Breaking Through the Line: Evaluating Running Back Contributions to Running Plays

December 2019

1 Introduction

In football, Running backs make plays that are great for highlight reels. Typically of average height, Running Backs can routinely be seen running into defensive brick walls and occasionally emerging on the other side still standing. From Marcus Allen's 74 yard run in Super Bowl 18 to Marshawn Lynch's 'Beastquake', a highlight reel seems incomplete without an iconic run. But what made each of these plays highlights worthy? One may suggest that these are Marcus Allen and Marshawn Lynch highlights specifically because they highlight the individual contribution of each running back. In Allen's iconic run, he meets multiple unblocked defenders in the backfield and is forced to make something out of the play himself. The Beastquake similarly deals with Marshawn Lynch's individual effort, shedding multiple Saints' defenders on his way to a historic playoff touchdown.

These two examples are simple to work with because the plays demonstrate immense individual effort. There is value to be assigned to the offensive line making blocks, for the quarterback possibly looking defenders off with his eyes, and many other things though they all seem less important given what each of these runners did.

It is not always the case, however, that it is trivial to separate the efforts of the running back from those of their teammates. This difficulty has become exasperated as communities go so far as to say that 'Running backs don't matter' without adequate ability to parse contributions. These are paraphrased of course but represent sentiments that exist within the football analytics community.

In this report we seek to evaluate the question of how to quantify the contribution of a running back or ball carrier on a given play. We make use of the tracking data provided in the 2018 NFL Big Data Bowl which took measurements of each player on the field at 10 measurements per second for the first 6 weeks of the 2017 NFL season. Considering only the running plays from this data set, we build out metrics that specifically address the importance of running backs. Making use of the established Expected Points (EP) model (Yurko et al., 2019) along with a Field Ownership model (Fernandez and Bornn, 2018), we propose a novel metric to measure the value added by running backs on a given play. We extend this through running back comparisons and provide our metric values for frequent rushers through the first 6 weeks of the 2017 season. Additionally, we provide an app to improve the usage of the Field Ownership model which was adapted to football Chu et al. (2019). The interactive Shiny app[1] based on the field ownership model allows us to hypothetically adjust players' positions and observe how the field ownership of the offense and defence would change at that moment. We also go on to extend our work further through the incorporation of blocking and running back's vision in our field ownership approach, again summarizing with a comparison of frequent rushers with respect to these new considerations.

[1] https://bigdatabowl.shinyapps.io/demo/

The end result of this work is a method that makes sense to a football fan as much as it does to an analyst: how much value did a running back add above and beyond what his offence provided for him. We conclude with shortcomings of our approach and future considerations to make use of as tracking data becomes more readily available.

2 Building Blocks

Understanding running back contribution is a multi-pronged approach. There are numerous factors to consider in building an evaluation tool. The first is what information is currently available to use. Firstly, Expected Points and Expected Points Added models (Yurko et al., 2019) provide a tool to evaluate the results of a play. Secondly, field ownership (Fernandez and Bornn, 2018) enables us to contextualize the movements of players within a play.

Additional considerations are made for blocking and the perspective of the ball carrier and are addressed separately in later sections of this paper.

2.1 Expected Points Model

One of the biggest advances in NFL analytics has been made possible through the work of Yurko et al. (2019) and the nflscrapR package (Horowitz et al., 2018). Their efforts resulted in the commonly used value metric known as Expected Points (EP) and Expected Points Added (EPA). The general acceptance of this method comes from its ability to account for numerous in game factors including time remaining, down, distance to first down, distance to endzone, score differential and more.

The result of Expected Points is a value that quantifies the worth of a given set of game factors. Naturally it allows for differing situations to be compared with respect to this value. For example, a first down with only one yard to go for a touchdown is clearly more valuable than a fourth down with 15 yards to go for a first down from your own 20.

We will be integrating the EPA model with ours as a way of establishing the expected resultant value of a play. As this is a result-based approach, it summarizes how much was gained or lost through the execution of the given play call. We will look to compare this value against our own calculated version of Expected Points done through field ownership.

