ALGORITHMS IN THE WAR ROOM

A Preprint

Alex Stern
Graduate School of Data Science
University of Virginia
`acs4wq@virginia.edu`

January 2, 2021

1 Introduction

Football is an extremely difficult sport to quantify. The inter-dependency on so many other players' decision making, execution, and physical abilities creates a uniquely nuanced challenge. A challenge met initially with stat lines, but one that most often requires trained eyes watching film to generate situational context. Player tracking data has now been collected every tenth of a second of every play for the last couple of years. Its level of granularity allows it to compliment traditional film analysis in previously inconceivable ways.

## 2 Quantifying the Previously Unquantifiable

Defensive Backs are some of the most important players on the field, especially as the NFL trends in a pass-heavier direction each year. Player tracking data presents the opportunity to design entirely new, holistic methods for evaluating the skill-set of a Defensive Back, enriched by context. The algorithms described below do just that, allowing scouts to ask sharper questions about how a DB performs relative to his peers, given assignments to receivers of varying skill levels.

In close coverage, player actions/results depend heavily on the DB-receiver matchup, presenting the perfect opportunity to quantify individual talent, relative to the other player and situation. The goal of every front office is to put players with the greatest chance of succeeding on the field, while also maximizing value given the league's salary cap. The key to designing scouting algorithms is properly accounting for a commonly overlooked, yet known variable in the equation: who the opposing receiver is. As large of a sample size we have on the average DB, we also have on the average receiver. The first four algorithms described below rely on comparing the results of a DB's actions to those expected when covering a certain receiver. This keeps us from docking good DBs who allow any level of production from the league's finest. Vice versa, the algorithm appropriately lowers the scores of DBs who allow generally unproductive receivers to be productive. Raw completion percentage allowed tells you very little when it doesn't take into account where everyone else on the field was and whether the receivers defended were DeAndre Hopkins and Davante Adams or some UDFAs recently pulled up off the practice squad.

This project relies on distributions rather than rankings. Flagged with the percentile a player's score lies in, these distributions introduce a visual aspect to player comparison which helps illustrate relative differences in abilities and the overall spread of the metric. While the discrepancy between players *ranked* 20th and 35th can vary by metric, percentile scores will **always** carry the same meaning. Additionally, each distribution will require DBs meet some minimum volume requirement for inclusion in order to eliminate outlying small samples.

