**Defending the Edge: Evaluating OT Performance through Euclidean Measurements**

Michael McCarron, Data Science Engineer (macciarain@protonmail.com)
Project repository: https://github.com/autonomous019/NFL_Big_Data_Bowl_23

Abstract:

NFL data used in the Kaggle competition NFL Databowl[1] is based in x,y coordinates, this allows for the measuring of player positions in relation to each other to be easily measured using Euclidean Distance (d). The problem of automated analysis of Offensive Lines remains a problem in NFL analytics. I develop a series of metrics to automate analysis using Euclidean Distances. I develop a proof of concept by focusing on the edge rush from Defensive Ends that are mitigated through the blocking by the Offensive Tackle (OT) position.

Like GIS applications the NFL tracks each game using visual recognition systems to locate the players on the playing field and marking their x.y locations 10 times per second. This data can be used to track player actions for each play. This typical graph representation allows for easy computational paradigms to be used to automate the interpretation of plays on the field. I develop a metric system based on six aspects of evaluating each play.

## Euclidean Distance in Metrics

First, a basic understanding of Euclidean Distance is presented. If we look at equation 1 below, we see that it is a straight forward line-of-sight measurement or 'as the crow flies' between two vectors, say $P^1_{x,y}$ and $P^2_{x,y}$ giving a value for the distance between the two vectors.

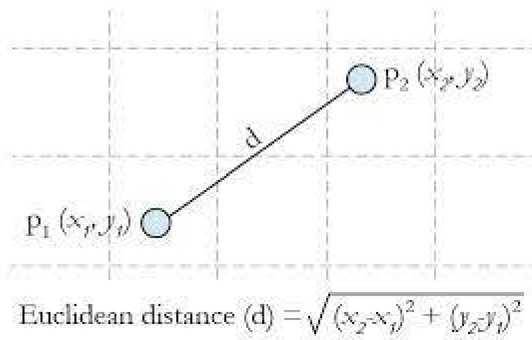
$$d = |\mathbf{x} - \mathbf{y}| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \,. \tag{1}$$

I use Euclidean Distance for several different measurements, measuring the distance (d) between the Defensive Rusher (DE) and the OT used for analyzing the style of block which is stored in the 'frame_metrics' table. The d between the DE and the Quarterback (QB), for comparison of starting d and ending d for measuring momentum. Another place that I use d is for measuring the d of straight line positions of the DE from one frame to the next and how effective the OT is in deterring the DE from proceeding in a straight line toward the QB, the shortest path to the goal. The straight line d is expressed as:

---

[1] Addison Howard, Ally Blake, Andrew Patton, Michael Lopez, Thompson Bliss, Will Cukierski , NFL Big Databowl 2023, ,(2002) https://kaggle.com/competitions/nfl-big-data-bowl-2023}

$$P^1_{x^{i-1},y^{i-1}} \text{ and } P^2_{x^i,y^i}$$




Where P1 is the predicted straight line next position. P2 is the next x,y coordinate of the actual path of the DE.

**Time Length of Plays in Evaluation Metrics**

It should be noted that there is a limit on the evaluation of plays, I do not evaluate past 4 seconds of a play, this affects a minority of the plays as the large majority of plays are less than 4 seconds, roughly 2/3rds of plays. 100ms equals 1 frame (10fps) in the NFL data. Also, one can see a distinction in the length of plays, one could query the database based on frames (i.e. where total_frames = 25, etc) to see results for a given time range in the plays and compare the scores for different time lengths. Plays become more chaotic the longer the play. I use 4s as a baseline since it provides enough of a sample size without encountering too much noise in the data due to the chaos of long plays.

QBI Avg by Length of Play (n is # of plays)*

| Play Length | QBI Avg |
|---|---|
| QBI < 2s[2] | 2.98405172413793 (n=580) |
| QBI >= 2.0s < 2.5s[3] | 2.97064531780689 (n=2061) |
| QBI i >= 2.5s < 3.0s | 2.93430656934307 (n=2603) |
| QBI >= 3.0s < 3.5s | 2.88750641354541 (n=1949) |
| QBI >= 3.5s <= 4.0s | 2.69439338235294 (n=3264) |

* Like the Richter scale there is an exponential decrease in success with length of time.

