TOWARDS QUANTIFYING INDIVIDUAL SPACE DOMINANCE IN FOOTBALL MATCHES

Alwin Tinneveld, 6807097
MSc. Thesis Applied Data Science

JULY 7, 2024

Utrecht University, Department of Information and Computing Sciences
Forward Football B.V.

Table of Contents

Introduction

In the world of competitive sports, individual performance often contributes in a critical way to the success of a team. Whether it is a football player dictating the pace of the game or a basketball player carrying the team through their scoring ability, understanding and quantifying individual player dominance can provide important insights for both coaches and the players themselves.

In a football match specifically, a team generally wants to use the space on the pitch such that advantageous positions or outcomes are more probable, both offensively and defensively. Therefore, ideally, a football team would strive for spatial control and cooperative movement, such that they are the more dominant team position-wise. To be spatially dominant as a team means to have a bigger dominant region. This concept was first introduced by Taki et al. in 1996. Taki et al. described a players' dominant region as a region that they can reach before any other player. When the dominant regions of all players on one team are combined, we talk about the team's dominant region. In their research, Taki et al. calculate the dominant region of a player with a Voronoi diagram approach, where each player has their own tessellation in which they are believed to be the dominant player. Other research has also shown that Voronoi diagrams can be used to characterize players' spatial interaction in Futsal (Fonseca et al. 2012) or how Voronoi diagrams have been used to find dominant player areas in different sports like Hockey (Fujimura & Sugihara, 2005). Other approaches to quantifying in-game space-control were done through machine learning (Gu et al., 2024) and through a space evaluation framework based on field weighting (Narizuka et al., 2021). Finally, there has also been research into the generation of space in football matches (Michalak, 2022). Despite much research done in the area of team dynamics and overall match performance, there remains a gap in the literature of individual player dominance in football.

In collaboration with Forward Football, this thesis aims to bridge the gap of individual player dominance in football by exploring football match data of individual players. Factors such as player coordinates, ball coordinates, angles, speed and distance to the ball will contribute to space control scores, which in turn will be used to quantify individual space dominance. The calculation of individual space dominance will be elaborated further in the method section of this thesis.

In this thesis, we seek to answer the following research question:
How can we quantify and assess the amount of space dominance exhibited by individual players at different timestamps during a match?

To address this question, this thesis will explore the established literature on space control and space dominance. Next to that, already existing code on space dominance at team level provided by Forward Football will be used to form the basis to determine space dominance at the individual level. Ultimately, a new, quantitative way of expressing space dominance at the individual level will be proposed.

Literature review

In the world of football, much research has been done in the last few years with regard to identifying key performance indicators (KPI) (Hughes et al., 2012). These are variables that have been associated with the performance and success of football teams (Jamil et al., 2021).

One of the KPIs that get a lot of attention in terms of game analysis in football is ball possession (Casal et al., 2017). To score a goal, a team is usually in possession of the ball. Much research has already been done to find a correlation between the possession of the ball and team success (Jones et al. 2004). And while some found positive outcomes (Hook and Hughes, 2001; Bloomfield et al., 2005), others found there was not enough evidence to support this (Stanhope, 2001).

It should be noted, however, that possession alone will usually not be sufficient enough. Possession is useful if utilized correctly. By utilizing possession, teams can advance into more beneficial scenarios, which in turn can yield goals. One such way of utilizing possession is through space (pitch) control. Space control is an interesting metric for off-ball qualities in football players that has been associated with both goalscoring and the probability of winning a match (Rein et al., 2017). In short, space control is the probability that a team (or player) will control the ball at location x, given the current configuration of players and the ball on the pitch. Before delving into space control, we should first investigate the player influence area, from which space control can be derived.

Player Influence Area

If we look at the individual level in football games, each team has players that play in certain positions (attack, midfield, defense, etc.). Based on their positions, they have a certain influence on the game. For example, a midfielder has high influence in the central part of the pitch, as their positioning is often close to the center. When the ball is close to this player, their chances of getting possession are much higher than when the ball is far away. Through this kind of reasoning, we can think of the area around this player as an area in which they have a certain probability of reaching the ball first. The closer the player is to the ball, the higher the odds of getting in possession. Apart from the distance to the ball, speed and direction are also factors that are important for this area in which they can reach the ball first. If the player is running towards the ball, their chances of getting the ball are much higher than if the player is moving slowly. If they run in a different direction, their chances shrink as well.

To model this phenomenon, Fernandez & Bornn (2018) defined the player influence area through a bivariate normal distribution, whose shape is adjusted for the player's location, speed, direction and distance to the ball. This leads to a distribution in which the full scale of the pitch is given a degree of influence between 0 and 1 for this particular player. Where an influence score close to 1 indicates a high probability of this player getting the ball, whereas an influence score close to 0 indicates it is nearly impossible for the player to get the ball. This could happen, for example, when we calculate the influence area of a defender while their team is attacking and they stay behind. The ball is too far away to be the first to reach it and therefore the probability of getting the ball from the current position, given the ball position, is nearly 0. The area in which a player has influence and therefore probability to get the ball is of importance when thinking about space control, as this area illustrates the control a certain player has in the area around them. The influence area is illustrated in figure **1**, which shows the influence areas of a stationary (left) and moving (right) player. The ball is illustrated as a green dot in this figure.

