

EFFICIENT SPEED USAGE AND THE IMPACT OF FATIGUE IN SPEED PERFORMANCE: AN EXPLORATORY STUDY

Calestini, Lucas Ha, Cathy

ABSTRACT

There are many complexities in evaluating individual speed performance in football. We explored one way to measure speed efficiency via the ratio of the vertical field gain to the average speed of the player with the ball in each play; from reception to first contact. We controlled for similar players x-y coordinates at ball reception by analyzing speed efficiency within five clusters. We also calculated the ball carrier's dominant region and distance to closest defenders to measure the impact of defensive pressure on offensive speed, but only distance was significant. Alvin Kamara, Todd Gurley, and Le'Veon Bell were the most efficient players in handoff plays whereas Melvin Gordon, Ty Montgomery and Christopher Thompson were the leaders in pass situations. Additionally, we explored a definition of fatigue using principal component analysis to group variables such as cumulative distance and time in the game, previous and maximum acceleration and speed dynamics of previous plays. Although no clear relationship was observed between our fatigue variables and speed, we found a significant relationship between cumulative time and rest to speed.

INTRODUCTION

In 2014, the NFL introduced RFID chips in player shoulder pads to track player statistics throughout the game. This has generated a wealth of data for analysis, including geolocation, speed, and acceleration. In the NFL's inaugural Big Data Bowl, detailed tracking data is provided to the analytics community to spur innovation through

crowd-sourcing. Our report will focus on the **first theme** provided in the competition, centered on speed. We will investigate: 1) which players are most effective at using speed on the field; and 2) understanding the factors that impact speed.

BACKGROUND

Speed is a recurrent theme in sports studies (Gudmundsson, 2016). In past studies, speed has been analyzed indirectly via player interactions (i.e. displacement) and in individual analysis of players' performances.

At its most basic definition, speed is a function of distance in a given time frame, and in American football both factors are of fundamental importance for teams. In mapping the space and displacement among players, Taki and Hasegawa (1998) modeled the *dominant region* for players, taking into account the sphere of influence in the field mapped as a function of the time for each player to reach locations in the field (using a speed and direction vector). This approach takes into consideration the Voronoi area (*dominant region*) for each player, but uses a time-function as opposed to the Euclidean distance function. This allows for a more realistic understanding of a true dominant region, taking into consideration not only the space around a player but also the time to reach each specific location. Similarly, Fonseca et al. (2012) used Voronoi diagrams and distance to closest players to look at the dynamics within the same team in futsal interactions.

Speed in football has also been studied from an anatomic perspective. Mayhew et al. (1989)

studied the relationship between speed, agility and body composition to anaerobic power output. Similarly, Gains et. al. (2010) analyzed the difference in speed between field turf and natural grass among football players.

While they help to understand factors controlling for the change in speed to the individual level, the challenge is in comprehending the dynamics of group and team interaction in relation to player speed. Although individual performance can be isolated in low-contact sports such as baseball, golf, and tennis, it does not behave so unidirectionally among full-contact sports such as football. For that reason, and the nature of high-contact sports, the advancements in football analytics related to speed have been more limited than in other sports.

CONTEXT

When is speed important?

Speed is important during **runs**, for both the offensive team which is trying to get the ball as far down the field as possible, and for the defensive team trying to catch the runner.

Speed is also important for **kick and punt plays**, where the kick/ punt returner is trying to get the ball and the gunner who is trying to tackle the kick or punt returner.

For our analysis, we'll focus on speed for offensive team runs as we can more clearly establish the quality of the outcome in offensive plays (yards gained).

Most "runners" are Tight Ends (TE), Running Backs (RB), and Wide Receivers (WR). These are the players for whom speed is most relevant in the offensive context.

What is "efficiency"?

Efficiency can be defined as the ratio of input to output. In our case, the input is speed, and the output is the **increase in game advantage**. Game advantage can be simply defined as **yards gained**; or, accounting for the game context, can be **Expected Points Added (EPA)**. Keeping in mind, however, that quality of outcome does not always best represent the ability of a player, we will also look at the **distance maintained from defenders** and the **vertical distance gained** (by an attacking player). Note that vertical distance gained does not take into account the line of scrimmage and shows any advancement from the moment of reception.

