An exploration of collective order in American Football using GPS tracking data

Introduction

This research paper aims to conduct an exploratory analysis of data from a set of American Football games provided by the National Football League (NFL). Professional sports in general have seen a rapidly increasing need for data science and analytical techniques in response to the increasing availability of detailed game data (Macdonald, 2020). In the US, professional baseball has a long history of statistical analysis thanks to the large volume of publicly available data, but American football has lagged behind (Yurko et al, 2019b). Recently however, the NFL started making selected portions of its detailed tracking data available through data analysis competitions on Kaggle's online platform, which has led to a surge in published research activity.

Historically, data analysis for American football was limited to published 'play-by-play' information which only included those players actively involved in the play's outcome (Dutta et al, 2020). This clearly limits the ability to conduct deep analysis of team performances. On the other hand, GPS tracking data allows for a thorough quantitative analysis of teams, groups of players and individual players (Herold et al, 2019). A side effect of this additional level of information is, of course, that the analysis becomes ever more complex as the volume of tracking data increases (Yurko et al, 2019a).

Obviously professional sports teams are primarily interested in data analysis in order to gain an advantage over their opponents and win their matches. There is a strong history of the use of key performance indicators (KPIs) in American football strategies but developing them was incredibly time consuming and could not give much consideration to subtle interactions or rapidly changing contexts (Herold et al, 2019). Hence, there had been some degree of stagnation in NFL team strategies, particularly on-field, as coaches brought up on these 'truisms' stuck with what they knew. However, largely thanks to the availability of high-quality data and analytical techniques, there is now a new wave of 'analytics' sweeping the league. Perhaps one of the most important aspects of these analytics is the nuance possible in probability-based methods, where the decision making can be assessed instead of the result – a good decision may not always work out, and occasionally bad decisions get lucky, but consistent results are more likely to come from consistently good decision making (Deshpande and Evans, 2019).

There has been substantial prior work done in other ‘field-invasion’ type team sports such as soccer and rugby. Welch et al (2021) conducted a study investigating collective order states and transitions in soccer and observed a number of differences in these states between the attacking and defending game phases. The concept of collective states is to look at the effect on the group of relatively small interactions between individuals and their environment. This paper will conduct a similar analysis of order parameters and collective state to ascertain any differences between American football, a discrete play-based sport, and soccer, a free-flowing sport.

Order parameters change over time, so they may also provide useful information to use in conjunction with the base data. In contrast to looking solely at play results, spatiotemporal information such as this is very useful for investigating the development of a play (Macdonald, 2020). Hence this paper will also investigate whether the order parameters at different stages of a football play can be used to map offensive movements in a way that can provide a quick classification system for offensive plays.

Offensive passing plays are defined by the combination of passing routes that are used. A single passing route, or pass pattern, is the predetermined path that an eligible receiver will run on a play (Chu et al, 2020). The combination of routes in each play is a ‘pass concept’. With up to five routes per play and around 70 plays per game, manually identifying the routes run on a play from video footage is a time consuming and arduous task but doing so is critical for making use of football analytical data (Chu et al, 2020). The final task in this paper will be developing a machine learning model to automate route detection from GPS data.

Materials and methods

Datasets and processing

This project utilises a dataset provided in a Kaggle competition titled “NFL Big Data Bowl 2021: Help evaluate defensive performance on passing plays”. The data was provided by the NFL and contains information for all passing plays that occurred in the 2018 regular season.

The focus of the competition was on the defensive 'coverage' with the winning entry being a model for classifying coverage and inherent pass defender capabilities on a frame-by-frame basis (https://operations.nfl.com/gameday/analytics/big-data-bowl/).

The data consists of game data, player data, play data and tracking data. The game and player data is high-level information, the most important of which are the unique `gameId` and `nflId` fields, which are used for identifying the games and players respectively, and the player's coordinate data in `position`. The play data contains more detailed summary information including a unique play identifier (`playId`) as well as which team is on offense, where they are on the field and a range of features that describe the game situation, such as the distance needed for a first down, current scores and time remaining. The tracking data contains the detailed information for each player in each play.

GPS tracking data was collected at a rate of 10Hz, or ten frames per second, and recorded the player's position, speed, acceleration, orientation and direction of movement. Distances are recorded in yards and angles in degrees. The tracking system consists of 20-30 ultra-wide band receivers, 2-3 radio-frequency identification tags (RFID) installed in each player's shoulder pads and additional RFID tags located in the ball, on game officials and on various field markers and instruments (Dutta et al, 2020). Position tracking is performed via triangulation of the RFID transmission from receivers spread around the stadium (Yurko et al, 2019a).

The complete dataset represents the passing plays from 32 teams in 512 games over a 16-game season. Each game consists of an average of 35 passing plays per team, with each play tracking the movement of 12-14 players for a period of up to 8 seconds. In the complete dataset there were 17,032,421 timesteps recorded. Filtering the dataset to only those offensive players involved in running a passing route reduces this to 4,884,074 timesteps representing 65,452 routes.

As found by Yurko et al (2019a) the data required pre-processing prior to using it for any analytical purposes. Firstly, errors in the data, such as incorrect formatting or incorrect data types, were corrected or, if the data was incomplete or otherwise unusable, removed. Several transformations were then applied to the tracking data to improve interpretability. The result of these transformations was to rotate the position data such that all offensive plays were heading to the right, and all directions were converted to radians with a zero-bearing pointing to the target endzone and positive and negative angles indicating left or right rotation respectively.

Ultimately the player's position is represented as a set of (x, y) coordinates within the field of play which has an origin of (0, 0) and extends to 120 yards along the x-axis and 53.3 yards along the y-axis. This is illustrated in Figure 1, showing the pre and post transformation field layout and player position parameters.

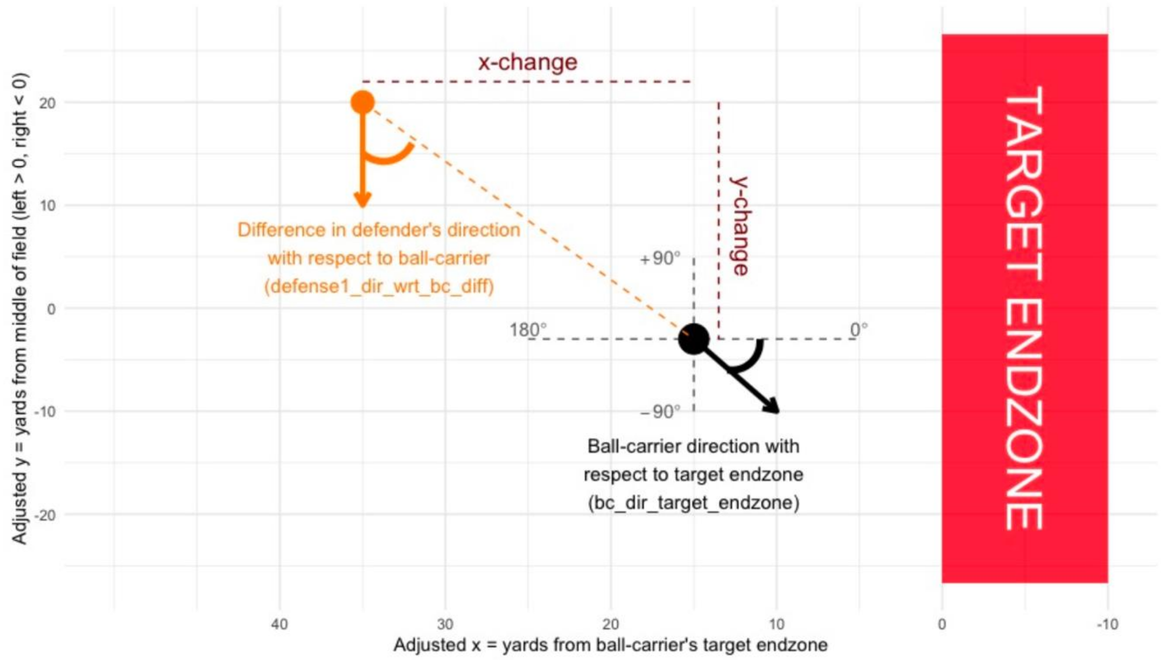

Figure 1: Visual representation of the supplied positional and directional data compared to the transformed version used for analysis
Note: From "Going Deep: Models for Continuous-Time Within-Play Valuation of Game Outcomes in American Football with Tracking Data," by R. Yurko, F. Matano, l.F. Richardson, N. Granered, T. Pospisil, K. Pelechrinis and S.L. Ventura, 2019, https://doi.org/10.48550/arXiv.1906.01760

The tracking data was prelabelled with various game events. The key events are the commencement of pre-snap motions, the snap of the ball, a pass being thrown, the pass arriving, and a range of play outcomes such as whether the ball was caught, knocked down or intercepted and at what point the ball carrier was tackled, went out of bounds or scored a touchdown. These labels help break each play down into key stages and are used as markers for potential transitions in play (Dutta et al, 2020). Any plays labelled with events that do not occur on standard passing plays (such as punting) were excluded from the analysis.

Order parameters and collective states

Order parameters are used to describe the collective movement of each team and indicate the relative order or disorder in their collective states (Tunstrøm et al, 2013). Two key order parameters, polarisation and angular momentum, are calculated and from those a classification of the state of the system can be derived.

