expected win rate in year n predict sacks in year n + 1 better than sacks themselves do (which have an r-squared of 0.202), but also better than their PFF counterparts alone (Tables 3.1.2 and 3.1.3). Most importantly when both of these rates are used together the r-squared values increase. In other words, charting data combined with tracking data do a better job than either one individually.

Metric	R-Squared (Pressure Rate)	R-Squared (Expected Pressure Rate)	R-Squared (Both)
Sack Rate	0.274	0.318	0.344
Pressure Rate	0.465	0.405	0.507

Table 3.1.2: The relationship between pressure rate (charting data) and expected pressure rate (tracking data) in year n and sack and pressure rate in year n + 1 for each NFL player with more than 150 pass-rushing reps in both years (n = 471 players, 2017-2020). As a baseline, sack rates in year n have an r-squared value of 0.202 with sack rates in year n + 1.

Metric	R-Squared (Win Rate)	R-Squared (Expected Win Rate)	R-Squared (Both)
Sack Rate	0.263	0.325	0.353
Pressure Rate	0.535	0.370	0.558

Table 3.1.3: The relationship between win rate (charting data) and expected win rate (tracking data) in year n and sack and pressure rate in year n + 1 for each NFL player with more than 150 pass-rushing reps in both years (n = 471 players, 2017-2020). As a baseline, sack rates in year n have an r-squared value of 0.202 with sack rates in year n + 1.

3.1.2. The Relationship Between Off-Field Athleticism and Those Measured by Tracking Data

Not surprisingly, these measures of pass-rushing "get off" map extremely well from traditional measures of athleticism (Table 3.1.4 and Figure 3.1.2), which are usually measured during a player's career at the NFL Scouting Combine and/or their "pro day", which is an event held at the player's university. There is some bias associated with pro days, which is written about in Eager 2021. Thus, we do make an adjustment to pro day data so that it is comparable to combine data before we compare it to a player's overall expected pressure and expected win rate from 2017-2020.

Athleticism Measure	R-Squared (Pressure Rate)	R-Squared (Expected Pressure Rate)
40-yard dash	0.380	0.567
3-cone drill	0.313	0.456
Short shuttle	0.334	0.430
Broad jump	0.239	0.369
Vertical jump	0.185	0.280

Table 3.1.4: The relationship with common measures of player athleticism and pressure rate (charting) and expected pressure rate (tracking data) for players with 150 or more pass-rushing snaps during 2017-2020 and a measurement from either the NFL Scouting Combine or a pro day.

The promise of the above work is apparent in that traditional measures of athleticism (at least in the public data space) are a one-time snapshot of information for a player, and while Table 3.1.4 shows that there is a significant correlation between it and on-field performance for pass rushers, being able to create a proxy for this information that is *time-varying* significantly enhances the ability to evaluate a player's current form, which may have changed due to things like an improvement due to maturation, decline due to age or injuries.
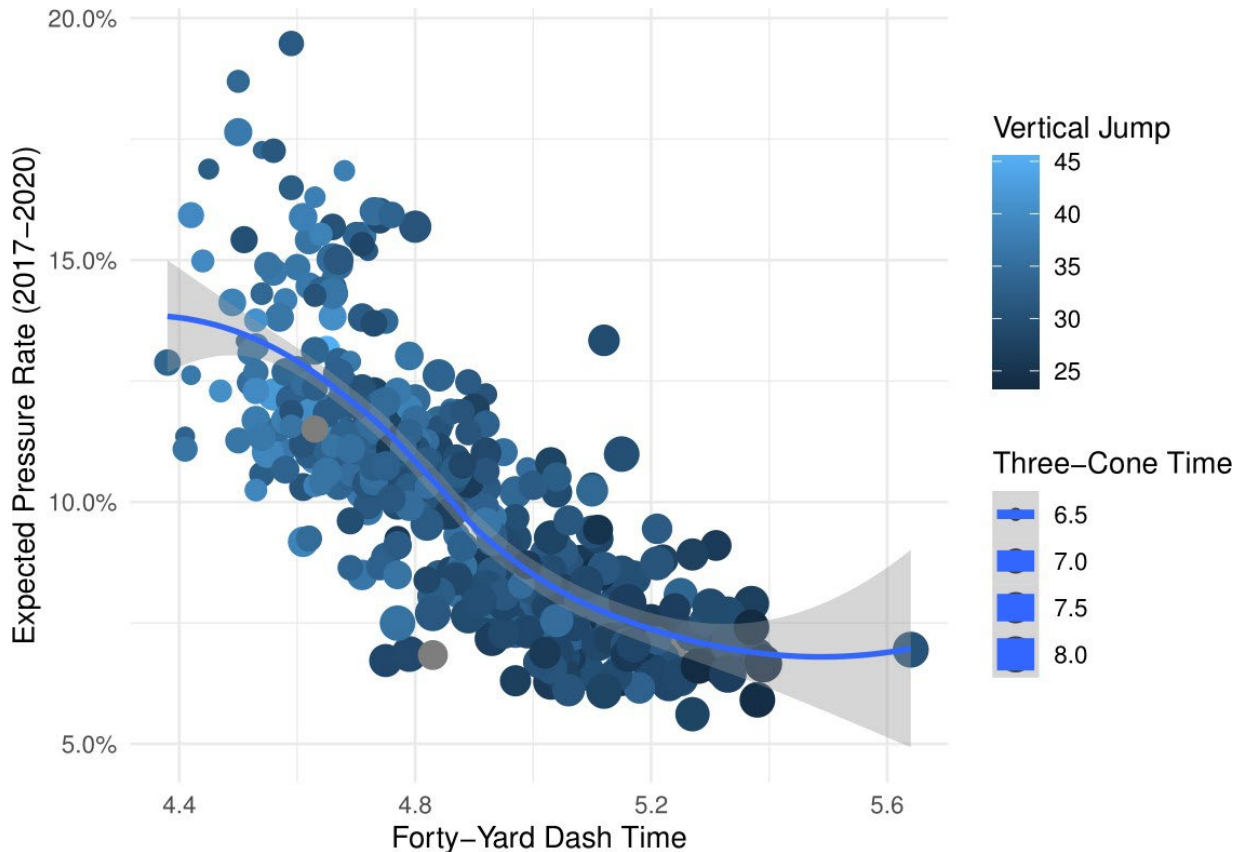

