MIT SLOAN SPORTS ANALYTICS CONFERENCE
MARCH 3 - 4, 2017 HYNES CONVENTION CENTER

Flow Motifs in Soccer: What can passing behavior tell us?

Joris Bekkers
Eindhoven University of Technology
The Netherlands
info@unravelsports.com

Shaunak Dabadghao
Eindhoven University of Technology
The Netherlands

In soccer, both individual and team performance is crucial to win matches. Passing is the backbone of the game and forms the basis of important decisions made by managers and owners; such as buying players, picking offensive or defensive strategies or even defining a style of play. These decisions can be supported by analyzing how a player performs and how his style affects team performance. The flow of a player or a team can be studied by finding unique passing motifs from the patterns in the subgraphs of a possession-passing network of soccer games. These flow motifs can be used to analyze individual players and teams based on the diversity and frequency of their involvement in different motifs. Building on the flow motif analyses, we introduce an expected goals model to measure the effectiveness of each style of play. We also make use of a novel way to represent motif data that is easy to understand and can be used to compare players, teams and seasons. Further, we exploit the relationship between play style and the pass probability matrix to support our analysis. Our data set has the last 4 seasons of 6 big European leagues with 8219 matches, 3532 unique players and 155 unique teams. We will use flow motifs to analyze different events, such as for example the transfer of Claudio Bravo to Pep Guardiola's Manchester City, who Jean Seri is and why he must be an elite midfielder and the difference in attacking style between Lionel Messi and Cristiano Ronaldo. Ultimately, an analysis of Post-Fàbregas Arsenal is conducted wherein different techniques are combined to analyze the impact the acquisition of Mesut Özil and Alexis Sánchez had on the strategies implemented at Arsenal.

1. Introduction

In today's world we see an increasing trend in the availability of data. This is particularly true for sports, and soccer is no exception. Firms such as OptaSports gather all sorts of data on matches played all over the planet. Gone are the old days of searching for statistics in the Sunday newspaper. Now a quick search on FourFourTwo.com or Squawka.com can provide you with in-depth information on any player – ranging from minutes played on the field to passes made, conversion rates, percentage of total possession held, fouls committed, and goals attempted and scored. Managers and fans have been using this data for many years to make strategies and comparisons between players. However, we are now also able to extract a second by second description of play from these websites. This opens doors to find a lot of hidden information on player and team behavior by devising new metrics and studying the data using models from network theory. We use the "network motif" concept shown in [Milo et al., 2002] to study patterns in the data and illustrate its use in sports analytics.

There are two approaches to studying passing behavior and team style. One is a static approach where analysis is done on aggregated data, while the other approach involves real time tracking of all players and the ball during play. Both approaches have their advantages and challenges and are more, or less, suitable for a specific sport. Spatio-temporal analysis has been studied extensively for the NBA. We refer the reader to the works of [Goldsberry, 2012, Shortridge et al., 2014]. In [Bialkowski et al., 2014] the authors apply this methodology to soccer and demonstrate how to accurately detect and visualize formations, as well as analyze individual player behavior.

Our focus is on the static analysis, where the players are not tracked unless they are involved in a ball event. The ball is also not tracked per se, but its position can be interpolated with the available data. A few articles have investigated this passing-possession network that the data allows us to construct. In [Gyarmati and Hefeeda, 2015], the authors estimate the maximal speed of soccer players by knowing their position on the field between two ball events (such as receiving a pass and then making a pass). In [Clemente et al., 2015a, Clemente et al., 2015b] the authors create an adjacency matrix with the passes made between the players with which they study network metrics like centrality to characterize the team, importance of players such as midfielders and study the style differences between two halves of the game. To study team passing behavior, [Gyarmati et al., 2014] use the motif analysis from [Milo et al., 2002] to obtain passing styles during the 2012/13 season for a few European leagues. A similar work by [Peña and Navarro, 2015] analyzes individual players on their styles in the English Premier League and the Spanish La Liga and show that Xavi has a unique style. In [Peña and Touchette, 2012] they analyze the motifs for passing behavior of countries in the 2010 FIFA World Cup. In [Peña, 2014], they use the possession data to show that a finite state Markov process is very accurate in approximating the distribution of passing sequences and chances of taking shots for English Premier League teams for the 2012/13 season.

In this research we will utilize the idea of flow motifs to find unique styles in both regular passing and the final attacking passes prior to a goal attempt of both teams and players. This is done by differentiating between possession and goal attempt flow motifs. Subsequently, these results are combined with multiple other techniques such as a coordinate based expected goals model adapted from [Macdonald, 2012], transition matrices of passes between players, and clustering techniques such as the mean shift algorithm from [Comaniciu and Meer, 2002] and the simple Euclidean distance. We analyze different events in six of the biggest European soccer leagues and illustrate that their implementation yields useful insight that might help teams and managers in buying or selling players, analyzing opponents or defining styles of play. We will apply these metrics to analyze the departure of Frank de Boer from Ajax to Inter Milan, the difference in flow motifs used between home and away games, the transfer of Claudio Bravo to Pep Guardiola's Manchester City, who Jean Seri is and why he must be an elite midfielder, the difference in attacking style between Lionel Messi and Cristiano Ronaldo and how Toni Kroos' play style changed over the past four years. Furthermore we try to answer the question posed by Lopez Peña "*Who can replace Xavi?*" [Peña and Navarro, 2015] and, ultimately, an analysis of Post-Fàbregas Arsenal is conducted wherein all different techniques are combined to analyze the impact the acquisition of Mesut Özil and Alexis Sánchez had on the strategies implemented at Arsenal from 2012/2013 onwards. Furthermore, in this analysis we look ahead to the 2016/2017 season to find possible strategic improvements, and hypothetical player and manager replacements based on individual or team flow motif styles.

2. Data

Data was obtained using a custom Python web crawler from www.squawka.com. It covers four seasons (2012/2013, 2013/2014, 2014/2015, 2015/2016), six different leagues (Dutch Eredivisie[1], English Premier League, Spanish Primera Division, Italian Serie A, French Ligue 1 and German Bundesliga) and 8219 matches[2]. This ultimately yields a vast data set containing 7412 players (of which 3532 are unique players) and 466 teams (of which 155 are unique) and their total time played per season.

The data set consists of chronologicalized vectors with the parameters: league, season, play type (pass, goal attempt, cross, tackle, clearance or interception), result of the play type (failed, completed or foul), team name, player name, total seconds expired since the start of the match, and the coordinates at which the play occurred on a (0,100) by (0,100) two-dimensional plain.

Due to time stamp structuring in the 2012/13 and 2013/14 season data points recorded during injury time show up as either 45:00 or 90:00 thus making it impossible to chronologicalize them. Therefor these data points, 1.5% and 5% of first half and second half data respectively, have been removed for these two seasons. Furthermore, failed tackles have been removed from the data set since they have no apparent influence on the flow of the game.

In order to reduce outliers when comparing individual players, those who played less than 900 minutes (the equivalent of approximately 10 matches) in a given season were omitted. In [Peña and Touchette, 2012], the authors proposed 19 matches, but we expect players that under perform for 19 matches during a season to be dropped from any squad, thus creating an intrinsic bias towards well preforming players. Moreover, a player that changes teams during the winter break would be dropped from the analysis with a threshold of 19 matches since it's nearly impossible to play 19 matches for a team in the first half of the season and subsequently play 19 matches for another team after the winter break. No such data treatment is made while analyzing team motifs; all players are considered irrespective of how much time they play in the season.

3. Methodology

3.1. Flow Motifs

Flow motifs, as shown in [Milo et al., 2002], are building blocks of the passing behaviour of teams. We differentiate between two flow motif types for both players (P) and teams (T):

- **Possession Motifs (PMs):** a sequence of at least 3 passes a team/player creates that does not lead to a goal attempt.
- **Expected Goal Motifs (xGMs):** a sequence of at least one pass that leads to a goal scoring opportunity with a certain expectation of being converted.

[1] No data available for the Eredivisie in ´12/13.

[2] Data for Herta BSC vs Frankfurt (Februari 3 2016), Herta BSC vs Hannover '96 (April 8 2016) and Herta BSC vs SV Darmstadt (July 5 2016) is missing.

We consider up to 3 passes as part of a motif[3] as long as they were made within 5 seconds individually. Any transition with an interval time greater than this upper bound is not considered to be a part of any motif. Furthermore, a passing sequence is terminated when a game ends at either half, a foul is committed, the ball goes out of play, or when an opposing team's player disrupts the flow by tackling, intercepting, passing or clearing away the ball.

The difference between the two motif types is illustrated in Figure 1 and Figure 2. For the expected goal motifs, the goal attempt is the definitive endpoint; the start is up to 3 passes before the attempt. In Figure 1 we show the ABACG motif. The motif would have been counted as BACG if the pass from A to B took longer than 5 seconds. The possession motifs are obtained by looking at passes during the whole game. Here we consider 3 passes to make a motif. From Figure 2, we can see the ABAB, BABC and ABCD motifs occur in the pass sequence. All motif analysis shown in the rest of the article is the average per 90 minutes off play.

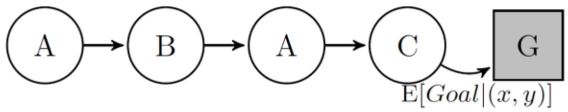

Figure 1: ABACG Expected Goal Motif

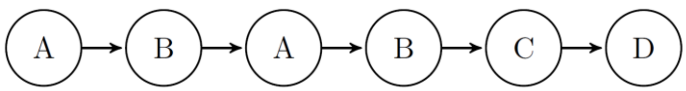

Figure 2: ABAB, BABC, ABCD Motifs

When looking at either Possession Motifs (PM) or Expected Goal Motifs (xGM), A is the player under consideration while B, C and D are other players on the same team that are involved in the motif. In Expected Goal Motifs (xGMs) a goal attempt (with a given expected value) is denoted by G.

Each player and team motif use is represented by the motif intensity vector:

$$v_j = \langle f_1, \dots, f_N \rangle \tag{1}$$

3.2. Expected Goals

Every xGM is valued by means of an expected goal (xG) model solely based of the coordinates of the final attempt. To obtain the expected value of a goal attempt the coordinates of the set of 204984 goal attempts are divided over a grid of 12 by 20 tiles (creating tiles of approximately 5.3 by 5.7 meters[4]). Subsequently, to establish the expected value of a passing sequence, each motif is weighted by the percentage of converted goal opportunities from within that tile.

3.3. Radar Graph

To be able to compare players or teams visually, we devised a novel way of representing their involvement in motifs, i.e. their style, by means of a radar graph. Players and teams' involvement in particular motif is compared against the maximum value for that motif from the whole data set. This makes it possible to compare players and teams across different seasons and leagues. This maximum value can also be specific to the season and league, however for the purpose of this article we have used the maximum value from the entire data set (unless stated otherwise).

---

[3] A maximum of four passes has 48 more combinations for players, thus making the problem too large.

[4] Given a median pitch of 105 by 68 meters [UEFA, 2016]

An example of a radar graph is shown in Figure 3. Each axis shown in the figure represents a motif. The extent of a players' or teams' use of this motif is a percentage of the maximum value from the data set. In this specific figure we show Paulo Dybala's expected goal motif performance in the '15/16 season at Juventus. The figure is constructed as follows: in the top right quadrant we depict all motifs ending with *-AG* indicating that Dybala was the final shot taker. Continuing counter clockwise, we have three motifs (*ABAG*, *BABAG*, and *BACAG*) that indicate a one-two combination with Dybala as the main protagonist. As we keep going counter clockwise, we have the assist motifs (*-AXG*) where Paulo Dybala was not the final shot taker, but played an assisting role. Again we can identify the one-two combinations *BABG*, *BCACG*, *ABABG*, and *ABACG*. In the latter two Dybala is the person with two touches in the motif. After the assist motifs the second assist motifs are shown (-*AXYG*), followed by the third assist motifs (*ABCBG*, and *ABCDG*). As it is seen in the graph, Dybala uses the *ABG* motif just more than half as many times as the player with the maximum use of that motif. We can see at a glance that Dybala was hardly involved in motifs with one-two combinations when creating goal attempts, but is well rounded in the use of most other motifs.

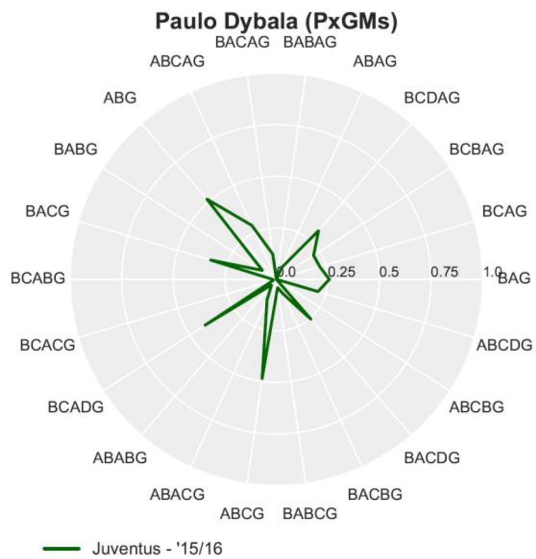

Figure 3: Dybala's Expected Goal Motifs '15/16
(player in question denoted as A)

3.4. Clustering

To find unique players or teams we make use of the unsupervised machine learning algorithm mean shift. This hierarchical clustering method is a centroid based algorithm which works by updating candidates for centroids to be the mean of the points within a certain radius (bandwidth) around the centroid [Fukunaga and Hostetler, 1975]. Every data point (v_j) is considered a centroid, until the mean of the data points within the bandwidth is stationary.