3 Defensive Back Evaluation

3.1 Sample Player Cards
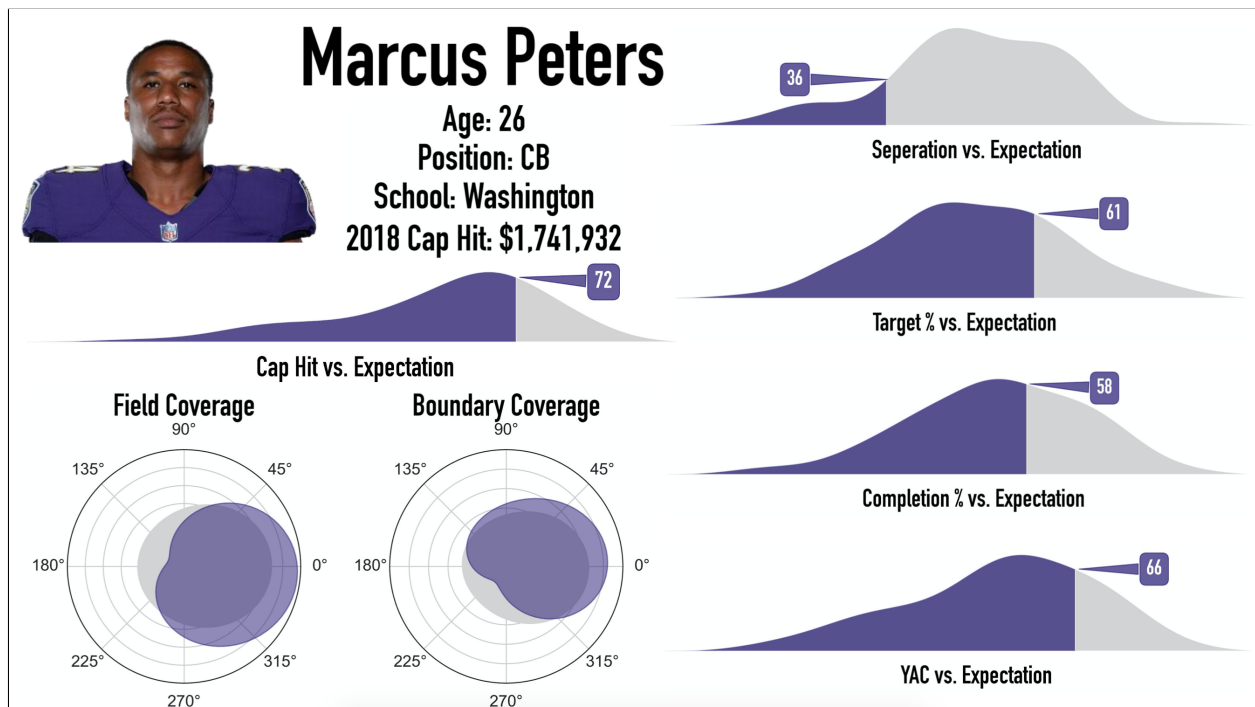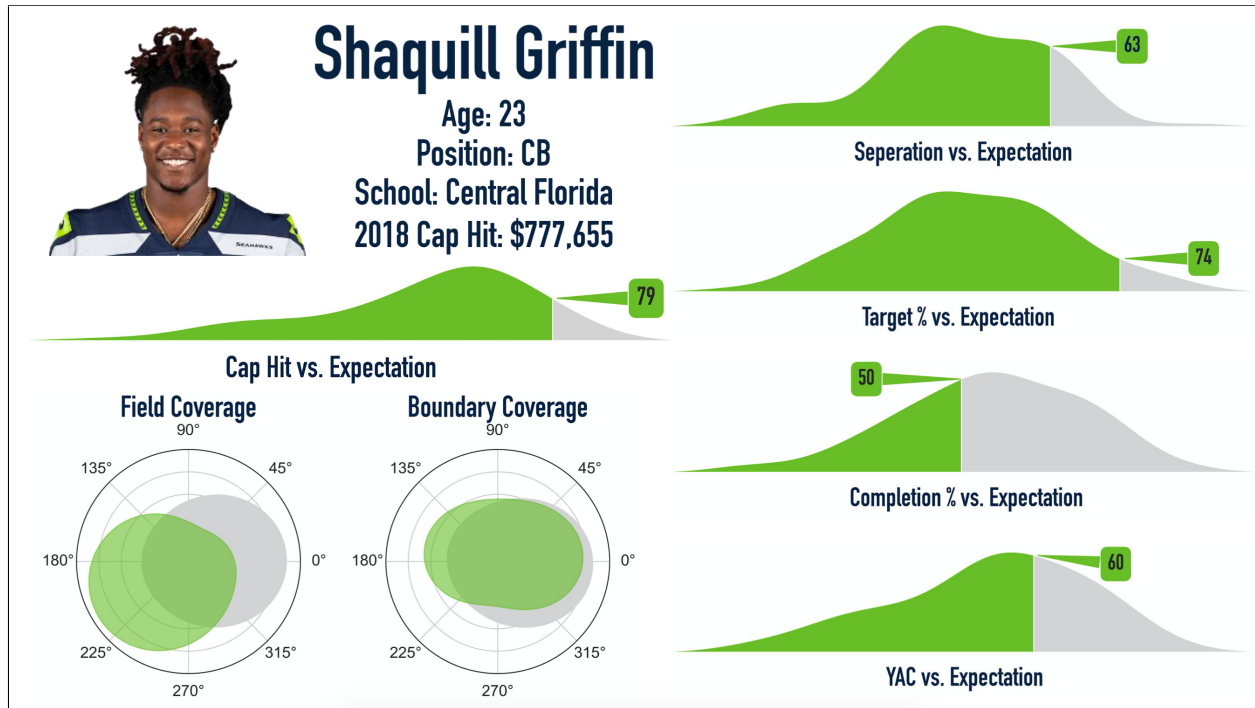
3.2 Separation

In close coverage, a Defensive Back's first job is to limit separation. Close coverage is defined as when a DB is the closest defender to a receiver for the vast majority of a play. From a burn-in time of 0.5 seconds (5 frames of tracking data) to the moment a pass is thrown, the distance between each DB and the receiver they're covering is calculated. At each frame, a DB's separation is compared to an expected separation given the receiver he's covering. Great receivers are going to create separation. What we can ask of great DBs is that they reduce this average value. The **Separation vs. Expectation** metric measures the extent to which a DB minimizes the separation we expect each individual receiver to generate. Since it captures all DB-receiver pairings instead of just those targeted, the sample size increases measurably, producing an incredibly robust metric.