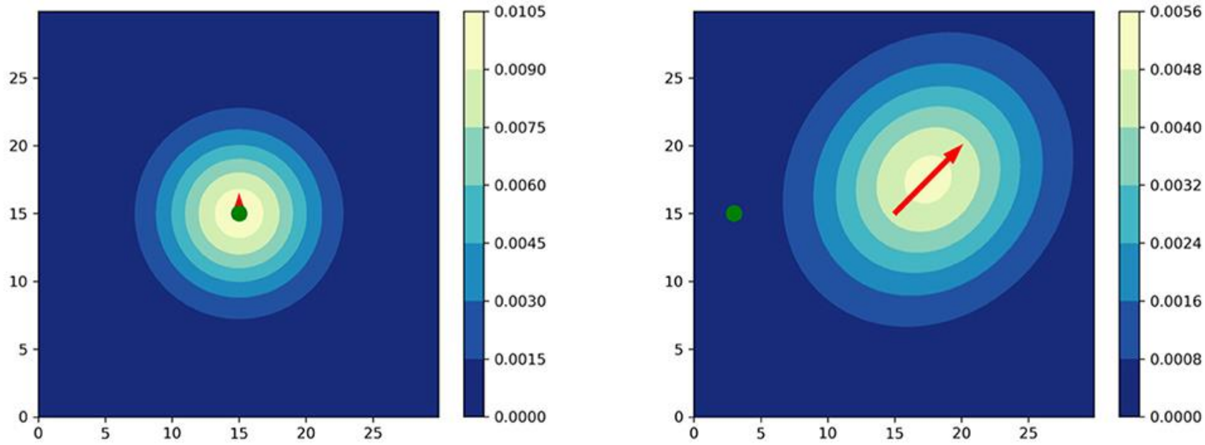

Figure 1: The Player Influence Area

Retrieved from "Space and Control in Soccer" by F. Martens, U . Dick & U. Brefeld, 2021, Frontiers in Sports and Active Living, 3 (https://doi.org/10.3389/fspor.2021.676179).

Space Control

In terms of a definition for space (or pitch) control, Fernandez & Bornn (2018) proposed the following:

Space control can be defined as the degree or probability of control that a given player (or team) has on any specific point in the available playing area.

The main difference between this and the player influence area is that the influence areas focus on individual impact, whereas space control deals with the broader concept of quantifying control over specific areas of the field by considering multiple players' positions and actions.

The authors also state that given the growing use of player tracking devices in sports, there are many possible models for space control. One of the widely used models of space control is a physics-based model that uses Voronoi diagrams, which uses every player's position and calculates their dominant space by comparing their position to each coordinate on the pitch and assigning the closest player to each coordinate. This way, the football pitch will be divided into regions such that any point within a region is closest to one player. A Voronoi diagram of a 4-3-3 football formation is shown in figure **2**, where each red dot represents a player and the blue lines represent the borders between each dominant area. Early applications using this approach were made by Taki et al. (1996), who did not only incorporate standard Euclidean distance, but also included a constant velocity function for reachability. This is an important difference, as it stresses the importance of running angles and speed, which play an important role in reaching a certain destination.

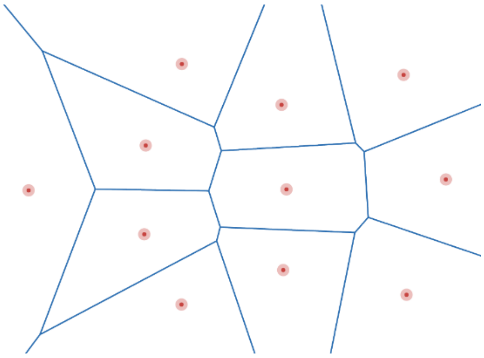

Figure 2: A 4-3-3 Formation as Voronoi Diagram

However, modeling space dominance by using solely Voronoi diagrams is not feasible, as it does not take into account many variables and outcomes. It does not, for instance, take into account the fact that these areas of dominance cannot specifically belong to one player, as there might be multiple players that have a chance at taking the ball in this area. Therefore, we should look at these areas as continuous values, representing probability distributions for different players, rather than assigning each area to one individual.

Space Dominance

Another thing space control using the previously discussed Voronoi technique does not account for, is the importance of the space that is occupied. Players can occupy different areas on the pitch, such as positions near their own goal or areas close to the opponent's goal. During attacking play, occupying areas near the opponent's goal is significantly more important because these positions offer better scoring opportunities. Ultimately, to win a game of football, one must score goals. Controlling one's own half might be beneficial from a defensive perspective, but does not necessarily do much in offensive play. However, having more spatial control around the goal area of the opponent will likely result in more offensive threat. This is where the expected threat comes in.

Expected Threat

Fujimura and Sugihara (2005) noted that it seems natural that the larger the player's dominant region, the greater this player's contribution to space dominance should be. This is not true though, as a player at a position far from the ball will not have anything to do with the ball at that specific time. The area around this player is likely sparsely populated, as the ball is not close to this player. Consequently, the dominant region of the player increases, as there is a lot of free space. Therefore it might seem as if the player contributes a lot to the space dominance, though this is not the case. Fujimura and Sugihara (2005) accounted for this by using weighted areas, where more important areas on the pitch are given higher weights than less important areas. In the dominant area weighted by the goal, for instance, the opponents' goal area contains high weights, such that space control in these areas is more impactful. This correlates to the approach of expected threat that is used in this thesis.