The k-means clustering algorithm does not automatically compute the optimal number of clusters given a certain bandwidth, and is biased towards equally sized clusters (making unique playing styles less apparent) given it's adherence to Euclidean distances, see [Georgescu et al., 2003]. However, we use the Euclidean distance to find the k-nearest neighbours with the most similar motif tendencies to a given node within the vast data set.

3.5. Pass Probability Matrix

To visually identify players who pass to each other more or less frequently during the course a of season a hollow probability matrix $P_{i,j}$ is created from all completed passes made by each player i to every other player j. An example for Feyenoord is displayed in Table 1. From this table we can see that Michiel Kramer made 21.2% of all his successful passes during the '15/16 season to Dirk Kuyt.

	Kramer	Elia	Kuyt	Rest
Kramer	-	7.0%	21.2%	...
Elia	3.0%	-	8.5%	...
Kuyt	4.3%	5.0%	-	...
Rest	...	...	...	...

Table 1: Part of Feyenoord's Pass Probability Matrix ('15/16)

4. Analysis of Team Motifs

In Table 2 we identify 5 different team passing motifs (TPM) and 8 different team expected goal motifs (TxGM). Team motifs are obtained from the sum of the motifs per 90 minutes of all players that belong to this team during a given season. Since substitutes are an integral part of a team, players who played less than 900 minutes will not be excluded from this analysis.

Passes	TxGMs	TPM
1	ABG	
2	ABAG, ABCG	
5	ABCDG, ABABG, ABACG, ACBAG, BABCG	ABCD, ABAB, ABAC, ACBA, ABCB

Table 2: All Possible Team Motifs

4.1. Uniqueness in Team Possession Motifs (TPMs)

To search for unique teams, we analyze their style by looking at how they use the five different motifs by employing a simple scatter plot. One of these five plots is shown in Figure 4. It represents the plot for the ABAC motif for all teams showing their intensity of use (per match) versus the popularity of use within the team (the percentage of time it is used). We find that some teams are consistent outliers in all five plots at either end of the spectrum, implying the presence of unique styles. To find these unique team motif tendencies, we can cluster the teams with the mean shift clustering algorithm using the vector of motif intensities per team as input to form these clusters (Formula 1). In Table 3 the results of the mean shift clustering are shown, indicating four unique clusters.

From these clusters we derive that Paris Saint-Germain has a unique passing style whereas the passing styles at FC Barcelona and Bayern Munich are closely related. The latter is not surprising considering Pep Guardiola integrated his specific style at Barcelona when he coached them between 2008 and 2012, and then at Bayern Munich after he joined there as a coach in 2013. In the third cluster with 29 teams we see teams that utilize intelligent possession based strategies.

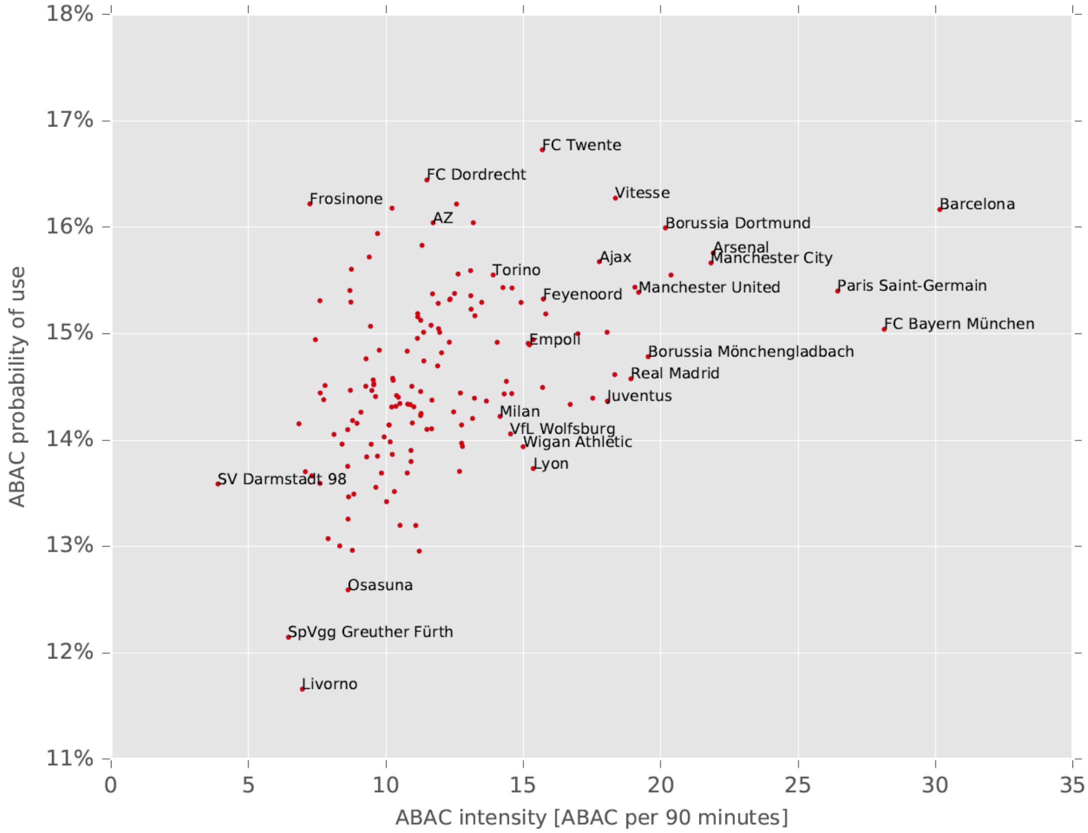

Figure 4: Use of possession motif ABAC as a percentage of all motifs used by a team, against the percentage of time it is used per 90 minutes. This figure includes all 155 unique teams.

Size	Cluster Member(s)
1	Paris Saint-Germain
2	FC Barcelona, FC Bayern Munich
29	Ajax, Arsenal, Borussia Dortmund, Borussia M'gladbach, Celta de Vigo, Chelsea, Empoli, Everton, Feyenoord, Internazionale, Juventus, Las Palmas, Lille, Liverpool, Lyon, Man. City, Man. United, Milan, Napoli, Nice, Real Madrid, AS Roma, Southampton, Swansea City, Tottenham Hotspur, VfL Wolfsburg, Vitesse, Wigan Athletic
123	All other teams

Table 3: Mean Shift Clustered by TPMs (estimated bandwidth)

Even the more suspect teams in this cluster seem to fit in rather well. Empoli managed by Maurizio Sarri in 2014/15, who is now head coach at Napoli; Nice managed by Claude Puel during the past four seasons, head coach of Southampton since the start of 2016/17 season; Vitesse coached by Peter Bosz during 2013-2016 and currently head coach at Ajax; Wigan Athletic, Swansea City and

Everton all coached by Roberto Martinez. Las Palmas, coached by Quique Setién, who finished 11th in their first year back in the Spanish top tier since 2002, also shows up in this cluster. This indicates that Quique Setién might be a suitable coach for other teams looking to play possession based soccer.

The main difference between the first three clusters and the remaining 121 teams seems their apparent lower probability of executing ABCD and ABCA motifs, and thus higher probabilities of utilizing ABAB, ABCB, and ABAC. Furthermore, all three clusters have overall higher average motif intensity per match.

4.2. Uniqueness in Team Expected Goal Motifs (TxGMs)

A similar analysis can be made for the team expected goal motifs (TxGMs). Applying mean shift on the team intensity vectors obtains three clusters (depicted in Table 4). The two unique cluster groups are mainly differentiated by their xGs created from the ABG, and ABCG. The fact that teams in the first two clusters create on average significantly more xGs than "all other teams" sets them both apart from the third cluster.

Size	Cluster Member(s)
9	Arsenal, Barcelona, Chelsea, Juventus, Manchester City, Napoli, Roma, Southampton, VfL Wolfsburg
12	Ajax, Bayer Leverkusen, Borussia Dortmund, Bayern Munich, Schalke 04, Feyenoord, Liverpool, Man. City, PSG, PSV, Real Madrid, Vitesse
134	All other teams

Table 4: Mean Shift Clustered by TxGMs (estimated bandwidth)

At the end of the 2015/2016 season Frank de Boer left Ajax after 5.5 seasons to become head coach at Internazionale. Ajax subsequently replaced him with former Vitesse manager (June 2013 – January 2016) Peter Bosz. Both these trainer changes can be explained by means of the TPMs and TxGMs employed by these three teams during the past four seasons. Ajax is the 5th nearest neighbour to Internazionale on TPMs by Euclidean distance. Despite this Inter does not rank amongst either of the two TxGM outlier clusters, although Ajax does (see Table 3 and Table 4). This indicates that Inter is lacking high potential goal attempts, and they appointed Frank de Boer to fix this. The fact that Peter Bosz replaced Frank de Boer is not surprising considering their styles are almost identical. Vitesse's TPMs and TxGMs are 2nd and 5th closest respectively to Ajax by Euclidean distance.

4.3. Team Motifs in Home and Away Games

Home advantage is a big part of soccer, with teams winning about 64% of their points in home games [Pollard, 1986]. This raises the question whether teams employ different styles or just increase the motif intensity during home games compared to their away games.

Separating the team possession motifs into home and away games for all teams and leagues shows that on average a team creates 7.6% more possession motifs, and 31.2% more expected goal motifs during home games. FC Twente (-9.0%), Borussia Monchengladbach (-7.8%) and Rayo Vallecano (-5.1%), consistently preform less possession motifs (TPMs) during their home games compared to their away games across three of the last four seasons recorded. SV Darmstad (-15.1%) and Siena (-12.1%) create significantly less expected goal motifs (TxGMs) when playing at home as compared to away. Pescara (+106.9%), Ajaccio (+101.9%) and Sporting de Gijon (+92.9%) have the biggest positive difference when comparing home versus away games on TxGMs.

Three time consecutive UEFA League winner Sevilla stands out when comparing home and away team possession motifs. In 2012/13 they created 22.5% more possession motifs during their home games. After Unai Emery took over the role as head coach in January 2013 he increased this number to absolutely extraordinary heights and consistency in the following three seasons. Sevilla respectively created 60.4%, 42.4%, and 49.3% more possession motifs during their home games when compared to their away games. Table 5 shows the teams closest to Sevilla when playing at home or away using Euclidean distance. The fact that Sevilla has a similar away style to bottom-of-the-league teams like Sunderland, Chievo and Cordoba suggests that Sevilla is using completely different passing strategies during games away from their own stadium.

Home	**Away**
Sevilla	Sevilla
Athletic Bilbao	Sunderland
Rennes	Chievo
Bordeaux	AZ Alkmaar
Valencia	Cordoba

Table 5: Four teams with styles closest to Sevilla during their respective home and away games

4.4. The ABABG Motif

It seems that almost every week a long range screamer is scored, and subsequently named to be a contender for goal of the season. Other goals such as Jack Wilshere's 1-0 against Norwich City in 2013 [Wilshire and Giroud, 2013] and Messi's 3-0 against Real Sociedad in 2010 [Messi and Alves, 2010] are also considered some of the most beautiful goals ever scored. Both these goals utilize the *ABAB* motif (between Giroud and Wilshere, and Dani Alves and Messi respectively) to create the final shot on goal. Why are these goals considered to be so beautiful and why do they not dominate the goal of the season lists? In Table 6 we see that of all goal attempts in the data set only 0.90% are created from an *ABAB* motif.

Furthermore we see that it is also more difficult to create a valuable goal attempt from the *ABAB* motif; every goal attempt from this motif results in only 0.064 expected goals, almost half the value of an *AB* goal attempt.

So, in the case of the *ABAB*-goals the beauty lies in the rarity of the event, and the inherent difficulty of turning the motif into an actual goal. Due to this they hardly ever appear in the goal of the season lists, but when they are scored they are surely considered.

Motif	Frequency	Mean Goal Attempt Value
ABG	39.2%	0.123 xG
ABCDG	25.2%	0.101 xG
ABCG	18.5%	0.109 xG
ABACG	6.0%	0.109 xG
ABCBG	4.6%	0.084 xG
ABCAG	3.4%	0.098 xG
ABAG	2.3%	0.081 xG
ABABG	0.9%	0.064 xG

Table 6: Goal Motif Frequency and
Mean Goal Attempt Value in Expected Goals

5. Analysis of Player Motifs

Like teams, we can also differentiate between players by analyzing their styles. In Table 7 all possible player possession motifs (PPMs) and player expected goal motifs (PxGMs) are shown. We can construct a scatter plot similar to what we saw in the previous section; Figure 5 shows all the 3532 unique players in our data set with respect to their involvement in the ABCD motif. We differentiate the players by their positions (goalkeeper, defender, midfielder or forward) and observe a clear link between player position and how they use a motif.