Figure 3.1.2: Expected pressure rate's (as measured using tracking data) relationship with forty-yard dash time (x-axis), vertical jump (point color) and three-cone time (point size).

3.2. Linebacker Coverage Grades and Play-Action Awareness

3.2.1 Model Diagnostics

Models for bite distance at the play and player level were compared with naïve models that included no features at all, and one that includes all the features stated in Section 2.2. The RMSE with the naïve model was 3.00, whereas with the features included it was 2.23. Variable importances for the full model are in Figure 3.2.1.
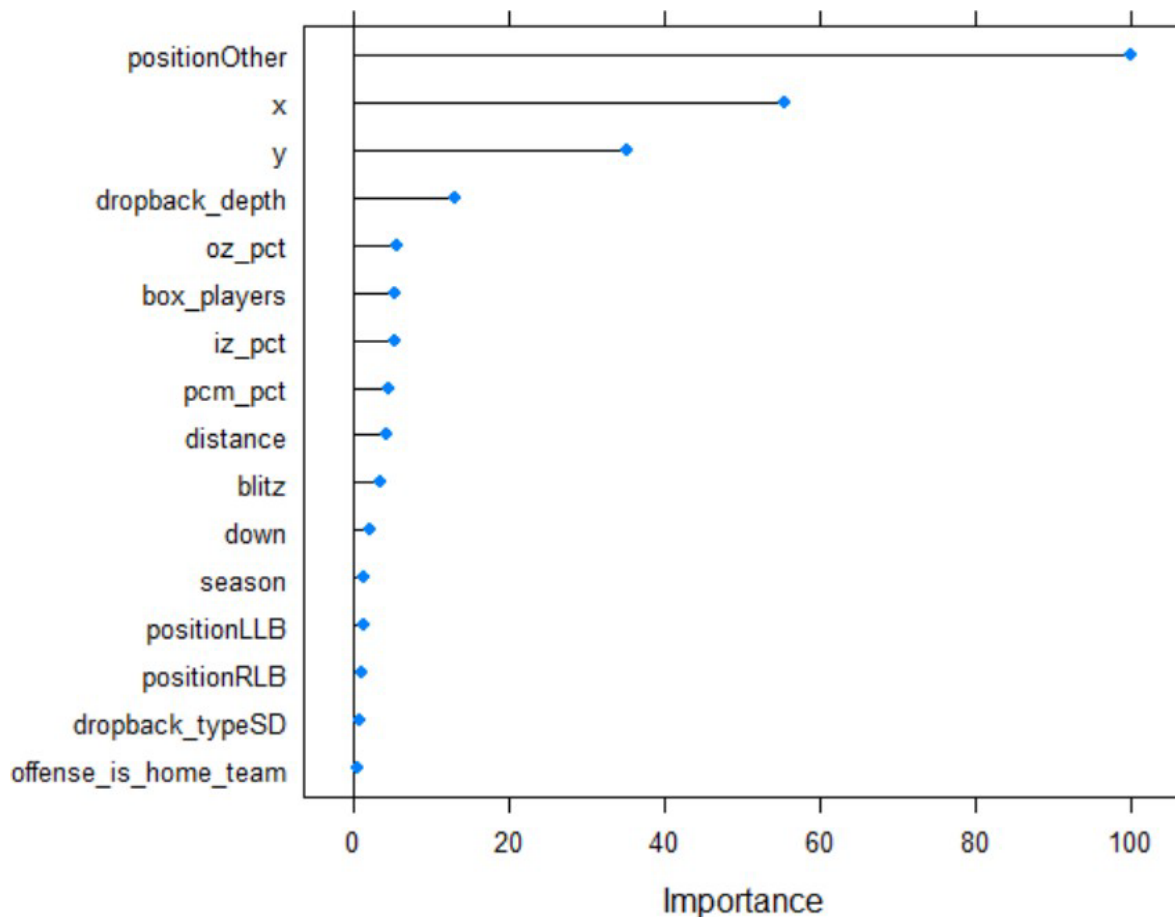
Figure 3.2.1: The highest variable importances for the XGBoost model for bite distance at the player and play level variable importances are on a scale of 0 (not important given the model construction) to 100 (the most important given the model construction). x and y are the x and y positions at the snap of the ball, respectively, and position is the charting-data position of the player. dropback_depth and dropback_type are the dropback depth and type, as measured using charting data. oz_pct, iz_pct and pcm_pct are the percentage of outside zone, inside zone and power/counter/man plays by the team's opponent coming into that game. blitz is whether there was a blitz on the play by the defense. down and distance, offense_is_home_team and season are self-explanatory.