Introduced by Singh (2018), expected threat is a metric that can be used to quantify the potential of ball possession to result in a goal. It is different from traditional statistical metrics like expected goal or assists, as it rewards individual actions that contribute to moving the ball to more threatening positions. The expected threat can be defined as a raster-like layer that is applied on top of the match pitch. Each cell in the raster corresponds to a combination of x and y coordinates. The value in the cell is characterized as the threat-level of this position, with the positions that are close to the opponents' goal containing the highest values, as scoring from such positions is more likely. Different approaches to expected threat have also been proposed. One of which, for instance, includes a different expected threat matrix for the defending team (Nordahl et al., 2023). Due to the scope of this project, however, we will stick to Singh's expected threat matrix. A visualization of the expected threat can be seen in figure **3** below.

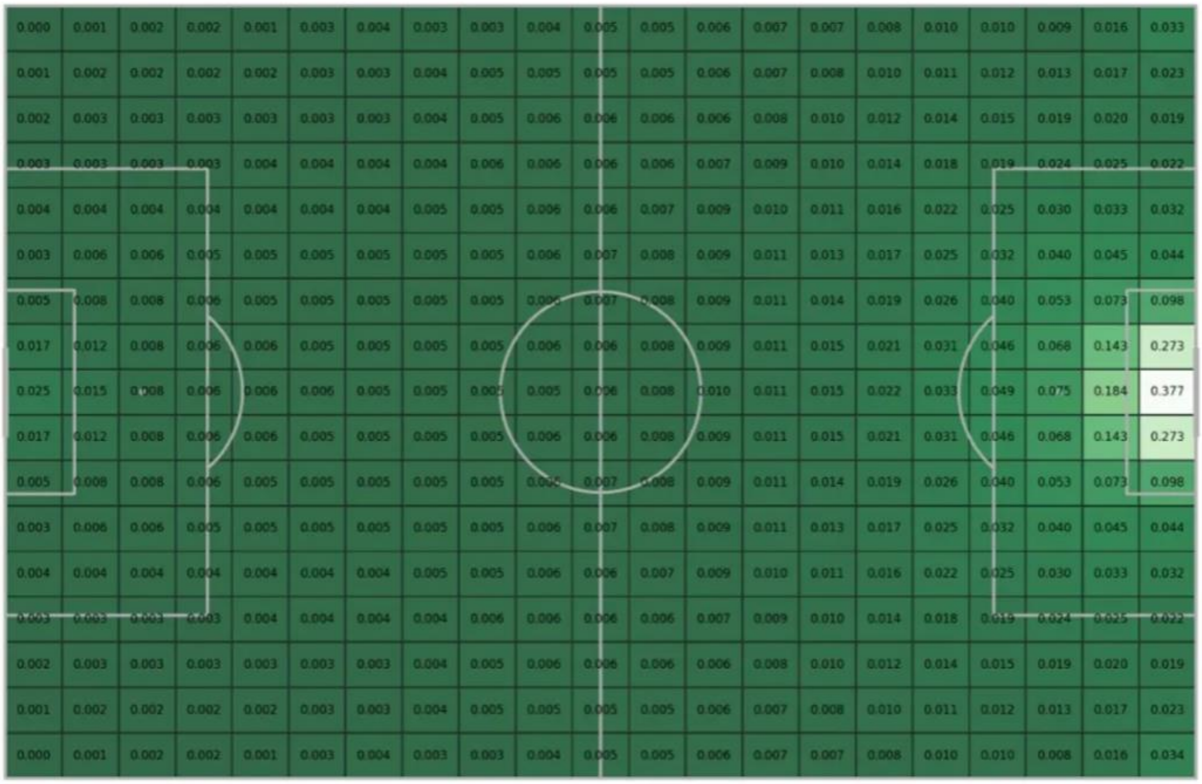

Figure 3: The Expected Threat Matrix.

Retrieved from "Explaining Expected Threat by D. Sumpter, 2021, Medium (https://soccermatics.medium.com/explaining-expected-threat-cbc775d97935).

Interpersonal distance

Next to Voronoi diagrams, other interesting behaviours occur in football matches that can help us quantify space dominance. For instance, throughout a football game, the attacking team will usually try to create space, whereas the defending team will try to keep space as limited as possible for the attacking team (Gréhaigne et al., 1997, McGarry et al., 2002). This means that in terms of spatial management, the attacking team will try to distribute their players further apart from each other over the pitch, where the defending team will try to play as compact and narrow as possible (Fonseca et al. 2012). Using this knowledge, one can imagine interpersonal distance between teammates as an important factor in space control and therefore space dominance.