The player possession motifs (PPMs) can be used to cluster similar players, compare players to one-another and scout potential replacements that employ a similar style.

Passes	PxGMs	PPMs
1	ABG, BAG	
2	ABAG, ABCG, BABG, BACG, BCAG	
3	ABCDG, ABABG, ABACG, ABCAG, ABCBG, BABCG, BACDG, BACAG, BABAG, BACBG, BCADG, BCACG, BCDAG, BCBAG, BCABG	ABCD, ABAB, ABAC, ABCA, ABCB, BABC, BACD, BACA, BABA, BACB, BCAD, BCAC, BCDA, BCBA, BCAB

Table 7: All Possible Player Motifs

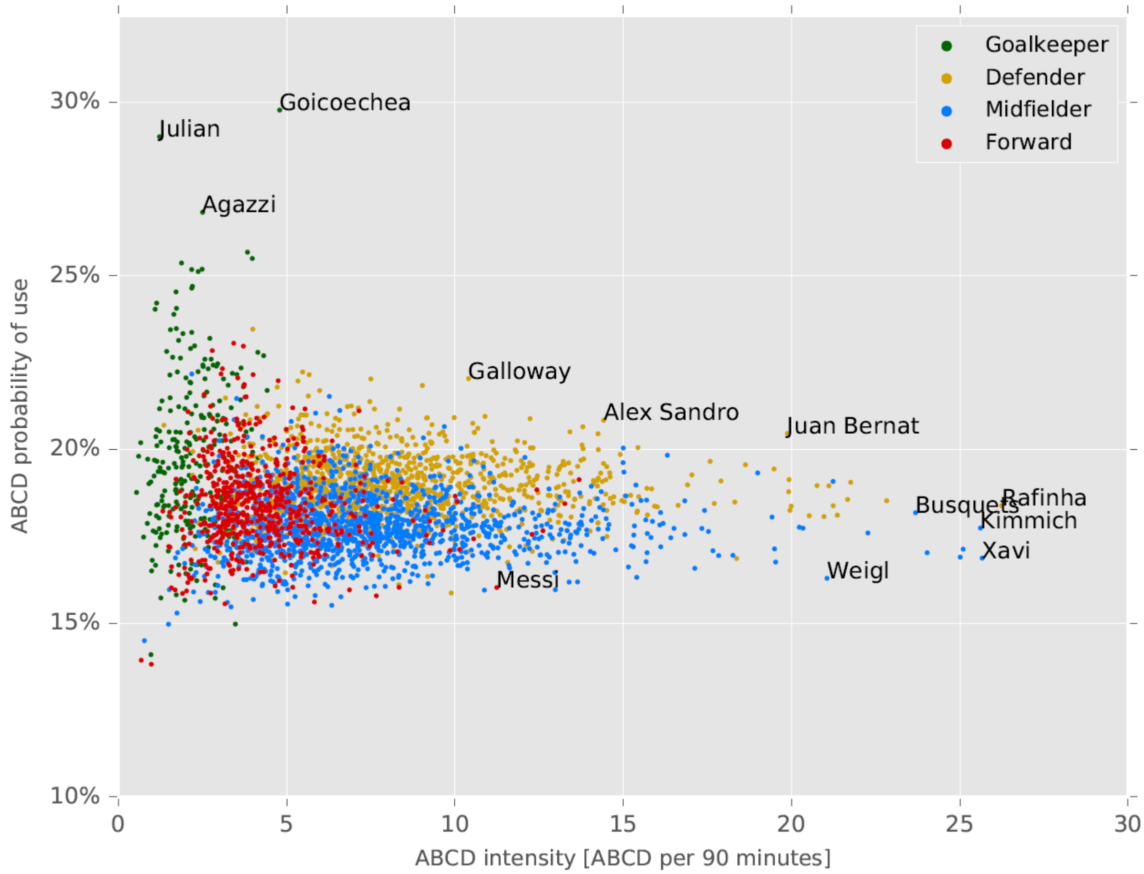

Figure 5: Involvement in possession motif ABCD as a percentage of all possible motifs, against the percentage of time it is used per 90 minutes. This figure includes all 3532 unique players.

5.1. Uniqueness in Player Possession Motifs (PPMs)

Applying the mean shift clustering algorithm with an estimated bandwidth identifies eight clusters. Clusters with less than 100 nodes are shown in Table 8 accompanied by a cluster classification for the multiple node clusters[5].

Size	Cluster Member(s)	Classification
1	Iniesta	
1	Rafinha (Bayern)	
1	Denswil	
3	Benatia, Busquets, Xabi Alonso	Central/Defensive
6	Kimmich, Weigl, Verratti, Thiago Alcantara, Thiago Motta, Xavi	Central Midfielders

Table 8: Mean Shift Clusters by PPMs (estimated bandwidth)

[5] Stefano Denswil, a central defender at Ajax during the 2012/13 season (before moving to Belgium side Club Brugge), is one of the outliers. In that season Denswil and other Ajax central defender Moisander were involved in an exceptional amount of possession motifs.

We also ran the mean shift clustering algorithm to give us 25 unique clusters. Table 9 shows all the small clusters (the next biggest cluster has 18 nodes) with the defining trait of the cluster. These classifications demonstrate the accuracy with which the mean shift algorithm is able to cluster players playing in similar positions.

Size	Cluster Member(s)	Classification
1	Xavi, Iniesta, Rafinha (Bayern), Kimmich, Verratti, Thiago Alcantara, Denswil, Thiago Motta, Ribery, Puyol, Seri, Dani Alves, Weigl	Individual Nodes
2	Ramsey, Pastore	Central Attacking Midfielders
2	Busquets, Xabi Alonso	Central Defensive Midfielders
4	Adriano, Alaba, Alba, Lahm	Full-Backs
4	Schweinsteiger, Matuidi, Fabregas, Y. Toure	Central Midfielders
5	van Buyten, Dante, Boateng, Arteta, Kroos	Central (Defence)
5	Vidal, Badstuber, Gundogan, Joringho, Strootman, Rabiot, Pjanic, Taddei	Central Midfielders (excl. Badstuber)

Table 9: Mean Shift 25 Clusters by PPMs (shown: size < 10)

Inspecting the players in Table 9, we see mostly (former) players from elite European teams such as FC Barcelona, Bayern Munich, Paris Saint-Germain, Real Madrid and Borussia Dortmund. An interesting player within this exclusive list of players with his own single node cluster is OGC Nice and Ivory Coast central midfielder Jean Seri. To see whether Seri is an outlier at the bottom or top end of the player spectrum we can find the nodes closest to him by Euclidean distance. The players closest to Seri are: Toni Kroos, Jérôme Boateng, Cesc Fàbregas and Bastian Schweinsteiger. This implies that Jean Seri must be a prolific central midfielder.

5.1.1. Joe Hart

Before the start of the 2016/2017 season, Pep Guardiola, after three seasons with Bayern Munich, was appointed manager at Manchester City. In Guardiola's unique strategies the goalkeeper is an integral part of the team when in possession of the ball. Therefore the goalkeeper is expected to be an exceptionally prolific passer. At Bayern Munich this role was executed by Manuel Neuer. In Figure 6 Neuer's PPM radar graph against all goal keepers in the data set is presented. Guardiola would expect his Manchester City goalkeeper Joe Hart to also be able to execute passing motifs frequently without error. However, from looking at his radar graph (Figure 7) it becomes evident that Hart would probably have a hard time executing this new role. He was subsequently loaned out to Italian side Torino. To find a suitable replacement for Joe Hart we look at the top 10 goalkeepers closest to an aggregate Manuel Neuer over four seasons by Euclidean distance, depicted in Table 10.

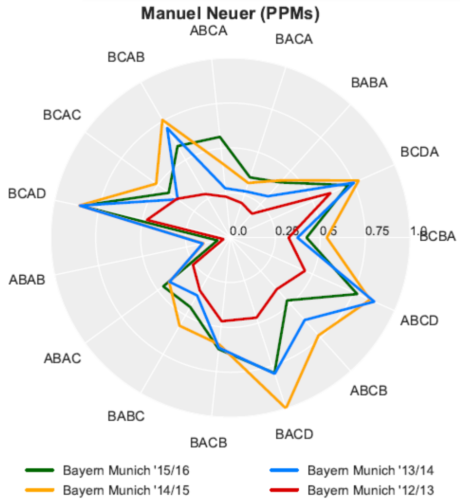

Figure 6: Manuel Neuer PPM style for all seasons; compared to all goalkeepers.

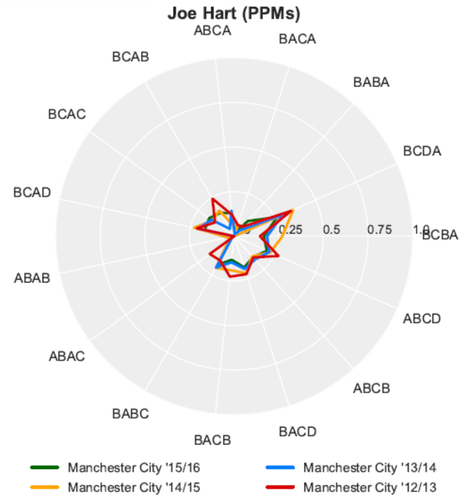

Figure 7: Joe Hart PPM style for all seasons; compared to all goalkeepers.

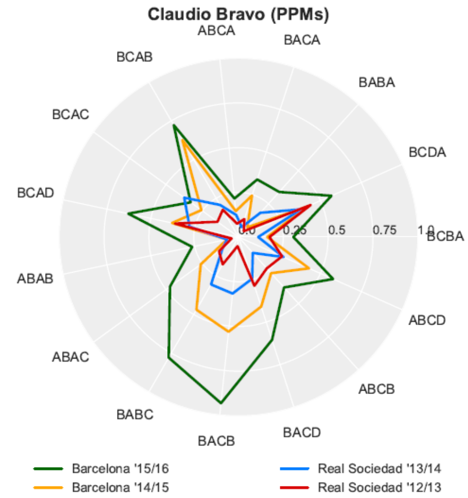

Figure 8: Claudio Bravo PPM style for all seasons; compared to all goalkeepers.

Season(s)	Goalkeeper	Club
All	Manuel Neuer	Bayern Munich
2013/14	Nick Marsman	FC Twente
2014/15	Yann Sommer	Borussia M'gladbach
2014/15	Ron-Robert Zieler	Hannover 96
2015/16	Claudio Bravo	FC Barcelona
2014/15	Nick Marsman	FC Twente
2013/14	Jasper Cillessen	Ajax
2015/16	Koen Casteels	VfL Wolfsburg
2015/16	Yann Sommer	Borussia M'gladbach
2012/13	Marc-Andre ter Stegen	Borussia Dortmund
2013/14	Piet Velthuizen	Vitesse

Table 10: Ten Goalkeepers Closest to Manuel Neuer on PPMs

Ultimately Manchester City bought Claudio Bravo (Figure 8) from FC Barcelona. Thereafter Marc-Andre ter Stegen was appointed first goalkeeper at FC Barcelona, and Jasper Cillessen was bought from Ajax to be ter Stegen's stand-in.

5.2. Uniqueness in Player Expected Goal Motifs (PxGMs)

Player Expected Goals Motifs (PxGMs) can be used to shed light on the mind of the individual, the way they personally shape their opportunities, create opportunities for their team mates and show how effective they are at finding the right spot for a goal attempt. Applying the mean shift clustering algorithm with an estimated bandwidth, we identify thirteen different PxGM clusters. Clusters with less than 20 nodes are shown in Table 11 accompanied by a cluster classification.

Size	Cluster Member(s)	Classification
2	Lewis Baker, Nicky Shorey	Central Attack
2	Jacob Mulenga, Slaon Privat	Center Forwards
4	Imbula, Weigl, Lanzini, Ibe	Defensive Midfielders (excl. Ibe)
9	C. Ronaldo, Adrian Ramos, Mitrovic, Diafra Sakho, Dzeko, Uche, Coda, Lewandowski, Necid	Center Forwards
16	Messi, Robben, Morata, Tevez, Sturridge, Lampard, Bale, Hernandez, Higuain, Luis Suarez, Benzema, Ibrahimovic, Depay, Vucinic, Muller	Wingers & Center Forwards

Tabel 11: Mean Shift Clustered by PxGMs (estimated bandwidth)

From Table 11 we derive that Messi and Ronaldo, the two best players in the world since 2008[6], have different goal scoring styles. In Figure 9 and Figure 10 the differences in expected goal scoring style between Messi and Ronaldo are easily seen. Messi is involved more in the build-up of his own chances utilizing *BACAG*, *ABAG*, *ABCAG* and *BABAG* to create most of his expected goals. Ronaldo is significantly less involved in his own chances prior to taking the final shot (*BCBAG* and *BAG*).