3.3 Target %

A rarely quantified skill of DBs is their ability to generate a lack of targets. The selection bias occurring when a QB decides *not* to target his favorite receivers provides information about a DB's coverage ability. Similar to the above method, each receiver was tagged with an expected probability of being targeted. The **Target % vs. Expectation** metric measures how infrequently the receiver a DB covered was targeted *relative* to the probability he's targeted against any DB. Players who score well are those who QBs are targeting significantly less than the target probability of the receiver they're covering suggests.

3.4 Completion %

First, a Convolutional Neural Network (CNN) was developed using a modified version of the structure proposed by last year's Big Data Bowl winners, Dmitry Gordeev and Philipp Singer [1]. The idea of mirroring every play on the y-axis in order to double the size of the training set and improve model performance proposed by Gordeev and Singer was implemented as well.
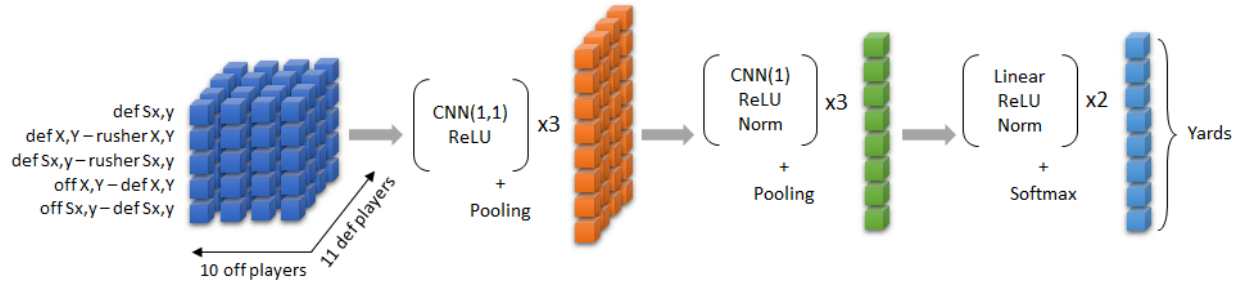
Due to the limited nature of the tracking data, this model only incorporated information from the closest 5 offensive and 6 defensive players to the targeted receiver at the time the pass is thrown. The convolutional layers remained the same structurally but were fine tuned for this problem. Instead of the softmax prediction required by last year's competition, this model has a sigmoid activation layer which predicts completion probability. Gordeev and Singer proved that this model construction is exceptional at generating a spatial approximation of multiple players relative to each other on a "blind" field, where the identity of each player is unknown. This prediction is then fed into a Bayesian hierarchical model where the identity of the targeted receiver is taken into account through Variational Inference and Hamiltonian Monte Carlo sampling.
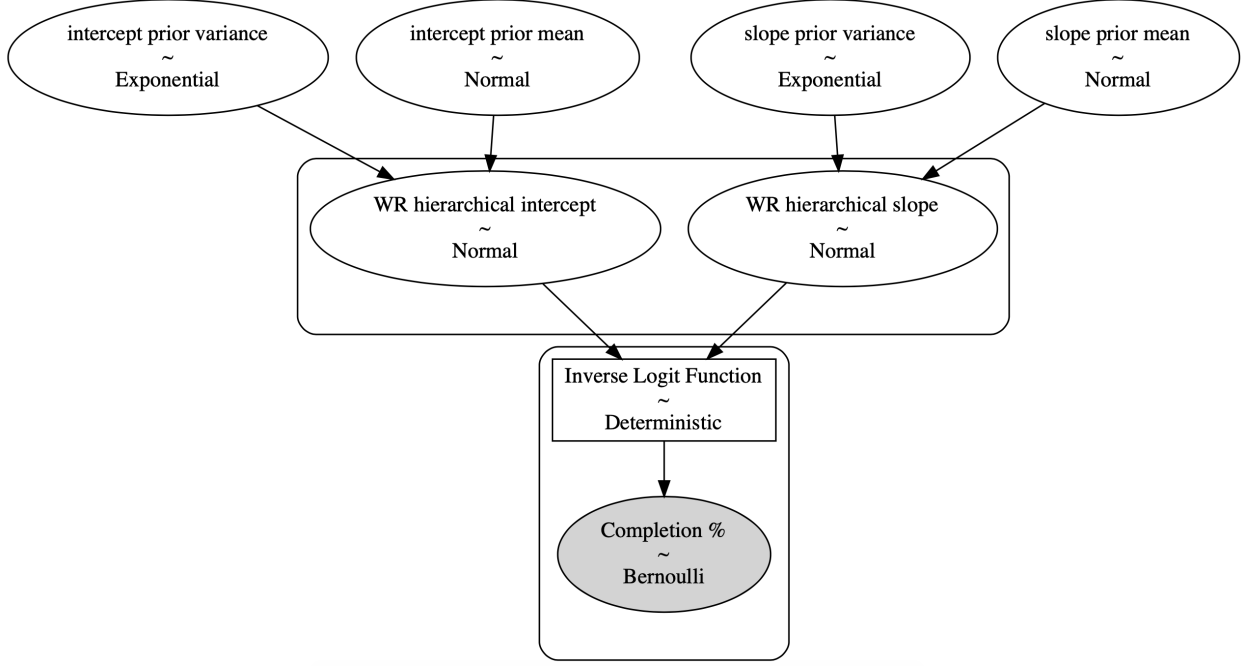
Statisticians have spent a great deal of time trying to identify the volume required for a player's sample average to be more predictive of their future results than simply the league average. This Bayesian adjustment discovers the precise blend of player and population influence over the course of $40,000$ samples to compliment the naked spatial completion probability calculated by the CNN. Comparing these predictions with real outcomes yields a **Completion % vs. Expectation** metric which scores a DB's influence over the expected probability a given receiver catches a pass, taking into account the position and movement of 11 other players.