Polarisation (P_{group}) is a measure of common alignment. If all players are running straight up-field they will be highly polarised (P_{group} approaches one), whereas if they were running in different directions across the field, they would have low polarisation (P_{group} approaches zero). Following the methodology of Tunstrøm et al (2013) the P_{group} value is calculated at each timestep using the following formula:

Equation 1: Calculating the polarisation of a group at time t

$$P_{group}(t) = \frac{1}{N}\left|\sum_{i=1}^{N} \boldsymbol{u_i}(t)\right|$$

Where $\mathbf{u_i}$ is the unit vector for the direction of player *i* and N is the total number of players involved in the play from the same team.

The angular momentum (M_{group}) of the group is defined by equation two as follows:

Equation 2: Calculating the angular momentum of a group at time t

$$M_{group}(t) = \frac{1}{N}\left|\sum_{i=1}^{N} \mathbf{u_i}(t)\ x\ \boldsymbol{r_i}(t)\right|$$

Where $\mathbf{r_i}$ is the unit vector from the centre of the group to player *i*. This measures the degree of uniformity in direction of rotation (clockwise or anti-clockwise) of the players around the group's centre. High levels of angular momentum (M_{group} approaching one) indicate a consistent direction of movement compared to an inconsistent direction at low angular momentum (M_{group} approaching zero).

Tunstrøm et al (2013) found that the degree of polarisation and angular momentum at each time point can be combined to classify the overall state as *polar* (P >0.65 & M<0.35), *swarm* (P <0.35 & M<0.35), *milling* (P <0.35 & M>0.65) or *transitional* otherwise. These definitions have been used for classifying the collective state in our data, which allows direct comparison to the results in Welch et al (2021). A transitional state indicates that a group can move from one state to another, or lose and regain a state. While in the transitional state the group will hold some combination of features from the major states.

As per Welch et al's (2021) methodology several additional parameters are calculated to provide further information relating to the group's behaviour. The group's mean speed (V_{group}), in yards per second, is defined by equation three:

Equation 3: Calculating the mean group velocity at time t

$$V_{group}(t) = \frac{\left|c_{group}(t) - c_{group}(t - \Delta t)\right|}{\Delta t}$$

where Δt is the constant time between datapoints (0.1 seconds in this case) and the position of the group's centre C_{group} is:

Equation 4: Determing the centre of a group at time t

$$C_{group}(t) = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{c_i}(t)$$

Where $\mathbf{c}_i(t)$ is the position of player *i* at time t. This calculates the speed of the group centre (centroid) at each time point.

The *Shannon Entropy* (H_{group}) is an indicator for the degree of predictability in the system (Welch et al, 2021). The H_{group} value is calculated across the whole play and several windows within the play: before the snap of the ball, from the snap until the ball is thrown, from the throw to the ball arriving at the target and from the arrival of the pass to the end of the play. This is similar to the breakdown of plays in Yurko et al (2019a) but maintaining discrete periods instead of combined periods. The square domain of -12<=x<=12, -12<=y<=12 (yards) is divided into a grid of square bins 1 yard wide. This results in a two-dimensional grid of 576 bins that are indexed as column *i*, row *j*. This configuration was found to be the most suitable balance between the resolution of the data and the maximal speeds and distances covered per timestep (approximately 1.2 yards per 1/10 second).

Following the method in (Welch et al, 2021), each player's motion vectors are overlaid on the grid with every vector's tail located at (0,0) so that the relative change in position per timestep can be assessed. We then count the number of vectors pointing into each bin and aggregate the counts across all players involved in the play. The number of vectors counted vary from as few as 100 to as many as 400, depending on the portion of the play being counted. The simple probability of a given change in direction pointing to a particular bin is calculated for each bin in the grid and then the H_{group} value is calculated via the equation below:

Equation 5: Calculating the Shannon-Entropy of a group at time t

$$H_{group}(t) = -\sum_{i=1}^{24}\sum_{j=1}^{24} P_{ij}\log_2 P_{ij}$$

Where P_{ij} is the probability of a change in direction pointing to bin *ij*:

Equation 6: Calculating the probability of a given change in direction

$$P_{ij} = \frac{f_{ij}}{\sum_i \sum_j f_{ij}}$$

Low values of H_{group} indicate the group state is more predictable and carries less information, whereas high values of H_{group} indicate the motion of the group is less predictable and changes in position provide more information. This is intended to provide another layer of information over polarisation and angular momentum by including the effect of each player's speed instead of solely directional differences (Welch et al, 2021).

Finally, the surface area (A_{group}) occupied by each team is calculated using methods available in Python's Scipy library. The surface area is determined by finding the convex hull surrounding the players and calculating the area of the polygon it forms.

The P_{group}, M_{group}, V_{group} and A_{group} values were calculated for each frame across the whole play and then averaged across the whole play and the sub-play windows as described previously. This results in play level values for all five parameters across various windows relating to specific events within the plays.

Unsupervised learning – clustering methods for 'type of play'

Several unsupervised learning methods were used to attempt classification of offensive plays using the order parameters described previously and other features from the dataset. The types of passing routes labelled in the data set were determined and a count of the number of times each route was run in each play was added to the play level summary data along with a count of the total number of routes run. Approximately two thirds of the data did not have labelled route data and was removed from the data set for these methods. Yurko et al (2019a) found a similar proportion of incomplete data, indicating some underlying issues with the dataset or the pre-processing pipeline in use by the NFL. The data was restricted to the order parameters for the offense's *pre-snap*, *to-throw* and *to-arrived* sections of the play, along with the aforementioned route counts. The *to-end* stage of the play, being a reaction to the play's execution, is discarded as it is not indicative of the play's essential structure.

The first clustering method investigated was a simple K-means. K-Means uses Euclidean distance between feature dimensions, so the dataset was first normalised to remove the effect of differing dimensional scaling. Silhouette and elbow methods were used to determine the optimal number of clusters, *K*.

The second method investigated was agglomerative clustering. Agglomerative clustering is a form of hierarchical clustering that works in a bottom-up manner, pairing the most related features together to form clusters of closely related features. Ward linkage with Euclidean distance was used to find the clusters. The search was tuned on number of clusters as well as distance threshold with results appraised using a silhouette score.

Supervised learning – machine learning methods for 'route identification'

Approximately 15% of otherwise eligible players did not have any route label recorded. Several methods have been used to attempt this previously, including using Bezier curves to find a best fit (Chu et al, 2020) or convolutional neural networks (Sterken, 2019). Alternative approaches have been to use unsupervised learning to find clusters of similar routes and then label the clusters manually (Chu et al, 2019). The supervised attempts have been able to achieve approximately 70% accuracy. Here we investigate multiple other supervised learning methods in an effort to find a simpler, yet equally effective, labelling agent.

The models were trained against the player's position and speed in each time point from the snap of the ball to the earliest end point of the play after the ball is thrown. Most routes develop in less than five seconds and pre-snap movements are not significant in running the actual route, hence constraining the data focuses the model onto the most representative portion of the play (Kinney, 2019). For each route run it was first determined whether they were positioned to the QB's left and, if so, their positional data was inverted so that all route positional data was based on the same orientation, the y-axis then having an 'inside-outside' context rather than 'left-right'. The player's speed was then mapped to their *x, y* coordinate position for each timepoint with the resulting grid flattened into a 2D vector of speed values which could be directly used by the desired algorithms.

The dataset was randomly split with 60% of data used for training and 40% retained for testing. All models were assessed using F1 scores and a confusion matrix plot.

The first algorithm used was the Support Vector Machine (SVM) via the Scikit-Learn SVC module. The 'rbf' kernel was selected as the best performing kernel using the default hyperparameters. Tuning of the hyperparameters, *gamma* and *c*, was abandoned due to the performance issues of the algorithm, in particular that it is single-threaded, which made it impractical for further experimentation. Though given the availability of other, more performant, algorithms this did not cause any great issue to the study.

The Scikit-Learn Random Forest Classifier was used, with a maximum depth of 10 and 1000 estimators.

Scikit-Learn's Multi-layer Perceptron (MLP) was used for an initial investigation into neural network performance. The L-BFGS algorithm was used as it is recommended for faster convergence in smaller datasets (https://scikit-learn.org/stable/modules/neural_networks_supervised.html). However, Scikit Learn does not support GPU processing, limiting the amount of tuning that can practically be performed. Hence, the neural network development was moved to the Pytorch framework, the Multi-Class Neural Network (MCNN) module specifically, to make use of CUDA based GPU processing.

The MCNN model utilised alternating linear and non-linear layers. The input layer comprised 9180 feature nodes, being the size of the (x, y) domain for player positions across all routes run. The output layer contained 11 classification nodes, each representing one of the possible routes run. The model used five hidden layers consisting of 256 nodes in each. The five total layers alternated between linear and non-linear. A Stochastic Gradient Descent (SGD) algorithm was used for optimisation, with a learning rate of 0.5, and Cross Entropy Loss for the loss function. The model was trained over 2000 epochs.

All models were assessed using the F1 score from the Classification Report from Scikit-Learn's Metrics module and a Confusion Matrix was generated using functions from the same module.

Plots and visualisation

Probability density plots were generated to visualise the relationships between various order parameters. The methodology for this analysis was as per (Welch et al, 2021), using 50x50 binning of parameter pairs and plotting the resulting heatmap of bin counts. The baseline of 50 bins was used for the P_{group} and M_{group} parameters while 60 bins were used for the V_{group} and A_{group} and 140 bins for the H_{group}. The distributions were smoothed by averaging the values of each bin and its adjacent bins.