The underlying assumption behind looking at the distance from defenders lies in the attacking objective of keeping the ball away from defenders for the longest period of time while moving the longest period of time in the horizontal axis (gaining yards).

Complications

There are factors which introduce a high amount of complexity into the analysis:

- **Play-specific factors** such as formation, ball location, play type (handoff vs. pass vs. kickoff vs. punt), and play routes
- **Game factors** such as playing at home vs. away, game surface, weather, distance traveled, field conditions, attendance.
- **Player fatigue**, measured in distance travelled, rest since last play, intensity of the last play, and the total amount of time played in the game

We have attempted to factor in these complications in various ways in our analysis in order to distill a "fair playing ground" for evaluating all players; as outlined in the "Methodology" section.

METHODOLOGY

Data Source

The data used in this analysis was provided by the NFL. It included game-level, play-level, player-level and play-frame level data with player coordinates for each frame.

Data Cleansing

For the present work we limited the plays to either **running** or **passing** plays. Kicks, punts, field goal and extra point plays have their own dynamics and fell out of the present scope.

We also limited the analysis to use **only frames from reception until first-contact** with the defense. As we are trying to understand speed efficiency, we need to be able to analyze plays in which the player has freedom to choose where to move, and how to move directionally. Before reception, we assume players would usually follow predetermined routes, and therefore the speed efficiency would be bound to route strategy. Additionally, we wanted to stop the play at the moment of the first contact, as it brings complexity given the impact of all contacts after the first.

Determining Player with Ball

Our work is concerned with the speed of players that are running with the ball, but this was not obvious from the tracking data. We thus derived the player with the ball by looking at the player that was closest to the ball (by euclidean distance, see Figure 1) for most of the frames from receiving the ball to first contact.

Visually validating plays

Aside from creating filters for specific events in the data, we also ran spot checks at random

plays to see if they represented what was expected, by animating the play in Python.

Determining Distance from Defenders

We determined the closest defenders from the player running with the ball by looking at the euclidean distance (see Equation 1) between the player and each defender at every single frame in the data. We kept the five closest defender in our analyses, even though the closest one seemed to represent much of the impact carried by the distance to other defenders.

Additionally, we mapped the Voronoi area for the player with ball in comparison to defensive players. However, in order to calculate the voronoi region, we cropped the boundaries to the player boundaries in the x-axis and y-axis, in a way that the area would be bounded to the football field area.

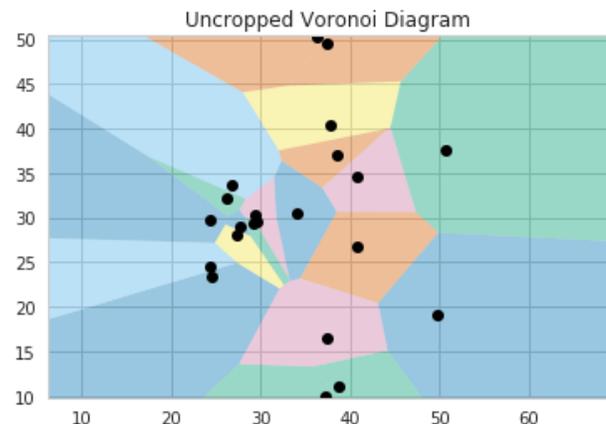
Equation 1: Euclidean distance between players i and j

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

Equation 2: Voronoi region for offensive player

$$R_k = \{x \in X \mid d(x, P_k) \leq d(x, P_j) \text{ for all } j \neq k\}$$

Figure 1: Voronoi diagram for player dominance



Clustering Players in the Field

One of the complications in analyzing speed and field progression is the distribution of

offensive and defensive players in the field for each play.

It is unfair, and perhaps inaccurate, to measure speed efficiency in different types of plays, mainly considering that speed depends on routes, potential gains and offensive support. Also, formation by itself was not enough to group plays, as we were interested in reception-to-first-contact frames, and by the time the ball is received, players could be in very different positions than when the ball was snapped.