Furthermore we can see that Messi gave assists with a higher probability of being converted into goals in the 2014/15 and 2015/16 seasons (*BCACG*, *BCADG* and *BACG*). This is most likely the result of Barcelona acquiring Luiz Suarez in the summer of 2014. Suarez managed to outscore Messi during the 2015/16 season by 14 goals (40 against 26). It also becomes evident that Ronaldo and Real Madrid evolved from a team that would simply pass the ball to Ronaldo to let him shoot on goal (*BAG*) to a team that would attempt a build-up and then pass to Ronaldo as *BCDAG* and *BCBAG* have clearly increased.

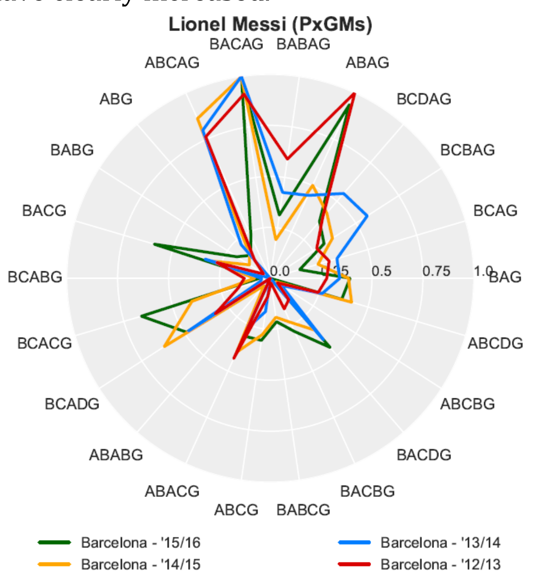

Figure 9: Messi PxGM style for all seasons

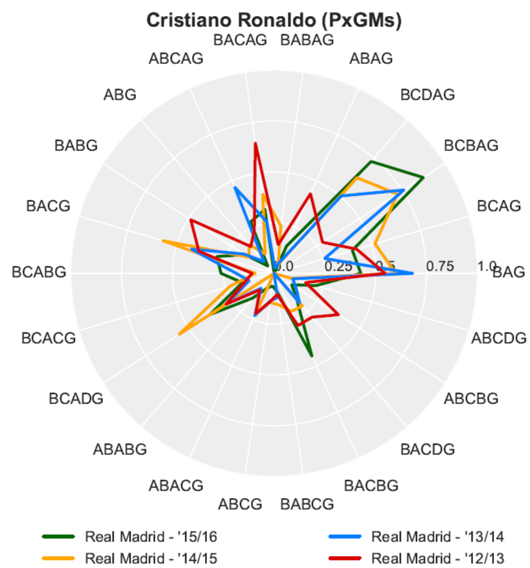

Figure 10: C. Ronaldo PxGM style for all seasons

[6] Either Ronaldo or Messi has won the FIFA Player of the Year award or its successor the FIFA Ballon d'Or since 2008.

PxGMs can be used to analyze players from every position (not only attackers). To showcase this we will take a closer look at Toni Kroos. Toni Kroos played for Bayern Munich from 2007 until 2014 (with a short loan period at Bayer Leverkusen during the early stages of his career) after which he moved to Real Madrid in the summer of 2014, for an estimated sum of €25 million. Looking at our data set this means that we have two seasons of data of him playing at Bayern (12/13 and 13/14) and two seasons of data playing at Real Madrid (14/15 and 15/16), making him an excellent subject for comparative analysis.

	Games	Goals	Assists	Pass Success %	Total xG Motifs per 90 mins	Total Goal Attempts per 90 mins
Bayern '12/13	24	6	8	90	0.86	6.42
Bayern '13/14	29	2	4	92	0.57	3.10
R. Madrid '14/15	36	2	7	92	0.48	2.62
R. Madrid '15/16	32	1	10	94	0.49	1.65

Table 12: Toni Kroos's Game Statistics per Season

From Toni Kroos's statistics in Table 12 we see a decline in goal per match ratio from 8 goals in 53 matches for Bayern, to just 3 goals in 68 matches for Real Madrid and an increase in passing accuracy and assists. Perhaps it is possible to paint a clearer picture of what happened with the help of the PxGMs. And perhaps we can explain why his stats changed so dramatically.

Table 12 also shows the total xGs he is involved in and the total average amount of goal attempt motifs he created per 90 minutes[7]. Kroos went from being involved in 0.856 xGs per 90 minutes to only being involved in just 0.57 xGs per 90 minutes, and from creating 6.42 goal attempts per 90 minutes to just creating 3.10 per 90 minutes after Bayern changed coaches between these two seasons. Even though Bayern as a team increased their expected goals per game from 1.59 to 1.73 and their total attempts created increased from 12.38/90 minutes to 13.99/90 minutes.

Figure 11 shows Kroos's attacking involvement drastically decreased after Guardiola took over at Bayern in 2013. In this season Bayern played primarily a 4-1-4-1 formation wherein Kroos was one of the two central (attacking) midfielders. While in 2012/13 Kroos played mostly as the central attacking midfielder behind Bayern's main striker Mario Mandžukic.

During his two seasons at Real Madrid Kroos played a more controlling role, leaving the attacking duties to players like Ronaldo, Benzema, Bale, James and/or Isco. This is indicated by a clear increase in the use of the ABCDG motif. Despite Real Madrid having the highest and eight highest expected goals respectively for a team during these two seasons, Toni Kroos saw a further decrease of the xGs he was directly involved in. The biggest chunk of his xGs involvement at Real Madrid comes from the motifs in which he is the 3rd assister (*ABCDG*, *ABCBG*).

[7] A goal attempt motif is the unweighted expected goals motif

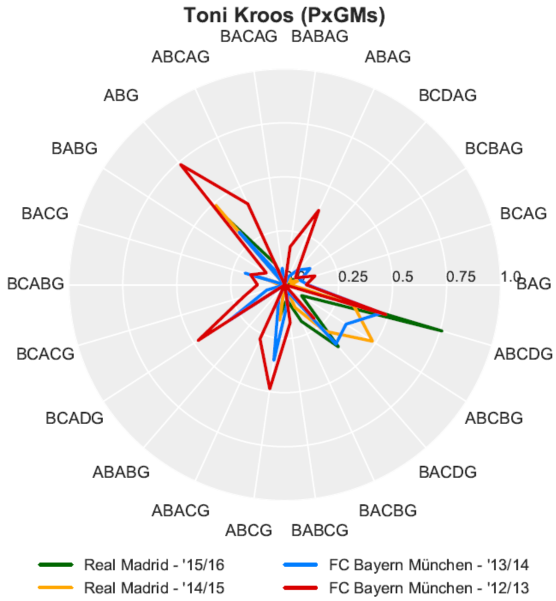

Figure 11: Toni Kroos PxGM style in all 4 seasons

5.3. Who *can* replace Xavi?

In his research appropriately titled "*Who can replace Xavi?*" Lopez Peña poses the question who could possibly replace Xavi after he departs from FC Barcelona [Peña and Navarro, 2015]. *La Computadora*, who moved to play for Al Sadd in 2015, is widely regarded as the puppet master pulling the strings in arguably the best soccer team the world has ever seen. In his analysis Peña shows that Xavi is a clear outlier when it comes to his extraordinary passing ability, raising the question who could possibly follow in his footsteps. In Figure 12 Xavi's radar graph is depicted for the season 2012/13 (by far his best season in this data set) in which he was involved in 228.6 PPMs per 90 minutes.

Xavi's successor by possession motifs can be found by investigating the players closest to Xavi on Euclidean distance. Not surprisingly, the five players closest to Xavi are Thiago Alcantara, Verratti, Thiago Motta, Kimmich and Weigl respectively over all their seasons in the data set (see Table 8). Taking a closer look into the data shows that Marco Verratti is an interesting prospect (see Figure 13) because in the 2015/2016 season he produced 237.8 PPMs per 90 minutes.

Subsequently we look at the players closest to Xavi by PxGMs. We find that Verratti is only the 520th player away from Xavi by Euclidean distance. However, looking at both Xavi and Verratti's PxGM radar graphs (Figure 14 and 15) we see that Verratti is involved in more expected goal motifs than Xavi, but they both create most of their xGs as the second and third assister.

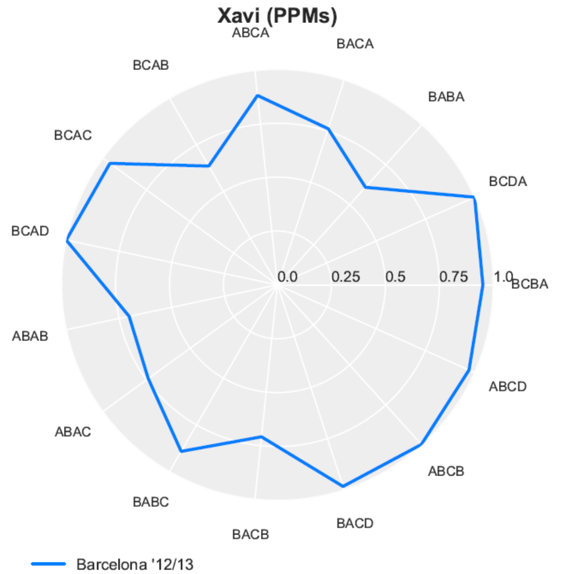

Figure 12: Xavi PPM style in '12/13

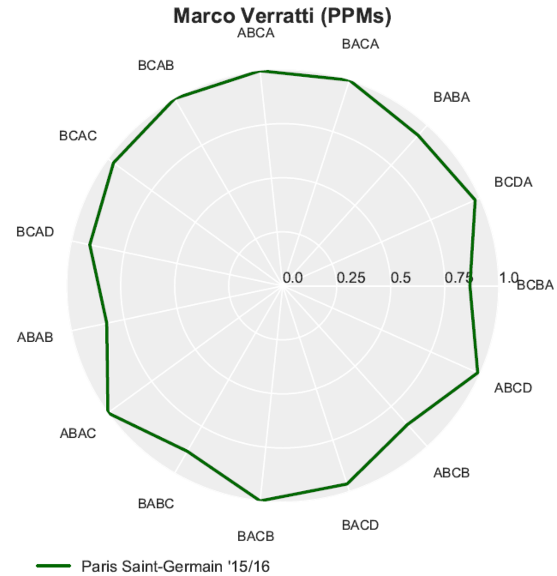

Figure 13: Verratti PPM style in '15/16

Does this make Marco Verratti the perfect replacement for Xavi had Barcelona still needed one? Possibly. Unfortunately, Marco Verratti only played the equivalent of 10.19 matches that season, due to seven different minor injuries[8], whereas Xavi played the equivalent of 35.24 matches. However, in the 2014/15 season Verratti played the equivalence of 28.3 matches wherein he reached a total average of 183.9 motifs per 90 minutes ranking him first that season.

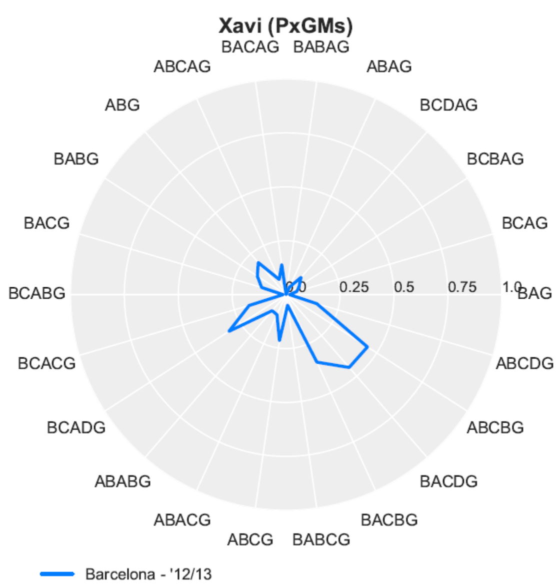

Figure 14: Xavi PxGM style in '12/13

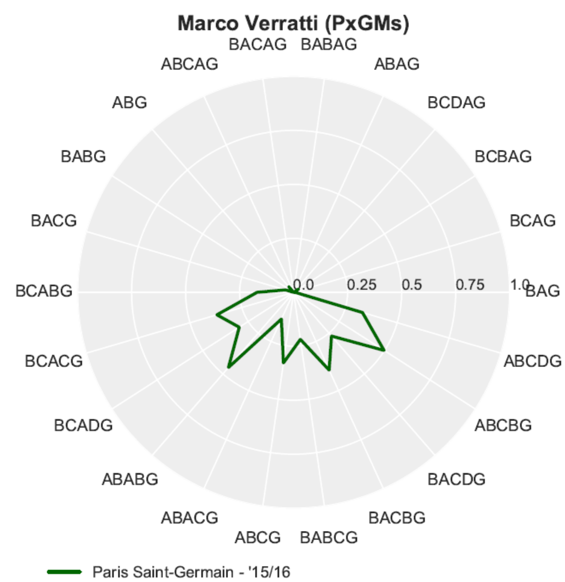

Figure 15: Verratti PxGM style in '15/16

[8] According to transfermarkt.com

6. Analysis of Post-Fàbregas Arsenal

With both the team motifs and player motifs at hand, it is possible to make a detailed analysis of the influence of specific players to the play style of a team. To demonstrate this we will analyze the impact of Mesut Özil and Sánchez on the team after they were transferred to Arsenal.