The density plots use either the entire eligible plays dataset (n=6053) or the subsets of plays classified as being in the polar, swarm or transitional states (n=403, 241 and 5409 respectively). There were zero incidences of the milling state.

To aid interpretation of clustering results, pairwise relationship plots were generated to visualise the distribution of clusters in each feature pair. For each pair a scatter plot of data points coloured according to the assigned cluster is accompanied by histograms for each axis illustrating the distribution of data points along that axis. A 3D scatter plot was then generated to visualise the separation of clusters in the set of most important features.

Statistical analysis

A paired t-test was used on clustered data to ascertain the degree of difference in route usage in each cluster. This used the independent T-test method from Python's Scipy library. Means for each route in each cluster were compared against each other and tallied if they were significant at the selected p-value (0.02).

Results and discussion

Global analysis

The complete dataset contained 17,907 plays however 11,854 plays contained incomplete or possibly erroneous data. These plays were removed from the analysis leaving only 6053 plays to assess. The remaining plays represented approximately half of the games played in the season. This proportion of usable plays is similar to that seen in other work, such as by Yurko et al (2019a).

(Welch et al, 2021) found a significant portion of play in soccer to be in an 'out of play' phase however unlike free-flowing field sports such as soccer there is no 'out of play' phase in American football. Teams may have a difference in time of possession, and hence the amount of time (or number of plays) they engage in offensive or defensive activities, but because it is possible to score from any position on the field, including by the defence, the importance of the length of a given play sequence is very dependent on the game context. The context of the game situation for each play is outside the scope of this investigation thus the analysis focuses on the aggregation of all offensive and defensive play.

The first set of plots examines pair wise relationships between order parameters over full play durations. Play durations may range from very fast (< 2 seconds) in the case of short and incomplete passes to relatively long (> 7 seconds) in the case of completed deep passes or short completions taken for long gains. Figures 2 through 8 summarise the relationship between various order parameters for the offense (top) and defence (bottom). Figure 2 (left) shows moderate rates of polarisation and angular momentum for both teams, with the offense tending towards higher levels of angular momentum. Some play types, such as floods, involve most of the offensive players flowing across the field to overload one side of the defence, which likely explains this subtle difference in the $P_{group} – M_{group}$ relationship.

Figure 2 (right) shows a similar, but less dramatic, positive relationship between polarisation and mean group velocity as found in soccer by Welch et al (2021). The defence generally operates at a broader range of speeds, utilising low speeds far more often. It is likely that there are similar effects on the offense's speed while crossing through the on-field traffic as has been found in pedestrian systems where a more ordered, polarised state enables higher average speeds (Helbin and Molnar, 1995, as cited in Welch et al, 2021). Defensive players generally have less distance to travel as they get to their assigned areas of coverage, hence do not need to reach maximal speeds until they need to 'make a play', whereas the offensive players generally have further to travel and need to travel at higher speeds in order to break through the defence's structure.

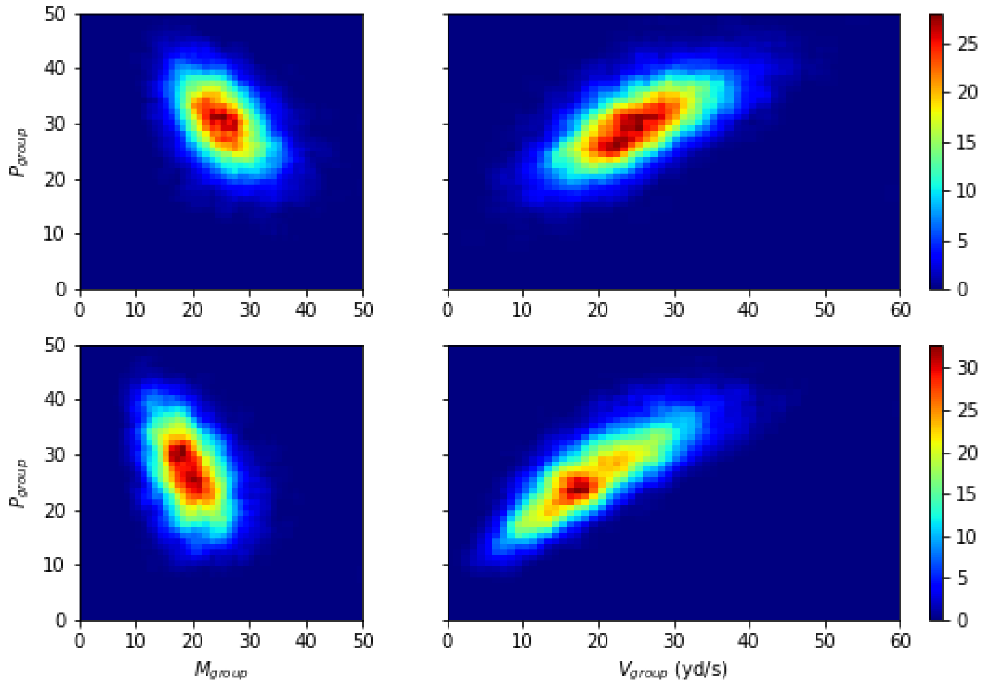

Figure 2: (Left) Density plots of paired polarisation (P_{group}) and angular momentum (M_{group}) and (Right) paired polarisation (P_{group}) and mean group velocity (V_{group}) for offense (top) and defence (bottom)

The chaotic nature of American football is demonstrated by the entropy distributions in Figures 3 and 4 (left). Both teams exhibit consistently high levels of entropy at all group speeds, illustrating the fact that teams disguise what they are doing for as long as possible (Dutta et al, 2020). This also contrasts with the moderate levels of entropy seen by Welch et al (2021) in soccer, where there was also a positive relationship with group speed. Where American football is played at a frenetic pace interspersed with stoppages for planning the next play, soccer is played at a more moderate pace with intervals of higher speed breakthrough plays.

Similar to Welch et al (2021) we find that the defence has a far more compact coverage of field area, as shown in Figures 3 and 4 (right). This is partly due to the offense's Quarterback (QB) providing an anchor point near the start of the play while the receiver's move up-field, whereas the pass-coverage team within the defensive unit generally moves in cohesion with the depth of the offensive attack. This aside, as per Brown (2017), the offensive strategy is generally based around stretching the defence horizontally, vertically or both ways simultaneously, therefore it makes sense to see the offense occupying a greater surface area of the field, especially at higher group velocities where they are likely attacking further upfield.

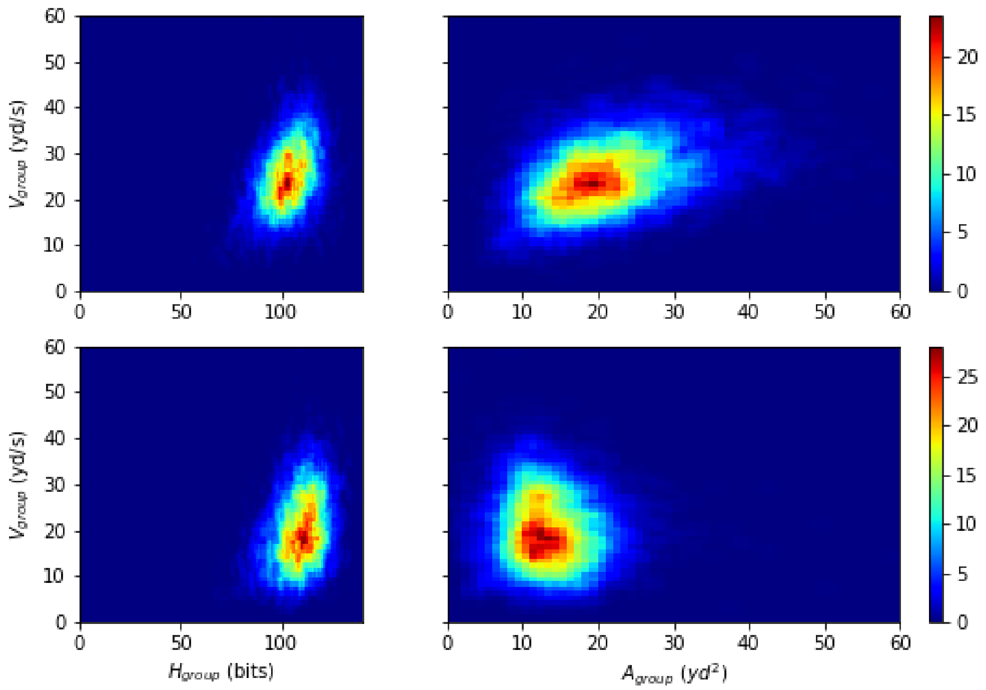

Figure 3: (Left) Density plots of paired group entropy (H_{group}) and mean group velocity (V_{group}) and (Right) paired surface area (A_{group}) and mean group velocity (V_{group}) for offense (top) and defence (bottom)

