Opleiding Informatica

Analysis of tactical behavior in soccer:

capturing and analysing space

Wessel van Zetten

Supervisors:
Arno Knobbe & Rens Meerhoff

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

24/07/2019

Abstract

While a soccer team is defending against an attack, they invariably leave spaces in their formation unoccupied, which the attacking team must use to the best of their ability. Due to widespread positional data, it is now possible to analyse the creation of these spaces and monitor them during pressure events. Pressure events occur when the defending team puts the ball carrier under pressure, in order to capture the ball or force a pass. This research attempts to capture these spaces, by creating different features for the space the attackers have at their disposal, and the space they actually use, and find if these features differ under several kinds of defensive pressure and successful and unsuccessful events.

The features for potential space and used space were combined in a list of different performance measures, out of which the one that fits the defensive pressure types the best was selected, using a combination of correlation and subgroup discovery. The value of this performance measure was then investigated to see if it differs per pressure type as defined by the KNVB. Following this, pressure events were marked successful or unsuccessful depending on if a pass was made to a teammate in an attacking role. After marking all events, the value of the performance measure was investigated to see if it differed between successful and unsuccessful events.

The results showed that a certain combination of a measure that expressed the available space in an area, and a measure for distances between midfielder and defender lines was the best performance measure. Furthermore, the analysis found this performance measure significantly differed over all the pressure types, as well as between two pressure types. The tests for successful and unsuccessful events found that there was no significant difference between the two groups for three of the four pressure types, except in events where the offensive team was positioned high on the pitch.

This research demonstrated the possible work flow to develop features in order to capture the flow of space during pressure events in matches. A similar approach might be applied in other team sports. Furthermore, future research could focus on the difference between successful and unsuccessful pressure events where the significant difference was found.

Contents

1 Introduction

Soccer is a game with many different aspects, all culminating into a single purpose; to score more goals than the opposing team. Although the performance of a team is shown in the amount of goals they score, there are much more tactical aspects to soccer. In the pursuit of a goal-scoring opportunity, teams fight over control of the ball and strategic positions on the pitch. Investigating the tactical approaches used during a game can bring unknown or previously unclear patterns to light, which could be useful in analysing an opponent or improving your own play.

Tactical analysis of your opponent can give you an important competitive edge in soccer. Historically, this analysis was based purely on event data recorded by observers, using variables that discarded most contextual information [MLS16]. Progress in game logs meant that more information with regards to passes, shots on goal and other game events became available for analysis. This analysis could connect the location of the players at the time of the event to the outcome of the event. This allowed for a spatial analysis by calculating, for example, the distance to the closest defender at the time of a shot on goal. However, the exact position of all players during the whole duration of the event was still unavailable. An advance in tracking technologies meant that detailed game logs containing positional data for all players and the ball over the duration of the whole game became available, allowing tactical analysis of soccer to take a leap [SMA+14]. These detailed game logs allowed for the spatial component of the data to be connected to the position of all the players during, for example, a shot on goal. This combination formed spatio-temporal data, which allowed us to measure certain spatial measurements during the whole duration of an event. Due to the importance of tactical analysis, the detailed game logs that have become available should be incorporated into tactical analysis of your opponent, and the analysis of soccer overall.

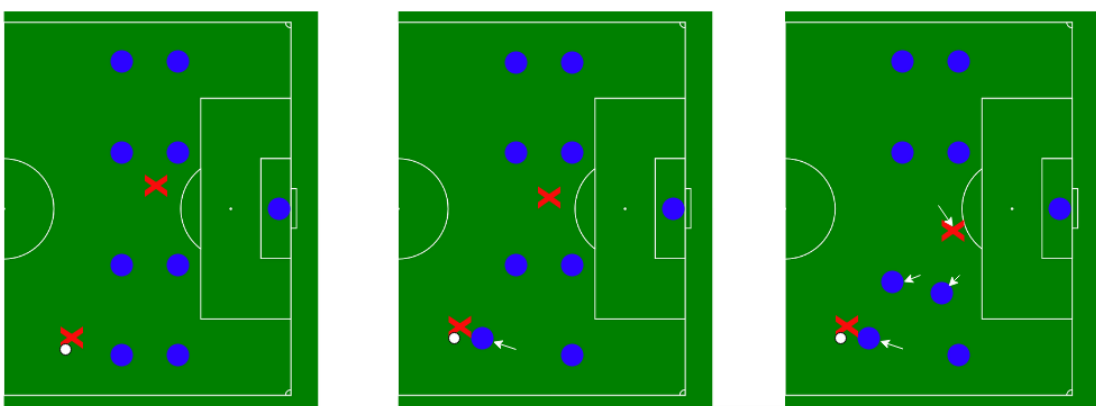

Figure 1: The first panel shows the starting situation, with the attackers (in red) in ball control, attacking the blue defenders. In the second panel, the first defender shifts to put pressure on the ball carrier. In the last panel, the other defender shifts to fill up the space left behind, leaving a second attacker more space within their ranks.

An important aspect to consider is the concept of defensive pressure, in short: how closely a player with the ball is being guarded with the aim of causing a turnover. The defending team must apply this pressure cautiously, as it forces them to move out of position, leaving valuable pitch space for the attackers to utilise. This process is visualised in Figure 1. The first panel shows a simplified situation, with only 2 attackers versus eight defenders (red X and blue dot in Figure 1, respectively). In the second panel, one of the defenders moves to the attacker in possession, leaving a space where he stood before. The third panel shows the other defenders moving to fill up the space left behind by the first defender, allowing a second attacker to occupy the spaces they left behind. This leaves the second attacker in a position to receive a pass from the attacker with the ball. The defenders must always consider the amount of space they leave to the attackers when they move out of position to exert defensive pressure on the ball carrier.

Several attempts at quantifying different aspects of pressure and the space it creates have been carried out before. Link's Dangerousity research [LLS16] aimed at measuring how likely a player was to score a goal at a given time, and incorporated pressure on the ball carrier to determine his dangerousity. Fernandez and Bornn [FB16] attempted to determine the space an offensive player possessed by looking at the influence of defensive players around him. However, these papers did not incorporate the space a player could possess, instead they only looked at the space he actually possessed. Combining the space a player could have with the space he actually has can provide insight in the performance of a player, following the reasoning that a good player possesses as much space he could potentially possess. This can be extended to the performance of a team and the space it possesses. When the defending team exerts defensive pressure, the performance of the attacking team could be measured by looking at the amount of space it theoretically has available, versus the amount of space it actually controls. The closer these two are together, the better the offensive team is performing in that situation. This thesis develops features to capture this ratio between the space a team could possess, called the potential space, and the space it effectively possesses, called the used space.

This thesis tests the numerous developed features to find the most appropriate combination of a potential space and used space measure. This combination is selected as the performance measure, which is then observed during a pressure event, simply a sequence where a player is in possession and an opposing player is within a certain distance. Subsequently, this performance measure will be used to investigate two questions. The first question will investigate whether the value of the performance measure differs between several predefined pressure types. The second question considers the value of the performance measure between successful and unsuccessful events.

2 Modelling approaches

Using the positional tracking data, a number of features to capture potential space and used space were derived. In order to find the combination of a potential space and used space feature that best describes the predefined pressure types, multiple different features and variants were derived, and later tested to find the most appropriate combination to use as performance measure in the rest of the research. This section explains the features developed for potential space, followed by the features for used space. Lastly, the pressure event and different pressure types are explained. Some features covered in this section (listed in Appendix A) have been adapted from existing methods, while others have been built from the ground up, factoring in expert advice provided by the Royal Dutch Football Association (KNVB). This advice has been combined with insights gained from several low level features developed in the early stages of the research, in order to devise features that accurately capture the potential space and used space. These features were all validated using an application that visualises the position of all players and the calculation of a selected feature. The features were checked during pressure events to decide if their values were logical with the intention and computation of the features. The development of the features and analysis of games was done using the TacticsPy Python pipeline, a pipeline developed by Leiden University in cooperation with the KNVB. The pipeline is explained further in section 3.2.