Therefore we decided to cluster similar players distributions in the field at the moment of reception based on their x-y coordinates on the field (either when the pass arrived or when handoff happened) and on the vertical ball coordinate (yball).

Two similar plays would be those where, at the time of reception, offensive and defensive players would be in similar positions in the field and the ball caught in a similar position as well. We centralized the x-location at the time of reception (to x= 0) and shifted all plays in the x-y coordinates for offense to be always in ascending x-direction (the higher the x, the higher the position down the field). This way plays would have the same reference position. Thus, for two plays Pa and Pb, we have that the distance between two plays D_{PaPb} is the maximum¹ between the players distance $PDist_{PaPb}$ and the yball distance $YDist_{PaPb}$ (see Equation 2). $PDist_{PaPb}$ is the maximum distance between offensive players $PDist_{OffPaPb}$ and defensive players $PDist_{DefPaPb}$ (see Equation 3). All distances used were Euclidean, as observed in Equation 1, providing us with a spatial shortest unblocked distance between two players. The rationale for using the maximum of offensive and defensive distances was to make sure that both distances would fall below a threshold

¹ We also explored different ways to combine offensive , defensive and ball distances, such as sum of square root, mean, and squared distances.

when used for comparison. We also test different ways to cluster, such as using the sum of square root, the mean of distances of the sum of squared distances (penalizing dissimilar plays).

Equation 3: Distance between plays and ball location

$$D_{PaPb} = \max(PDist_{PaPb}, YDist_{PaPb})$$

$$YDist_{PaPb} = \sqrt{(yball^{Pa} - yball^{Pb})^2}$$

Equation 4: Distance between players

$$PDist_{PaPb} = \text{Max}[PDist_{Off^{PaPb}}, PDist_{Def^{PaPb}}]$$

$$PDist_{Off^{PaPb}} = \sum_{i=1}^{11} \min_i(Off_i^{PaPb})$$

$$PDist_{Def^{PaPb}} = \sum_{i=1}^{11} \min_i(Def_i^{PaPb})$$

$$Off_i^{PaPb} = [d_{i1} \quad d_{i2} \quad d_{i3} \quad \dots \quad d_{i11}]$$

$$Def_i^{PaPb} = [d_{i1} \quad d_{i2} \quad d_{i3} \quad \dots \quad d_{i11}]$$

For all plays in P, we have then a vector-form distance vector that can be used to cluster plays with similar distances in the two-dimensional field.

Equation 5: Pairwise distance among all plays

$$D = |D_{PaPb}|$$

$$Pa, Pb \in P$$

Using hierarchical clustering with single-linkage criteria, we are able to identify five clusters of plays that fall close together. The result can be observed in Figure 2 and the clusters formed are present in Figure 3. As we can see, Cluster 1 represents passes with the ball on the left-side, Cluster 2, long passes (small number of plays), Cluster 3, mid-field passes in the center, Cluster 4 rushing and Cluster 5 passes with the ball thrown on the right-side.

Equation 6: single-linkage clustering

$$\min \{ d(a, b) : a \in A, b \in B \}.$$

Figure 2: Dendrogram of play distances

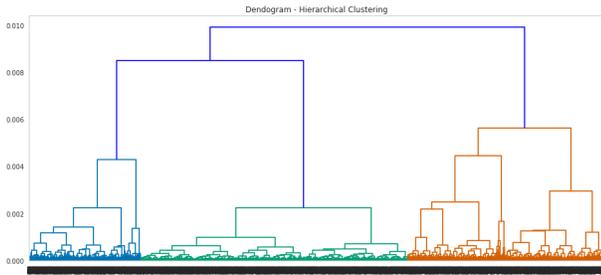


Figure 3: Cluster result sample plays (n=30), ball in black.