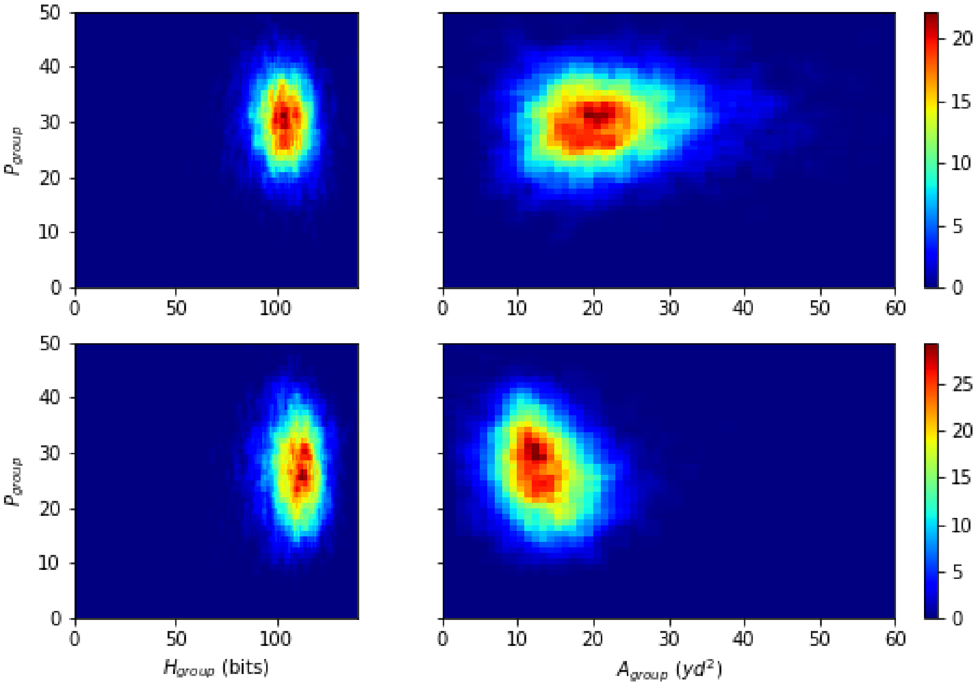

Figure 4: (Left) Density plots of paired group entropy (Hgroup) and polarisation (Pgroup) and (Right) paired surface area (Agroup) and polarisation (Pgroup) for offense (top) and defence (bottom)

We know that collective dynamics arise from the interactions between individuals within a group with other individuals and their environment (Tunstrøm et al, 2013). As stated in Dutta et al (2020) offensive and defensive players react to other players throughout the course of the play. Given the nature of plays in football these interactions are necessarily of extremely short durations. We have also seen that both teams hide as much of their on-field strategy as possible resulting in a constant high-entropy state. These factors result in a low level of collective order for both teams, represented by the predominance of the transitional state and the high average entropy shown in Figure 4 (left).

Collective movement analysis by game phase

The second set of plots show the order parameter relationships across different time windows within the plays. The density plots were generated for the 'snap to throw', 'throw to arrival of pass' and 'arrival of pass to end of play' windows. The 'presnap' window was excluded as both teams are in a generally static position during this phase of the play (Chu et al, 2019).

Figure 5 shows interesting differences across the play phases when compared to the aggregate result in Figure 2. Over the course of the play the offense's state of order becomes much less concentrated with a slight general trend towards the polar state. Contrarily, the defence demonstrates a much stronger move towards an ordered polar state. The likely explanation for this is that the offensive players aim to spread across the field and provide an open target for the QB to throw to, while the defensive players must take a more reactive approach, waiting for the offense to show their hand before moving. However, once the defence knows where the ball is going, they can rapidly converge to a single point, as it is exceedingly rare for an offensive player to attempt passing off to another player. This movement pattern is very well illustrated by the move towards the polar state towards the end of the play.

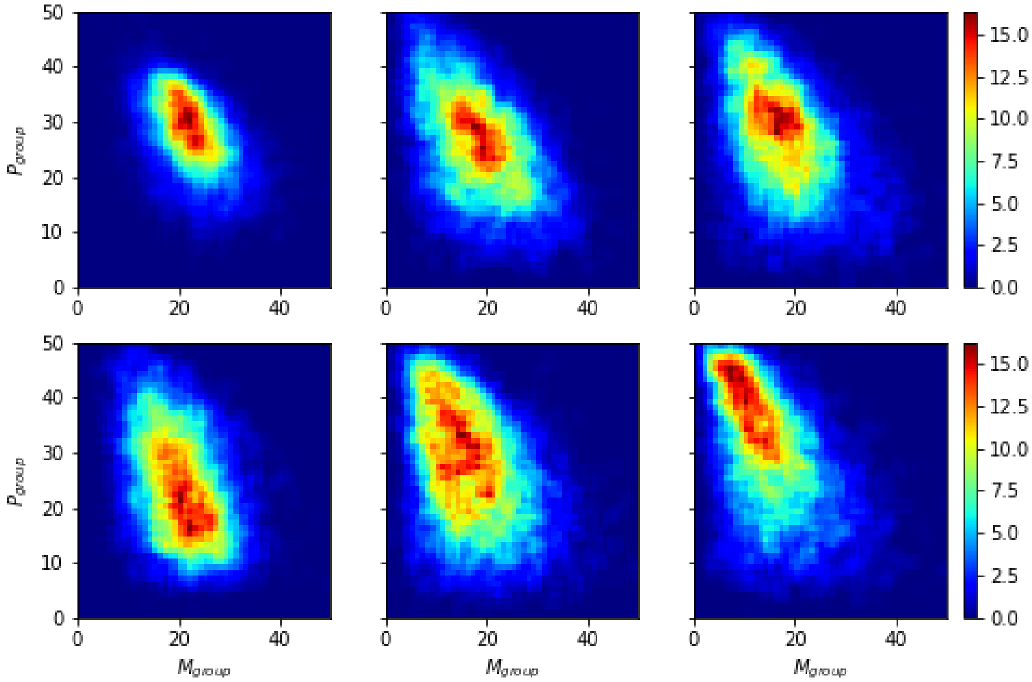

Figure 5: Density plots of paired polarisation (Pgroup) and angular momentum (Mgroup) across the 'snap-to-throw' (left), 'throw-to-arrival' (middle) and 'arrival-to-end' (right) play phases for offense (top) and defence (bottom)

The main observation when looking at the field area coverage throughout the course of play (Figure 6) is that the defence trails the offense in field area covered until equalising towards the end of the play. Again, we see both teams maintaining a largely transitional state of order, with only the defence showing a marked degree of polarisation towards the end of the play. There does not appear to be any particular relationship between the degree of polarisation and area coverage, which is as expected given the ability of the offense to stretch the field using a wide variety of play designs (Brown, 2017).

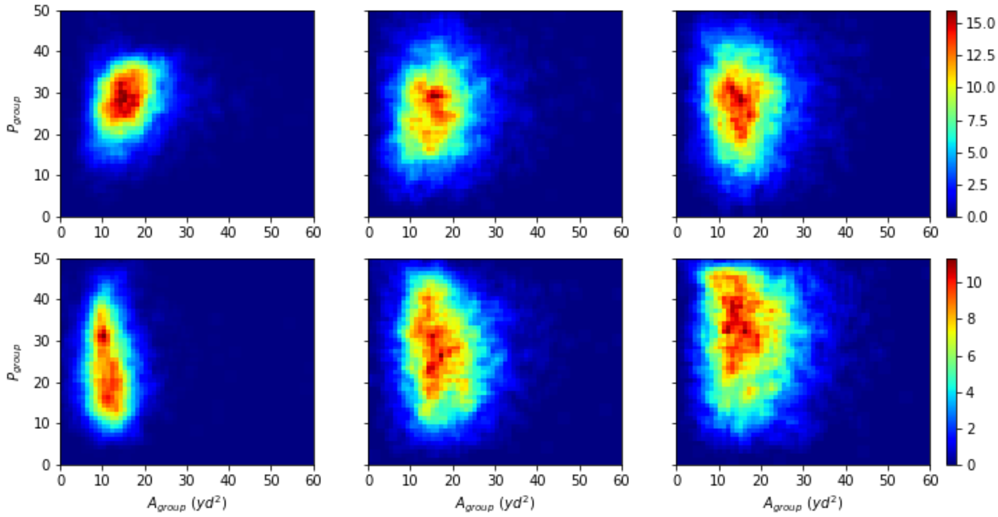

Figure 6: Density plots of paired polarisation (Pgroup) and surface area (Agroup) across the 'snap-to-throw' (left), 'throw-to-arrival' (middle) and 'arrival-to-end' (right) play phases for offense (top) and defence (bottom)

Offensive players largely dictate their own movements while defenders generally react to the offensive players, and it is also common for defensive players to orientate themselves towards the nearest offensive player (Dutta et al, 2020). As such we can expect defensive players to have quite a low degree of polarisation, particularly in the early stages of a play. This is illustrated in Figure 7, where the defence moves from a low velocity, non-polar state to a high velocity, high polarisation state towards the end of the play. The offense on the other hand maintains a transitional state throughout as previously described, but interesting shows a general reduction in mean velocity through the course of play. This is likely due to the pre-throw phase of play being when potential receivers are trying to 'get open' and then once the ball is thrown, or caught, they slow down due to defenders converging on the ball carrier. It is likely that players on extreme edges of the play slow down as they either have no effective role in the play or take on a blocking role, which also explains why the defence maintains a wide spread of velocities instead of showing a concentration of high speed at the end of the play.