2.1 Potential space

The potential space a team can occupy with its attackers is the space between the defenders and midfielders of the defending team. This is where the attackers must build up the last part of their attack. There is no existing feature to capture the potential space here, so one must be developed. The obvious choice when describing space between the midfielder and defender line would be looking at the centroids of both of these lines and the distance between those, but we found that this approach was flawed, as it does not take into account the position of all the players in a line. The players at the extremities of a line can be positioned in different places and still yield the same centroid.

Instead, we chose to develop a feature called PotentialSpace that captures the area the midfielder and defender lines occupy. The computation of this feature is shown in Figure 2, the red players forming the defending team, the area they cover represented with the dotted line. A second version of this feature, PotentialSpace2, computed the convex hull of all players in the midfielder and defender lines, to deal with players that are positioned inside their line. Variants of both these features, PotentialSpaceNorm and PotentialSpace2Norm respectively, were developed in order to overcome a possible problem, namely that the number of players in the defender and midfielder lines is of importance. We hypothesised that, the higher this number of players, the lower the potential space. Therefore we normalised the PotentialSpace and PotentialSpace2 features for the number of players in the midfielder and defender lines. These four features were incorporated into the TacticsPy pipeline and used during the selection of the performance measure as a possible feature to capture potential space.

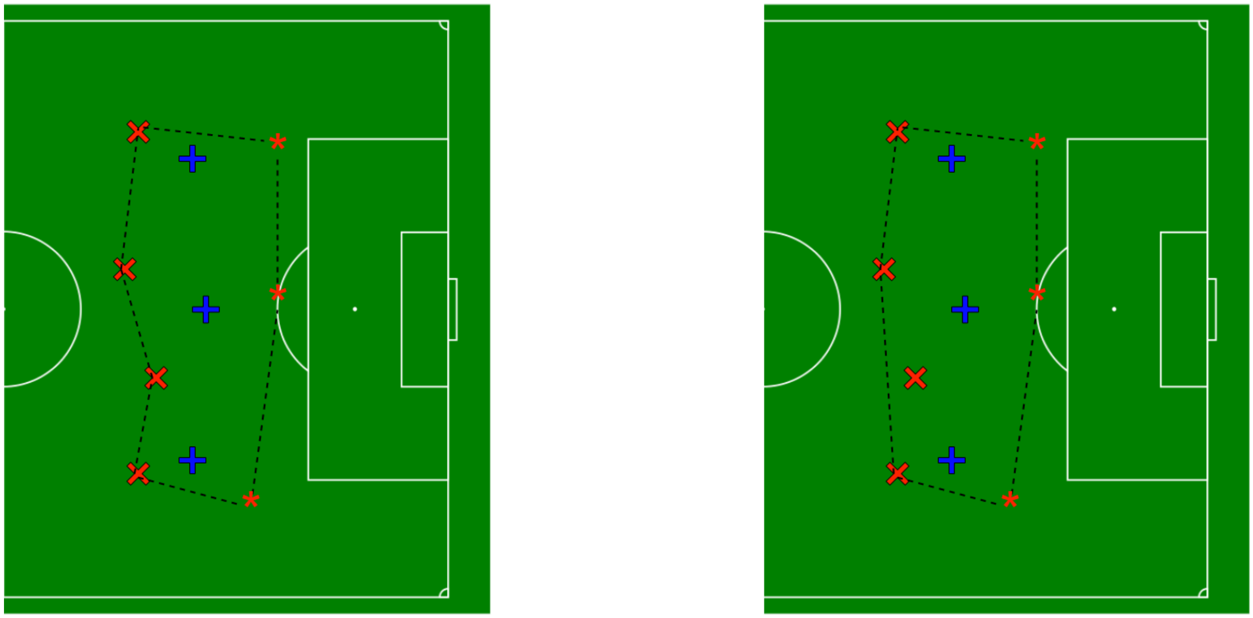

Figure 2: PotentialSpace for the blue attackers (+) is the area within the dotted line in the left scenario, formed by connecting red midfielders (X) and red defenders (*). The right scenario shows PotentialSpace2, calculated using a convex hull to negate the effect players have when they are positioned inside their line. The second midfielder from the bottom is not used to create the area within the dotted line, because he is positioned inside his line.

2.2 Used space

The used space of a team is the amount of space it actively controls. The obvious way to capture this used space is by defining an area around an attacking player and regarding that area as used. Then this area might be added up for all attackers to compute the used space for the whole team. However, it is very difficult to prove that the player actually controls all this space around him, and reason how far this control should reach. Therefore, two other features to capture the used space of a team were chosen, and variants of these were developed to be able to select the best fit for the performance measure later in the research. The first of these two approaches uses aspects of Link's Dangerousity [LLS16], and is explained in section 2.2.1. The second approach uses the distance between an attacker and the centroid of the midfielder or defender line of the defending team, whichever is closest, and is further explained in section 2.2.2.

2.2.1 Pressure

The first approach to capture the space used by a single player is to use the inverse of pressure. When an attacker is under a large amount of pressure, the amount of space they use is very low. Conversely, when the pressure on an attacker is low, they must control (use) a large area. The TacticsPy pipeline incorporates the dangerousity model as defined by Link. This model aims to calculate the chance of the player currently in possession scoring a goal [LLS16]. One of the four parts of this model is pressure, which is calculated using the pressure zone shown in Figure 3.

This pressure zone covers four areas, some bigger than others. These areas result from the angle between the player (P in Figure 3) and the goal. The difference in size of the areas is based on the assumption that a defender (D in Figure 3) located between the player and the goal is more likely to prevent a scoring opportunity than when he is located towards the side of the player with the ball. Within the zones, there is a linear increase in pressure, the closer the defender is to the player.

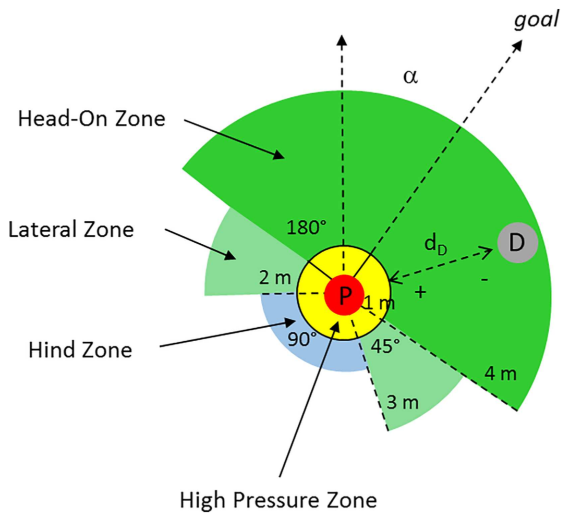

Figure 3: The different pressure zones surrounding a player (marked P). The pressure on the player depends on the distance to a defender (marked D), and the pressure zone this defender is in. The orientation of the pressure zones depends on the angle to the goal. [LLS16]

The pressure model is used to define two features. The first feature, UsedSpaceAvg, is the average pressure value across all attackers, designed to take the individual pressure value of a player and use it to develop a team measure. The second feature, UsedSpaceMax, is the highest pressure value between all attackers, following the reasoning that the player under the highest pressure is the deciding factor in the amount of space a team uses.