It's instructive to see that bite distance depends the most on a player's charting data position in rare circumstances, likely due to the fact that while (x,y) position – which are the second- and third-most important variables, matter, the "role" of a player – which is probably still better judged through charting, determines the propensity for a player to bite or not to bite on certain plays. The proportion of time the opposition runs certain run concepts is also very important, as is down and distance.

To evaluate players and teams using this model, we subtract the expected bite distance from the actual bite distance to create bite distance under expected (BDUE), which we analyze below.

3.2.2   Analysis of Bite Distance Below Expectation (BDUE)

Bite distance below expectation (BDUE) is the difference between actual and expected bite distance for a player. BDUE values > 0 mean the defensive player bit less than expected, and < 0 more than expected. For players who face more than 50 play-action snaps in a season at an off-ball linebacker spot (determined by charting data), the year-to-year r-squared value is 0.350 (Figure 3.2.2). The correlation between average BDUE among a team's linebackers and a team's EPA allowed was small (r-squared equal to 0.046) but due in large part to the fact that a team's play action rate against is not correlated with their average BDUE value (r-squared values ~ 0).
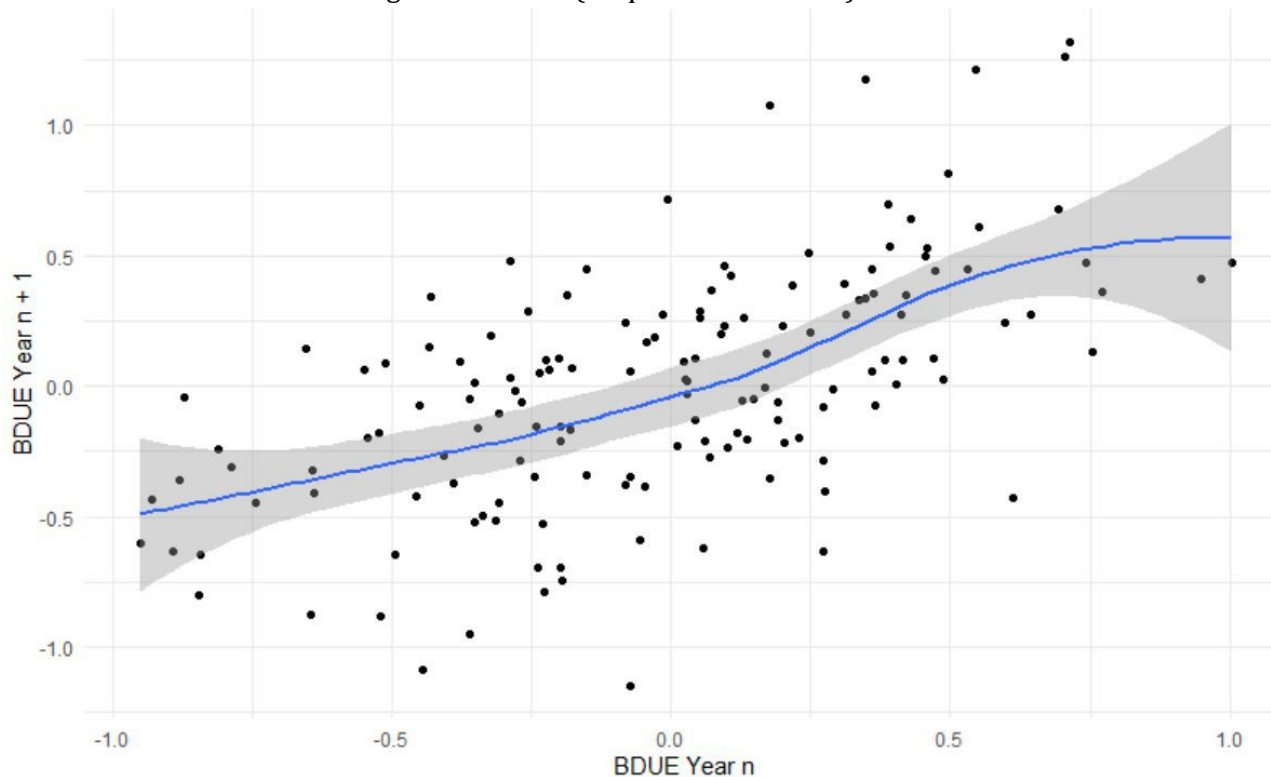

Figure 3.2.2: The year-to-year relationship between BDUE for off-the-ball linebackers in the NFL from 2017-2020. A minimum of 50 coverage snaps against play-action in both years were required.