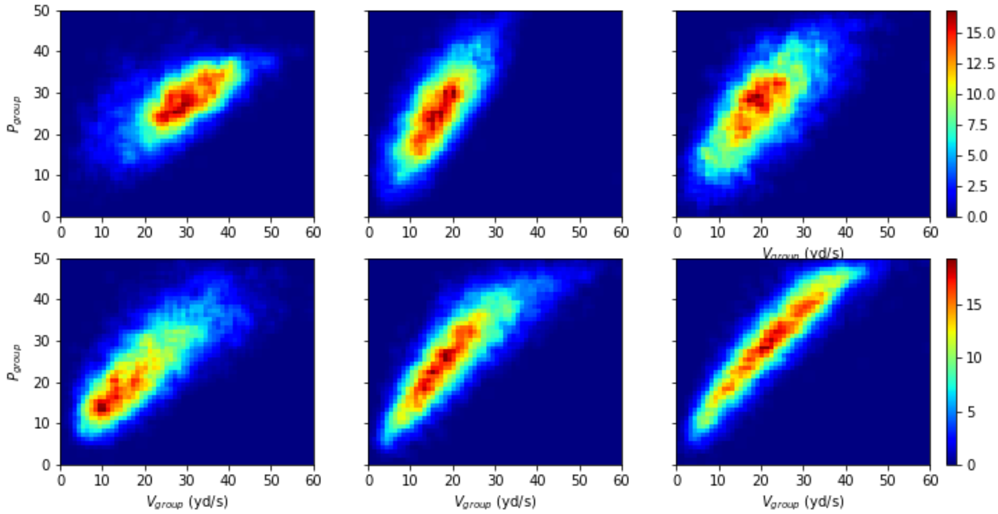

Figure 7: Density plots of paired polarisation (Pgroup) and mean group velocity (Vgroup) across the 'snap-to-throw' (left), 'throw-to-arrival' (middle) and 'arrival-to-end' (right) play phases for offense (top) and defence (bottom)

Figure 8 builds on the pattern of a high entropy levels seen previously in Figure 4. Entropy is maximised in the initial stage of the play, reduces slightly during the play's development, becoming more widely distributed by the end of the play. At all stages of play development entropy levels are maintained at higher levels than seen in the soccer data from Welch et al (2021). This reinforces the previous statement that American football is a tactical and chaotic sport where teams have a short time window to learn and react to what the other team is doing before commencing another play and starting the cat-and-mouse game over again.

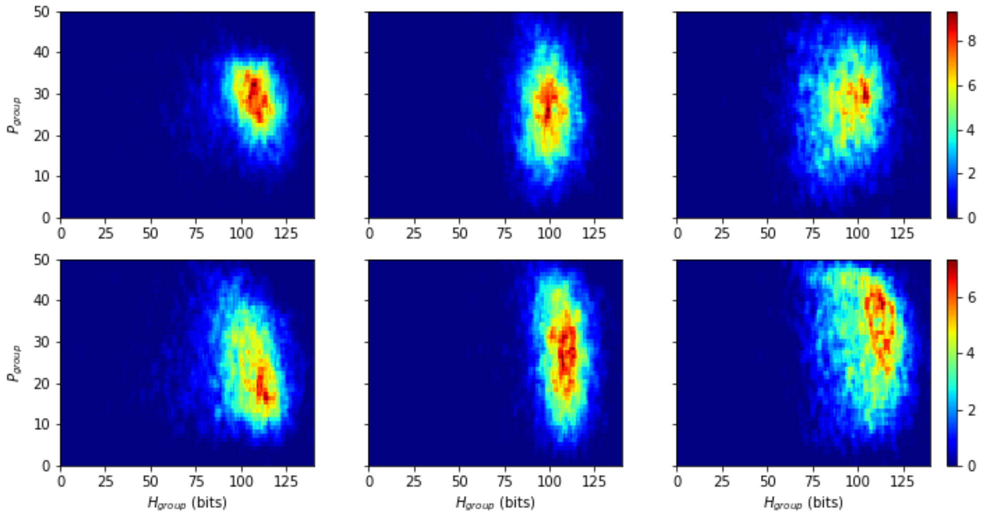

Figure 8: Density plots of paired polarisation (Pgroup) and group entropy (Hgroup) across the 'snap-to-throw' (left), 'throw-to-arrival' (middle) and 'arrival-to-end' (right) play phases for offense (top) and defence (bottom)

Collective state transitions

Tunstrøm et al (2013) proposed a three-state model for summarising collective order with an additional 'transitional' state for moving between the steady-state behaviours. Welch et al (2021) found that the team in their soccer sample showed a much higher level of collective order when playing defensively, where the team was generally in a polar state, compared to offensively, where the team tended to move between polar and transitional states. In their out-of-play sample they found a general spread between polar and swarming states, with the balance shifting towards swarming. The soccer example demonstrates a generally ordered state within the team. Conversely, we find that American football teams, on offense and defence, operate almost entirely from the transitional state, as shown in Figure 9.

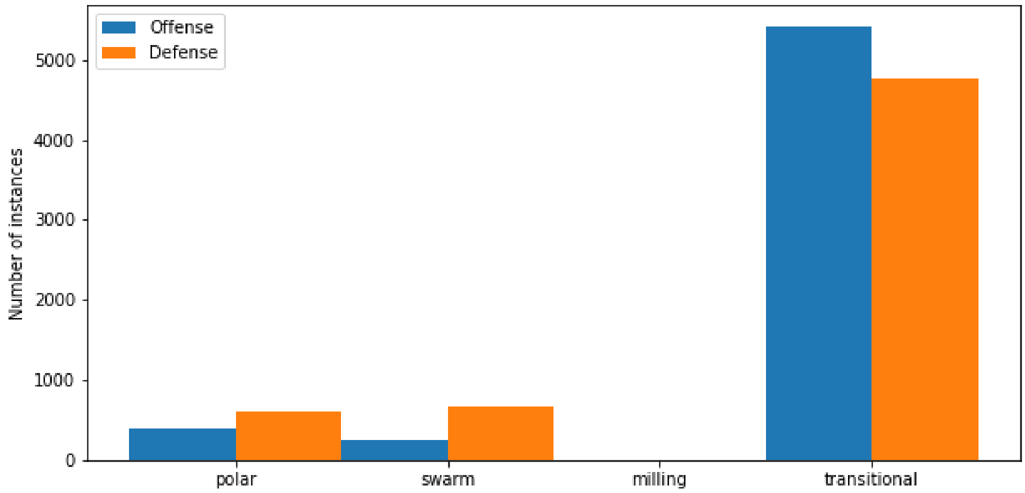

Figure 9: Distribution of collective states for offense (blue) and defence (orange)

No plays exhibited characteristics of the milling state, where movement circles a central point, for offense or defence, which is as expected in a sport requiring forward progression. The transitional state was by far the most dominant state for both teams. The defence had similar proportions of polar and swarm states (610 and 672) whereas the offense was almost twice as likely to be in a polar state than swarming (403 vs 241) but more likely overall to be in the transitional state.

The dominance of the low-order transitional state agrees with the short chaotic nature of plays in NFL football.

Use of collective state features in classification of play types

Given the clear differences in order parameters throughout a developing play they may provide a useful tool for helping classify offensive play types. Dutta et al (2020) use unsupervised clustering methods for identifying pass coverage, using an array of parameters determined from relationships between offensive and defensive players in each frame of data. Their attempts were successful in identifying whether the defence was in man or zone coverage (two clusters) but relied on the fact that the defence is reacting to the offense, so similar methods would not be of use in identifying offensive plays. Other work has been done by Chu et al (2019), Sterken (2019) and Haar (2019) to find optimal two and three route combinations but not for broad classification of the entire play concept.

A K-Means algorithm was used to scan for up to 30 clusters. The maximum silhouette score (Figure 10 - Upper) was found with only two clusters but we know that there are more than two base concepts to identify. Dutta et al (2020) remind us that clustering is an imprecise activity and that there can be more than one 'true' set of clusters, therefore we must use context and domain understanding to rationalise clustering attempts. Using the elbow method we find a value of around six to be the optimal value of *k* (Figure 10 - Lower), which remains a valid option based on optimising with the silhouette score.

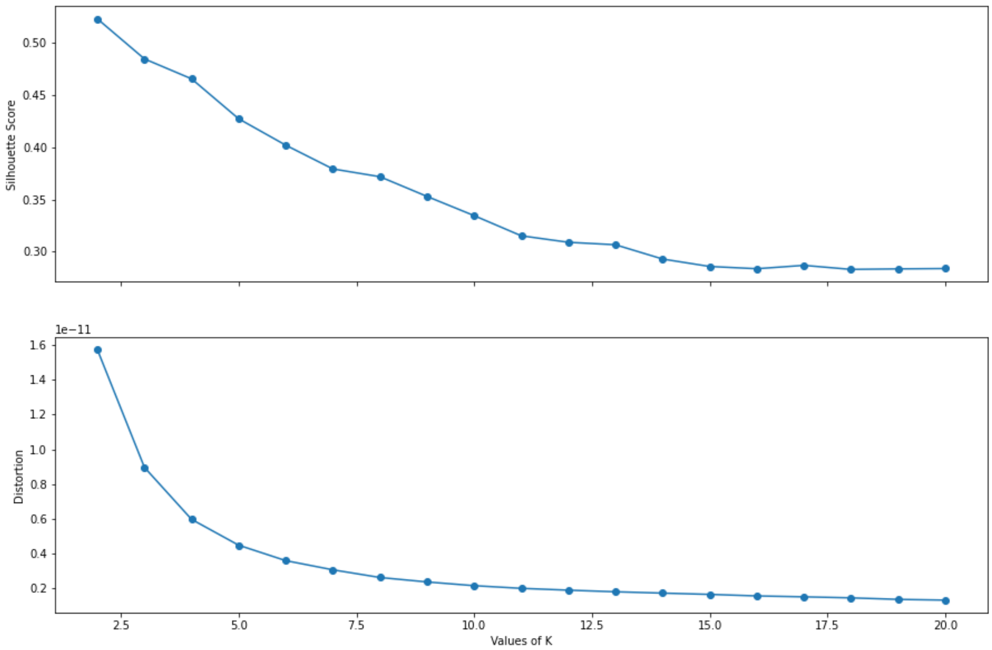

Figure 10: Optimisation of the K-Means clustering algorithm using Silhouette Score (Top) or the Elbow method (Bottom)

Using an Agglomerative Clustering algorithm again finds the optimal number of clusters to be in the range of two to seven. Scanning for clusters by distance threshold finds the silhouette score plateaus at a threshold of 1.1×10^{-6} which generates six clusters (Figure 11).