[2] select AVG(qbi_rating) from play_results WHERE total_frames < 20
[3] select AVG(qbi_rating) from play_results WHERE total_frames >= 20 AND total_frames < 25

Also related to the time length is that of the concept of 'interface length' which I define as when a Defender and Blocker and/or QB are within 1.5yds of each other. Several metrics are only calculated during this 'interface length' as a limiting factor on data analysis. This is important for measuring momentum, straight lines, block efficiency, and QB-DE metrics, It is viewed as the length that an OT can reasonably affect the DE.

**Frame Metrics Aggregation**

As mentioned above d is measured between the Blocker, Defender and QB, these measurements are stored in the 'frame_metrics table in the db. Frame metrics gathers relevant x,y coordinates of QB, OT, and DE or other rusher, measures the required euclideans and writes these values to the 'frame_metrics' table in the db[4] giving 3 different sets of euclideans to do data analysis on to formulate the basis of the evaluation metrics. The code that drives this aggregation is in the OT_Defending_The_Edge_Metric_Aggregation.ipynb notebook file.

**Six Aspects of Evaluation**

There are six different metrics generated by me to use in evaluating each block. I view 1 as quantitative and 5 as related to individual style. The QBI is quantitative and I use it in comparison to other systems for evaluating blocks such as those by PFF (pff.com). The other 5 are more subjective in understanding and do not directly relate to success or failure as some blockers are less prone to close proximity blocking and use their arms but generate as good results as those that only use arm blocks.

- Quality of Block Index (QBI)
- Block Efficiency
- Block Momentum
- QB-DE Rating
- QB-DE Euclidean
- Straight Lines Euclideans

The above metrics are presented in more detail below.

**-QBI: Quality Block Index** is a rating system that penalizes based on the OT either being beaten or allowing himself or the rusher to enter three zones around the QB: buffer (1.5yds), danger (1.0yds) or sack (<.7yds). It is based on a scale of 0-3 and comprises the primary metric for overall evaluation akin to PFF ranking of OL. If a blocker or defender is in the buffer zone a .5pt penalty is applied, if in the danger zone a further .5pt penalty, if in the sack area they are zeroed out, score=0. If they are beaten, the DE is closer to the QB then the blocker, a .5pt penalty, if they regain control, the beaten penalty is .25pts. These are strictly speaking position based or coordinate based ratings relative to the QB. Qualitative metrics are noted below. I use two scales: the QBI and the QBI+Style Metrics for

---

[4] The db, 40_databowl.db is available at https://www.kaggle.com/autonomous019/defending-the-edge-game-results/ in the data directory, /inputs/4-seconds/

evaluation. QBI data is held in the qbi_metrics table in the db. QBI is what I use to rank players although other more quantitative measures from the styles metrics such as block momentum and QB-DE Rating could be combined to augment the results. According to QBI alone rankings the top 10 results were:

| QBI Rank | Name | nflId in db | AVG(QBI) all plays |
|---|---|---|---|
| 1 | Andrew Whitworth | 30869 | 2.934584 |
| 2 | Tristan Wirfs | 52421 | 2.934543 |
| 3 | Andrew Thomas | 52412 | 2.926037 |
| 4 | Charles Leno | 41475 | 2.921698 |
| 5 | Trent Williams | 35443 | 2.920399 |
| 6 | Rob Havenstein | 42400 | 2.919004 |
| 7 | Riley Reiff | 38553 | 2.915470 |
| 8 | Justin Herron | 52603 | 2.907897 |
| 9 | Donovan Smith | 42377 | 2.902957 |
| 10 | Brian O'Neill | 46131 | 2.893011 |