Table 3.2.1 shows the best linebackers in terms of BDUE since tracking data has been made available. Some interesting names emerge, with a couple of New England Patriots in Dont'a Hightower and Jamie Collins (although he earned his high mark in Cleveland), along with 2020 Super Bowl champion linebacker Devin White. Fred Warner signed a contract in the 2021 offseason that made him the highest-paid linebacker in the history of football at the time (since surpassed by Colts' linebacker Darius Leonard). Luke Kuechly will likely make the Pro Football Hall of Fame.

Player	Season	Play Action Snaps	BDUE
Neville Hewitt	2020	155	1.32
Dont'a Hightower	2019	75	1.26
Jatavis Brown	2018	86	1.21
Thomas Davis	2019	117	1.17

Devin White	2020	165	1.08
Jamie Collins	2018	96	1.00
Luke Kuechly	2017	139	0.95
Fred Warner	2020	140	0.81

Table 3.2.1: The top BDUE values in the NFL from 2017 to 2020. A minimum of 50 coverage snaps against play action were required.

As far improving upon our existing metrics, BDUE has promise as well. Currently, play-by-play PFF grades for players playing off-the-ball linebacker (before any adjustment for situation) correlates year-to-year with an r-squared value of 0.082 (from 2017 to 2020) for players with more than 150 coverage snaps off the ball, and 50 off-the-ball snaps against play action (n = 147 players). This improves to 0.112 when adding in BDUE into a linear model. Interestingly, BDUE and PFF raw coverage grade are less correlated with each other in a given season (r-squared of 0.038) than they are from one year to the next (r-squared of 0.052), suggesting that measuring this trait is important in and of itself.

3.3. Wide Receiver and Tight End Deep Speed

3.3.1. Modeling PFF Grade on Deep Passes Using Tracking Data

The model for PFF grade earned on deep routes were trained naively to all variables, naively to all non-speed variables and using all variables, with model performance in Table 3.31. Variable importances can be found in Figure 3.3.1.

Model	RMSE (Naïve)	RMSE (Naïve to Tracking)	RMSE (Full Model)
PFF deep receiving grade	0.194	0.193	0.188

Table 3.3.1: Root Mean Squared Errors for a model for PFF receiving grade on deep routes. One model is completely naïve to all factors, the second naïve to speed-related factors (using tracking data) and the last a full model.