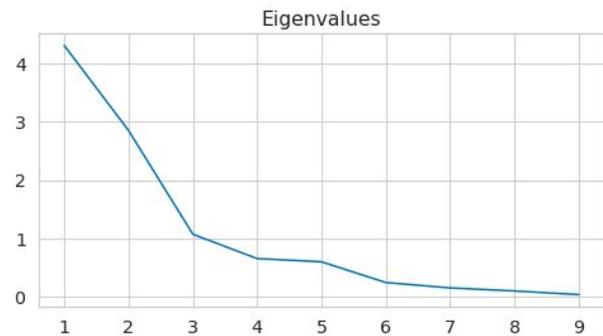


Calculating Player Fatigue

We wanted to explore the effect of player fatigue on player performance. There are many factors which may play into player fatigue, such as the intensity of the last play (average speed, maximum acceleration, total duration, total distance), rest since the last play, and cumulative time / distance run on the field. Since the number of variables were numerous and the relationship among them were not linearly defined, we opted to reduce

dimensionality of the data using principal component analysis. This analysis yielded two principal components which absorbed most of the variance present in the other variables, as we can observe in Figure 4 of the descending eigenvalues. While the first component was correlated to intensity of the previous play, the second was related to cumulative workload during the game.

Figure 4: Eigenvalues for fatigue PCA



RESULTS

I. Most Effective Players

As aforementioned, we looked at efficiency as the ability of maximizing game advantage with the best usage of speed. In other words, efficiency is a function of the vertical gain in the field between reception to first contact divided by the speed needed to achieve such gain.

$$Efficiency = \frac{Vertical\ Gain}{Mean\ Speed\ during\ Play}$$

Controlling for similar plays at reception, and types of play, we are able to aggregate the median player efficiency within each cluster and retrieve the total weighted average for each player. This means that we looked at the efficiency for different types of plays and judged them separately before combining all for an overview of a player performance.

Due to the high number of plays with Running Backs (RB), most of the results with a

minimum significant sample (20 for runs and 15 for passes) had mainly RB and some Tight-Ends or Wide-Receivers.

The top three players who scored with the highest efficiency in handoff runs were Alvin Kamara, Todd Gurley, and Le’Veon Bell - players who caught the highlights for good performances in the 2017-2018 season (see Figure 5). In situations where there was a pass involved, Melvin Gordon, Ty Montgomery and Christopher Thompson were the leaders, respectively (see Figure 6).

What is also important to note is that the average / max / median speed had no direct relationship with yardage gains, or with the overall performance of players measured by displacement in the field. There seem to exist a positive relationship between long-yardage plays and speed, but possible because of the time allowed for acceleration, which is also reflected in greater distance from defenders and in dominance region.

Figure 5: Most efficient players (Handoff)

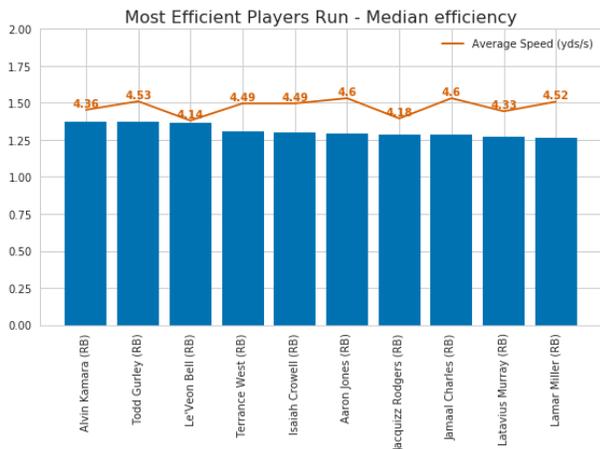
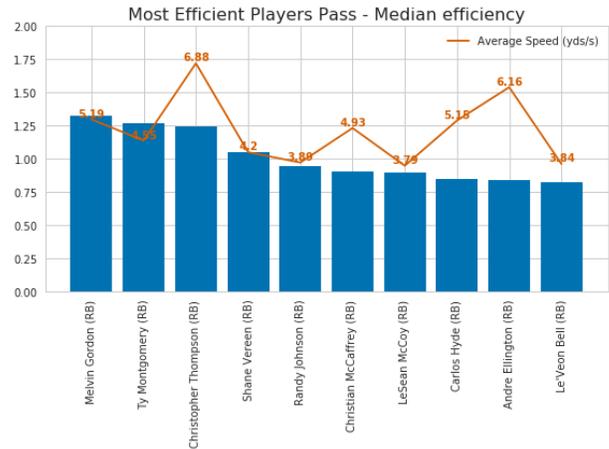