Comparing the distribution of clusters for both methods shows some consistency in identifying three major clusters, combining for almost 75% of the distribution, and three minor clusters making up the remainder (Figure 12).

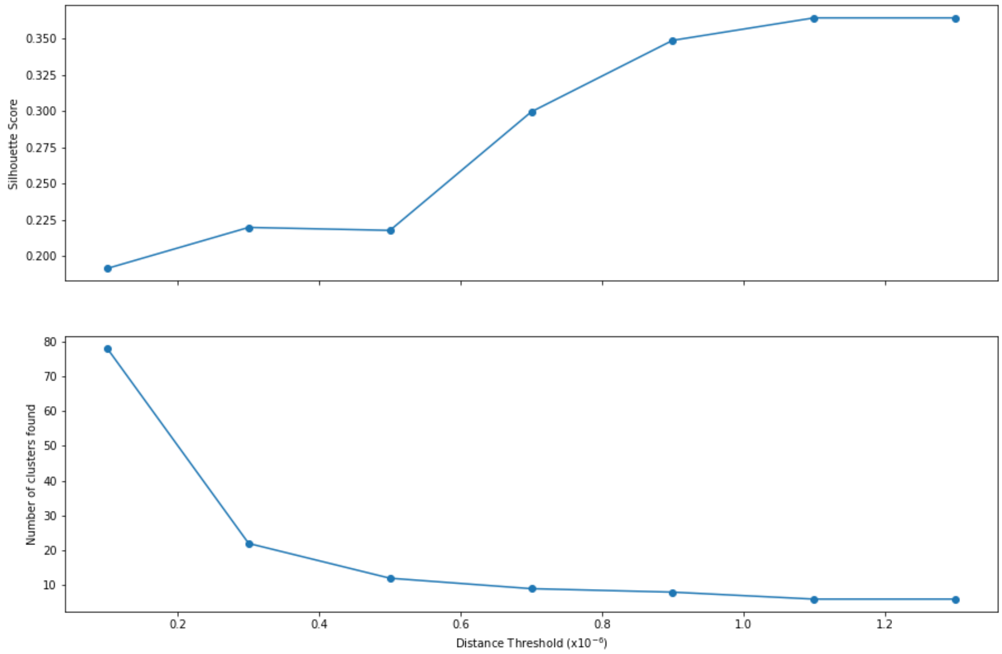

Figure 11: Optimisation of the Agglomerative Clustering algorithm using Silhouette Score (Top) or the Elbow method (Bottom)

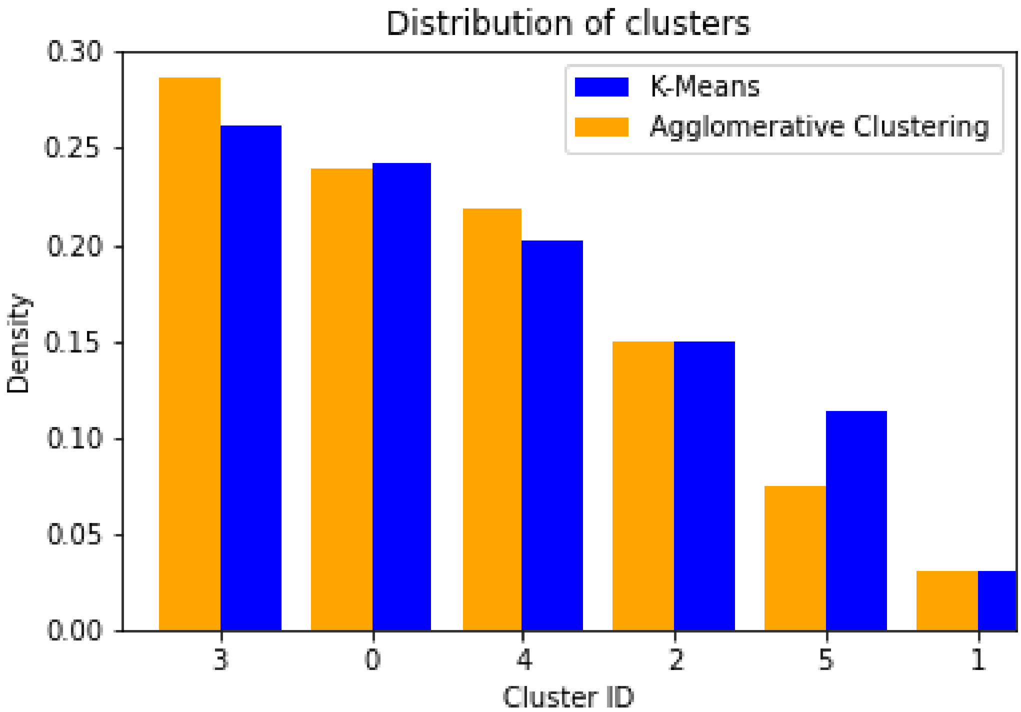

Figure 12: Comparison of the distribution of clusters found through K-Means (Blue) or Agglomerative Clustering (Orange)

Using these clustering methods to label the dataset and then train a Random Forest Classifier enables finding the most important features. The two clustering methods produce clusters with very similar features (Figure 13), with the offense's 'area-to-arrived' and 'area-to-throw' explaining more than 85% of the clustering. The 'entropy-to-throw' and 'entropy-to-arrived' features complete the set of most important features, though they pale in significance compared to the former pair.

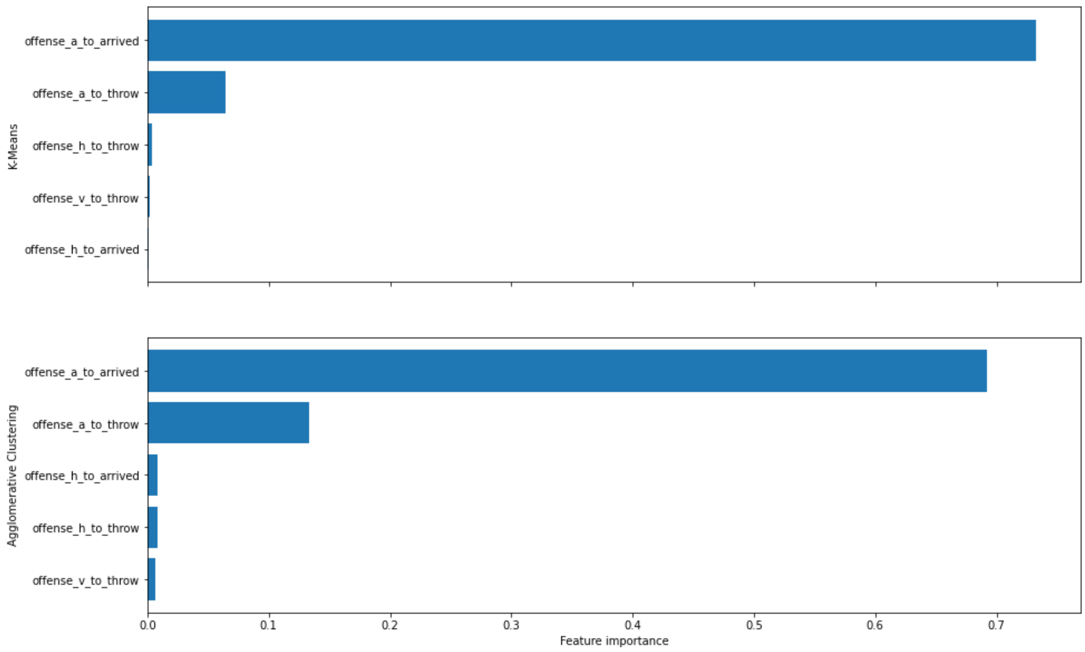

Figure 13: Comparison of the Top-5 most important features in the K-Means (Top) and Agglomerative Clustering (Bottom) clustering methods

Plotting cluster relationship plots for each of these pairs (Figure 14) shows the strong linear grouping based on increasing area in the period between the throw and the ball arriving at the target, with a similar but less distinct pattern forming according to area prior to the throw. A passing play concept is essentially defined by the combination of routes used (Brown, 2017). As the area of field covered will map closely to the pass routes run on the play it appears that the measure of field coverage provides a useful proxy or generalisation of the routes for means of clustering.