It's reassuring that the speed variables, as measured by the tracking data, are the most important in this model, and it's not surprising that a player's maximum speed is the most important variable here, with median (not outlier influenced) speed the second-most important. The order of the remaining, contextual variables makes sense, and while we did not display them in Figure 3.3.1, different stadiums did influence the model in a small way, due likely in part to small differences in the way stadiums had installed the NGS system.
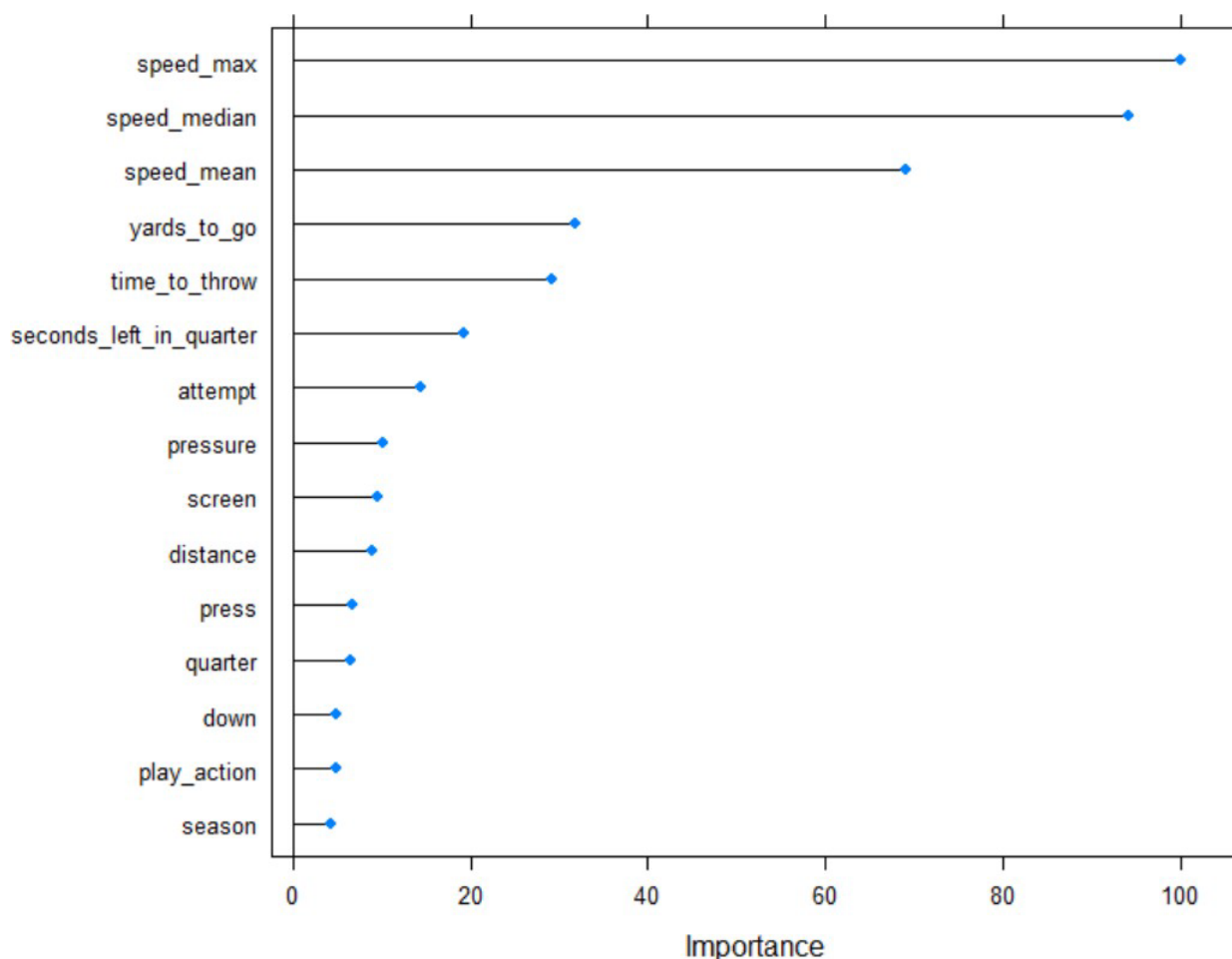
Figure 3.3.1: The highest variable importances for our model for PFF grade on deep routes. Variable importances are on a scale of 0 (not important given the model construction) to 100 (the most important given the model construction). speed_max, speed_median and speed_mean are the max, median and mean speed for the receiver during the play. attempt is whether a pass was attempted on the play, pressure was whether there was pressure, time_to_throw was the time in the pocket for the passer, press is whether the receiver of interest was pressed on that play. screen is whether there was a screen on the play. down, distance, yards_to_go, seconds_left_in_quarter, quarter and season are self-explanatory.

On each play a player runs an "8" or "9" route from a wide receiver position, either outside or in the slot, we applied this model to assign them a tracking-data-adjusted PFF deep receiving grade (TDA-PFF grade). These grades were then aggregated, and for players with more than 150 deep routes in a season from 2017-2020, the top names are not surprising (Table 3.3.2):

Player	Season	Deep Routes	Deep Targets	TDA-PFF
Antonio Brown	2018	299	46	15.4
Tyreek Hill	2018	236	39	15.0

Antonio Brown	2017	271	66	14.9
Curtis Samuel	2019	251	31	14.4
Tyreek Hill	2017	215	27	14.3
Brandin Cooks	2017	248	36	14.1
John Brown	2018	248	36	13.8
Julio Jones	2018	165	42	13.1
Tyreek Hill	2020	231	28	13.1
Marquez Valdes-Scantling	2020	180	36	12.6

Table 3.3.2: TDA-PFF grades for players with 150 or more deep receiving routes in a season from 2017-2020.

These top aggregated grades give the names of some of the truly elite receivers in the NFL (Antonio Brown, Julio Jones, Tyreek Hill, Brandin Cooks) as well as some of the elite deep threats (Hill, Cooks, John Brown). It also gives some surprising names in Samuel and Valdes-Scantling, players who both ran sub 4.4 seconds in the forty-yard dash coming out of college (4.31 and 4.37, respectively), and have had their moments as NFL players as deep receivers.

While these grades are useful, translating them back into the language that most analysts, media members and fans already have familiarity with – 40-yard dash times – might allow adoption of this metric to be quicker. As such, we took all players that had more than 150 deep receiving snaps from 2017 to 2020, their TDA-PFF grades from Table 3.3.2, divided by the number of deep receiving snaps, and used a simple linear model to regress to their 40-yard dash time coming into the league (with the same adjustments as in 3.1). Figure 3.3.2 displays the relationship between these two variables, which has an r-squared value of 0.2671. As a comparison, 40-yard dash time has an r-squared value of 0.04 with raw PFF grades on deep routes.