Figure 6: Most efficient players (Passes)



II. Factors that Impact Speed

1) Play Type

We hypothesized that the use of speed would differ depending on the type of play, which seems to be true for pass plays vs. rush (i.e. handoff) plays. The distributions of average speed are very different (tested with 2-sample Kolmogorov-Smirnov statistic) for pass and rush plays (standard deviation of 2.02 and 1.39 respectively). This is intuitive considering the similarity of most running plays versus the variety of distribution of pass tactics.

It also appears that players reach much higher maximum acceleration during handoff plays, which also makes intuitive sense given that they are much closer to defenders when they receive the ball and must use acceleration to get away from the defenders.

Figure 7: Distribution of average speed and max acceleration for pass plays vs. handoff plays

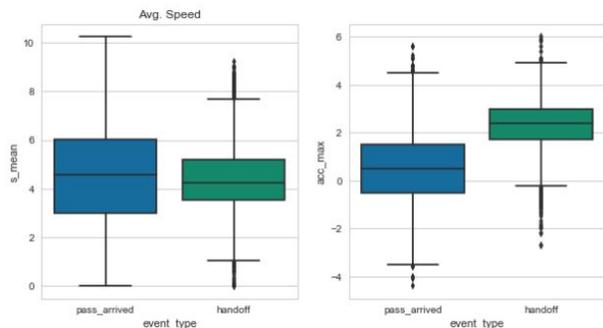
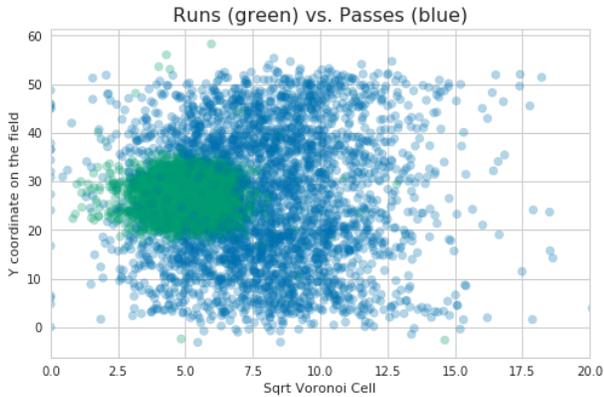


Figure 8: Square-root of voronoi cell for player with the ball, for pass plays(blue) and handoff plays(green)



2) Game Clock Time

How do players maintain speed and acceleration throughout a game? Plotting the average speed and maximum acceleration of each play against the time that has lapsed, it seems that players maintain the same level of speed throughout a game, for both pass and handoff plays. We were not able to reject the hypothesis that average speed was the same throughout the game duration.

Figure 9: Game time lapsed vs. average speed by play type

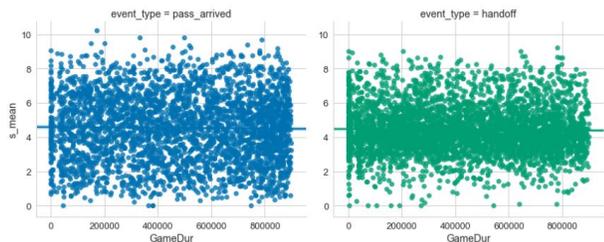
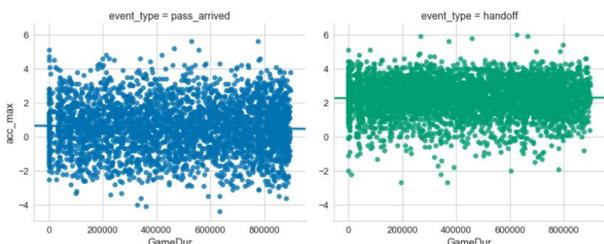


Figure 10: Game time lapsed vs. max acceleration by play type



3) Cumulative Game Time

Do players get tired the more that they're on the field? For each match, we added up the

total time that the player has been involved in plays so far (regardless of type of play), and plotted this cumulative game time against the average speed and maximum acceleration for the play that the player was involved in. It appears that players do get tired, but the effect is more evident on speed than on acceleration.