2.2 Field Ownership Model

The work of Fernandez and Bornn (2018) have helped to conceptualize the idea of pitch control in soccer. Through their work, we were successfully able to translate the ideas from soccer into a football context. Field ownership then is expressed as the amount of total influence being exerted by the offence or defense, with a given portion of the field being owned by the team that exerts more influence at that point. The field ownership model allows us to quantify the space controlled by offence and defense.

For the technically interested, field ownership is measured as a difference of sums at each point on the field. For each player we define a bivariate gaussian distribution centred at a location on the field dictated by the player's current position, velocity, and direction. A team's influence at that point is then the sum of all player's influence from the same team at that location. Ownership of this point is then ultimately awarded to the team with the largest sum, assigning the magnitude of ownership through a sigmoid transformation of the difference of team influence. Further details can be found in (Fernandez and Bornn, 2018).

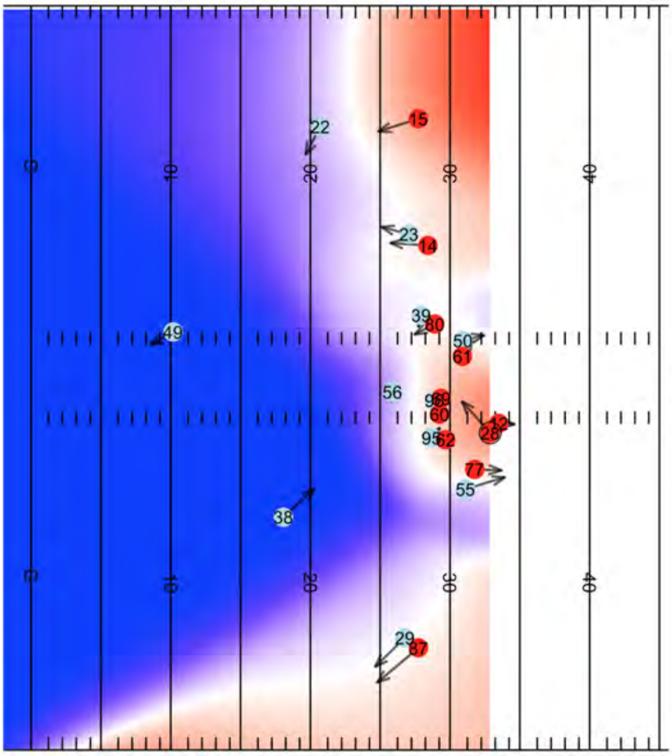

Figure 1: Sample Evaluation of Field Ownership. The owned spaces of offence and defence teams are plotted in red and blue respectively. The arrows indicate players' moving directions and the length of arrow indicates how fast they are moving.

Using field ownership, we look to explore a new measure of Expected Points. Here we are looking to capture the amount of space that the offence provides its ball carrier. By only considering the spots in our influence plot that are controlled by the offence (red spaces in Figure 1), we can calculate the proportion of offensively owned cells that exist at each yard marker. For convenience, we utilize this proportion as a probability of running to that yard line. Note that this doesn't capture the assumed behaviour of the ball carrier aiming down the field with the objective of maximizing yards gained. This isn't always the objective of a run play and as such we did not aim to incorporate it extensively.

With these proportions in mind, we can now investigate the new situation we would be in at the given yardline. Take for example a 3rd & 2 at the 40 yard line. Suppose our range of outcomes exist on $[-2, 5]$ yards gained. Any outcome on $[-2, 1]$ puts us in 4th down while any outcome in $[2, 5]$ pushes us into a new set of downs. We can then calculate our Expected Points by using the Expected Points associated with the new state multiplied by the probability of arriving in that state, denoted as $\sum_{i=-2}^{5} p_i * (EP)_i$. Since this is calculated with respect to the area controlled by the offence at exactly the time the ball is handed off it can reasonably be interpreted as the locations a ball carrier should be able to get to with limited difficulty. Essentially, we are trying to weight the expected points by the field ownership at each yardline. We will then label this value as Expected Points Ownership as it denotes a version of expected points related specifically to field control.

The idea of field control is our first big step in quantifying ball carrier contributions as it represents a measure that is otherwise unattainable through play by play data.