2.2.2 Distance to line

The second approach to capture used space uses the centroids of the different player lines. For each time stamp, the centroids of the midfielder and defender lines of the defending team are calculated (denoted A and B in Figure 4, respectively). The `X` position of every player in an attacker role is then compared to the `X` position of these two centroids, and the minimum distance is chosen (shown by the arrows marked C and D for each attacker in Figure 4). The reasoning behind this is that the larger this distance, the more space the attacking team controls. We chose to measure the smallest distance because we wanted to know how much space the team possibly had at its disposal.

An illustration of the calculation for distance to line is shown in Figure 4. Two variants using distance to line have been developed and added to the pipeline. The first one, DistToLineAvg, takes the average distance to line for each attacker. The second feature, DistToLineMax, selects the largest distance to line between all attackers.

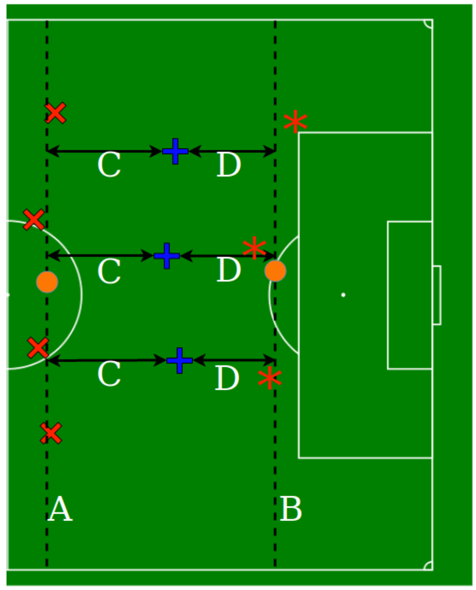

Figure 4: Attacking players are denoted with a blue +, while the defending team's midfielders are shown using a red X and the defending team's defenders are shown with a red *. The line centroids of the defender and midfielder lines are calculated (orange dot) and their X position is shown using the dotted lines. A denotes the X coordinate line for the midfielder line and B denotes the X coordinate line for the defender line. For each attacker, distances to both these lines are calculated, shown with the arrow lines marked C and D. To calculate the DistToLine feature, for each attacker, the shortest arrow (either C or D) is added up.

2.3 Pressure event

Now that several possible features to measure both potential space and used space have been defined, it is time to clearly define the events during which the performance measure is monitored. As this thesis focuses on the progression of the potential space and used space ratio in pressure situations, an event that marks these situations is needed. This event must mark the pressure being exerted by the defending team on the attacking team. Experts at the KNVB have classified four different ballstarts, to which four different pressure types will be connected. Each pressure event will have a certain pressure type, which is derived from the ballstart of that event. The ballstart is decided based on the position of the centroid of the last two defenders of the attacking team (denoted by A and B in Figure 5). A higher ballstart indicates that the offensive team is positioned higher on the pitch, in which case the defensive pressure will also be higher as the offensive team is close to the defending team's goal.

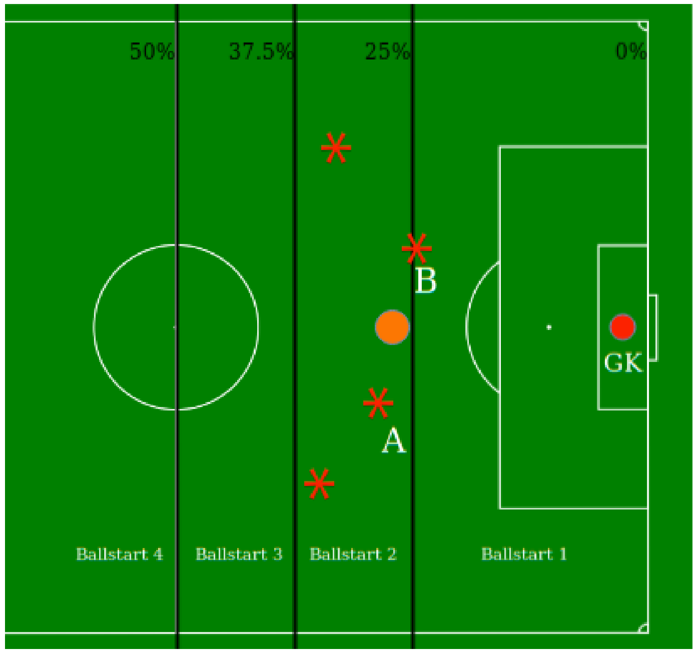

Figure 5: This figure shows a situation in which the red team is attacking towards the left of the pitch. It illustrates the ballstarts, each black line is set on a certain percentage of the pitch length. Ballstart 1 is the first 25%, ballstart 2 is 25% to 37,5%, ballstart 3 is from 37,5% to 50% and ballstart 4 is from 50% upwards. The orange dot denotes the centroid of the last two defenders, which are denoted by A and B. In this situation, the centroid is located in ballstart 2, giving this event pressure type 2.

The borders between zones are expressed in percentages of the pitch length, seen from the defending goal. In Figure 5, the centroid of the last two defenders is illustrated using the orange dot, therefore this situation is ballstart 2, which also means pressure type 2. If the two last defenders were positioned higher, for example when the team is attacking, their centroid might lay in ballstart 3, meaning the attack would be pressure type 3.

3 Methods

This section covers the dataset used in the research, followed by a brief explanation of the Python pipeline used for the analysis of matches. The data mining part of this research is explained, starting with the filters used to select only relevant pressure events, followed by an explanation of the selection of the performance measure that will be monitored during a pressure event. Then, the methods to answer the first main question, namely if the selected performance measure differs per pressure type as defined in section 2.3, will be explained. Finally, the methods for the second question, namely if the performance measure differs for successful and unsuccessful events, are clarified.

3.1 Data

The dataset used in this research contains 106 halves of matches played by the Dutch soccer team. These matches were monitored using the ChryonHego static camera system, after which InMotio software was used to extract player and ball positions with a frequency of 10 Hz. Every match was split into two files containing one half of the match each. Every player and the ball has `X` and `Y` coordinates, acceleration, distance to closest opponent and teammate, and heading.

3.2 Python pipeline

All matches were analysed using the TacticsPy pipeline, an existing Python pipeline that imports positional data from soccer matches, and implements several standard features such as the length and spread of a team. It was further extended with pressure, based on Link's Dangerousity and other new features to describe potential space and distance between lines, as described in section 2. The pipeline allows for a spatio-temporal analysis, where the average values and standard deviations for a feature are created during a given event.

3.3 Data mining

After computing the features and running all matches through the TacticsPy pipeline, a dataset was obtained containing 46992 pressure events as defined in section 2.3. Each of these events contained an average value for all of the features, calculated over the total time span of the event. This average value was used in the analysis.

3.3.1 Filtering

The large number of pressure events obtained from the pipeline had to be filtered further in order to get rid of incomplete or unwanted events in the match. First off, events where either team had less than 11 or 10 players on the pitch were filtered out. Of course, 11 players on both teams would be ideal as it indicates a normal playing situation, but 10 players is also a possibility. In that case, it is possible a team received a red card, or a player was temporarily absent due to an injury. It is important to still take these situations into consideration for our analysis, as they can normally occur in a game.

The second filter stated that, in order to qualify as a real pressure event, a player of the opposing team had to be within 9 meters of the player in ball possession. According to Andrienko et. al. [AAB+18], a player is put under pressure when an opposing player is within this distance of 9 meters. This filter also helps to take out corners and direct free kicks, as there is a minimum distance to the ball of 9.1 meters (10 yards) that must be observed by the players.

The third filter removed all pressure events that lasted longer than 20 seconds. Analysis of the data showed that most events already fell into this rule. When a pressure event took longer than 20 seconds, it was likely a free kick, a goal kick or a moment in time where normal play was halted in order to treat an injury, for example.