Individual Dominance Score

The concept of space dominance, introduced by Taki et al. (1996), was first explored on the team level. Taki et al. defined team space dominance as the sum of all individual dominant regions. Building upon Taki et al.'s idea, Martens et al. (2021) also defined space control on the team level as the sum of all individual player influence areas. In their research, they follow up by multiplying the space control values with the so-called pitch values, referred to as expected threat in this thesis. Ultimately, the result of these calculations make up for what is known as space dominance. It seems therefore only logical to assume individual space dominance as the fraction of dominance exhibited by one player, contributing to the overall team space dominance. These percentages can then be interpreted as the amount of dominance a player contributes to the overall team dominance. These percentages can be seen as a score, ranging from 0 to 1, where a score close to 1 represents a very high share in the team space dominance.

Next to this, an additional measure can be introduced to further refine the individual dominance score. That is by implementing the previously discussed interpersonal distance into the individual dominance score. This can be done by taking the interpersonal distances and normalizing them to an interpersonal distance score (IDS) between 0 and 1. The IDS for the defending team is flipped to *1 - IDS*, reflecting the defending team's tendency to play more compactly rather than being distributed.

Methods

All calculations, data and code will be combined in a Python 3.11.8 interactive notebook (.ipynb) file. This includes loading the data, data exploration, preprocessing steps, calculations for individual space dominance scores and the final visualization of which a few plots will be included in the results section of this thesis. The notebook will not be publicly available due to privacy sensitive information and will only be shared with Forward Football and Utrecht University.

Data

The data that was used for this thesis is structured in excel (.xlsx) format. For each football match, there are three excel files. The first file contains sheets including overall match statistics, heart rate statistics, coordinate statistics and player-specific events like turns, passes and acceleration. The second and third excel files contain the teams that played the match, along with basic information about the match (duration, date, match ID). Next to this, these files also contain player names, player IDs and the time they played.

Each file contains specific information that can be used to assess individual player dominance on the pitch. This thesis will mostly use coordinate statistics of both players and the ball, but some other variables like match duration, timestamps, player IDs and team IDs will be of importance as well.

The coordinate statistics of each player are collected through portable stances, armbands worn by all players, and a special ball with a chip inside. The portable stances act as anchor points, receiving signals from the armbands and the ball. The armbands track each player's movements, while the chip in the ball monitors the ball coordinates. The coordinate statistics are captured at a rate of 5Hz, meaning five coordinates per player per second will be captured.

The data is challenged with much heterogeneity, as every match can contain a different number of players, a different length and different pitch size. Therefore, it is of importance to know what kind of teams are playing (e.g. youth or adult matches), as this affects both coordinate statistics, as well as the preprocessing techniques used to convert the data. Next to this, the data can also contain false readings or empty entries as players or coaching staff might have turned the measuring software off or on at the wrong time. Other issues that came up is that if a player gets substituted, the coordinate device gets turned off, when they get back on, the device is turned on, but the previous location of this player is different from when it was turned off. This way, we cannot calculate things like the change in x- and y-coordinates, which are important for the speed and angle variables. It can also occur that players or the ball get out of bounds of the pitch.

Handling Missing Data

After importing the data into a Python 3 environment, it became evident that certain columns contained missing values that needed to be replaced in order to successfully do the calculations. The first problem that arose was that of missing timestamps, the data frames that contained player coordinates and ball coordinates overlapped partially on the timestamp column, but there were some cases in which the player positions data frame contained timestamps that the ball coordinate data frame did not.

Another issue that had to be taken care of is the problem of missing values in the ball position columns. These ball coordinates are very important as they will be used in the calculations described in the next section of this chapter. To account for this, linear interpolation was used to replace the missing x- and y-coordinates of the ball. Linear interpolation is a method used to estimate unknown or missing values that fall between two known values. It makes the assumption that the change between values is linear. The formula for linear interpolation can be found in appendix A of this thesis.

The expected threat matrix that was adopted from Singh (2018) also requires linear interpolation, as the original matrix is smaller than all possible locations on the pitch.

Finally, one of the things that will be used in the calculations for speed, angle and ultimately space control is the difference in x-and y-coordinates, Δx and Δy. The problem that arose with these is that once players are substituted, their tracking devices are turned off. When the tracking device is turned on again, there are no previous coordinates to refer to for the calculations of Δx and Δy and therefore the value is set to NaN. Having these values set to NaN makes the calculations for space control impossible. The NaN values were set to 0 because there is no difference with the previous coordinates, as there are none.

Calculations

In this section, we delve into the calculations required to determine individual space dominance as a score for every player on the pitch. For the first part of these calculations, we will closely follow the model for space dominance on the team level, defined by Fernandez & Bornn (2018). Note that every calculation is made for each player individually and that at each timestep *t*, these calculations will be done for all players on the pitch.