3 A Metric for Ball Carrier Contribution

Our proposal for a ball carrier contribution metric follows naturally from the aforementioned building blocks. Each play in our data set has an Expected Points Added value from the modeling in nflscrapR. Each play also has a value of Expected Points Added by Ownership value from the modeling of offensive field ownership. This means that for each play we are able to calculate the difference between the observed (nflscrapR EPA) and the baseline (ownership EPA). The difference in these two values then measures roughly the contribution above the baseline observed on a given play. We will call this the **Earned EPA** of a play and use it closely in tandem with the value of the ball carrier.

Note that we mention the value roughly measures the contribution of the ball carrier. We make this note due to recognizing that the result of a play includes many other factors beyond just the space available at the time of handoff and the capabilities of the ball carrier. On many plays we see injuries, mistakes, and outstanding individual efforts that play into the ultimate evaluation of this metric. As more running plays are added for each ball carrier we expect the value calculated to further stabilize.

3.1 Ball Carrier Value through Weeks 1-6 of 2017

We implemented our work on the data set we have been using throughout to investigate ball carrier contributions through the first 6 weeks of the 2017 season. We first looked to answer which ball carrier generated the most Earned EPA per touch among ball carriers with at least 20 touches.

Rank	Name	Average EPA Earned	Standard Dev. EPA Earned
1	James White	1.68	1.34
2	Alvin Kamara	0.98	2.06
3	Duke Johnson	0.97	1.88
4	Jamaal Charles	0.87	1.70
5	Jalen Richard	0.83	1.19
...			
61	Chris Carson	-0.26	1.49
62	Ameer Abdullah	-0.29	1.47
63	Orleans Darkwa	-0.38	2.12
64	Paul Perkins	-0.39	1.37
65	Isaiah Crowell	-0.44	1.34

Some interesting other players to note that just missed the table above are Dion Lewis (7), Buck Allen (8), Ezekiel Elliot (11), Adrian Peterson (12), Melvin Gordon (58), Joe Mixon (59), and Alex Collins (60).

The table results tend to be digest-able. James White and Alvin Kamara, along with the others in the top 5, had exceptionally efficient early weeks to the 2017 season. On the other hand, the bottom five is populated with runners that were often inefficient, turnover prone, or both. For better or worse, EPA as a metric severely punishes turning the ball over and we see that reflected here in our results.

From a statistical point of view, it is clear that although the values are not equivalent we do not have any statistical significance in our pairwise comparisons. The variation in the EPA earned on a play is currently too large to say with any statistical certainty that ball carrier A earns more EPA per touch than ball carrier B.

A point of interest to be considered from this table is the style of ball carrier at the top. Each of the running backs in the top 5 have been commonly viewed as pass catching or secondary backs in this early 2017 window. The Patriots had Dion Lewis ahead of James White, the Saints had Mark Ingram ahead of Alvin Kamara, and similar stories with the rest. This may suggest a slight bias of our approach towards these secondary backs. Possible reasons for this slight bias could include the style of handoff delivery as a pitch to the ball carrier frequently incurs a later handoff time than a standard handoff and therefore changes the underlying ownership plot to a later point in the play. It may also be from the threat of these running backs in the receiving game, opening up spaces on the field that are not yet fully accounted for by our field ownership. Even with these possible minor biases, perhaps these ball carriers are better performers. We see both Duke Johnson (3) and Isaiah Crowell (65) on opposite ends of our list despite both playing for the Cleveland Browns in 2017. This may be the closest we get to a controlled environment comparison between two running backs.

Following up on the ideas from above, we wanted to explore our new metric with respect to the EPA from nflscrapR. Figure 2 shows the results of this, considering only the lead running backs from each team determined by number of carries.