The last filter helped to select only pressure events where the defending team was organised in their defensive formation, as events when it is disorganised, for example right after a turnover, are not representative for normal play. These events were filtered out by discarding all pressure events after a turnover, until the attacking team has played the ball in the direction of its own goal. After playing the ball back once, it is assumed the defending team is in their defensive formation again.

3.3.2 Performance measure

The performance measure will be a combination of a potential space measure as defined in section 2.1, and a used space measure, as defined in section 2.2. The exact combination that fits all four pressure types will be decided using a combination of Spearman's Rho correlation and Cortana, a subgroup discovery tool [Dat]. Spearman's Rho was selected because the possible performance measures are all ordinal variables.

For Cortana, two smaller datasets were extracted from the dataset used in the research. The first dataset included all events with potential pressure events using DistToLine measures, while the second dataset included all events with potential pressure events using UsedSpace measures. This was done because of missing data in some events, where the features for DistToLine were available but those for UsedSpace were not, or the other way around. Each potential pressure event had its pressure type listed. Cortana was used to find subgroups of depth 1, with a single nominal target using `WRAcc` as its quality measure. To deal with the multiple labels, we ran Cortana multiple times, each time considering one of the four labels as `true`. For each pressure type, the threshold was calculated with $\alpha < 0.01$. The search strategy used strategy type `beam`, with numeric strategy `best-bins` and number of bins `128`.

3.3.3 Difference between pressure types

The method to analyse the first question of the research, namely if the selected performance measure differs per pressure type, will be answered first using ANOVA or Kruskal-Wallis to test for a significant difference in the performance measure between the four different pressure types. If this first test shows a significant difference between the groups, an independent t-test or a Mann-Whitney test is done between each of the four pressure types, in order to further explore the differences between the groups.

3.3.4 Successful pressure events

The second part of the analysis focuses on the differences between successful pressure events and unsuccessful pressure events. From the perspective of the attacking team, a successful pressure event results in a pass to a teammate in an attacking role, positioned between the midfielder and defender lines of the defending team. Conversely, an unsuccessful pressure event does not result in a pass to an attacking teammate, but in a pass to a teammate in a defender or midfielder role. The defender, midfielder and attacker roles used in the research are based on a k-means clustering with $k = 3$, on the positions of all players excluding the goalkeeper. This clustering is done for each time stamp in the dataset.

For this part of the research, we had to mark each pressure events either successful or unsuccessful according to the definition above. Following explicit advice from experts at the KNVB, pressure events resulting in a turnover were marked neither successful nor unsuccessful, but were discarded. This was done because in this case, it is impossible to know if the turnover occurred as a result of a duel between two players of opposing teams or as a result of a failed cross. If it was the result of a duel, it should be marked unsuccessful. However, if it was the result of a cross (which are naturally risky) that took good advantage of the available space, it should be marked successful. Because it was impossible to correctly determine the nature of such an event, they were discarded. Using the filtered and marked events, it is possible to divide the successful and unsuccessful events per pressure type. Then, using a Mann-Whitney U test for each pressure type, it is possible to compare the successful and unsuccessful events in each group to see if the value of the performance measure differs. This will give information on the importance of creating pressure to achieve a pass to a teammate in an attacker role.

4 Results

This section starts by covering the results of the filters meant to filter out unwanted and irrelevant pressure events. Subsequently, the final performance measure is selected following the methods described in the previous section. This final performance measure will be used to answer the two questions of this research, namely if the value of the performance measure differs for each of the four predefined pressure types, and if there is a difference in the value between successful and unsuccessful events. In order to answer the first question, a test will be done to investigate the value of the performance measure over the four pressure types together, followed by post-hoc tests between pairs of pressure types. For the second question, successful and unsuccessful events are marked, after which tests between successful and unsuccessful events per pressure type are done.

4.1 Filtering

As discussed in section 3.3.1, several filters were applied to the dataset obtained from the TacticsPy pipeline. These filters covered the number of players, the distance to the closest opponent, the maximum duration of the pressure event, and turnover situations. Table 1 shows the number of pressure events in the dataset after each filter was applied.

Table 1: Table displaying the filters used to filter out pressure events, applied in order, with the number of pressure events remaining after applying the filter and the percentage of events of the total that fall under that filter are also listed.

Filter	**Events remaining**	**% of total discarded**
Full dataset	46992	
Opponent within 9m	38959	20.60
No turnover situations	28669	20.49
Maximum duration of 20 seconds	28261	1.47
11 or 10 players per team	28213	0.42

After applying all the filters, the dataset counted 28213 events. This filtered dataset was further used in both parts of the analysis.

4.2 Selecting the performance measure

As mentioned in section 3.3.2, the performance measure will be a combination of one of the measures developed for potential space, and a measure developed to capture used space. After filtering the pressure events to ensure the dataset only contains relevant events, Cortana Subgroup Discovery and Spearman's Rho correlation were used to find the performance measure that fits the four different pressure types.

All possible combinations between potential space and used space measures were calculated. Cortana was used to discover subgroups of depth 1 that best described a certain pressure type. Spearman's Rho correlation was also used, to calculate the correlation between all possible combinations and the four pressure types. Tables displaying the results of both these test are in Appendix

B. By ranking the performance of all possible features based on their Quality for each pressure type, and their correlation, it was possible to combine the Cortana and Spearman results in order to find the performance measure that fit the four pressure types best. Table 2 shows this ranking.

Table 2: Table showing the ranking of the potential performance measures, based on their combined ranks from Cortana and Spearman's Rho correlation rankings.

Rank	**Performance measure**	**Cortana rank**				**Spearman rank**
		Type 1	Type 2	Type 3	Type 4	
1	PotentialSpaceNorm / DistToLineMax	1	3	3	11	1
2	PotentialSpaceNorm / DistToLineAvg	3	5	2	9	2
3	PotentialSpace / DistToLineMax	2	7	1	10	3
4	PotentialSpace / DistToLineAvg	4	8	4	7	4
5	PotentialSpaceNorm × UsedSpaceMax	5	12	6	5	5
6	PotentialSpace × UsedSpaceMax	6	13	10	8	6
7	PotentialSpace2 / DistToLineMax	10	1	5	2	11
8	PotentialSpace2Norm / DistToLineMax	9	4	8	4	10
9	PotentialSpaceNorm × UsedSpaceAvg	7	14	11	6	7
10	PotentialSpace × UsedSpaceAvg	8	15	12	12	8
11	PotentialSpace2 / DistToLineAvg	14	2	7	1	14
12	PotentialSpace2Norm / DistToLineAvg	11	6	9	3	13
13	PotentialSpace2Norm × UsedSpaceAvg	13	9	-	-	15
14	PotentialSpace2 × UsedSpaceAvg	12	10	-	-	16
15	PotentialSpace2Norm × UsedSpaceMax	16	11	14	13	9
16	PotentialSpace2 × UsedSpaceMax	15	16	14	-	12

The formula used to calculate the final ranking uses the Cortana rank for each of the four pressure types, divided by the number of pressure types that particular performance measure was ranked in, and adds the Spearman rank. The lower this result, the better the ranking. The ranking shows that the combination *PotentialSpaceNorm / DistToLineMax* has the best combined rank. *PotentialSpaceNorm* is a potential space feature that calculates the area that the midfielders and defenders of the defending team together cover, divided by the amount of defensive players that make up this area, and *DistToLineMax* is a DistToLine variant that takes the maximum distance of one of the attacking players to either the defender or midfielder centroid. Further discussion of the performance measure results and the meaning behind the selected measure can be found in section 5.1.