Delta x and delta y

The first step required to calculate space dominance is obtaining the coordinates of all players. The coordinates will be compared to the previously known coordinates of the same player, the differences between them will be stored in Δx and Δy variables, denoting the change in x- and y-coordinates from timestep *t* to timestep *t+1*.

$$\Delta x = \text{current x} - \text{previous x}$$
$$\Delta y = \text{current y} - \text{previous y}$$

Euclidean distance

Two important variables are the distance between the ball and a player and the distance between a player’s current position and their previous position. Both distances were calculated using the Euclidean distance. The Euclidean distance between points (x1, y1) and (x2, y2) is calculated as follows:

$$\text{Euclidean Distance} = \sqrt{(x2 - x1)^2 + (y2 - y1)^2}$$

Angle
The angle in degrees of a player is necessary in the following few steps. It is defined as follows:

$$\theta = \arctan\left(\frac{\Delta y}{\Delta x}\right) * \frac{180}{\pi} + 360 \quad if \arctan\left(\frac{\Delta y}{\Delta x}\right) * \frac{180}{\pi} < 0$$

$$\theta = \arctan\left(\frac{\Delta y}{\Delta x}\right) * \frac{180}{\pi} \qquad otherwise$$

Where Δx and Δy are the difference in x- and y-coordinates, arctan is the arctangent function returning the angle in radians and 180/π converts the angle from radians to degrees. If the result of the computation is below 0, 360 is added to ensure the angle stays in the range of [0, 360].

Radius
To determine the influence area of a player, we need to calculate the radius around the player as well. It is calculated by taking the minimum of $(3/200 * \mathrm{E})^2 + 4$ and 10. Where E is the Euclidean distance from the specific player to the ball. We choose the minimum out of 10 and $(3/200 * \mathrm{E})^2 + 4$ because these are deemed as the minimum and maximum distances in meters of the player's space control (Fernandez & Bornn, 2018). The calculation is captured in the following formula:

$$\mathrm{Rad} = \min\left(\left(\frac{3}{200} * E\right)^2 + 4, 10\right)$$

R and inverse of R
To get the rotation matrix R, we have to use the angle theta (θ). The following formula is used to calculate R:

$$\mathrm{R}(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

Consequently, the inverse of R, is calculated using the following:

$$\mathrm{R}(-\theta) = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

Mu (μ)
To find the probability distribution from the multivariate gaussian distribution, we need both mean mu (μ) and the covariance matrix *COV*, the calculations for the latter will be discussed further into this section. Mu is calculated by adding the players current position and their speed vector, which is multiplied with a factor of 0.5. This yields the following formula:

$$\mu = \ Player\ position + \underset{S}{\rightarrow} * 0.5$$

In which speed vector $\vec{s}$ is defined as:

$$\vec{s} = \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}$$

Speed ratio
$Srat$ is the ratio of the speed at which a player is moving. The player's current speed is compared to the maximum speed in the game, this approach differs from the established approaches like the one Fernandez & Bornn (2018) did. In their research, the maximum speed was defined to be 13m/s as this is a good indication of a maximum speed a professional football player can move. However, as this research deals with very different data, namely that of both amateur and professional clubs, with players in different age categories, it seems only logical to use the maximum speed of a game. The maximum speed in the match will be squared and divided by the current speed of the player, which is also squared.

$$Srat = \frac{(player\ speed)^2}{(max\ speed)^2}$$

Scale matrix
S is the scale matrix, which serves the purpose of adjusting the covariance matrix to account for the player's speed and distance to the ball. It is captured in the following matrix:

$$S = \begin{bmatrix} \frac{Rad + (Rad * Srat(\vec{s}))}{2} & 0 \\ 0 & \frac{Rad - (Rad * Srat(\vec{s}))}{2} \end{bmatrix}$$

Where Rad is the radius of a player at time t and $Srat$ is the ratio of this player's speed compared to the maximum speed.

Covariance matrix
The covariance matrix COV is particularly important for the calculations of space dominance as we have to use a multivariate gaussian distribution to determine values for our probability density function. Having explained all variables that are needed for the covariance matrix up until this point, the formula for the covariance matrix COV is the following:

$$COV = RSR^TS$$

Where R is the previously discussed rotation matrix, R^T is the transposed version of this and S is the scale matrix.

Influence area
All previously discussed variables and matrices are only necessary for one thing; the player influence area. The player influence area will be used as an indication of space control. It captures the probability distribution of how likely the player will get the ball at timestep *t,* given the player's position, speed and direction.

$$I(\mu, COV, L, p) = \frac{\mathrm{PDF}(\mu, \mathrm{COV}, \mathrm{L})}{\mathrm{PDF}(\mu, \mathrm{COV}, \mathrm{p})}$$

Where influence area I is calculated using the previously discussed mean μ, the covariance matrix COV and all locations on the pitch, which is denoted as L. Lastly, p is the player vector, indicating the x- and y-coordinates of the player.

Expected threat
The expected threat matrix that was used is adopted from Singh (2018), the size of the matrix, however, is not accurate as the matrix used in this thesis is slightly bigger. To account for this, linear interpolation was used to extend the threat matrix, leading to the following threat distribution when visualized.