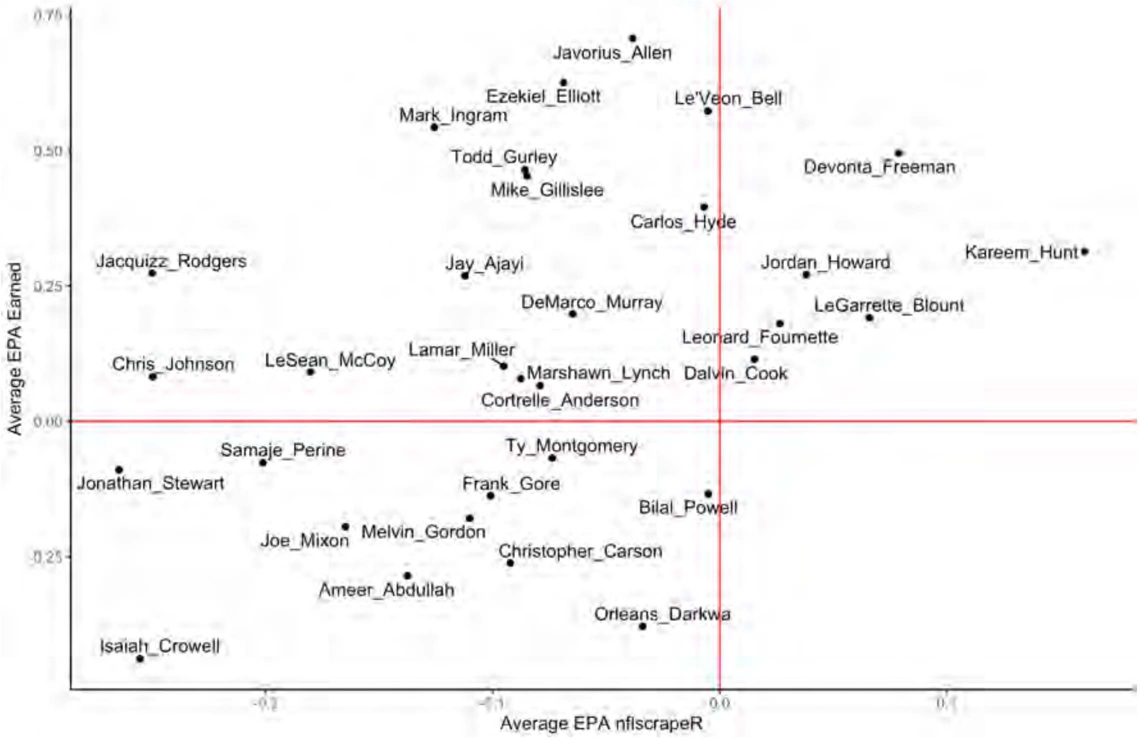

Figure 2: Comparison of lead running backs per team between EPA Earned and EPA nflscrapR.

The quadrants that exist on this plot are separated by positive or negative EPA plays with respect to each metric. Players in the upper right quadrant generate on average both a positive Earned EPA and a positive nflscrapR EPA. Here we see common names like Devonta Freeman, Kareem Hunt, and Dalvin Cook. Players in the lower left have the opposite, generating on average negative values of both metrics. Here we find Isaiah Crowell, Samaje Perine, and Ameer Abdullah among others.

The upper left quadrant is then the zone of interest as it is the area in which our metric disagrees slightly with EPA. This quadrant represents the ball carriers that on average generated a positive Earned EPA while scoring on average a negative observed EPA. Housed here are players like Le'Veon Bell, Ezekiel Elliot, and Mark Ingram. Players in this quadrant were able to add more expected value to their plays despite being put in generally unfavourable positions, whether by play call or by design.

For full disclosure, a known feature of EPA is that a running play on average generates a negative value. In 2017 that average value was -0.0973. Adjusting the coordinate grid in Figure 2 to this average, we get the following in Figure 3:

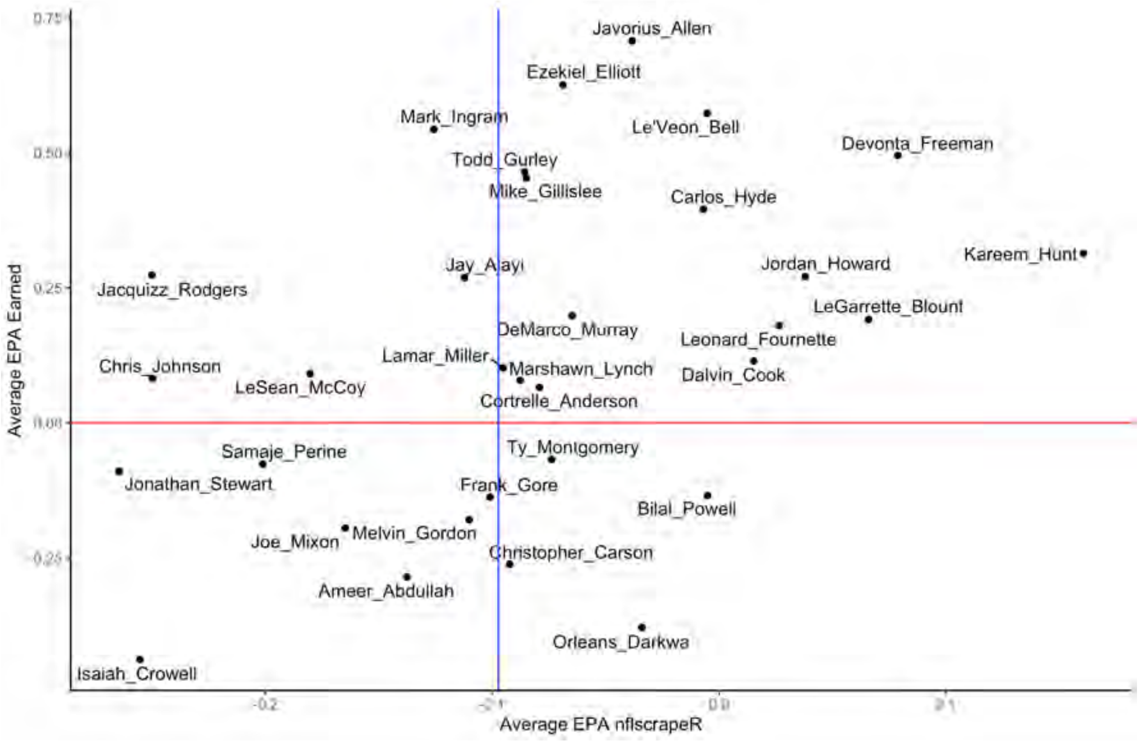

Figure 3: Comparison of lead running backs per team between EPA Earned and EPA nflscrapR.

The shift moves a few players into the above average nflscrapR EPA category but does not overly impact the other observations.

4 Further Extensions

Ownership in the context that we have applied it has been derived from other sports with the primary source being soccer. Unfortunately for the modeling process, these sports are largely different and each have unique challenges. Two major challenges that don't exist to the same extent in soccer as in football are i) blocking and ii) visibility of players. We will discuss our work on these two areas as we continue to expand our earlier metric.

4.1 Blocking

Using the soccer methodology for ownership runs into a small challenge with respect to blocking. In soccer blocking represents interference and is penalized. In football this act is encouraged (mostly). When we implement our formulas exactly as they are in soccer, we lose some of the football context in our analysis.

In an attempt to calculate blocking, we define a block with respect to the velocity of a player and their proximity to the nearest player on the other team. Since a block is designed to slow down and impede an opposing player, we define a block to be a player moving slower than a cutoff value and within a cutoff distance from the nearest opponent. For the purposes of our analysis, we currently treat a defender as being blocked if they are moving at less than 3 yards per second and are within

1.5 yards of the nearest offensive player.