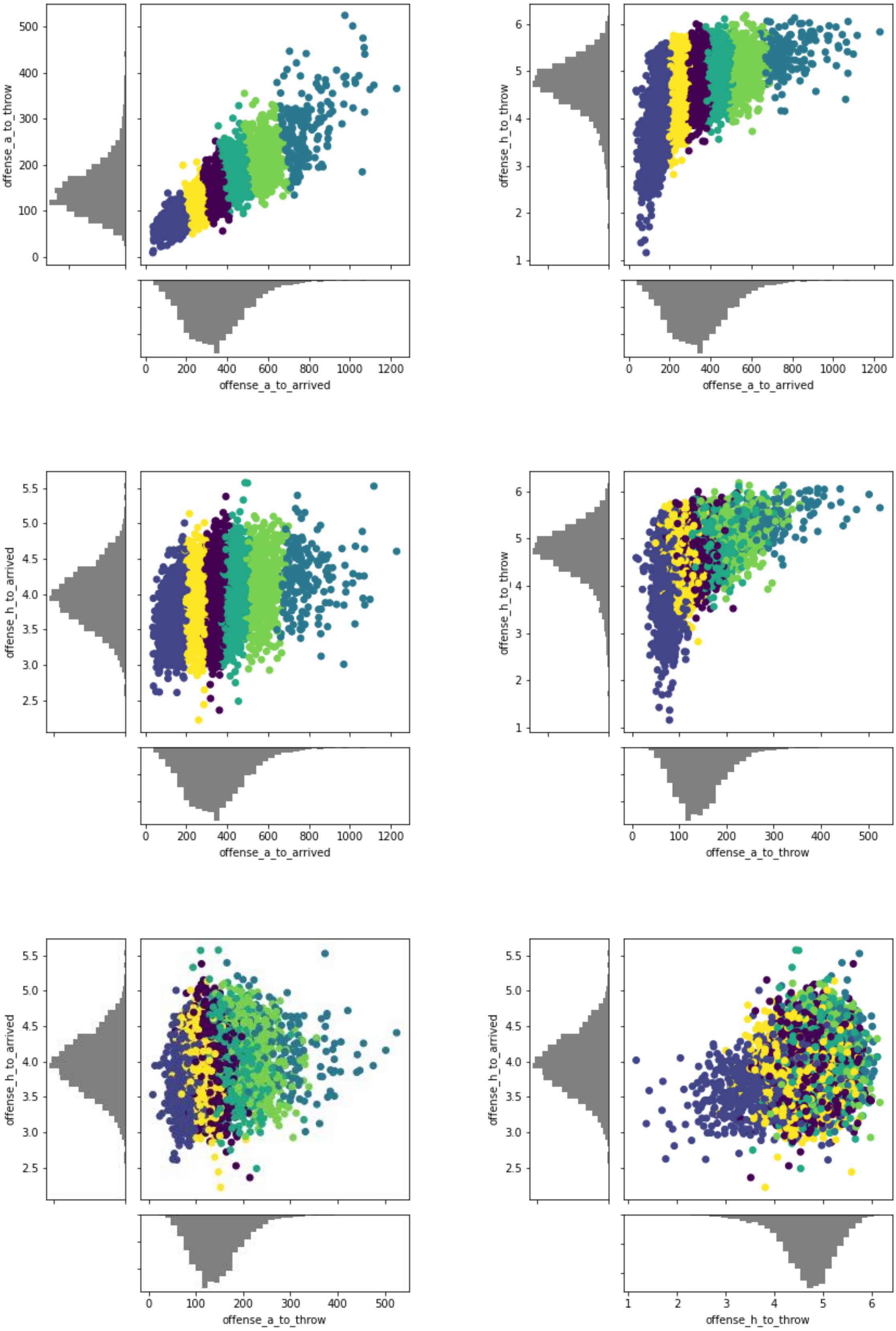

Figure 14: Relationship plots showing the separation of groups in each pair of the most important features for clustering

Table 1 summarises the degree of difference in route usage across the six groups of offensive play types. All routes are used in all clusters but to varying extents. Figure 15 illustrates the variation in the level of individual route usage across the clusters.

Table 1: Number of routes with significant (p <= 0.02) difference in usage between clusters

		Cluster A					
		0	**1**	**2**	**3**	**4**	**5**
Cluster B	**0**	N/A	11	6	5	7	8
	1		N/A	8	10	11	10
	2			N/A	5	4	6
	3				N/A	3	8
	4					N/A	7
	5						N/A

These results indicate that clustering of offensive plays is possible and that the number and type of routes utilised in the plays within a given cluster have a small direct effect on cluster selection, but likely have a much greater indirect effect through their association with field area and player velocity.

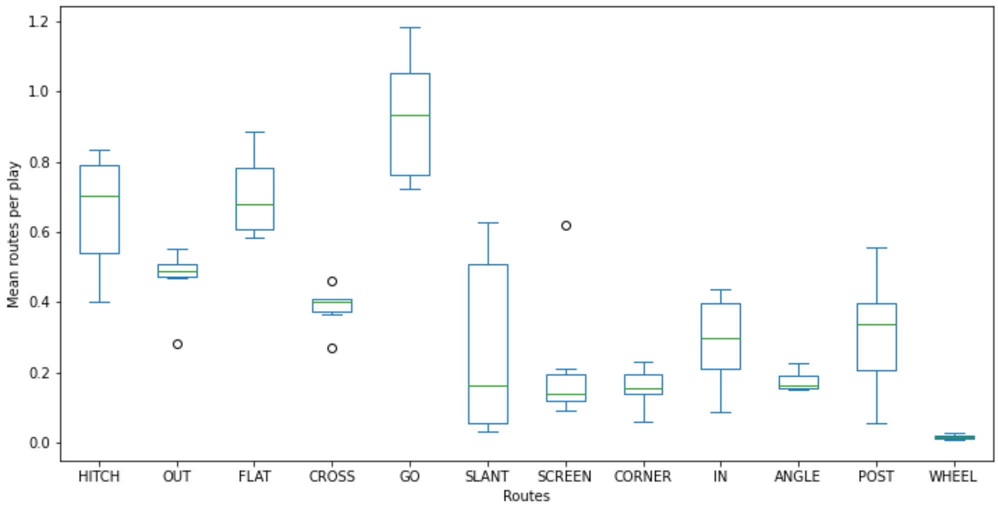

Figure 15: Side-by-side boxplots showing the distribution of various routes across the six groups of offensive plays

Unfortunately, the supplied data was missing route labels for approximately one third of plays, although GPS tracking data otherwise appeared correct. As the route labels appear to have some significance for interpretation of the clustering results it was next investigated whether these routes could be labelled through machine learning techniques.

Use of ML methods for determining routes from GPS data

Figure 16 provides an example of a basic passing tree. Although a given route can be run with a high degree of variation depending on the offensive player's physical and mental abilities as well as the defence being run against it, all routes retain a few key features, the most important being the number, depth and angle of the breaks (Kinney, 2019). Generally, an offensive player will want to maximise their speed coming out of the final break to provide the greatest opportunity to separate from the defender. In principle the sequence of a player's position and speed at each point should provide a sufficient approximation for training a model to identify the unlabelled routes.

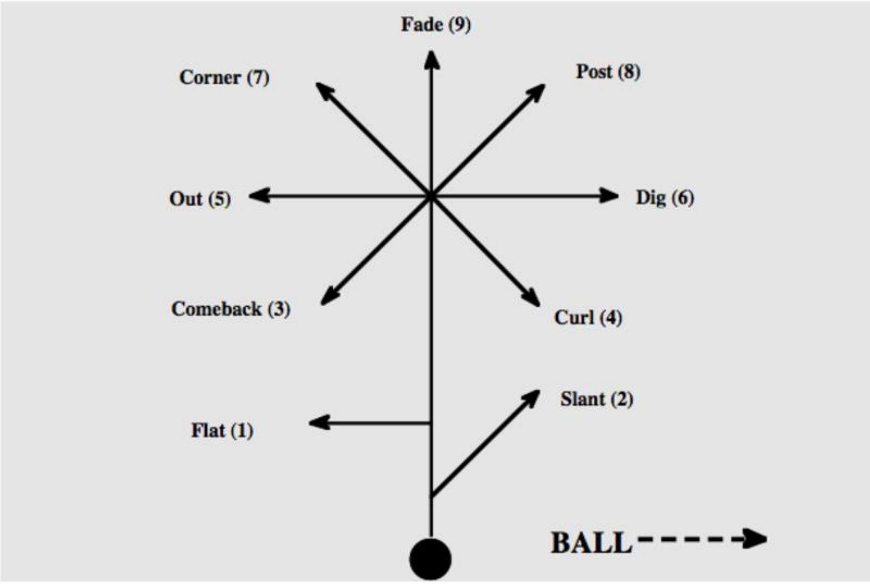

Figure 16: Basic NFL passing route tree (https://bleacherreport.com/articles/2016841-nfl-101-breaking-down-the-basics-of-the-route-tree)

The F1 score from Scikit-Learn's Classification Report is a more useful assessment tool for model reliability than simple accuracy as it takes into account both precision (the proportion of routes correctly predicted) and recall (the proportion of route predictions that are correct) (www.muthu.co/understanding-the-classification-report-in-sklearn).

Initial investigation using a small portion of the dataset indicated that techniques such as the Support Vector Machine (SVM), Random Forest and Multiple Layer Perceptron would be able to achieve about an F1 accuracy score of around 50%, which is much better than random guessing (the most common route, 'GO', has an incidence of below 20%) but not sufficiently accurate to provide a reliable value. Further development of these models would require extensive running times as they are purely CPU based, and only the RF and MLP models in Scikit-Learn can make use of multiple processors.