4.3 Differences between the four pressure types

After selecting *PotentialSpaceNorm / DistToLineMax* as the performance measure in the previous section, it is now time to investigate the first question, namely whether the value of this performance measure differs between the four pressure types. The go-to test for comparing a performance value over 3 or more different groups is a one-way ANOVA [HN09]. ANOVA has three assumptions, and our data failed the normality assumption and the homogeneity of variance assumption, as tested with the Shapiro-Wilk test [SW65] and Levene's test [Lev60] respectively. Appendix C goes into further detail on these tests and their results. As two of the three assumptions for ANOVA failed, it was decided to instead choose the Kruskal-Wallis H test [KW52], the nonparametric alternative to ANOVA.

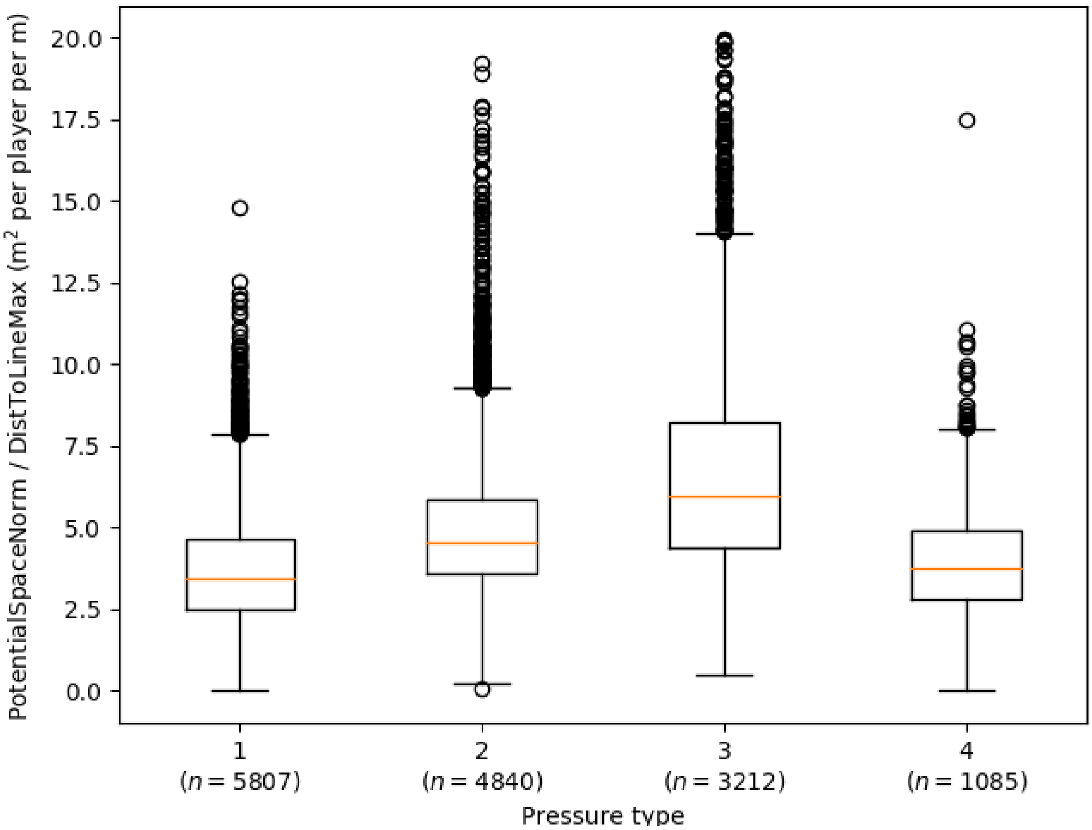

Figure 6: Boxplot displaying the performance measure per pressure type. The pressure type is based on the ballstart, and represents the location on the pitch of the centroid of the last two defenders of the attacking team (See also Figure 5). The performance measure is a ratio consisting of the surface area formed by the midfielders and defenders of the defending team, normalised for the amount of players (m^2) and a feature that captures the maximum distance between an attacker to either the defender or midfielder line of the defending team (m). 71 outliers in pressure type 2 and 3 are not displayed.

Figure 6 displays a boxplot, showing the *PotentialSpace / DistToLineMax* performance measure, divided over the four pressure groups. Pressure type 3 is the highest, followed by pressure type 2. Types 1 and 4 are relatively similar. It also shows an upward trend throughout pressure types 1, 2 and 3. However, this trend falls off with pressure type 4. The Kruskal Wallis H test resulted in a H value of 2921.57, which gives a p value $< .001$. The test showed that there is a significant difference in the performance measure between the four pressure types. The next step is finding out if this difference exists between every combination of two pressure types, or only a few. The usual post-hoc test for Kruskal-Wallis is a Mann-Whitney U test [MW47] to give us further details by comparing all the groups separately. All the tests between different pressure types resulted in U values leading to a p value of $< .001$, meaning all pressure types differ significantly. A table with the precise results can be found in Appendix D.

4.4 Differences between successful and unsuccessful events

This section covers the results gathered while investigating the second question of this research, namely whether the value of the performance measure, *PotentialSpaceNorm / DistToLineMax*, differs between successful and unsuccessful events. In order to be able to investigate this question, first we had to label pressure events as successful or unsuccessful. This was done by investigating all events in the filtered dataset, the same dataset used to investigate the first question in the previous section. All pressure events involving a midfielder or defender being put under pressure were analysed. If the next pressure event involved a player in an attacker role, it was assumed a pass had occurred. These pressure events where thus marked successful. If a pass was made to another midfielder or defender, it was marked unsuccessful. If a pressure event was the last in a certain team possession sequence, the pressure event was marked neither successful nor unsuccessful, because it could not be ascertained if the ball was lost in a duel or a failed cross.

This resulted in a filtered dataset that contained 1799 successful events and 11106 unsuccessful events. The remaining 2247 events were excluded because their nature could not be determined. The corresponding percentages of successful and unsuccessful events are shown in Table 3.

Table 3: Percentage of successful and unsuccessful passes per pressure type.

Pressure type	**% successful**	**% unsuccessful**
1	13.6	86.4
2	15.3	84.7
3	13.5	86.5
4	11.5	88.5

To investigate whether the performance measure differs for successful and unsuccessful pressure events within one pressure type, further investigation using Mann-Whitney U tests between the successful and unsuccessful groups was carried out. These tests showed that the performance measure did not differ significantly for pressure types 1, 2 and 4. However, for pressure type 3 the Mann-Whitney U test showed that the value of the performance measure *PotentialSpaceNorm / DistToLineMax* was significantly higher for unsuccessful events (Mdn = 5.97) than for successful events (Mdn = 5.59), U = 380869, $p = 0.024$.

5 Discussion

This section first discusses the observations found in the selection of the performance measure, and what the selected performance measure might indicate. Then, the results of the statistical test and post-hoc tests for the differences in the performance measure between the four pressure types are discussed. Subsequently, the results of the selection of successful and unsuccessful events and subsequent statistical tests for differences in the performance measure between successful and unsuccessful events are investigated. Lastly, future work is discussed, proposing improved features and new interesting research possibilities.

5.1 Performance measure

The performance measure that was eventually decided on was *PotentialSpaceNorm / DistToLineMax*, found using a combination of Spearman's Rho correlation ranks and Cortana ranks. This combination of two techniques used means that both the Spearman's Rho results and the Cortana results must be examined in this section.

Judging from the correlation values obtained from Spearman's Rho in Table 6, one of the combinations involving PotentialSpace or PotentialSpaceNorm, and either DistToLineMax or DistToLineAvg would be the performance measure that fits the four pressure types best. The fact that PotentialSpaceNorm variants perform slightly better than PotentialSpace variants indicates that the individual potential space for midfielders and defenders is more important than the potential space when looking at the team. The fact that DistToLineMax performs better than DistToLineAvg means that the maximum space an individual attacker has is more important than the average space all attacking players have.