With the departures of Cesc Fàbregas to FC Barcelona in August 2011 and Robin van Persie to Manchester United in August 2012, Arsenal lost a great piece of their creative and attacking prowess. From an Arsenal point of view it was necessary to recoup some of these lost powers. After the departure of Fàbregas, Mikel Arteta was signed from Everton. In August 2012, Olivier Giroud was acquired from surprising French champion HSC Montpellier, alongside Santi Cazorla from Málaga CF and Lukas Podolski from FC Köln.

During this first season after both Fàbregas and van Persie left (2012/13), the attacking force of Walcott, Podolski and Giroud together only received a shocking 11.3% of all passes. As becomes clear from Figure 16 Arsenal were lacking a creative attacking mind to supply the necessary creativity, especially in the *ABAG*, *ABABG* and *ABCAG* motifs. This missing piece would not necessarily be needed for creative possession play, since Arteta was ranked the 9th player behind eight FC Barcelona players and Ramsey 12th behind only Martín Montoya (also FC Barcelona) and Marco Verratti (PSG) in average total pass motives per 90 minutes in the 2012/2013 season.

In the summer of 2013 Mesut Özil, arguably one of the best assist makers in the game, ranking third in the 2012/2013 season on expected goals assisted, made a surprise move from Real Madrid to Arsenal for an estimated €50 million.

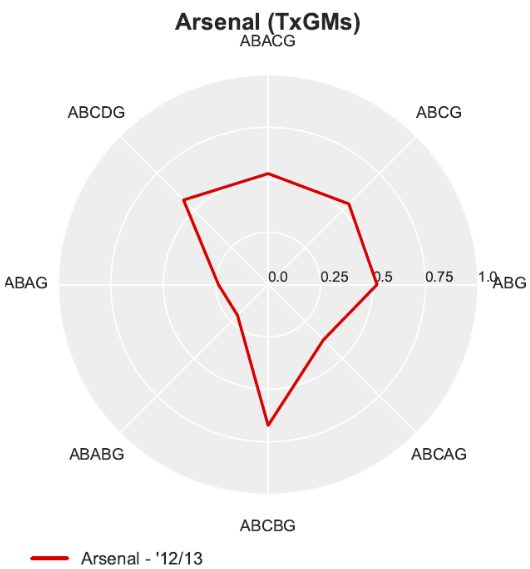

Figure 16: Arsenal´s TxGM style in ´12/13

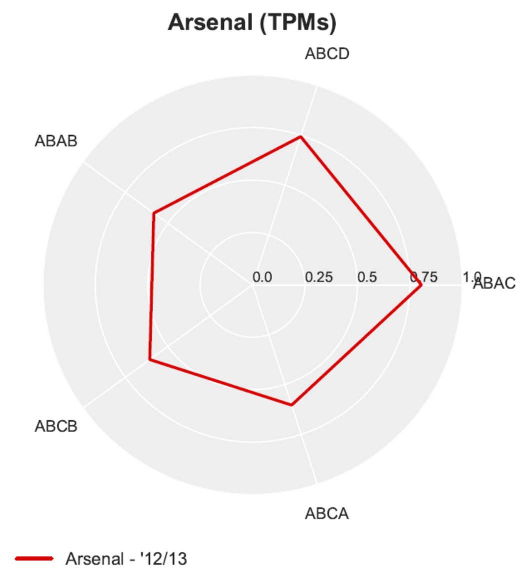

Figure 17: Arsenal´s TPM style in ´12/13

6.1. Season 2013/14

In his first season at Arsenal, Özil used his vision to make direct and indirect link-ups with Olivier Giroud (Arsenal's main striker) in order to try and improve the attacking impulse of *the Gunners*. We can see this by looking at the pass probability matrix shown in Table 13.[9]

[9] The table shows the goalkeeper and 10 outfield players with the most minutes played, and 'rest' indicating all other players that were selected by Arsène Wenger during the season.

7.6% of Özil's passes were made to Giroud (2.6 percentage points more than any player in the previous season). Furthermore Özil reached Giroud plenty of times indirectly by passing to players who pass to Giroud with a high probability (Gibbs 7.4%, Sagna 6.0%, and Wilshere 5.9%). This helped elevate the overall pass percentage towards Giroud from 3.8% in the '12/13 season to 5.4% in the '13/14 season, and his goals scored went up from 11 to 16.

Özil's style (see Figures 18 and 19) resulted in a big increase in xGs from one-two type combinations (*ABAG* and *ABABG*, and *ABACG*) and a minor decrease of *ABCG* and *ABG* motifs. This is not surprising considering he was ranked 7th on assisting one-two combinations (*BABG*, *ABABG*, *ABACG*, and *BABCG*) and 16th on total xGs from one-two combinations in 2013/14.

	Szczesny	Gibbs	Koscielny	Mertesacker	Sagna	Wilshere	Cazorla	Ramsey	Özil	Arteta	Giroud	Rest
Szczesny	-	1.9%	13.7%	21.1%	20.5%	0.9%	1.6%	0.6%	1.2%	2.2%	20.2%	16.1%
Gibbs	3.9%	-	10.4%	2.2%	1.1%	8.0%	13.8%	10.9%	9.0%	9.8%	7.4%	23.6%
Koscielny	8.1%	10.9%	-	22.5%	6.1%	3.4%	5.1%	5.7%	4.8%	11.4%	1.2%	20.8%
Mertesacker	8.8%	1.7%	11.5%	-	14.8%	4.0%	3.8%	8.6%	5.4%	17.5%	2.9%	21.0%
Sagna	3.1%	0.3%	4.2%	16.2%	-	6.0%	9.7%	9.2%	9.0%	9.2%	6.0%	27.1%
Wilshere	0.3%	8.4%	2.9%	3.7%	8.5%	-	11.5%	8.8%	12.4%	10.2%	5.9%	27.3%
Cazorla	0.2%	6.4%	1.7%	2.5%	10.3%	8.2%	-	6.8%	12.1%	11.7%	3.9%	36.2%
Ramsey	1.0%	7.7%	5.3%	6.7%	8.9%	7.5%	10.1%	-	13.9%	7.6%	3.9%	27.5%
Özil	0.2%	5.8%	2.6%	2.8%	11.1%	8.3%	13.9%	13.3%	-	8.0%	7.6%	26.6%
Arteta	0.5%	5.4%	6.4%	8.2%	13.5%	4.8%	13.7%	6.4%	8.6%	-	2.4%	30.0%
Giroud	0.2%	6.4%	0.6%	2.1%	7.3%	8.3%	13.9%	8.9%	14.0%	7.3%	-	30.9%
Rest	1.1%	3.8%	5.7%	4.4%	11.6%	6.0%	8.7%	5.7%	9.2%	9.7%	3.7%	30.4%
percentage of total passes received:												
	2.3%	4.9%	5.4%	7.7%	9.5%	5.5%	8.8%	7.1%	8.3%	8.7%	5.4%	26.5%

Table 13: Arsenal Pass Probability Matrix '13/14 (n=16298)

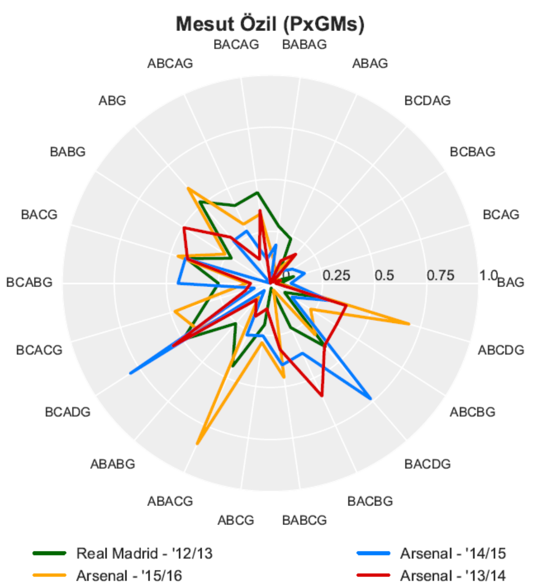

Figure 18: Mesut Özil PxGM style

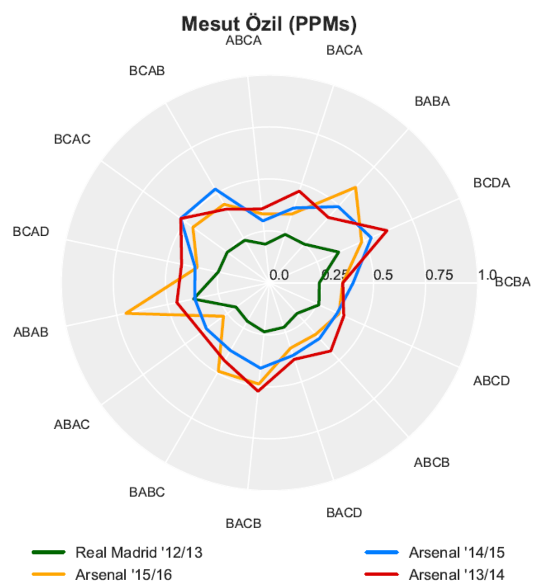

Figure 19: Mesut Özil PPM style

6.2. Season 2014/15

Despite the presence of both Mesut Özil and Olivier Giroud, another player had to be attracted to fill the gap between the creative midfield and this lone striker in order to improve on two consecutive

4th places that Arsenal reached during these past two seasons. To find a suitable player to fill this gap, we looked at the top 25 attackers by PPMs and their PxGMs in the 2012/13 and 2013/14 seasons of our data set.[10] After omitting all highly unobtainable players and players that only excelled in one of the two seasons we are left with seven suitable players, see Table 14 (notable omitted players include: Messi, Totti, Ibrahimovic and Eden Hazard).

	Season '12/13		Season '13/14	
Player	**PPM**	**PxGM**	**PPM**	**PxGM**
Pedro	103.3	0.38	83.0	0.63
Robinho	67.1	0.34	67.1	0.50
A. Sánchez	65.1	0.49	66.5	0.80
Bryan Ruiz	62.1	0.39	56.0	0.57
Adem Ljajic	61.6	0.56	54.6	0.81
Cassano	58.3	0.49	53.4	0.74
Insigne	54.2	0.51	59.4	0.70

Tabel 14: Potential Attacking Additions to Arsenal at the start of '14/15 (motifs per 90 minutes played).

Ultimately in July 2014, Arsenal announced that Alexis Sánchez would be added to their squad for £31.7 million. This is not surprising since the statistics presented above and his radar graphs (see Figures 20 and 21) clearly show his goal scoring, assisting and positional capabilities; and the fact that Arsenal aspire to play a style of soccer closely related to Sánchez's previous club FC Barcelona (as seen in the Tables 3 and 4).

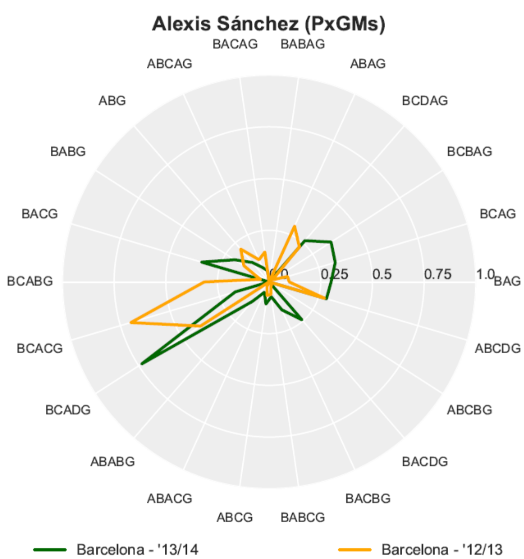

Figure 20: Sánchez PxGM style at FC Barcelona

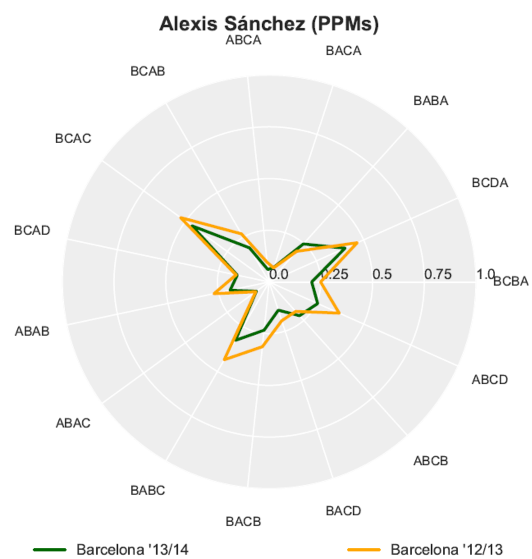

Figure 21: Sánchez PPM style at FC Barcelona

10 It's reasonable to assume that Arsenal is capable of attracting players of this caliber considering they are a consistent member of the top 5 most valuable teams in the world, according to Forbes.com

So, how did the acquisition of Alexis Sánchez affect the general attacking style of Arsenal, and the linkup between Giroud and Özil? In Arsenal's 2014/15 pass probability matrix (Table 15) we see that Giroud received 5.6% of all passes, Özil 8.0% and Sánchez 8.7%. Özil sent 14.1% of his passes to Sánchez, and only 4% to Giroud (as opposed to 7.6% in the previous season) and 13.2% of passes went from Sánchez to Özil.