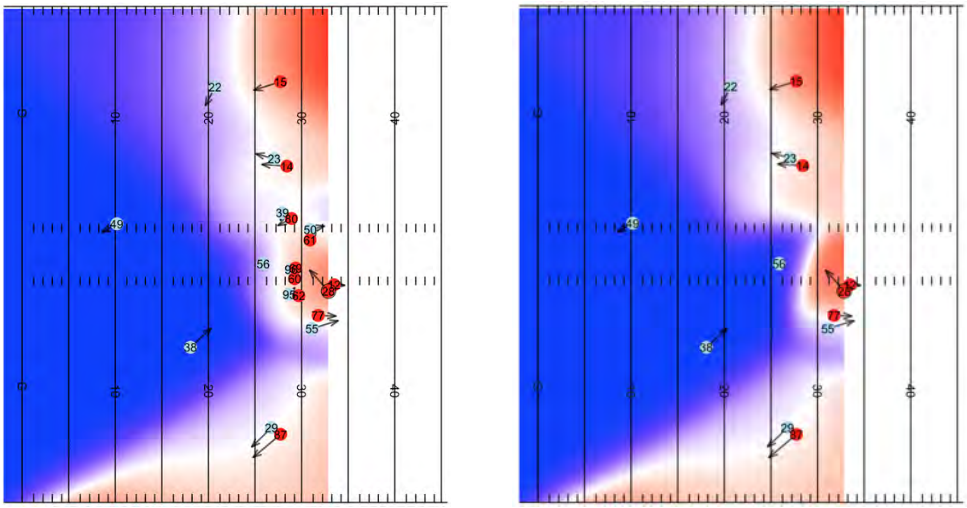

Figure 4: The figure demonstrates 2 images comparing the impact of blocking on field ownership of a run play. Offense is in red and defence is in blue. Field ownership without adjusting for blocking (left) vs. with adjusting for blocking (right). Player 28 has a higher density of ownership (darker red) in the direction he's moving towards after accounting for blocking from his teammates.

For our blocked players, we then modify the value of ownership as shown in Figure 4. A well executed block should essentially remove a player from the path of the ball carrier. We incorporate this idea by removing the players involved in blocking (offense and defense) from the ownership calculation. We suggest that this removal improves upon the basic soccer approach as players can not actually get to the places that their directions and velocities would otherwise suggest and therefore is more appropriate in the football context. Again, this idea of blocking can be easily explored using our Shiny app[2].

4.2 Vision of running backs

A standard plot of field ownership contains all of the space on the field that is owned by each team and that has a non-negligible total influence exerted upon it. This plot is viewed from a top down perspective. In terms of a passing play, this may make reasonable sense as players can move to the open spaces relatively unimpeded and have relatively clear lines of sight. For running, however, another consideration needs to be made as there are many obstacles obscuring vision.

We will approach this problem from the perspective of the ball carrier as demonstrated in Figure 5. Our attempt will aim to properly map what the carrier can see and calculate their expected points accordingly. Calculating this view is non-trivial and can be computationally intense. We will outline the algorithm briefly that we used to calculate ball carrier vision. Do note that due to time constraints the algorithm may be more inefficient than possible. We will resolve this in the coming weeks.

[2] https://bigdatabowl.shinyapps.io/demo/

Figure 5: The left image is an example of quarterback's vision cone from Robinson (2012), where a QB is trying to scan open receivers on the field. Imagine a similar vision plot can be generated for running backs. In the right image, players are represented as squares and the green lines demonstrate all possible paths of RB's vision in a 2D plot using a ray casting algorithm.

Ray casting is commonly used in computer graphics to detect intersection of objects by a light or ray (Wikipedia, 2019). Though complicated, the below algorithm allows us to identify each coordinate that is visible to the ball carrier at the given frame. Presently we assume the given frame is the frame at which the handoff occurs though this methodology and infrastructure easily generalizes to an arbitrary frame.

Result: Visible points from the view point of the ball carrier
RB $\leftarrow$ Coordinates of ball handler;
Ray Origin $\leftarrow$ RB;
Ray $\text{Endpoints}_{\text{x}} \leftarrow \text{RB} \pm 10$ yards;
Ray $\text{Endpoints}_{\text{y}} \leftarrow [\text{sideline}_1, \text{sideline}_2]$;
for *Ray Endpoints* **do**
    $\text{Ray}_{\text{i}} \leftarrow$ (Ray Origin $\rightarrow$ Ray $\text{Endpoint}_{\text{i}}$);
end
for *Coordinates in Influence Plot* **do**
    for *Ray* **do**
        if *Ray Intersects Coordinate* **then**
            Remove Coordinate;
            Go next;
        end
    end
end
Return Coordinates;

Figure 6 demonstrates what a play, where players are drawn as squares, would roughly look like through the perspective of the ball carrier. This run specifically is by James White and the New England Patriots and is a run that has been routinely demonstrated throughout this work. To clarify the perspective, coordinates visible in the influence plot are overlaid in Figure 6.