There is also a very big difference between potential performance measures using the PotentialSpace variants and PotentialSpace2 variants, when combined with a DistToLine feature. This is possibly because PotentialSpace2 does not always select all relevant players, simply due to the nature of a convex hull. This problem cannot be tackled using the current dataset, because it only contains the three basic player roles. For example, the current approach works best if a defending team is using a very linear formation, where every player stays in its own line. However, when this is not the case, selecting the correct players automatically becomes more difficult. This problem could be solved in two ways. The simpler way would involve adding more player roles, so formations with multiple lines might also be detected easier. The second approach would consider the complete defensive formation of the defending team at the time of a pressure event. It would be easier to select the players that actually form the space we are trying to capture, instead of simply selecting them based on their player role.

The correlation values also clearly show how ineffective the UsedSpace features were. This could simply be because they were combined with PotentialSpace features, but a more likely option is that pressure based features are not a good way to capture the space a player is occupying.

The Cortana results in Table 5 support the observations gathered from Spearman's Rho correlation values. Again, there is a big gap between PotentialSpace and DistToLine variants and the other potential performance measures. UsedSpace features again bring up the rear, with a big difference in quality from the top potential performance measures.

An interesting observation is that the quality of the performance measure in the case of type 4 pressure was much lower compared to the quality of the performance measures for the other pressure types, especially type 1 and type 3. This could simply be due to the fact that, compared to the three other types, there were relatively few observations of type 4 pressure in the dataset. However, further research would have to investigate this. Another interesting issue is that PotentialSpace variants perform much better in pressure type 1 than in pressure type 4, where PotentialSpace2 measures show better performance. This could be due to the fact that in pressure type 4 situations, the attacking team is much more bunched up. The convex hull approach used in PotentialSpace2 performs better in these situations.

Based on the combined rankings from Spearman's Rho correlation and Cortana, as shown in Table 2, the performance measure eventually used in the analysis was *PotentialSpaceNorm / DistToLineMax*. This feature captures the potential space per opposing player, divided by the largest distance from any player in an attacker role to the centroid of the midfielder or defender line of the opposing team. The lower the performance value, the better the attacking team utilises the available space.

5.2 Difference between pressure types

The question whether the value of all performance measure differs between the four pressure types was investigated by performing a Kruskal-Wallis H test to test the overall differences, followed by Mann-Whitney U tests serving as post-hoc tests.

Figure 6 displays a box-plot containing the performance measure per pressure type. Pressure type 3 has the highest value, followed by pressure type 2. Types 1 and 4 look roughly similar. At first glance, there is a clear upward trend through pressure type 1, 2 and 3. However, this quickly ends at pressure type 4. The upward trend makes sense, as higher values for the performance measure mean the attacking players are not utilising their potential space very well. In pressure types 1 and 2 they perform relatively well, but in pressure type 3 situations they are not very efficient at utilising the potential space. This could be because, in these situations, the players in an attacker role are positioned very close to the defender line of the opponent, often waiting for a pass beyond the offside line. This means there is a lot of potential space between the defender and midfielder lines that is not utilised. If the attacking team left some players more centered in this potential space, they could possibly use that position to then further their attack. The low values for the performance measure in type 4 situations mean that the attackers are good at utilising their potential space. In these situations, the centroid of the last two defenders is positioned above the centre of the field, meaning the rest of the team is bunched up towards the opponent's goal. These situations do not leave a lot of potential space, meaning a low performance measure is easily acquired here.

The Kruskal-Wallis H result concludes that there is a significant difference between the four groups, as expected by simply observing the box-plot. The post-hoc tests involved Mann-Whitney U tests for each pair of pressure types. As expected, these tests also concluded that all groups were not possibly related, as shown in Appendix D. The U value between pressure types 3 and 4 was much smaller than the results between other pressure types. Again, this becomes clear when looking at the box-plot, which also shows a large difference between the two types.

5.3 Difference between successful and unsuccessful events

The second question that was investigated was whether the performance measure differed between successful and unsuccessful events. When marking a pressure event, we looked at the next player in control, and if this player was in an attacking role or not. In this research, we disregarded all events where we could not be sure if possession was lost in a duel or in a failed cross. It could be argued that any case where possession is lost would be marked unsuccessful, but it is also possible that the ball was lost following a cross to a player who would then have attempted a shot on goal. In this case, making this cross attempt was the only option, meaning the team did everything right up until that point and it would be incorrect to mark the event as unsuccessful. To avoid this problem, events leading to a turnover were disregarded in this research.

As illustrated in Table 3, the pressure type was not indicative of the pressure event being successful or unsuccessful. This could be because, in the lower pressure types when both teams are spread out, it is easier to make long passes to your attackers, as they have more room to position themselves. When there is a situation with pressure type 3 or 4, both teams are bunched up, making all passing lines shorter, which can also mean passes are easier to make.

However, using Mann-Whitney U tests to analyse the performance measure for successful and unsuccessful events per pressure type, pressure type 1, 2 and 4 seem to be similar, in that the difference in performance measure for these pressure types was not significant when considering if an event was successful or unsuccessful. The result of this same test for pressure type 3 however, shows that there was a significant difference between the successful and unsuccessful events when considering the performance measure. This indicates that the value of the performance measure is indeed important in pressure type 3 situations, meaning that future research into successful and unsuccessful pressure events should focus on pressure type 3 situations, to find out what aspect could further explain whether an event is successful or not. For example, whether an event is successful or not could depend on a player centred in the potential space area on the pitch, as suggested in the previous section.

5.4 Future work

Several subjects can be explored further in order to gain more detailed knowledge or discover new areas of space in team-sports. This involves further exploration of the current features, improving new features and looking at other aspects, such as considering the area of the whole team or the defensive formation of a team.

5.4.1 Describing space in team sports

The increased availability of positional data across multiple fields of sport could mean the modelling approach described in this thesis could be very useful to describe space in any team sport. The approach of developing a measure for potential space, one for used space and investigating this ratio at a certain event might form a good basis to investigate the influence of pressure or any other factor that can be captured in an event, on the utilisation of potential space by the attacking party.

5.4.2 Improving features

The results showed that there was a big difference between the effectiveness of DistToLine features and UsedSpace features. While the DistToLine features proved the best option in the end, the UsedSpace features were discarded, because of their low correlation to the four pressure types.
The idea behind UsedSpace was designing a feature that accurately captured the personal room a player possessed, using a proven element from Dangerousity [LLS16]. Another possibility, one that was explored early in the research but discarded later on, was using a method to divide an area in subareas for each player, using a Voronoi approach [RRPM15]. This would provide a clear division of the available space for each player in an attacker role, and could give great insight into the transitioning of space possession between the two teams during a pressure event. Using a Voronoi approach would also give insight into the influence of individual players, which is an area not investigated in this thesis.

5.4.3 Other factors

If the transition to more individual space oriented features is to be made, another area that might be of importance is the area of the whole team. During pressure type 4, both teams are very bunched up, meaning the current features might be less effective. As the area of the whole team will not suffer from this effect, it might be worth further research.

During a pressure event, the formation of the defensive team influences the amount of space it leaves. Further research might aim at creating a 'fingerprint' of a defensive formation, and compare this fingerprint in different situations in an effort to explore the importance of the defensive formation for the amount of space a team potentially leaves open.

5.4.4 Successful events

The results showed a significant difference between the performance measure in successful and unsuccessful events for pressure type 3, as noted in section 5.3. Future work might dive deeper into these events with pressure type 3, in order to find out what the concrete differences between successful and unsuccessful events in this area are. Ideally, this would be combined with a new set of features, in order to capture more information of the team as a whole in these situations.

6 Conclusion

This research set out to create several features in order to capture the potential space an attacking team has to set up attacks and play in, and to accurately capture the space the players in an attacker role are actually using. The aim was to create a performance measure using two of these features and use this performance measure to investigate two questions, namely whether the performance measure differed between four predefined pressure types, and if the performance measure differed between successful and unsuccessful events.