3.5 Yards After Catch

The **YAC vs. Expectation** algorithm was crafted exactly like the previous one. The CNN structure was only adjusted so the final layer now predicts expected yards gained after catch at the time the pass is completed. This structure continues to provide a fantastic spatial approximation of nearby players and their possible influence on the play. The YAC prediction was fed into another Bayesian hierarchical model modified for a continuous variable. Convergence of both this and the previous model was confirmed across $40,000$ samples using trace and ELBO plots.

Each colored curve represents a different receiver's partially pooled estimate
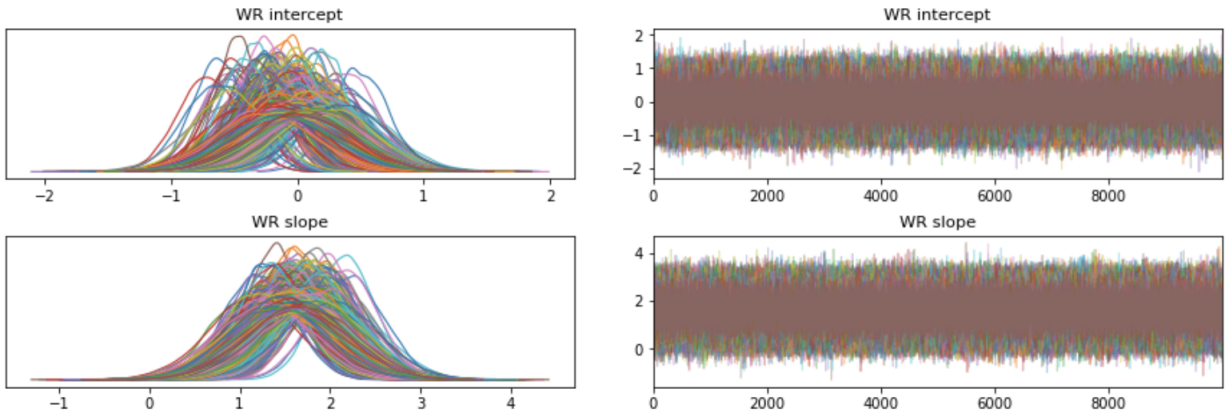
Let's say, given the distances and direction of movement of the closest players to a receiver at the moment they catch a ball, they're expected to gain 5 yards. They gain 6 and the DB covering them is scored negatively. You ask, but who was he covering? 5 yards is some average of an expectation for all receivers. Knowing the receiver is Julio Jones, the

4

Bayesian hierarchical model corrects its prediction to 7 yards. Our DB is no longer docked simply for having to cover a talented player but properly rewarded for limiting Julio's production relative to what we expect *him* to do.