MIT SLOAN
SPORTS ANALYTICS CONFERENCE
presented by ESPN 42 ANALYTICS
MARCH 4-5, 2022
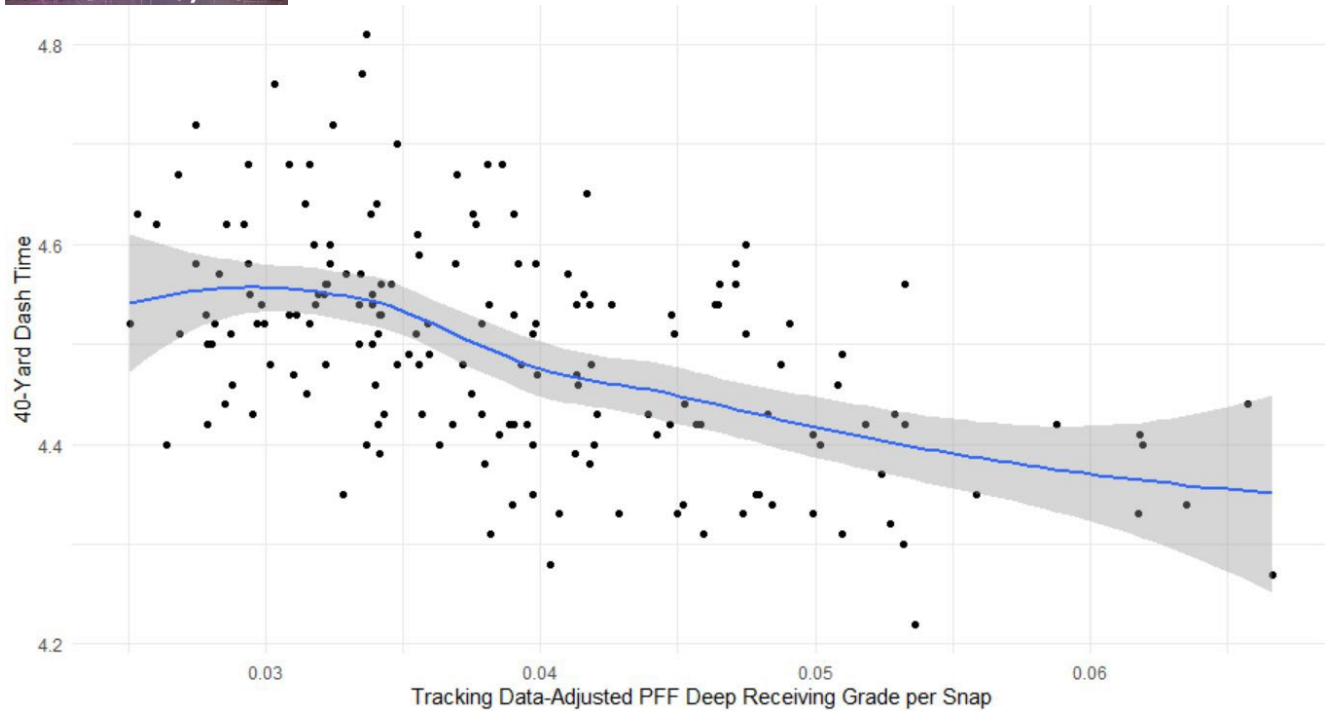
KAGR
KRAFT ANALYTICS GROUP

Figure 3.3.2 : The relationship between our TDA-PFF grade on a per-snap basis and a player's 40-yard dash time coming into the league. A minimum of 150 receiving snaps required to qualify.

3.3.2.   Who Plays Faster or Slower than Their 40 Time?

Looking at speed this way can help us quantify the players who are "faster in pads than in shorts", or players who maybe have been overdrafted because of their performance in Indianapolis during the scouting combine, despite their relative lack of receiving ability. It can also be used to find players that are on the decline; provide a leading indicator that a player's production might slip in the future due to a decrease in physical skills.

Player	Deep Routes	Deep Targets	40-Yard Dash	TDA-40
J.J. Nelson	244	33	4.28	4.49
Damiere Byrd	331	26	4.31	4.50
Deonte Thompson	211	26	4.35	4.53
John Ross	282	44	4.22	4.40
Corey Coleman	170	23	4.40	4.57

Table 3.3.3: Players whose career TDA-40 times are slower than what they timed coming into the NFL.

MIT SLOAN
SPORTS ANALYTICS CONFERENCE
presented by ESPN 42ANALYTICS
MARCH 4-5, 2022

KAGR
KRAFT ANALYTICS GROUP

Player	Deep Routes	Deep Targets	40-Yard Dash	TDA-40
Jarvis Landry	571	110	4.77	4.53
Devin Funchess	356	68	4.70	4.52
Preston Williams	170	37	4.65	4.48
Diontae Johnson	305	57	4.6	4.44
Antonio Brown	679	132	4.56	4.40

Table 3.3.4: Players whose career TDA-40 times are faster than what they timed coming into the NFL.

In the first group you have two players, John Ross and Corey Coleman, who were high draft picks based almost entirely on how fast they ran at the NFL Scouting Combine, neither of whom amounted to much in the NFL, while in the second group you find two of the better receivers in the NFL over the last few years in Jarvis Landry and Antonio Brown, with an emerging star in Dionte Johnson. This metric year-to-year is stable relative to other receiving metrics, with an r-squared value of 0.309 for players with more than 50 deep receiving snaps in a season.