The four pressure types were set up by KNVB analysts. In cooperation with the KNVB, a pressure event was defined using distances to opponents and several other criteria, such as the duration of an event. The different features that captured potential and used space were combined and tested, in order to find a performance measure that fit all four pressure types. The performance measure that was selected based on these tests was *PotentialSpaceNorm / DistToLineMax*. The fact that a normalised version performed better than the normal PotentialSpace feature indicates that the individual potential space per player is more important than the total potential space for the whole team. Similarly, DistToLineMax performing better than DistToLineAvg indicates that the maximum space an attacker has is more important than the average space all attackers have. These two factors can be important for training and coaching matches, in order to allow your team to use as much space as possibly available.

When investigating the value of the performance measure to see if it differed over all four pressure types, we found that there was a significant difference between these four groups. Using post-hoc tests, we concluded that the values of the performance measure between the four individual pressure types also differed significantly. The research found that the value of the performance measure was highest for pressure type 3 situations, indicating that in those cases, attackers were often not utilising all the space they potentially had. This could be due to the fact that the attackers are often positioned very closely to the opposing defender lines in these situations. It is possible that the attackers could improve their play in these situations by positioning a player away from the defender line, and use him as a receiver or playmaker as he has the most room available.

The second part of the research focused on investigating the value of the performance measure in successful and unsuccessful events. During this research, it became clear that the percentage of successful pressure events did not differ per pressure type, but greatly remained the same. Then, the successful and unsuccessful events per pressure type were compared. This showed that there was only a significant difference in the value of the performance measure between successful and unsuccessful events in situations with pressure type 3. The difference between a successful event and an unsuccessful event could be the positioning of a player away from the defender line to receive a pass, for example.

Further research might focus on developing features largely based on the ones in this research, or might be based on the area of the whole team, using an implementation similar to Voronoi diagrams. The events with pressure type 3 could also be subjected to further investigation, in an attempt to find which factors are responsible for the significant differences between successful and unsuccessful events.

References

[AAB+18] Gennady Andrienko, Natalia Andrienko, Guido Budziak, Tatiana von Landesberger, and Hendrik Weber. Exploring pressure in football. In *Proceedings of the 2018 International Conference on Advanced Visual Interfaces - AVI 18*. ACM Press, 2018.

[Dat] Data Mining Group, LIACS, Leiden University. Cortana subgroup discovery.

[FB16] Javier Fernandez and Luke Bornn. Wide open spaces: A statistical technique for measuring space creation in professional soccer. *MIT Sloan Sports Conference*, 2016.

[HN09] Richard M. Heiberger and Erich Neuwirth. One-way ANOVA. In *R Through Excel*, pages 165–191. Springer New York, 2009.

[KW52] William H. Kruskal and W. Allen Wallis. Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260):583–621, December 1952.

[Lev60] H. Levene. Robust tests for equality of variances. *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, pages 278–292, 1960.

[LLS16] Daniel Link, Steffen Lang, and Philipp Seidenschwarz. Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLOS ONE*, 11(12):e0168768, December 2016.

[MLS16] Daniel Memmert, Koen A. P. M. Lemmink, and Jaime Sampaio. Current approaches to tactical performance analyses in soccer using position data. *Sports Medicine*, 47(1):1–10, June 2016.

[MW47] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *The Annals of Mathematical Statistics*, 18(1):50–60, March 1947.

[RRPM15] Robert Rein, Dominik Raabe, Jürgen Perl, and Daniel Memmert. Evaluation of changes in space control due to passing behavior in elite soccer using voronoi-cells. In *Proceedings of the 10th International Symposium on Computer Science in Sports (ISCSS)*, pages 179–183. Springer International Publishing, November 2015.

[SMA+14] Hugo Sarmento, Rui Marcelino, M. Teresa Anguera, Jorge CampaniÇo, Nuno Matos, and José Carlos LeitÃo. Match analysis in football: a systematic review. *Journal of Sports Sciences*, 32(20):1831–1843, May 2014.

[SW65] S. S. SHAPIRO and M. B. WILK. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4):591–611, December 1965.

Appendix

A Developed features

This section lists all relevant features added during the project, with a short description.

Table 4: All relevant features added to the TacticsPy pipeline.

Feature	**Description**
PotentialSpace	Area of all the players in the defender and midfielder roles (m^2).
PotentialSpaceNorm	PotentialSpace divided by the number of players it consists of (m^2).
PotentialSpace2	Convex hull of all the players in the defender and midfielder roles (m^2).
PotentialSpace2Norm	PotentialSpace2 divided by the number of players it consists of (m^2).
UsedSpace	Summed Dangerousity pressure of all players in attacker role (0 to 1).
UsedSpaceAvg	UsedSpace divided by the number of players in attacker role (0 to 1).
UsedSpaceMax	Maximum pressure between all players in attacker role (0 to 1).
DistToLine	Smallest distance to the centroid of either the midfielder or defender line.
DistToLineAvg	Average DistToLine between all players in attacker role (m).
DistToLineMax	Maximum DistToLine between all players in attacker role (m).

B Selecting the performance measure

This section contains two tables, listing the full results for Cortana and Spearman's Rho correlation for finding the best performance measure. The **Prior** for any DistToLine based performance measures shown in Table 5 are: 5662 for pressure type 1, 4840 for pressure type 2, 3212 for pressure type 3, and 1085 for pressure type 4. The **Prior** for the UsedSpace based measures are: 5662 for pressure type 1, 4356 for pressure type 2, 2804 for pressure type 3 and 868 for pressure type 4.

Table 5: Overview of potential performance measures found using Cortana ($p < 0.01$), ranked based on the Quality (*WRAcc*).

Rank	**Coverage**	**Quality**	**Posterior**	**Performance measure**
				Pressure type 1
1	6072	0.0727	56.75%	PotentialSpaceNorm / DistToLineMax
2	7589	0.0721	53.07%	PotentialSpace / DistToLineMax
3	7940	0.0715	52.32%	PotentialSpaceNorm / DistToLineAvg
4	8056	0.0699	51.82%	PotentialSpace / DistToLineAvg
5	6520	0.0411	50.02%	PotentialSpaceNorm × UsedSpaceMax
6	5558	0.0339	50.01%	PotentialSpace × UsedSpaceMax
7	7696	0.0288	46.51%	PotentialSpaceNorm × UsedSpaceAvg
8	6627	0.0255	46.68%	PotentialSpace × UsedSpaceAvg
9	2569	0.0251	53.45%	PotentialSpace2Norm / DistToLineMax
10	3386	0.0224	48.73%	PotentialSpace2 / DistToLineMax
11	2336	0.0223	53.13%	PotentialSpace2Norm / DistToLineAvg
12	9192	0.0201	44.38%	PotentialSpace2 × UsedSpaceAvg
13	8764	0.0201	44.60%	PotentialSpace2Norm × UsedSpaceAvg
14	2336	0.0167	49.53%	PotentialSpace2 / DistToLineAvg
15	11115	0.0114	42.79%	PotentialSpace2 × UsedSpaceMax
16	11757	0.0111	42.68%	PotentialSpace2Norm × UsedSpaceMax
				Pressure type 2
1	11208	0.0363	37.23%	PotentialSpace2 / DistToLineMax
2	11675	0.0333	36.65%	PotentialSpace2 / DistToLineAvg
3	9691	0.0321	37.32%	PotentialSpaceNorm / DistToLineMax
4	10975	0.0307	36.57%	PotentialSpace2Norm / DistToLineMax
5	10508	0.0305	36.72%	PotentialSpaceNorm / DistToLineAvg
6	11909	0.0295	20.98%	PotentialSpace2Norm / DistToLineAvg
7	10158	0.0254	36.12%	PotentialSpace / DistToLineMax
8	10041	0.0230	35.81%	PotentialSpace / DistToLineAvg
9	6734	0.0161	35.05%	PotentialSpace2Norm × UsedSpaceAvg
10	7909	0.0144	34.25%	PotentialSpace2 × UsedSpaceAvg
11	6734	0.0110	34.01%	PotentialSpace2Norm × UsedSpaceMax
12	10474	0.0095	33.01%	PotentialSpaceNorm × UsedSpaceMax
13	11222	0.0091	32.88%	PotentialSpace × UsedSpaceMax
14	11971	0.0087	32.76%	PotentialSpaceNorm × UsedSpaceAvg