More extensive development, utilising the full valid dataset (65,364 routes), was completed using Pytorch's CUDA enabled MultiClass Neural Network. After tuning the model's various hyperparameters (learning rate, number of epochs, number of hidden units, number and type of layers) an optimised model was developed. The final model was able to achieve an F1 score of 65%. Figures 17 and 18 show the classification report summary and confusion matrix respectively. This level of performance compares well with more complicated models, developed by Chu et al (2019), Sterken (2019), Chu et al (2020) and Kinney (2019), that were able to achieve around 70% accuracy.

This implies that the speed trace of a player's movement accounts for a large degree of a route's identifying features but that there may be other features, such as orientation, acceleration and direction, that could be added to the training data to improve the predictive ability. Additionally, information from the defence and other offensive players, in particular the movement vectors of nearby players and the distance between those players and the route runner, could help explain some of the variability due to offensive players frequently being unable to run a 'perfect' route due to the impact of these other players.

```
Classification report for classifier MultiClassModel(
  (linear_layer_stack): Sequential(
    (0): Linear(in_features=9180, out_features=256,
bias=True)
    (1): ReLU()
    (2): Linear(in_features=256, out_features=256, bias=True)
    (3): ReLU()
    (4): Linear(in_features=256, out_features=256, bias=True)
    (5): ReLU()
    (6): Linear(in_features=256, out_features=12, bias=True)
  )
):
              precision    recall  f1-score   support

           0       0.62      0.69      0.65      2124
           1       0.75      0.68      0.71      1378
           2       0.76      0.79      0.77      1837
           3       0.69      0.68      0.69      1230
           4       0.63      0.73      0.68      2493
           5       0.68      0.45      0.54       745
           6       0.45      0.59      0.51       478
           7       0.60      0.40      0.48       493
           8       0.68      0.57      0.62       961
           9       0.48      0.38      0.42       452
          10       0.57      0.53      0.55       828
          11       0.40      0.26      0.31        54

    accuracy                           0.65     13073
   macro avg       0.61      0.56      0.58     13073
weighted avg       0.65      0.65      0.65     13073
```

Figure 17: Classification report for the MCNN based route predictor

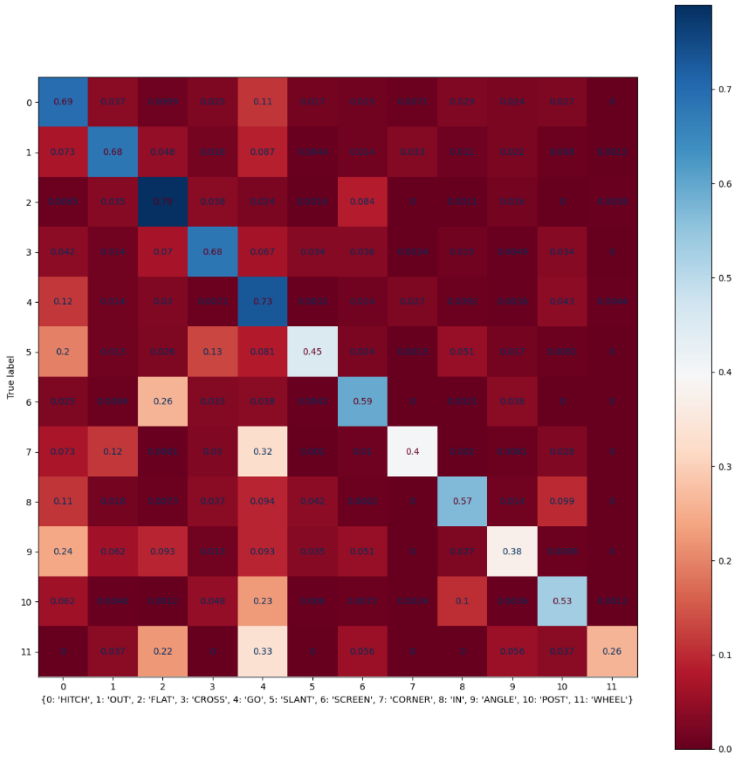

Figure 18: Confusion matrix for the prediction results of the MCNN route identifying model

Conclusion

This work presents an exploration of collective states in NFL football games and how teams transition between states across the duration of plays, as well as whether the associated order parameters can be used to identify classes of offensive plays being run. Further to this an investigation was conducted into the ability to identify routes run by offensive players using their GPS tracking data.

Unlike most field sports, which have a continuous ebb and flow of play, NFL football is broken into discrete plays of very short duration. Comparing against Welch’s (2021) soccer analysis we found some similarities between American football and soccer, in particular the relationship between polarisation and group velocity, but many differences. One major difference is in the predominance of the transitional state in American football, where teams only achieve a polar state once the ball has arrived at its target and players can converge on a known location. Related to this is the high level of disorder in the system, due to both teams trying to disguise what they are doing and the offenses ability to send the ball to the receiving player with the greatest advantage over the defence (Brown, 2017).

Classification of offensive plays can be aided through use of these order parameters, the most important for this function being the field area covered, entropy and mean group speed. Different combinations of passing routes are what define an offensive play call and, excluding some minor redirection by defensive players, the positioning of the offensive players. We found that the field area occupied acts as a proxy for families of play calls which aim to attack various parts of the field. There are significant differences in the types of routes run in each cluster group, but further work should be done to aid interpretation through a more context relevant label. This would require manual sampling of animations and play diagrams from each cluster to ascertain the dominant concepts in each group and provide a meaningful label.

A 7-layer deep-learning neural network was able to achieve 65% accuracy when attempting to identify pass routes based on the player's position and speed recorded in the GPS tracking data. The highest incidence rate within the 12 labelled routes in the dataset was less than 20%, while the lowest incidence was around 0.5%. Although a 65% accuracy is slightly worse than methods proposed by other researchers (Chu et al, 2019; Sterken, 2019; Chu et al, 2020; Kinney, 2019) it is significantly higher than random guessing, needs comparatively little pre-processing of the data and could be a useful aide for automating the detection of routes run in a given play, particularly in a real-time game scenario where processing speed may be more critical than absolute accuracy.

The GPS tracking data was limited to a 10Hz resolution which, given the short duration of plays and their sub-phases, provided limited data in certain windows for assessing order parameters through transition phases. Higher resolution data would be useful for further work to investigate differences in the order states and transitions within the play-type clusters. Similarly improving the ability to identify routes from unlabelled data through more sophisticated neural network techniques, higher resolution GPS data and possibly even automated video analysis, would provide a larger dataset to aggregate over and provide a more robust analysis.

References

Brown, C. (2017, November). *The Science of the Post: Going Deep with "Mills"*. https://www.smartfootball.com/offense/the-science-of-the-post-going-deep-with-mills

Chu, D., Reyers, M., Wu, L. and Thomson, J. (2019). Routes to success. *NFL Big Data* Bowl. URL https://danichusfu.github.io/files/Big_Data_Bowl.pdf

Chu, D., Reyers, M., Thomson, J and Wu, L. (2020). Route identification in the National Football League. *Journal of Quantitative Analysis in Sports*. 16(2), 121-132.

Deshpande, S. K. and Evans, K. (2019). Expected hypothetical completion probability. *arXiv*. https://arxiv.org/abs/1910.12337

Dutta, R., Yurko, R. and Ventura, S. (2020). Unsupervised methods for identifying pass coverage among defensive backs with NFL player tracking data. *arXiv*. https://doi.org/10.48550/arxiv.1906.11373

Herold, M., Goes, F., Nopp, S., Bauer, P., Thompson, C. and Meyer, T. (2019). Machine learning in men's professional football: Current applications and future directions for improving attacking play. *International Journal of Sports Science & Coaching*. 14(6), 798-817.

Kinney, M. (2020). Template matching route classification. *Journal of Quantitative Analysis in Sports*. 16(2), 133-142.

Macdonald, B. (2020). Recreating the game: Using player tracking data to analyze dynamics in basketball and football. *Harvard Data Science Review*. 24. https://doi.org/10.1162/99608f92.6e25c7ee

Schaerf, T. M., Dillingham, P. and Ward, A. J. W. (2017). The effects of external cues on individual and collective behavior of shoaling fish. *Science Advances*. 3(6). https://doi.org/10.1126/sciadv.1603201

Sterken, N. (2019). RouteNet: a convolutional neural network for classifying routes. *NFL Big Data Bowl*. URL https://operations.nfl.com/media/3671/big-data-bowl-sterken.pdf

Tunstrøm, K., Katz, Y., Ioannou, C. C., Huepe, C. Lutz, M. J. and Couzin, I. D. (2013). Collective states, multistability and transitional behaviour in schooling fish. *PLOS Computational Biology*. 9(2). http://doi.org/10.1371/journal.pcbi.1002915

Haar, A. V. (2019). Exploratory data analysis of passing plays using NFL tracking data. *NFL Big Data Bowl*. URL https://operations.nfl.com/media/3672/big-data-bowl-vonder-haar.pdf

Welch, M., Schaerf, T. M. and Murphy, A. (2021). Collective states and their transitions in football. *PLoS ONE*, 16(5), https://doi.org/10.1371/journal.pone.0251970

Yurko, R., Matano, F., Richardson, L. F., Granered, N., Pospisil, T., Pelechrinis, K. and Ventura, S. (2019a). *arXiv*. https://doi.org/10.48550/arXiv.1906.01760

Yurko, R., Ventura, S. and Horowitz, M. (2019b). nflWAR: a reproducible method for offensive player evaluation in football. *Journal of Quantitative Analysis in Sports*. 15(3). https://doi.org/10.1515/jqas-2018-0010