4   Discussion

The current revolution in football analytics began with play-by-play data and charting data, but it will be accelerated with tracking data. In this paper we built three evaluation metrics for players using both charting and tracking data and showed that in all three cases tracking data added either predictive power to the charting data and/or new insights all together.

In the case of pass rushers and receivers, tracking data can help us enhance or replace the traditional means we have for measuring athleticism, which is the NFL draft and pro days. We have already seen this make its way into the league through a discussion of Rams wide receiver Cooper Kupp [32], who's slow 40-yard dash at the combine was overlooked in favor of GPS data obtained at the Senior Bowl in Mobile, AL. The combination of him coming from a small school and Los Angeles looking at this measurement, as opposed to traditional ones, gave them an edge in acquiring him when others would not.

Our analysis of him using charting and tracking data reinforce his story, as Table 4.1 shows his tracking data-adjusted 40-yard dash during the first four years of his career, which are all faster than his time at the NFL Scouting Combine (4.62). He is now arguably the best wide receiver in the NFL.

Player	Season	Deep Routes	Deep Targets	TDA-PFF
Cooper Kupp	2017	120	16	4.53
Cooper Kupp	2018	38	6	4.38
Cooper Kupp	2019	91	8	4.50

Cooper Kupp	2020	64	6	4.46

Table 4.1: Cooper Kupp's tracking data-adjusted 40-yard dash times for his 2017-2020 seasons. His combine time was 4.62, which was slower than all of his on-field data.

Future research in this area is working to emulate other combine-like measures like three-cone drill and vertical jump, using on-field performance as our gauge, so that we can have a robust, time-varying measure of player athleticism that can sharpen player prediction models well into a player's career. Additionally measures of player athleticism that are sharper than pre-career combine numbers likely will map better to things like draft position or future contract values for a player. Tracking data at the college level and additional years of NFL tracking data, respectively, will be necessary to this cause, but are not far away in terms of availability.

In the case of linebackers, we are finally able to measure the impact that play action – the most-studied area of deception in football [2], [3], [22] has on the linebacker position, and if that trait is stable year-to-year and maps to other things we care about (like PFF receiving grades). In the case of our BDUE metric, we get both. Future work will look at other means of deception – run-pass options, read options, pre-snap, and at-the-snap motion – to understand how these tactics displace defenders and how that displacement affects defensive performance. There have been studies that have alluded to the "gravity" of players, for example running quarterbacks [15]. but to get to the mechanism of these outcomes is the promise of tracking data. Research like that in [20] has been very fruitful in basketball and soccer, but people have only just begun using it in football.

The football data revolution has been encouraging to this point, with our knowledge of the game of football growing with each year. The availability of tracking and charting data, along with the non-stationarity of the sport, means that the breadth of problems we can solve will be limited only by our imaginations.

Acknowledgements

References

[1] Baldwin, B. (2018) Rushing success and play-action passing.
https://www.footballoutsiders.com/stat-analysis/2018/rushing-success-and-play-action-passing

[2] Baldwin, B. (2018) Further research on play-action passing.
https://www.footballoutsiders.com/stat-analysis/2018/further-research-play-action-passing

[3] Baldwin, B. (2021) nflfastR, EP, WP, CP, xYAC, and xPass models
https://www.opensourcefootball.com/posts/2020-09-28-nflfastr-ep-wp-and-cp-models/

[4] Burke, B. (2010) Expected points (EP) and expected points added (EPA) explained.
http://www.advancedfootballanalytics.com/2010/01/expected-points-ep-and-expected-points.html

[5] Burke, B. (2013) How coaches and the NYT Bot compare.
https://www.nytimes.com/newsgraphics/2013/11/28/fourth-downs/post.html

[6] Burke, B. (2019) DeepQB: Deep learning with player tracking to quantify quarterback decision-making & performance. 2019 MIT Sloan Sports Analytics Conference.
https://www.sloansportsconference.com/research-papers/deepqb-deep-learning-with-player-tracking-to-quantify-quarterback-decision-making-performance

[7] Burke, B. (2019) We created better pass-rusher and pass-blocker stats: How they work.
https://www.espn.com/nfl/story/_/id/24892208/creating-better-nfl-pass-blocking-pass-rushing-stats-analytics-explainer-faq-how-work

[8] Burke, B. (2020) Introducing new NFL run-blocking and run-stopping stats: How run block win rate and run stop win rate work. https://www.espn.com/nfl/story/_/id/29813062/introducing-new-nfl-run-blocking-run-stopping-stats-how-run-block-win-rate-run-stop-win-rate-work

[9] Burriel, B., Buldu, J.M. (2021) The quest for the right pass: Quantifying player's decision making.
https://statsbomb.com/2021/11/statsbomb-conference-2021-research-papers/

[10] Chu, D., Reyers, M., Thomson, J., Wu, L. (2019) Route identification in the NFL. Journal of Quantitative Analysis in Sports. 16(2): 121-132

[11] Cohen, A. (2020) Sportlogiq and Telemetry Sports team up to track college football players.
https://www.sporttechie.com/sportlogiq-telemetry-sports-partnership-college-football-players-tracking