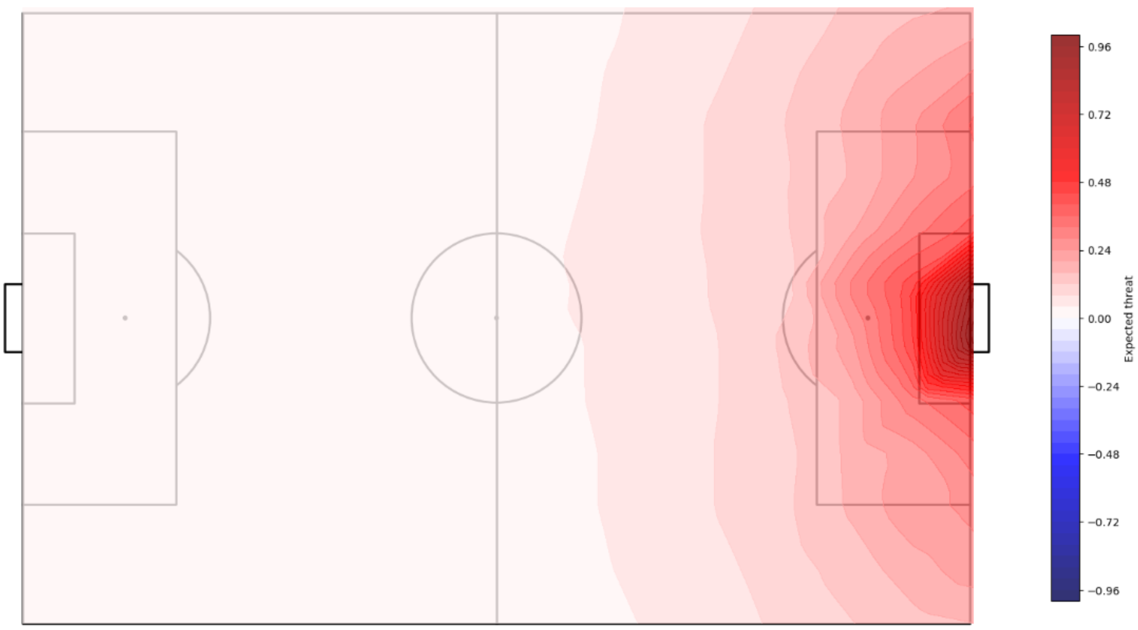

Figure 4: Expected Threat After Linear Interpolation Visualized on a Football Pitch

In this figure (**4**), the red zone indicates a large threat value, whereas the whiter zones indicate lower values for the expected threat. We see that the area around the opponent's goal is highly valued, whereas the team's own half does not have high threat values. Obviously, depending on the team, the matrix will be flipped. At half time, the matrix will be flipped again to make sure the expected threat matrices do not get mixed up, as this will greatly impact the calculations for space dominance.

Space control and space dominance scores
Current research about space control and space dominance mainly focuses on the team level rather than on the individual level (Taki et al., 1996; Fernandez & Bornn, 2018). To quantify this on an individual level, we will first look at the team level of these metrics. In their research, Fernandez & Bornn (2018) had a simple approach to quantifying space control on the team level. They would sum up the influence areas of all players on team A and all scores of the players of team B would be subtracted. The matrix that remains is the space control matrix for team A.

To quantify the player's contribution to the team's space control, an interesting metric to look at would be the fraction that the player contributed to the team's score. To do this, we sum up all influence areas, meaning all matrices get summed and put into one matrix. To know the overall space control at that moment, we sum all the values in this matrix to get one final space control score. The player's individual contribution is then calculated by taking the sum of the individual influence area matrix and dividing it by the total space control score.

$$\frac{\sum_{i=1}^{n} I_i}{\sum_{j=1}^{m} I_j}$$

Where I is the player influence area. The individual player that is evaluated is i and m denotes all players on player i's team.

To quantify space dominance on the individual level, the same strategy is applied to space dominance scores. In team dominance, we would calculate space dominance by multiplying the space control matrix for team A with the corresponding expected threat matrix. Implementing the aforementioned idea of taking the fraction of the total score, we get the individual contribution to the team's spatial dominance.

Interpersonal distance
The interpersonal distance between teammates was defined as the Euclidean distance from a given player to their closest teammate.

In terms of feature scaling, the values were then normalized between 0 and 1 using min-max normalization. After normalizing, the final interpersonal distance score (IDS) was calculated based on the player's respective team. For players in the attacking team, the IDS value was directly used.

For players in the defending team, the score was adjusted to 1 - IDS as discussed in the previous chapter. To determine which team is the attacking team, we defined the attacking team as the team that is closest to the ball in terms of Euclidean distance. After inspecting the results, however, it became clear that the adjusted 1 - IDS for the defending team would not be sufficient. The defending players received extremely high scores even in areas where they were clearly not dominant. Various formulas were experimented with to reduce the high scores for the defending team. These include:

- $\frac{1}{1+IDS}$
- $(1 - IDS) * 0.1$
- $(1 - IDS) * 0.2$
- $(1 - IDS) * 0.5$
- $(1 - IDS)^2$
- $(1 - IDS)^4$
- Setting the IDS score for the defending team to 0

After careful examination of the results, the formula for the defending team that was chosen is $(1 - IDS) * 0.1$, mainly because of the good balance between adding value to a compact formation, without neglecting the offensive positioning of the opponent. The factor of 0.1 also reduces the extreme variation between defending and attacking team, which is preserved in other formulas like $\frac{1}{1+IDS}$ or $(1 - IDS) * 0.5$.

Individual scores

In this section, different individual scores were defined, including the fraction of dominance score and the interpersonal distance score. These scores capture different dynamics in terms of spatial distribution on the pitch. In this context, the interpersonal distance score measures the spatial relationships between individuals on the pitch, reflecting aspects of cooperation, coordination, or avoidance. On the other hand, the fraction of dominance score quantifies the proportion of spatial dominance an individual asserts on the pitch.