Vision is then used to modify the Expected Points Ownership calculation referred to previously. The Expected Points associated with each yard marker remains the same but the proportion of points owned by the offence varies. Changing the proportions changes the Expected Points

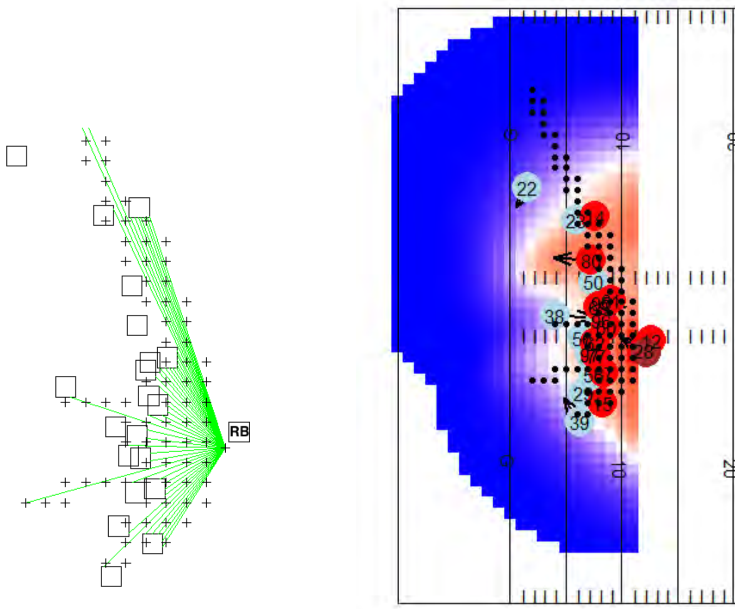

Figure 6: The image on the left is an example of RB vision cone and the plus signs indicated the possible locations through his vision on the field. The right image represents an overlay of the running back's vision and our field ownership plot. The calculation of field ownership can be weighted based on RB's vision at a given frame.

Ownership to reflect only what the running back can see and represents a different take on a similar calculation. Again the infrastructure exists for this though we have yet to do a full scale analysis due to time constraints. This, tied together with blocking, seem to be promising avenues of exploration to consider!

5 Concluding Remarks

Quantifying individual contributions in football is hard. The availability of player tracking data makes it possible. Our method utilizes tracking data, leveraging it to capture spatial-temporal effects that are otherwise invisible to play by play data. We use a single frame of measurements and from it are able to calculate values for field ownership, an expected points added associated with ownership, and an expected points added for the ball carrier. From these values we can then compare the contributions of ball carriers across the first 6 weeks of the 2017 season (and eventually all of 2017-2018 with the new data from this contest) with respect to ball carrier contribution and average play results.

Use of this new metric hinges on the reliability of our underlying ownership metric. With reference to the existing methodology our approach is built from, we claim that our approach is reasonable in the given context of football. We do also recognize possible improvements in ownership for football through blocking and ball carrier vision as these don't exist to the same extent in other sports. We go further by implementing these methods and performing cursory analyses for general proof of concept level intuition.

Understanding the contributions of ball carriers to running plays will not be entirely solved by this work. Rather, this work aims to build a solid foundation to further expand upon. Many infrastructures and summary results have been provided to allow for both justification and improvement.

We hope that these results demonstrate the incredible usefulness of tracking data in football and represents a novel approach to measuring ball carrier contributions.

References

Chu, D., Reyers, M., Wu, L., and Thomson, J. (2019). Routes to success. https://operations.nfl.com/media/3670/big-data-bowl-sfu.pdf. Accessed: 2019-12-22.

Fernandez, J. and Bornn, L. (2018). Wide open spaces: A statistical technique for measuring space creation in professional soccer. In *Proceedings of the 2018 MIT Sloan Sports Analytics Conference*.

Horowitz, M., Yurko, R., and Ventura, S. (2018). *nflscrapR: Compiling the NFL Play-by-Play API for easy use in R*. R package version 1.8.1.

Robinson, J. (2012). Qb vision cones. https://ca.ign.com/articles/2005/07/19/qb-vision-cones. Accessed: 2019-12-22.

Wikipedia (2019). Ray casting. https://en.wikipedia.org/wiki/Ray$_{c}asting.Accessed: 2019-12-$22.

Yurko, R., Ventura, S., and Horowitz, M. (2019). nflwar: a reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*, 15(3):163–183.