In this season the in-field players with the highest probability of passing to Giroud were Sánchez with 5.9%, Welbeck with 5.9% and Nacho Monreal with 5.5%. Despite the lesser connections between Özil and Giroud this did not cause a lack of TxGMs for Arsenal. On the contrary, because Sánchez functioned as a link between Özil and Giroud while also serving a second target-man, Arsenal saw an increase in TxGMs created (from 1.42 xGs to 1.56 xGs per 90 minutes played).

Overall in 2014/15, the one-two type motifs decreased because Rosicky played significantly less than the season before. However the *ABAG*, *ABCG* and *ABCAG* saw an increase in xGs produced. The increase in *ABAG* is explained by both Sánchez and Giroud ranking 9th and 3rd respectively (compared to the whole data set) during this season for this motif.

So, although Özil only passed 4% to Giroud, Giroud actually played 14% of his total passes to Özil, indicating that the necessary combinations involving Giroud, Özil and Sánchez were still utilized a lot. Arsenal finished 3rd place that season for the first time since 2011/12.

	Özil	**Sánchez**	**Giroud**
Özil	-	14.1%	4.0%
Sánchez	13.2%	-	5.9%
Giroud	14.0%	15.5%	-
percentage of total passes received:	8.0%	8.7%	5.6%

Table 15: Part of Arsenal's Pass Probability Matrix ('14/15)

	Özil	**Sánchez**	**Giroud**
Özil	-	16.3%	4.9%
Sánchez	22.6%	-	3.8%
Giroud	26.5%	7.0%	-
percentage of total passes received:	13.4%	6.8%	4.9%

Table 16: Part of Arsenal's Pass Probability Matrix ('15/16)

6.3. Season 2015/16

In the 2015/16 season, Santi Cazorla endured a knee injury and an Achilles irritation eliminating him from November 30th 2015 until April 20th 2016, missing 27 consecutive league matches.[11] Due to this unfortunate event, Özil became not only the go-to player for creating expected goal motifs, he was now also in charge of creating possession motifs. This event sparked a change in the lay-out of the Arsenal team, making them even more attack oriented. In that season, Özil ranked 8th and Sánchez ranked 21st on their involvement in total xGs per 90 minutes played.

Due to Cazorla's absence, Özil's total percentage of balls received sky rocketed from 8.0% in the '14/15 season, to 13.4% in the '15/16 season (see Table 16), with both Sánchez and Giroud passing to Özil over 20%. And because Özil played 16.3% of all his passes to Sánchez it comes as no surprise that Arsenal saw another increase in expected goal motifs (from 1.56 in '14/15 to 1.69 in '15/16). Furthermore, in this season Özil assisted 4 goal attempts per 90 minutes, making him the player with the most assists per 90 minutes in the whole data set. This resulted in Arsenal's first 2nd place finish since the 2004/05 season.

The evolution of Arsenal after the acquisition of Özil and Sánchez can ultimately be summarized in Figure 22 and 23 in which we see a per season increase in both TxGMs and TPMs.

[11] Positions according to transfermarkt.com

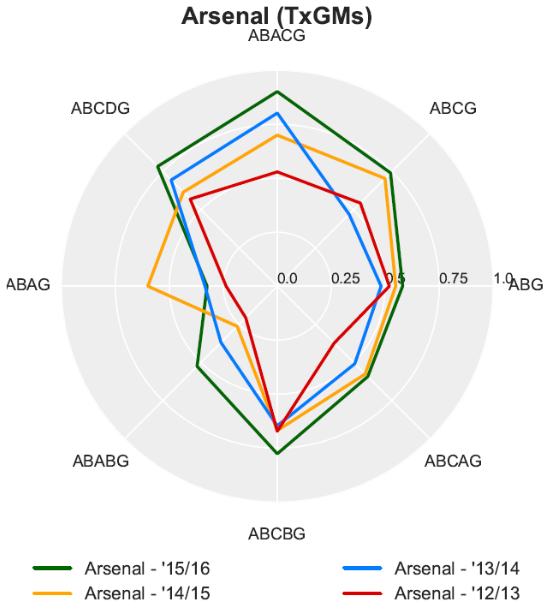

Figure 22: Arsenal's TxGM style

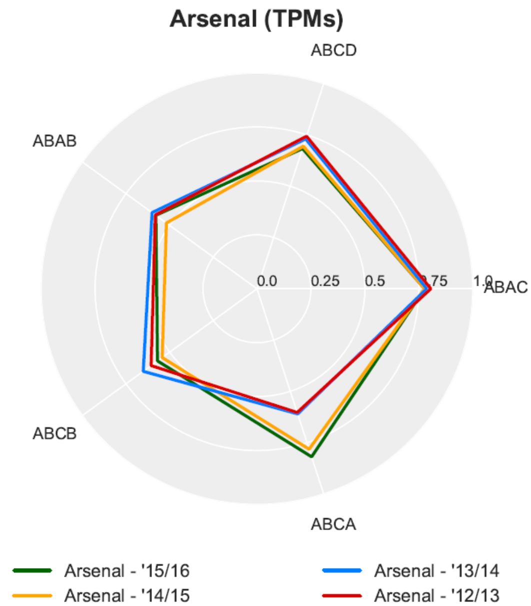

Figure 23: Arsenal's TPM style

6.4. Season 2016/17

In addition to analyzing past events, the PMs and xGMs can also be used to select possible replacements for departing players and coaches in the near future and finding more optimal strategies than those currently utilized by a team.

6.4.1. Hypothetical Player Replacements

Besides Özil, another player who was often involved in passing motifs with Alexis Sánchez in 2015/16 was left-back Nacho Monreal. In this section we will look at possible replacements for these players in the hypothetical case that either of them leaves Arsenal during the summer of 2016. In this case it is possible to execute a nearest neighbour search for the departing player to look for a replacement with a similar PPM and PxGM style, and similar position.[12] This search will be conducted on the aggregate last two seasons of both players - the seasons in which they played with Sánchez. These possible replacements, identified with the help of the Euclidean distance nearest neighbours for PxGMs and PPMs, are shown in Table 17 for Mesut Özil, and Table 18 for Nacho Monreal.

Player	Team	Season(s)
Mesut Özil	Arsenal	'14/15 & '15/16
David Silva	Man. City	'15/16
Isco	Real Madrid	'15/16
James Rodriguez	Real Madrid	'14/15
Mathieu Valbuena	Olympique Lyon	'15/16

Tabel 17: Central Attacking Midfielders closest to Özil by PPM and PxGM using Euclidean Distance

[12] According to www.transfermarket.com

Player	**Team**	**Season(s)**
Nacho Monreal	Arsenal	'14/15 & '15/16
Cesar Azpilicueta	Chelsea	'14/15 & '15/16
Miquel Nelom	Feyenoord	'15/16
Jonny Castro	Celta de Vigo	'15/16
Jeremy Morel	Olympique Lyon	'15/16

Tabel 18: Left-Backs closest to Monreal
by PPM and PxGM using Euclidean Distance

6.4.2. Hypothetical Manager Replacement

2016 marked the 20th anniversary of Arsène Wenger as Arsenal's head coach. Judging from the void left by Alex Ferguson at Manchester United when he retired in 2013 after a 26 year reign, it would be very difficult to replace a manager that shaped a team for such a long period of time. To be able to make a smooth transition whenever Wenger leaves Arsenal, finding a manager that coached a team with a similar play style to that of Arsène might help smooth this eventual transition.

To find such managers we conduct a search for coaches that managed teams with a similar style in one season, when compared to the aggregate style of Arsenal over 4 seasons. In Table 19 the coaches that are close to Arsenal on both TPMs and TxGMs are depicted.

Coach	**Team**	**Season(s)**
Arsène Wenger	Arsenal	All
Lucien Favre	Borussia M'gladbach	'14/15
Rudi Garcia	AS Roma	'13/14
Massimiliano Allegri	Juventus	'15/16
Antonio Conte	Juventus	'14/15
Luciano Spaletti	AS Roma	'15/16

Tabel 19: Coaches/Teams with a Style Closest
to Arsène Wenger by TPMs and TxGMs

6.4.3. Strategic Improvements

In Section 4.4 it was shown that some goal attempt motifs occur more frequently than others and that the average value of a goal attempt differs significantly depending on the type of motif executed. When looking at these two aspects for Arsenal it is possible to find xG-motifs that are less effective and might hinder Arsenal in creating even more valuable opportunities.

Table 20 shows that *ABAG*, *ABABG*, *ABCBG* are executed 11.3% of the time. These motifs, ending in a one-two combination, are 3 of the 4 motifs with the lowest average goal attempt value. By introducing a new player after the one-two combination *ABAG* becomes *ABACG*, *ABABG* becomes *BABCG* (=*ABACG*), and *ABCBG* becomes *BCBAG* (=*ABACG*). Extending the motif by one player (within five seconds) to make the final attempt can create a great amount of extra value, because ABACG generates an average goal attempt value of 0.124 xG as opposed to 0.110, 0.098 and 0.094. For instance a reduction of just 1% for both *ABAG*, *ABABG* and *ABCBG*, and thus an increase of 3% in *ABACG* would help Arsenal from a weighted average of 0.110 xG per goal attempt motif to 0.120 xG per goal attempt motif - an increase of 9%.

Motif	Frequency	Mean Goal Attempt Value
ABCDG	30.2%	0.117 xG
ABG	29.5%	0.131 xG
ABCG	14.9%	0.122 xG
ABACG	8.7%	0.124 xG
ABCBG	8.0%	0.094 xG
ABCAG	5.3%	0.091 xG
ABAG	2.2%	0.110 xG
ABABG	1.1%	0.098 xG

Table 20: Arsenal's Goal Motif Frequency and
Mean Goal Attempt Value in Expected Goals (across 4 seasons)

7. Conclusion

We proposed a quantitative method to evaluate the styles of soccer teams and players through their possession and goal attempt flow motifs. Players can be analyzed with the help of their personal radar graph which depicts their flow motif performance against all other players in the dataset. Furthermore, we find that players with similar styles can be scouted with the use of the Euclidean distance between a pre-determined player and all other players' motif intensities. For example, we use this to explain the purchase of Claudio Bravo by Manchester City and it aided in the search for the heir to Xavi's throne.

Unique styles for both teams and players can be found by applying the mean shift algorithm to cluster them based on motif use intensities. We find that Paris Saint-Germain is in a cluster by itself indicating its unique style, and FC Barcelona and Bayern Munich are in their own cluster, being heavily influenced by Pep Guardiola. Furthermore we identify three players with unique styles; Iniesta, Rafinha and Denswil. In the final chapter we showcase the possibilities the expected goal and possession motifs in combination with the pass probability matrices offer in regards to analysis of team structures, the search for new players, as well as the strategic influence they hold.

These lines of thought can be extended to identify differences in styles between the two halves of the game, whether a change in the style of a team follows a certain trend and even analyze two players simultaneously by studying their joint passing behaviour and impact on team performance. Powered by the availability of data and this analysis, it is now possible to explore many more questions about the sport providing us with valuable insight that has strategic implications on the game.

References

[Bialkowski et al., 2014] Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., and Matthews, I. (2014). Large-scale analysis of soccer matches using spatiotemporal tracking data. *In 2014 IEEE International Conference on Data Mining*, pages 725–730. IEEE.

[Clemente et al., 2015a] Clemente, F. M., Couceiro, M. S., Martins, F. M. L., and Mendes, R. S. (2015a). Using network metrics in soccer: A macro-analysis*. Journal of human kinetics*, 45(1):123–134.

[Clemente et al., 2015b] Clemente, F. M., Martins, F. M. L., Wong, D. P., Kalamaras, D., and Mendes, R. S. (2015b). Midfielder as the prominent participant in the building attack: A network analysis of national teams in fifa world cup 2014. *International Journal of Performance Analysis in Sport*, 15(2):704–722.

[Comaniciu and Meer, 2002] Comaniciu, D. and Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619.

[Fukunaga and Hostetler, 1975] Fukunaga, K. and Hostetler, L. (1975). The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Transactions on information theory*, 21(1):32–40.