Figure 11: Cumulative player time on field (on a game basis) vs. average speed by play type

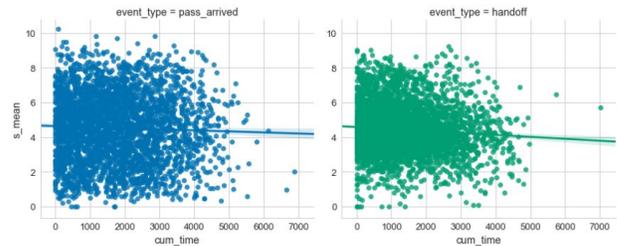
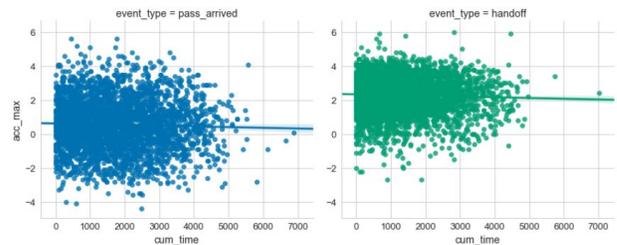


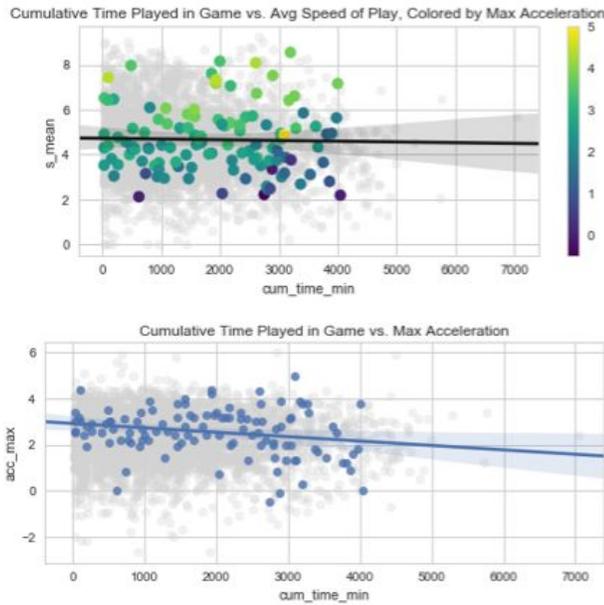
Figure 12: Cumulative player time on field (on a game basis) vs. max acceleration by play type



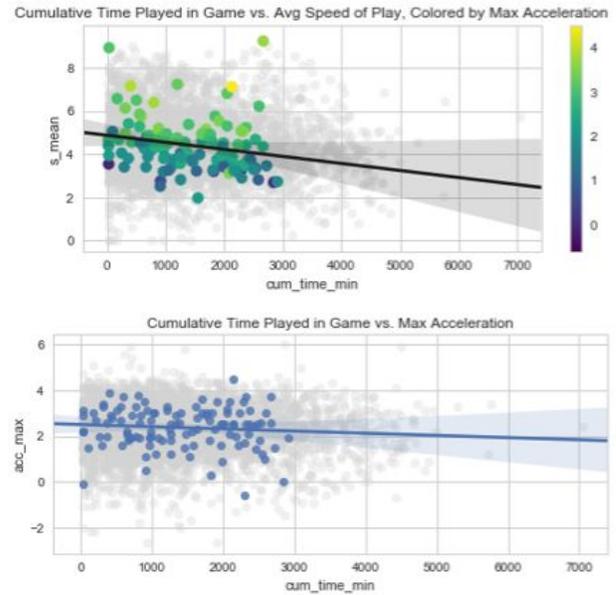
Although there seems to be a downward trend in speed relative to cumulative time that the player has played in a game, the trend differs on a player level, possibly because some players need time to warm-up to their top speed, while others start with their top speed and become more worn-out as they stay on the field for longer.