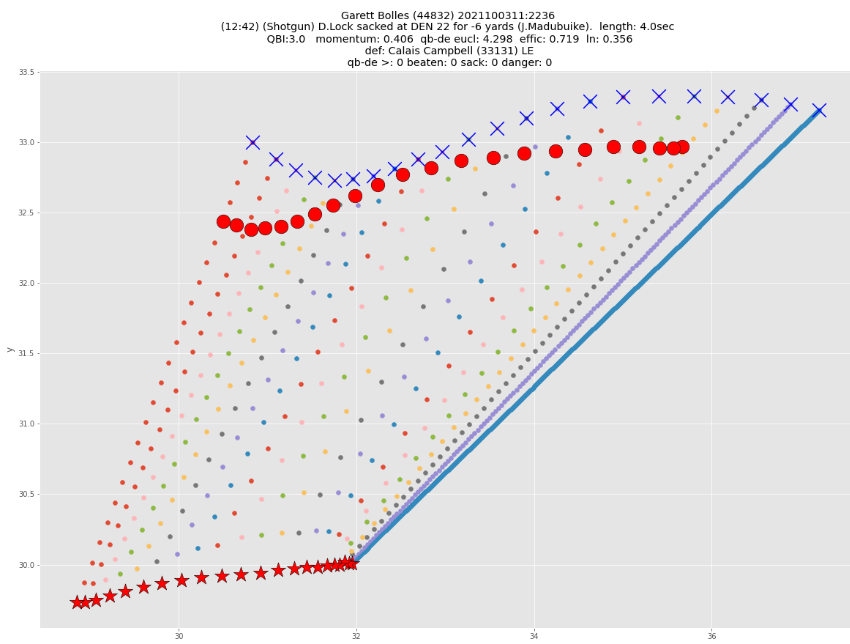
As is seen some of these players are usually included in most top 10 lists of OTs, such as PFF’s. It should be noted that these rankings are based on partial game plays not a full game, which gives different results depending on the plays included in the partial sample of the game. If by random chance there is a negative or positive bias in the plays dramatically affects overall rankings, such as the results of Riley Reiff and Brian O’Neill. See discussion below regarding rectifying results with other evaluation systems, whereas the majority of the top 10 are also in many lists of top 10 OTs here.

**-Block Efficiency:** Is a measure based on d that gives a score for the percentage of blocks that are close to the body, viz arm blocks, with arm close to fully extended. A positive score is given if the close body block is above 25% of the overall interaction length.

**-Block Momentum:** Is a measure based on whether the blocker is able to slow the momentum from the start of the interface length to the end of the interface length, if a blocker is able to slow the momentum in more frames then not slowing the defender (>=50%) a positive score is given. Momentum is a measure of whether the blocker slows the rate of closure on the QB, measured on a frame by frame basis, score reflects total percentage of frames where momentum decreased during interface.

**-QB-DE Rating and Euclidean**: Is a measure that looks at the d between the DE and the QB. If the d of the DE increases, increases between the QB and DE, from the start of the interaction to the end of the interface length a positive score is given and the QB-DE Rating is given as 1, otherwise it is scored as 0 (0,1).  If the euclidean average value is above 4yds then a positive score is given. It is important to note that this value usually increases on edge rushes when the arc is affected by the OT to push the rusher away or outward from the QB. This can be seen in the plots provided at the link below.

Plot showing Metrics and Straight Lines to QB from Defender[5]




**-Straight Lines Divergence:** Is an evaluation based on the concept of divergence of the defender from taking a straight line toward the quarterback, the ability of the blocker to deter the rusher from the shortest path toward his target, the QB, it is also based in measuring the euclidean. This euclidean is a measurement from the defender's position compared to his predicted position on a straight line to the QB from his previous position. The higher the euclidean distance the better the result, the higher one is off the straight line to the QB. Straight Lines is held in the 'lines_metrics'

[5] Straight Line plots as well as 5 vs 5 animated plots are coded in the notebook at https://colab.research.google.com/drive/1C6aM5tLTj6luCqj951OdMvQhbH_IHBQE?usp=sharing

table in the db. If the line_rating avg for the play is above .3yds (1ft) a positive score is given. There is a divergence in all metrics between completed and incompleted passes. It is interesting to look at the straight line averages in all the data as an example of this, see below.

Metrics such as Momentum, the euclidean relative to start and end of play during interface between blocker and defender, the QB-DE rating and euclidean scores are aggregated in the 'block_metrics' table. So that 3 metrics are created through the block metrics algo: qb-de rating (0,1), qb-de euclidean (float), momentum (0-100) and efficiency (0-100). The code for the aggregation algorithm for the block metrics table is in the QBI and Block Metric Algorithm Notebook[6]

**-Play and Game Results:** As mentioned the QBI Rating, used for a more objective evaluation and the QBI+Style (QBI+) evaluation which includes measures that are a question of blocking styles and more subjective but informative on an individual level of blocking style. The final QBI+ is on a scale of 0-5.5 with QBI 3pts and QB-DE Rating (.5pts), QB-DE Euclidean (.5pts), Lines Rating (.5pts), Momentum (.5pts) and Efficiency (.5pts). Scores are 0 or .5 based on a threshold value. Completed Passes v. Incompleted Passes is one interesting comparison to look at as far as the total results go based on averages for all play results in all games for all blockers. Though caution is given not to make too broad of conclusions as we are only analyzing one persons contributions to a completed pass which involves the syncronicity of 11 players in all, even in terms of pass blocking we are only dealing with 1 out of 5 players involved in blocking, thus more often then not the individual player is successful in their assignment but someone else may not be thus giving a incompletion. However, as a broad measure it is an interesting comparison nonetheless.