These two scores were integrated into a final combined score to provide a comprehensive view of individual performance. By combining the interpersonal distance score with the fraction of dominance score, the final score captures effects of possession related spatial distribution, as well as open space and threatening positions. This integrated approach not only enhances the depth of analysis but also facilitates a deeper understanding of spatial positioning behaviour on the pitch.

Results

The dataset used in this thesis contained many detailed game metrics. The most important ones include player coordinates, ball coordinates, timestamps and both team and player IDs. Preprocessing of the data involved removing missing values, removing invalid data readings and replacing missing values through linear interpolation. After the necessary preprocessing steps, calculations for the player influence areas were done. Following the player influence areas, we calculated space dominance through the expected threat matrix. Ultimately, individual contribution was characterized by taking the fraction of the total dominance score, interpersonal distances were calculated and normalized to add to the combined score that depicts the spatial dominance of players individually.

The results showed significant variation in individual space dominance across different matches and players. One of the main aspects of the results is the visual representation of individual dominance scores on the pitch. By plotting player positions and individual dominance scores on the pitch, we can observe the behaviour of the dominance scores in various positions and situations. By observing this behaviour in key events like goals in matches, we can see how certain movements or positioning by individual players impact their dominance on the pitch. These insights can give better understanding to the players contributions to spatial dominance on the pitch, which in turn can improve the spatial distribution of players in a team.

Heatmaps

An important element of our findings is the visual representation of dominance scores on the football pitch. By plotting the team dominance over the whole pitch and putting the individual player scores at the player positions, we can see which positions and players contribute significantly to the team dominance compared to others. Intuitively, players with higher dominance scores should have larger and more strategically positioned dominant regions, especially near the opponent's goal as these areas are highly valued. By using this visual approach we can determine not only which players actively control space, but also how their positioning and movement aligns with strategic objectives such as scoring goals. The results will be split up in two sections: one for the individual dominance scores by fraction and another for the combined score that includes interpersonal distance scores.

Individual dominance as fraction

When plotting individual dominance contributions as a fraction of the total dominance score by a team, we get some interesting plots. We plot the team dominance score on the pitch, with the individual contributions as a fraction printed over the player dots to see which player contributes the most to the team dominance score at time *t*.

The match event we inspect is the first goal of a friendly match between youth players of the Dutch football club Zeeburgia. We inspect a few moments leading up to the goal by looking at the corresponding dominance plots. Note that there is one player on the blue team that stands on the sideline; this is probably due to an injury treatment.

The first timestamp depicted in figure **5**, shows the situation on the pitch a few seconds before a goal. The ball is in the middle of the pitch and we can see a few players that contribute to their team's dominance. We see the striker (**S**) of team 1 (the red team) in a potentially threatening position, if the ball reaches this player, the path to the goal is essentially open. We also see that the more central positions contribute significantly less than the positions on the side or in the front like the right winger (**RW**) on team 2 (the blue team), for instance. This is due to the fact that space is harder to control in crowded, central positions than on the sides of the field where there is more open space. Overall, it also seems like defenders have less impact on the team's dominance score, as the four defenders on the red team make up for only 12% of the dominance score for their team.