Following are some diagrams that dive into player-specific patterns of speed and acceleration based on the amount of time that the player has been on the field.

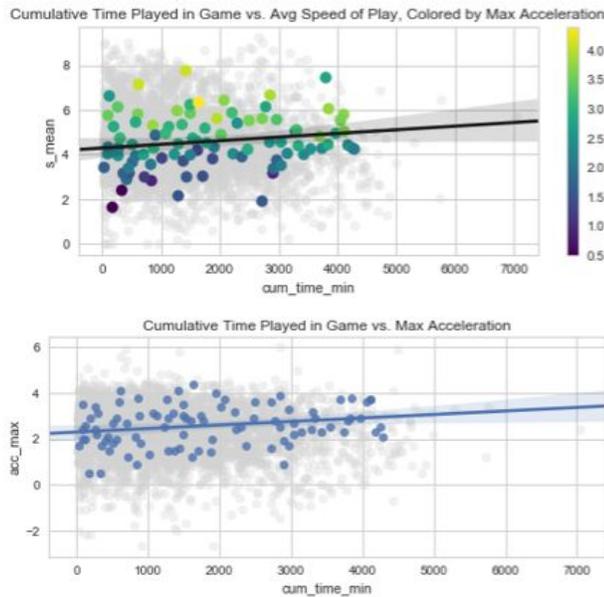
Todd Gurley (RB for LA); 113 plays for 6 games
 Todd Gurley seems to maintain speed well through fatigue, but not acceleration.



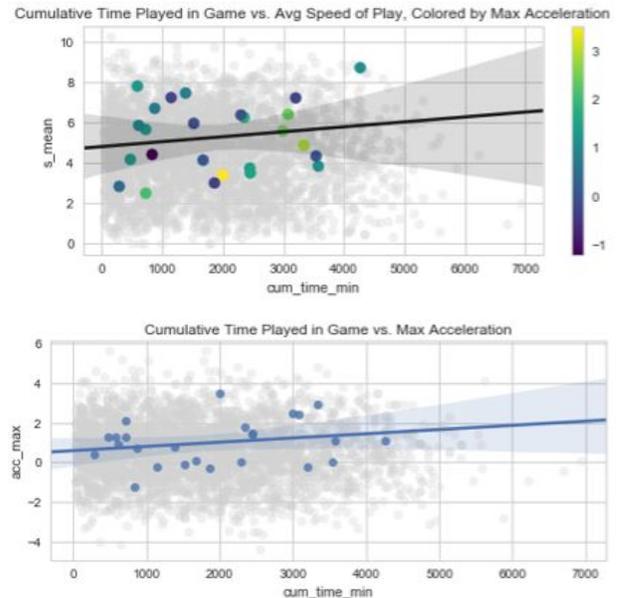
Leonard Fournette (RB for JAX); 113 handoff plays for 6 games
 Leonard Fournette's average speed for plays seems to follow the generally observed trend for handoff plays, where the average speed declines as his playing time accumulates.



Ezekiel Elliot (RB for DAL); 91 handoff plays for 5 games
 Ezekiel Elliot seems to get faster and accelerate more the longer that he's on the field.



Melvin Gordon (RB for LAC); 25 pass plays for 5 games
 Melvin Gordon seems to follow the same pattern as Ezekiel Elliot, where both average speed and max acceleration increase as he spends more time on the field.



Modelling Speed

As our final step for understanding the factors that impact speed, we ran an ordinary least squares linear model, using backward elimination to arrive at significant variables.

Note that the independent variables were scaled by removing the median of the dataset and scaling the data according to the quantile range, in order to manage outliers.