[Georgescu et al., 2003] Georgescu, B., Shimshoni, I., and Meer, P. (2003). Mean shift based clustering in high dimensions: A texture classification example. In Computer Vision, 2003. Proceedings. *Ninth IEEE International Conference on*, pages 456–463. IEEE.

[Goldsberry, 2012] Goldsberry, K. (2012). Courtvision: New visual and spatial analytics for the nba*. In 2012 MIT Sloan Sports Analytics Conference*.

[Gyarmati and Hefeeda, 2015] Gyarmati, L. and Hefeeda, M. (2015). Estimating the maximal speed of soccer players on scale. *In Proc. Machine Learning and Data Mining for Sports Analytics Workshop*.

[Gyarmati et al., 2014] Gyarmati, L., Kwak, H., and Rodriguez, P. (2014). Searching for a unique style in soccer*. arXiv preprint arXiv:1409.0308.*

[Macdonald, 2012] Macdonald, B. (2012). An expected goals model for evaluating nhl teams and players. *http://www.sloansportsconference.com/wp-content/uploads/2012/02/NHL-Expected-GoalsBrian-Macdonald.pdf.*

[Messi and Alves, 2010] Messi, L. and Alves, D. (2010). FC Barcelona vs Real Sociedad. https://www. youtube.com/watch?v=Aqp3ggl_hzI8.

[Milo et al., 2002] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827.

[Peña, 2014] Peña, J. L. (2014). A markovian model for association football possession and its outcomes. *arXiv preprint arXiv:1403.7993.*

[Peña and Navarro, 2015] Peña, J. L. and Navarro, R. S. (2015). Who can replace xavi? a passing motif analysis of football players*. arXiv preprint arXiv:1506.07768.*

[Peña and Touchette, 2012] Peña, J. L. and Touchette, H. (2012). A network theory analysis of football strategies. *arXiv preprint arXiv:1206.6904.*

[Pollard, 1986] Pollard, R. (1986). Home advantage in soccer: A retrospective analysis*. Journal of Sports Sciences*, 4(3):237–248.

[Shortridge et al., 2014] Shortridge, A., Goldsberry, K., and Adams, M. (2014). Creating space to shoot: quantifying spatial relative field goal efficiency in basketball. *Journal of Quantitative Analysis in Sports*, 10(3):303–313

[UEFA, 2016] UEFA (2016). UEFA Guide to Quality Stadiums.
http://www.uefa.org/MultimediaFiles/
Download/EuroExperience/competitions/General/01/74/38/69/1743869_DOWNLOAD.pdf.
Accessed: 2016-11-03.

[Wilshire and Giroud, 2013] Wilshire, J. and Giroud, O. (2013). Arsenal vs Norwich. https://www.youtube.com/watch?v=NmQfhkGPrM8.

Appendices

A. Team Possesion Motif Clustering

Team involvement in the other four possession motifs as a percentage of all possible motifs, against the use intensity per 90 minutes of these motifs are shown in Figure A, B, C and D. These figures include 155 unique teams.

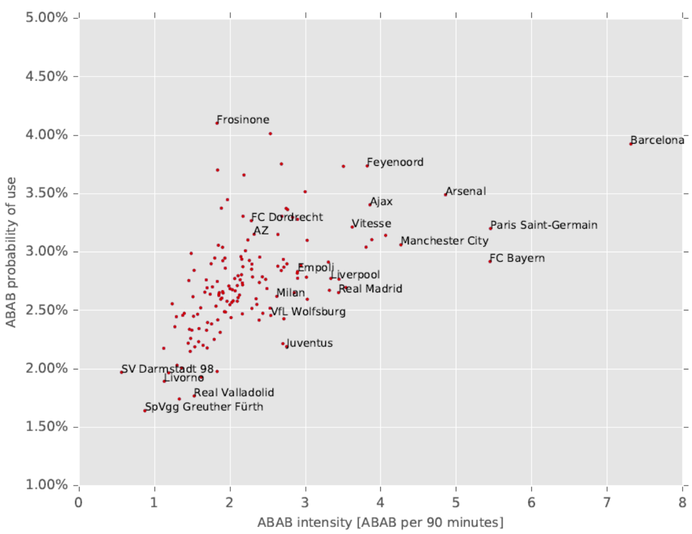

Figure A: Team involvement in ABAB vs use intensity

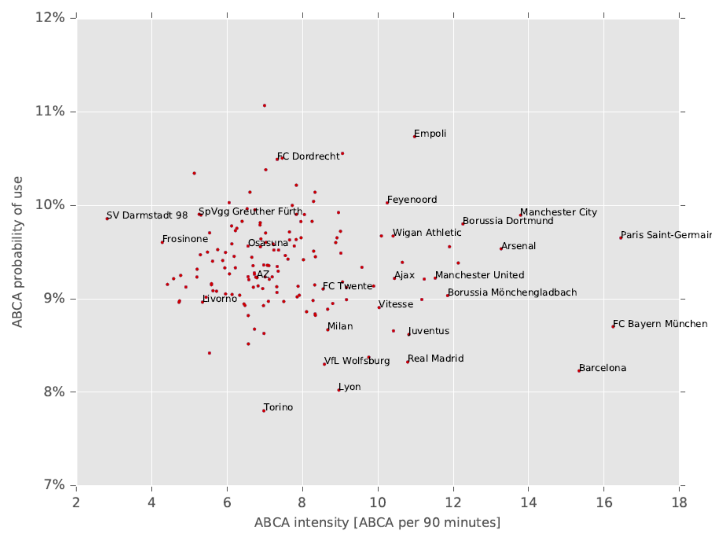

Figure B: Team involvement in ABCA vs use intensity

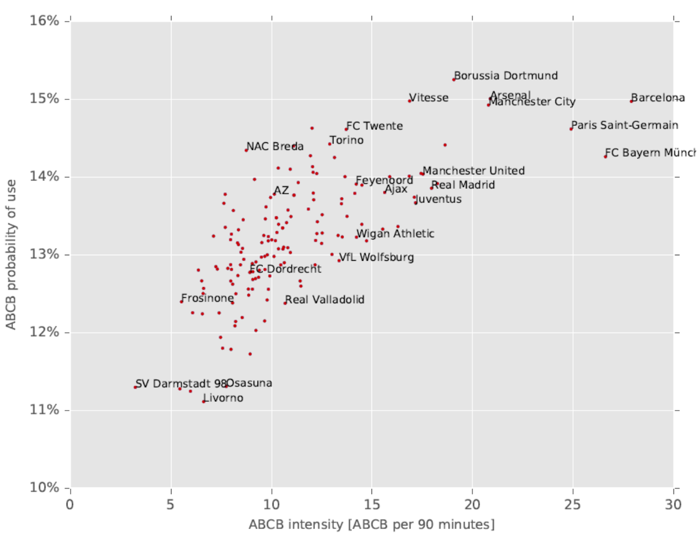

Figure C: Team involvement in ABCB vs use intensity

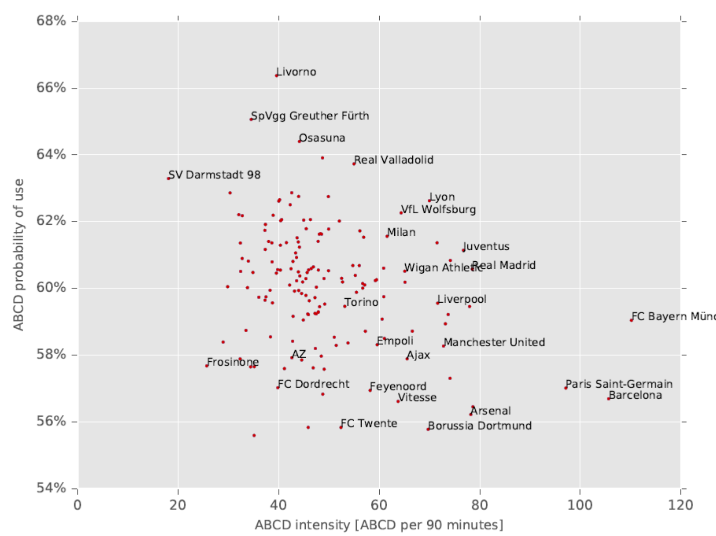

Figure D: Team involvement in ABCD vs use intensity

B. **Player Possession Motif Clustering**

Player involvement in the other fourteen possession motifs as a percentage of all possible motifs, against the use intensity per 90 minutes of these motifs is shown in Figure E until R. These figures include 3532 unique players.