15	11436	0.0085	32.79%	PotentialSpace × UsedSpaceAvg
16	5237	0.0081	33.89%	PotentialSpace2 × UsedSpaceMax

Pressure type 3

1	5488	0.0627	38.58%	PotentialSpace / DistToLineMax
2	4670	0.0622	41.41%	PotentialSpaceNorm / DistToLineAvg
3	5838	0.0622	37.43%	PotentialSpaceNorm / DistToLineMax
4	5604	0.0608	37.71%	PotentialSpace / DistToLineAvg
5	4437	0.0297	31.51%	PotentialSpace2 / DistToLineMax
6	7161	0.0293	26.10%	PotentialSpaceNorm × UsedSpaceMax
7	4203	0.0288	31.74%	PotentialSpace2 / DistToLineAvg
8	4787	0.0279	30.21%	PotentialSpace2Norm / DistToLineMax
9	5721	0.0257	28.21%	PotentialSpace2Norm / DistToLineAvg
10	7802	0.0251	24.90%	PotentialSpace × UsedSpaceMax
11	5665	0.0224	25.93%	PotentialSpaceNorm × UsedSpaceAvg
12	6974	0.0195	24.34%	PotentialSpace × UsedSpaceAvg
13	1924	0.0087	26.66%	PotentialSpace2Norm × UsedSpaceMax
14	2672	0.0068	23.99%	PotentialSpace2 × UsedSpaceMax

Pressure type 4

1	4904	0.0334	17.43%	PotentialSpace2 / DistToLineAvg
2	5721	0.0311	15.38%	PotentialSpace2 / DistToLineMax
3	6188	0.0298	14.45%	PotentialSpace2Norm / DistToLineAvg
4	5955	0.0281	14.32%	PotentialSpace2Norm / DistToLineMax
5	4489	0.0135	10.47%	PotentialSpaceNorm × UsedSpaceMax
6	5130	0.0109	9.26%	PotentialSpaceNorm × UsedSpaceAvg
7	10508	0.0108	8.79%	PotentialSpace / DistToLineAvg
8	4917	0.0107	9.33%	PotentialSpace × UsedSpaceMax
9	8990	0.0107	9.03%	PotentialSpaceNorm / DistToLineAvg
10	10391	0.0099	8.68%	PotentialSpace / DistToLineMax
11	10508	0.0099	8.67%	PotentialSpaceNorm / DistToLineMax
12	6092	0.0090	8.36%	PotentialSpace × UsedSpaceAvg
13	1069	0.0037	11.13%	PotentialSpace2Norm × UsedSpaceMax

Table 6: Full Spearman's Rho correlation results for DistToLine and UsedSpace measures.

Rank	**Performance measure**	**Spearman's Rho correlation**
1	PotentialSpaceNorm / DistToLineMax	0.339
2	PotentialSpaceNorm / DistToLineAvg	0.337
3	PotentialSpace / DistToLineMax	0.322
4	PotentialSpace / DistToLineAvg	0.319
5	PotentialSpaceNorm × UsedSpaceMax	0.223
6	PotentialSpace × UsedSpaceMax	0.175
7	PotentialSpaceNorm × UsedSpaceAvg	0.171
8	PotentialSpace × UsedSpaceAvg	0.139
9	PotentialSpace2Norm × UsedSpaceMax	0.001
10	PotentialSpace2Norm / DistToLineMax	0.000
11	PotentialSpace2 / DistToLineMax	0.00
12	PotentialSpace2 × UsedSpaceMax	-0.010
13	PotentialSpace2Norm / DistToLineAvg	-0.013
14	PotentialSpace2 / DistToLineAvg	-0.017
15	PotentialSpace2Norm × UsedSpaceAvg	-0.044
16	PotentialSpace2 × UsedSpaceAvg	-0.049

C ANOVA assumptions

In this section, the two failed assumptions for a one-sided ANOVA are discussed.

Normality check

The normality check was performed using a Shapiro-Wilk test. The null-hypothesis for this test states that the data is drawn from a normal distribution. The test statistic was 0.770, with a p value of $< .0001$. This means that the null-hypothesis can be discarded, which means the data was not drawn from a normal distribution. The QQ plot in Figure 7 illustrates this.

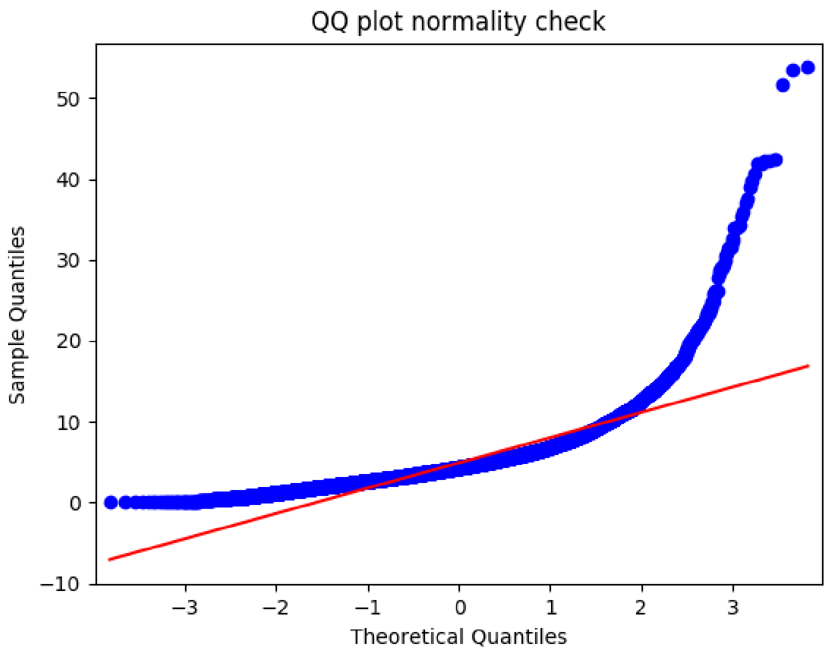

Figure 7: QQ plot showing the distribution of the data.

Homogeneity of variance check

In order to test the homogeneity of variance, Levene's test was used. This test was chosen as an alternative to Bartlett's test because of the non-normal distribution of the data.
The test statistic was 371.299, with a corresponding p value of 1.326e-232. Thus, the null-hypothesis that all four pressure type groups are from populations with equal variances can be discarded.

D Post-hoc tests for differences between the pressure types

This table features the results of the Mann-Whitney U tests between all four individual pressure type groups, listing their median, the resulting U value and corresponding p value. All tests resulted in a p value $< .001$.

Table 7: Results of Mann-Whitney U tests between different groups and their p values.

Groups	**median group X**	**median group Y**	**U**	**p value**
1 and 2	3.47	4.57	8888466	$< .001$
1 and 3	3.47	6.03	3520995	$< .001$
1 and 4	3.47	3.76	2860999	$< .001$
2 and 3	4.57	6.03	5019190	$< .001$
2 and 4	4.57	3.76	1863021	$< .001$
3 and 4	6.03	3.76	740519	$< .001$