3.6 Contract Value

The previous four metrics each evaluate an isolated skill, adjusted for the receiver covered. Combined, they tell a comprehensive story of a DB's ability to limit separation, how rarely he's targeted, his ability to force incompletions when he is targeted, and his ability to limit YAC when completions occur. The **Cap Hit vs. Expectation** algorithm takes every player's normalized scores and clusters them. A Gaussian Mixture Model (GMM) was selected over more common techniques, like K-Means. In addition to cluster means, GMMs account for the variance, or *spread*, of values for each metric in each cluster. This makes the clustering more sophisticated and provides a probability estimate that each player belongs to each cluster. Players are ultimately assigned to the cluster they have the highest probability of belonging to. Multiplying each player's cluster probability by the average salary of all players assigned to that cluster and adding those up generates an expected salary [2]. This is the salary a player with similar metric scores is expected to earn on average. Comparing this with true cap hits in 2018 provides an estimate of how over- or under-paid each player is given their skill-set, assigning a level of value to the contract. A number of the top 20 most valuable DB contracts in 2018 were outstanding young players like **Tre Flowers**, **Tre'Davious White**, **Marcus Peters**, **Jaire Alexander**, and **Shaquill Griffin**. The value in this metric includes discovering cheaper, impactful free agents, identifying valuable contracts to trade for, and recognizing players deserving of an extension.

3.7 Shadow Sonars

One thing that differentiates Defensive Backs is the way in which they shadow receivers. The sonars measure the angular difference between the direction the DB is moving and the expected position a receiver will be at 0.5 seconds in the future. The latter is calculated using the receiver's position and velocity vector, a combination of direction of movement and speed. One interpretation of this would be: the smaller this angle is, the better a DB can anticipate the movements of the receiver he's covering. In his article, *Manifold Nonparametrics*, renowned basketball statistician Justin Jacobs describes a "circular kernel density estimator" for the von Mises distribution which runs over angles -180 to 180 degrees [3].

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{e^{h \cos(x - x_i)}}{2\pi I_0(h)} \tag{1}$$

This is used to estimate the distribution angular differences between present DB directionality and future receiver movement in close coverage. In the sonars, gray represents the density estimation for all DBs. *Boundary* and *Field* designations were made due to the size of the relative areas of coverage and determined by location at the time of the snap. $0°$ represents a DB moving exactly in the direction of where we expect the receiver to be in 0.5 seconds. Any movement between $90°$ and $270°$ likely indicates movements within a zone match scheme. With this in mind, **Shaquill Griffin's** sonar makes sense when you watch some of his tape from 2018. Conversely, **Marcus Peters** displays an exceptional ability to track his target because his personal distribution pulls so much stronger towards $0°$ than average.

4 Discussion

While far from a replacement for traditional techniques, the additional information provided by algorithmic scouting cards is unparalleled and would provide any team with a significant competitive advantage. With information about coverage types and schemes, these algorithms could be modified to infer how joining a specific team would change in a player's score for each metric. With multiple years of data, these algorithms could find precise historical comparisons to aid in projecting future growth. As well, incorporating information about passers would help account for varying levels of QB talent as receivers make their way around the league.

References

[1] Dmitry Gordeev and Philipp Singer. "1st place solution The Zoo." 2019. https://www.kaggle.com/c/nfl-big-data-bowl-2020/discussion/119400.

[2] Kent, Brendan. "Sam Goldberg, Soccer Data Scientist." Episode 48. https://www.measurablespod.com/podcast.

[3] Jacobs, Justin. "Manifold Nonparametrics: Which Way Do Passers Pass?" October 9, 2019. https://squared2020.com/2019/10/09/manifold-nonparametrics-which-way-do-passers-pass/.