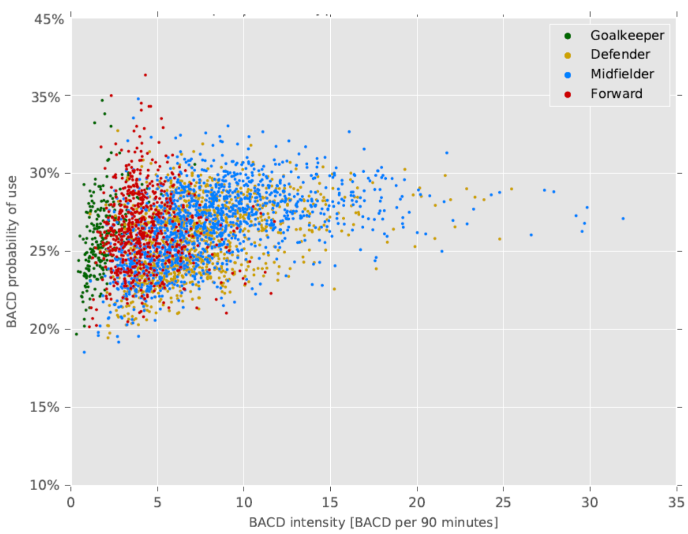

Figure E: Player involvement in BACD vs use intensity

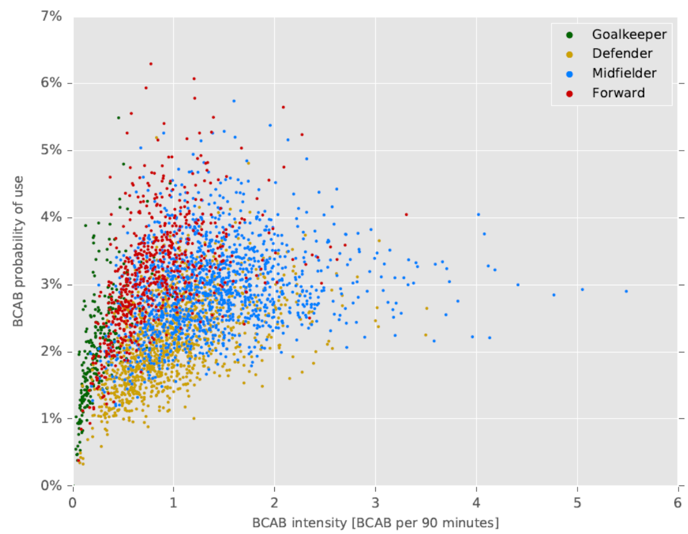

Figure F: Player involvement in BCAB vs use intensity

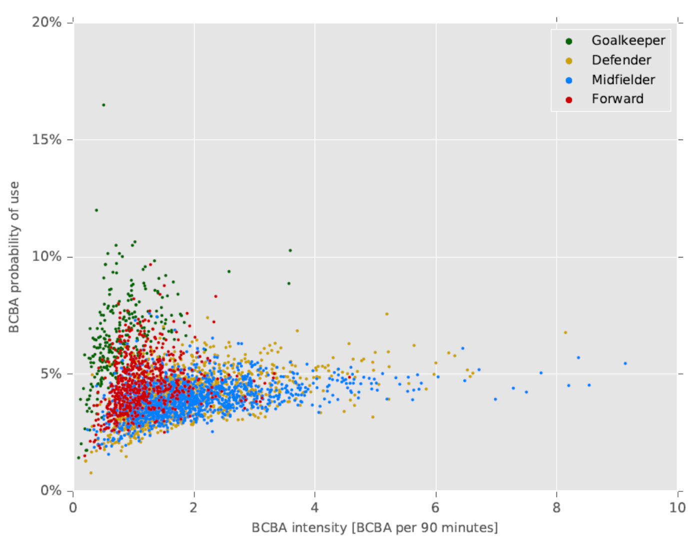

Figure G: Player involvement in BCBA vs use intensity

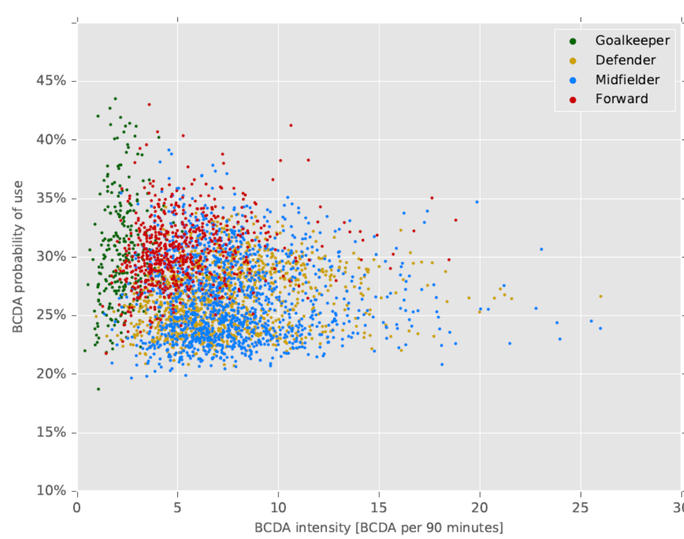

Figure H: Player involvement in BCDA vs use intensity

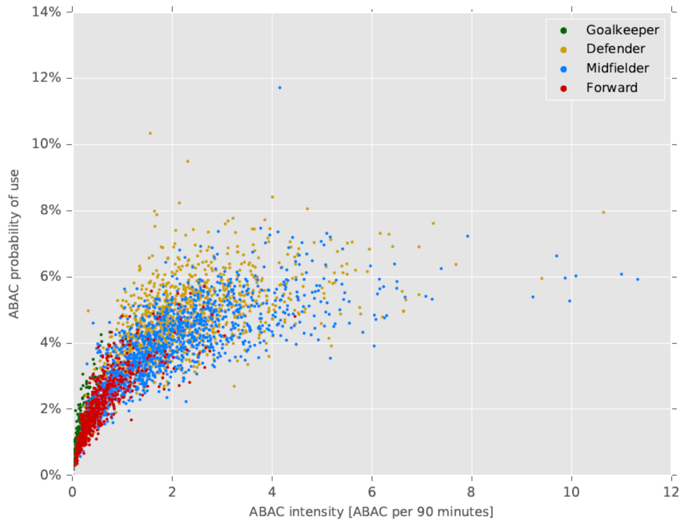

Figure I: Player involvement in ABAC vs use intensity

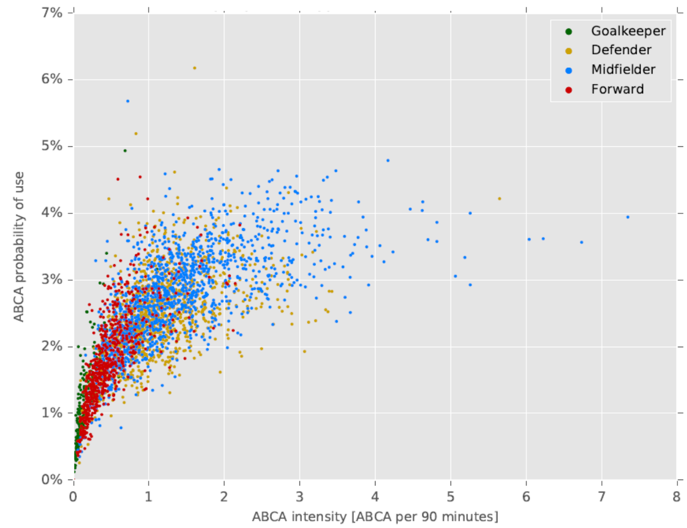

Figure J: Player involvement in ABCA vs use intensity

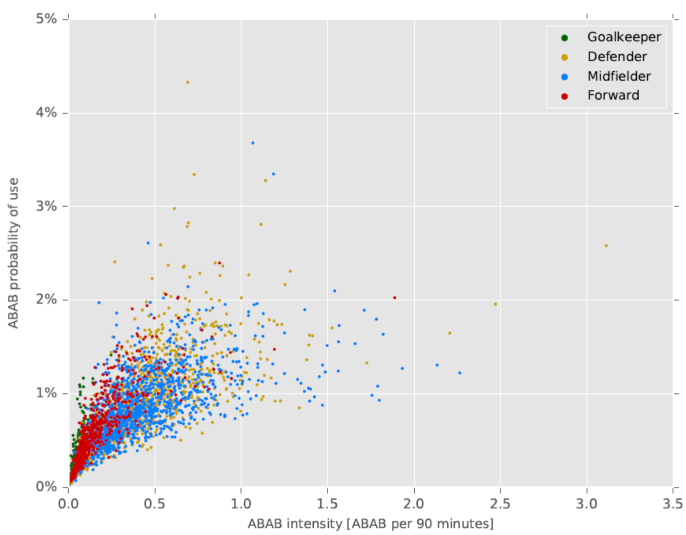

Figure K: Player involvement in ABAB vs use intensity

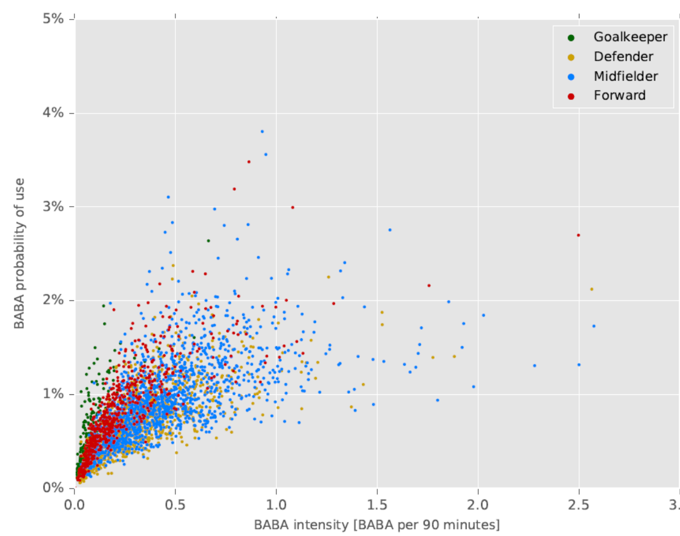

Figure L: Player involvement in BABA vs use intensity

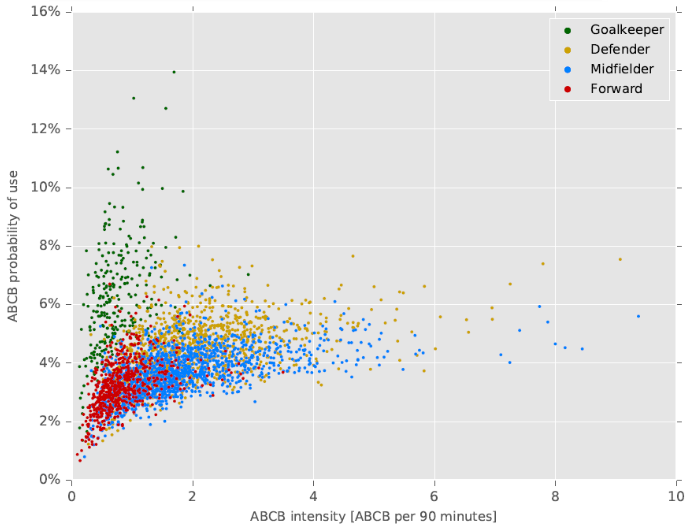

Figure M: Player involvement in ABCB vs use intensity

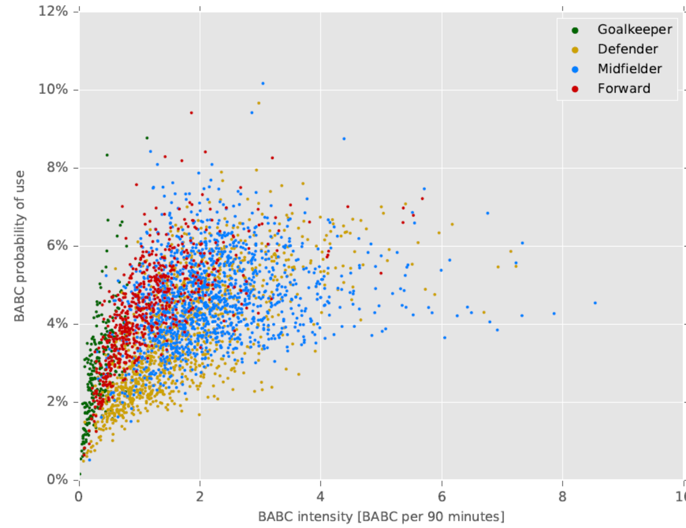

Figure N: Player involvement in BABC vs use intensity

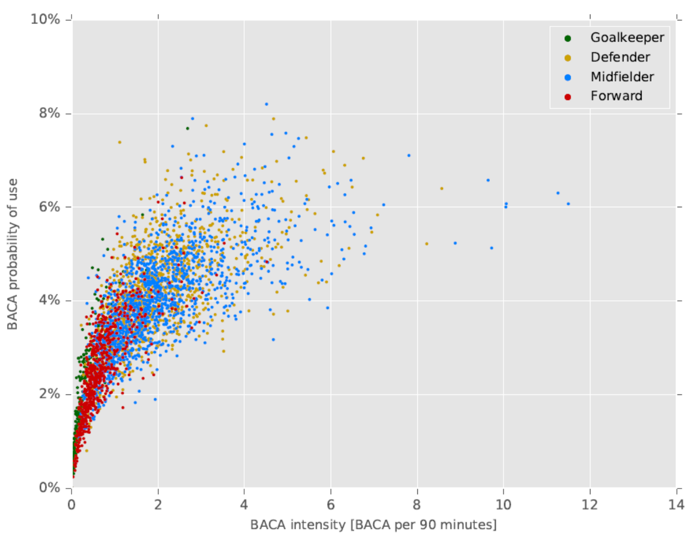

Figure O: Player involvement in BACA vs use intensity

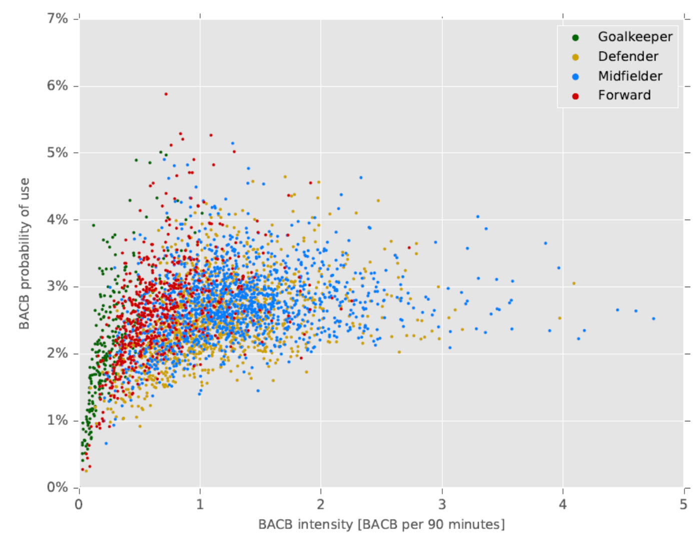

Figure P: Player involvement in BACB vs use intensity

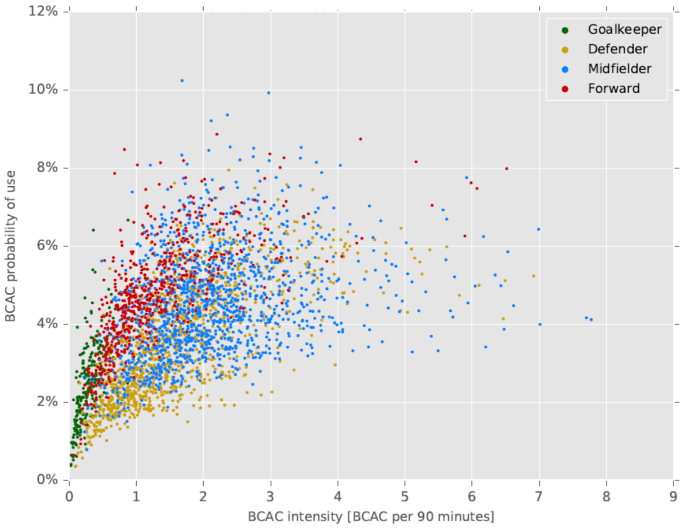

Figure Q: Player involvement in BCAC vs use intensity

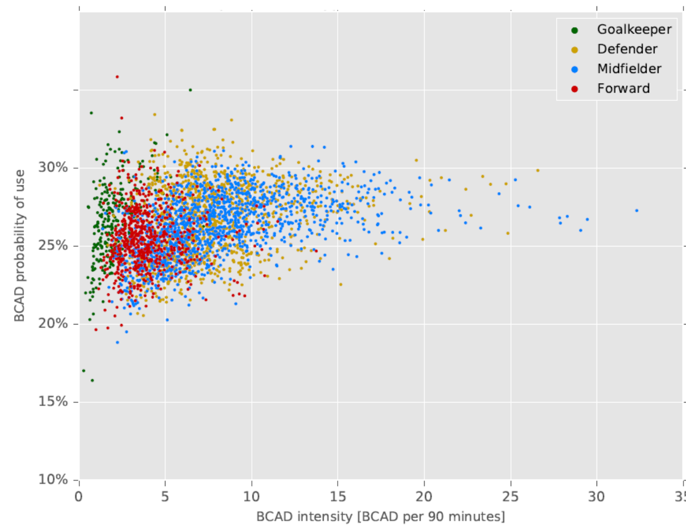

Figure R: Player involvement in BCAD vs use intensity