| **Metric** | **Completed** | **Incompleted** |
|---|---|---|
| QBI | 2.91670564508992 | 2.79130113132339 |
| QBI+Styles | 4.30825119043233 | 4.20102202731276 |
| Straight Lines | 0.302775380602603 | 0.29456476156998 |
| Efficiency | 0.257506123423736 | 0.282060508846367 |
| Momentum | 0.534847609731337 | 0.558647465270838 |
| QB-DE Euclidean Distance | 5.35071634105951 | 5.21878901320947 |
| QB-DE Rating | 0.223161053178863 | 0.247030742322318 |

[6] https://colab.research.google.com/drive/1sGJMSsW4gVAAPmYzmH0JQ_VeP-vQbQOp?usp=sharing

These comparisons suggest that some measures are more objective than others, since they match with the outcomes such as QBI, Straight Lines, and QB-DE Euclidean. Due to computation time lengths the next version will include all 5 blockers in the data rather then just one player per play, which may shed more light on what metrics are more informative in building a model of evaluating OL performance. To go over play and game results in more detail see the notebook https://www.kaggle.com/autonomous019/defending-the-edge-game-results

**Rectifying Differences Between this Evaluation Model and Other Systems of Evaluation**

As mentioned above, depending on the bias in the random sample of the plays involved, being an incomplete set, most evaluations use a full game or every play rather than a subset, and this has given a couple surprising results both positively and negatively. In contrast to such consistent results between this evaluation and PFF[7] where Andrew Whitworth finishes in the top spot in both, and players like Charles Leno, Dion Dawkins finishing almost in identical positions in both, others move up or down depending on the type of data in the partial data set used in this evaluation. A deeper look at the data used to model these results explains this divergence. Looking at Jordan Mailata, who is rated by PFF as the no. 7 Pass Blocker but here finished 49 of 50. Inversely we look at Riley Reiff who finished in the top 10 in QBI but is rated by PFF as the 99th Pass Blocker. As well as look at Tyron Smith who slides down from no. 3 on PFF PB ratings to 16 in QBI score here.

First, for Jordan Mailata it is important to note that in the 6 games rated he actually participated in 243 plays, but this dataset only has 161 of those plays. If you consider PFF sacks, hits, hurries and pressures you get a negativity rate of .1069. In this dataset when considering negative plays such as sack area, danger area, buffer area, beaten we have a negativity rate of .260 which is 160% divergence. In both systems the sack is weighted heavier but in PFF dataset there are 2 sacks, in this dataset there are 5. So we see the different results are explained by the different sample size and the negative bias of that sample set.

On the opposite side of things is the rapid rise of Reilly Reif in these results compared to PFF's up from PFF's 99th PB OT to 7 out of 50 here. Again, looking at the sample size we see that he participated in 7 games in weeks 1-8 for a total of 287 plays, in this dataset 245 are rated, the negativity rate in the PFF rated plays is .1045, but in this dataset it is only marginally greater .1306, in this evaluation there are more negative chances then in PFF data. Most error rates double between PFF and this evaluation, so this error rate of Reif is very low. Notably in the PFF ratings there are 4 sacks, but in this evaluation dataset there are 0 in the sack area, so we see how this divergence is explained.

Finally, there is the large slide of Tyron Smith PFF No.3, QBI 16 of 50, again this can be explained in the sample versus actual datasets. There are 264 plays, in this dataset there are 188 plays. The PFF negativity rate is .0378, the QBI negativity rate is .1329, so 3 fold greater. in QBI data has 6

[7] PFF data is at https://premium.pff.com/nfl/positions/2021/REGPO/offense-pass-blocking?position=T, file is behind a pay-wall subscription service. (Accessed 12/19/22)

plays with a score <= 2.0, 2 in sack area, in PFF data only 1 sack and very low numbers for other negative plays.

This gives us a clearer picture for interpreting the results. To truly test the results of this evaluation we would need to be able to directly compare them to all the plays for all the blockers. It does suggest that QBI is a good way to measure automatically whether a blocker is doing a credible job or not.