The model showed small, but significant relationships between several variables and speed of the play:

1. **rest**: The time lapsed since the last play in the game
2. **cum_time**: cumulative time that the player has been on the field, for the game that they're playing
3. **sum_def_dist_first**: sum of the player's distance to the closest defenders
4. **d0_first**: the player's distance to the closest defender
5. **event_handoff**: whether it's a handoff event
6. **cluster_1**, **cluster_2**, **cluster_3**: the formation when the player received the ball, determined as explained in the "methodology" section

Figure 13: Ordinary Least Squares Regression results: Adj. R-squared of **0.052**

	coef	std err	t	P> t
const	4.5095	0.031	144.969	0.000
rest	0.0513	0.023	2.193	0.028
cum_time_mins	-0.1678	0.033	-5.100	0.000
sum_def_dist_first	0.3722	0.037	10.083	0.000
d0_first	0.1276	0.024	5.323	0.000
event_handoff	-1.1278	0.172	-6.544	0.000
cluster_1.0	-1.6026	0.170	-9.421	0.000
cluster_3.0	-1.2326	0.171	-7.193	0.000
cluster_5.0	-1.5516	0.170	-9.153	0.000

LIMITATIONS AND FURTHER RESEARCH

There are a few limitations to the present analysis that, due to the scope and time constraints of the current project, were not incorporated into the findings. There are a few observations of exploratory character that were not statistically validated, such as differences in speed decay throughout the game on a player level and impact of acceleration on yardage gains.

Firstly, we would recommend a further study into the distribution of speed for each play in order to enhance the understanding of speed dynamics within plays. There is plenty of changes in speed distribution within plays and those changes might be more indicative of speed efficiency than the aggregate value of the distribution (hereby the mean/median).

Secondly, the more specific we dive into plays, the smaller the number of observations for the analyses. This is a particular challenge to football that has been observed in the past, given the small number of plays per season, and it remains a constraint for a better understanding of the games through a statistical lens. It is fundamental to balance specificity with sample size to be able to drive generalizable results.

A few other considerations concerning future analysis and the limitations of the present findings are below:

1. Defense strength was not taken into consideration when analyzing plays. It would be interesting to see how specific players on the field impact the outcome of a play and game;
2. There are other approaches to clustering plays that were explored for this work. One of them comprised of splitting the field into bins and matching the similarity of players in each bin among all plays. Although this method could help identify players density

within play, it failed to match similar plays to an interpretable level;

3. There seems to exist non-linear relationships among multiple variables in question. More advanced techniques can be used to estimate the impact of players speed in a given scenario, but they could be black-boxes in explaining what entails in speed efficiency;
4. A time-function voronoi mapping of dominant region could be used to understand how players maximize the distance from defensemen, given direction and velocity vectors.

CONCLUSION

The understanding of speed in the game of football is far from clear and deciphered from an analytical standpoint. The challenges in understand the intricacies of the players usage of speed are mostly related to the dynamics of group interactions and to the rules governing football.

We tried to shed light on one definition of efficiency when it comes to the usage of speed, by looking at the usage of speed for vertical gains, from reception to first contact, while the offensive player with the ball has most freedom and the largest responsibility towards its team.

We also looked what impacts the average speed across plays and throughout games. While generally we do not observe a clear linear relationship between our fatigue variable (derived from principal component analysis) and speed performance, the cumulative time that the player has spent on the field during the game does seem to have a statistically significant negative effect on the average speed of the play from reception to first contact. This effect, however, seems to vary from player to player. A deeper dive into the data may be warranted to understand

player-level performance, and gain a better understanding of speed.

REFERENCES

Taki, T., & Hasegawa, J. I. (1998, December). Dominant region: a basic feature for group motion analysis and its application to teamwork evaluation in soccer games. In *Videometrics VI* (Vol. 3641, pp. 48-58). International Society for Optics and Photonics.

Fonseca, S., Milho, J., Travassos, B., & Araújo, D. (2012). Spatial dynamics of team sports exposed by Voronoi diagrams. *Human movement science*, 31(6), 1652-1659.

Mayhew, J. L., Piper, F. C., Schwegler, T. M., & Ball, T. E. (1989). Contributions of speed, agility and body composition to anaerobic power measurement in college football players. *The Journal of Strength & Conditioning Research*, 3(4), 101-106.

Gains, G. L., Swedenhjelm, A. N., Mayhew, J. L., Bird, H. M., & Houser, J. J. (2010). Comparison of speed and agility performance of college football players on field turf and natural grass. *The Journal of Strength & Conditioning Research*, 24(10), 2613-2617.