[12] Deshpande, S.K., Evans, K. (2020) Expected hypothetical completion probability. Journal of Quantitative Analysis in Sports. 16(2): 85-94

[13] Eager, E (2018) Just how important are sacks for a defense? https://www.pff.com/news/pro-just-how-important-are-sacks-for-a-defense

MIT SLOAN
SPORTS ANALYTICS CONFERENCE
presented by ESPN 42 ANALYTICS
MARCH 4-5, 2022

KAGR
KRAFT ANALYTICS GROUP

[14] Eager, E., Chahrouri G. (2018) Do PFF college grades translate to the NFL for pass-rushers? https://www.pff.com/news/draft-pff-college-grades-and-nfl-correlation-for-pass-rushers

[15] Eager, E. (2019) the value of rushing quarterbacks in the NFL. https://www.pff.com/news/nfl-pff-data-study-value-of-rushing-quarterbacks-in-the-nfl

[16] Eager, E., Chahrouri, G. (2019) PFF Data Study: Coverage vs. pass rush. https://www.pff.com/news/pro-pff-data-study-coverage-vs-pass-rush

[17] Eager, E. and Chahrouri, G. (2020) PFF WAR: Modeling player value in American football. 2020 MIT Sloan Sports Analytics Conference. https://www.sloansportsconference.com/research-papers/pff-war-modeling-player-value-in-american-football

[18] Eager, E. (2021) How text analytics correlate with draft position, and why it might be a clue for Justin Fields' fall down boards. https://www.pff.com/news/draft-text-analytics-correlate-with-draft-position-justin-fields-fall-down-boards

[19] Eager, E. (2021) Pro day data is not the same as combine data: A comprehensive analysis. https://www.pff.com/news/nfl-pro-day-data-is-not-the-same-as-combine-data-a-comprehensive-analysis

[20] Fernandez, J., Bornn, L. (2018) Wide Open Spaces: A statistical technique for measuring space creation in professional soccer.

[21] Football Outsiders (2021) https://footballoutsiders.com

[22] Hermsmeyer, J. (2019) Can NFL coaches overuse play-action? they haven't yet. https://fivethirtyeight.com/features/can-nfl-coaches-overuse-play-action-they-havent-yet/

[23] Hermsmeyer, J. (2021) The NFL has a new way to measure the explosiveness of pass rushers. https://fivethirtyeight.com/features/the-nfl-has-a-new-way-to-measure-the-explosiveness-of-pass-rushers/

[24] Hernandez, T.J. (2021) Most predictable wide receiver stats (2021 update). https://www.4for4.com/most-predictable-wide-receiver-stats

[25] Lopez, M. (2020) Bigger data, better questions, and the return to fourth down behavior: an introduction to a special issue on tracking data in the National Football League. Journal of Quantitative Analysis in Sports. 16(2): 73-79

[26] Mallepalle, S., Yurko, R., Pelechrinis, K., Ventura, S. (2020) Extracting NFL tracking data from images to evaluate quarterbacks and pass defenses. Journal of Quantitative Analysis in Sports. 16(2): 95-120

[27] NFL (2020) Next Gen Stats: Intro to Expected Rushing Yards. https://www.nfl.com/news/next-gen-stats-intro-to-expected-rushing-yards

[28] NFL (2020) NFL Operations. 2020 Big Data Bowl results.
https://operations.nfl.com/updates/the-game/2020-big-data-bowl-results/

[29] PFF (2018) How we grade offensive lineman. https://www.pff.com/news/pro-how-we-grade-offensive-and-defensive-linemen

[30] PFF (2021) https://pff.com

[31] R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[32] Rodrigue, J. (2021) Rams WR Cooper Kupp is inventing routes, blocking D-linemen, settling old scores and leading the NFL in production. https://theathletic.com/2855472/2021/09/30/rams-wr-cooper-kupp-is-inventing-routes-blocking-d-linemen-settling-old-scores-and-leading-the-nfl-in-production/

[33] Seth, T. (2021) Defining PFF's new metrics for plays over expectation.
https://www.pff.com/news/nfl-plays-over-expectation-pff-research-stats-grades

[34] Sze Yui, L. (2021) Running backs don't matter: A dive back into the popular NFL analytics phrase.
https://www.pff.com/news/nfl-running-backs-dont-matter-a-dive-back-into-the-popular-nfl-analytics-phrase

[35] XGBoost (2021) https://xgboost.readthedocs.io/en/latest/

[36] Yurko, R., Ventura, S.L., Horowitz, M. (2019) nflWAR: a reproducible method for offensive player evaluation in football. Journal of Quantitative Analysis in Sports. Journal of Quantitative Analysis in Sports. 15(3): 163-183

[37] Yurko, R., Matano, F., Richardson, L.F., Granered, N., Pospisil, T., Pelechrinis, K., Ventura, S.L. (2020) Going deep: models for continuous-time within-play valuation of game outcomes in American football with tracking data. Journal of Quantitative Analysis in Sports. 16(2): 163-182