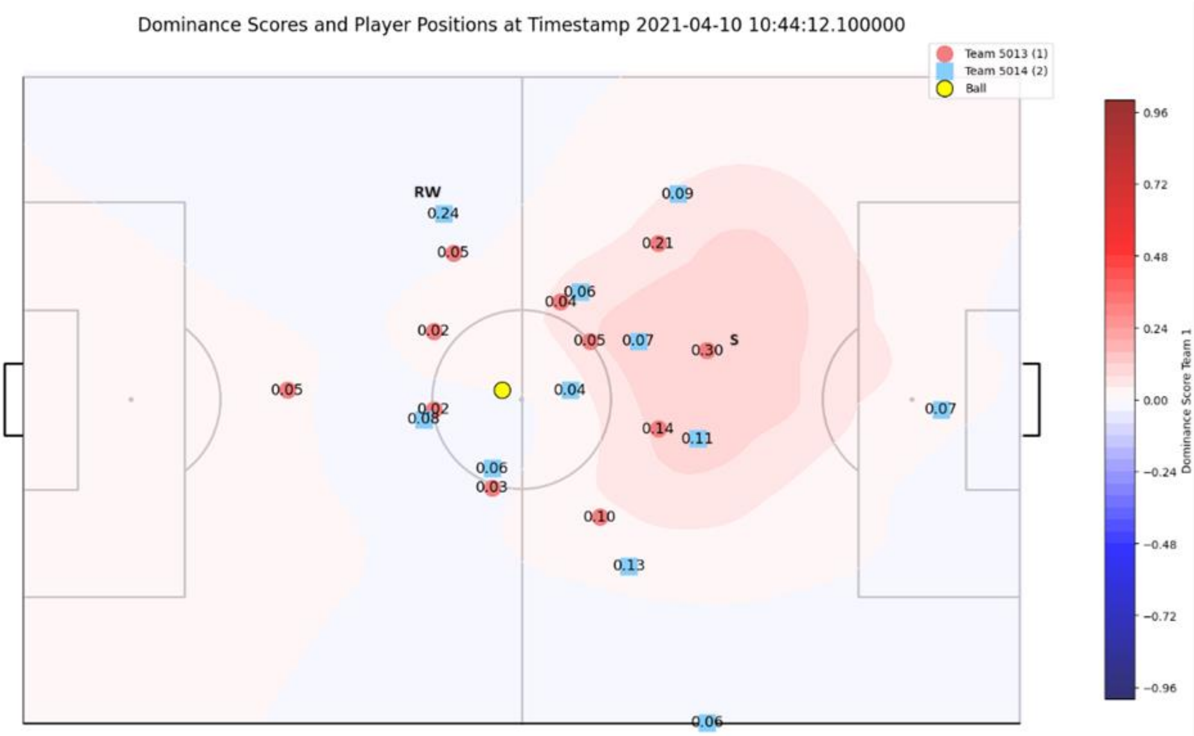

Figure 5: The fractional dominance scores at the first timestamp

In figure **6**, we see the situation just four seconds later. The ball has moved up to the goal of the red team, which is in a dangerous situation as there is a scoring opportunity for the blue team. We see the players close to the ball having relatively low values, this is likely due to the fact that both players are likely to control the ball when they are both this close. Therefore a duel between the players is more likely to decide the outcome of this chance, rather than their positioning. On the wing, however, we see a dangerous position for the right winger (**RW**) of the blue team, who has a lot of free space close to the goal in front of him. Next to this, we see that the scores for the defenders are still very low.

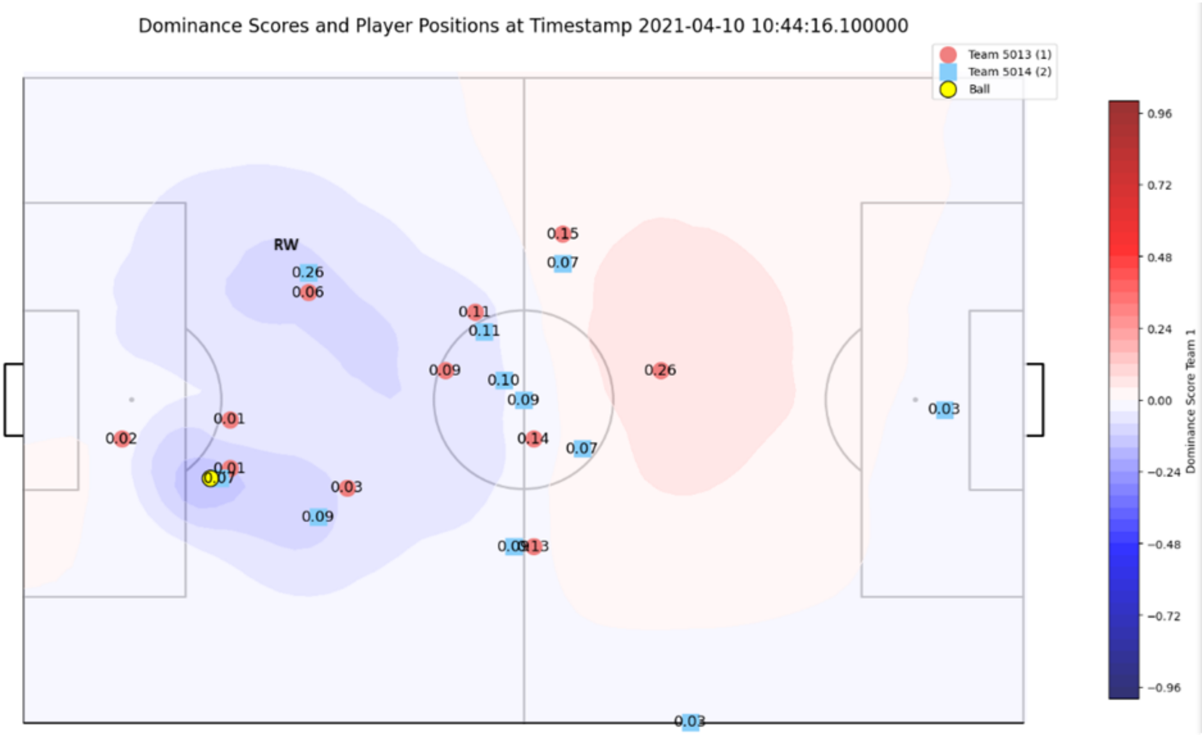

Figure 6: The fractional dominance scores at the second timestamp

Finally, the goal is scored and we can pause the game at this very moment. What is noticeable in figure **7** is that all attackers on the blue team have high scores for dominance contribution, this is mainly because there is a lot of free space to the side, as well as them being relatively close to the goal. This makes sense as their team just scored a goal. The defending team is positioned quite centrally and therefore does not really control the side. Aside from this, the values for the expected threat are very high for the blue team as they are close to the goal, which is why the dominance scores for the red team are way lower than those of the blue team. We still see the striker (**S**) of the red team having high space dominance values, mainly because this player is in a large open space with no direct opponents close, it is also the player on the red team that is closest to the opponent's goal and therefore the most threatening area.

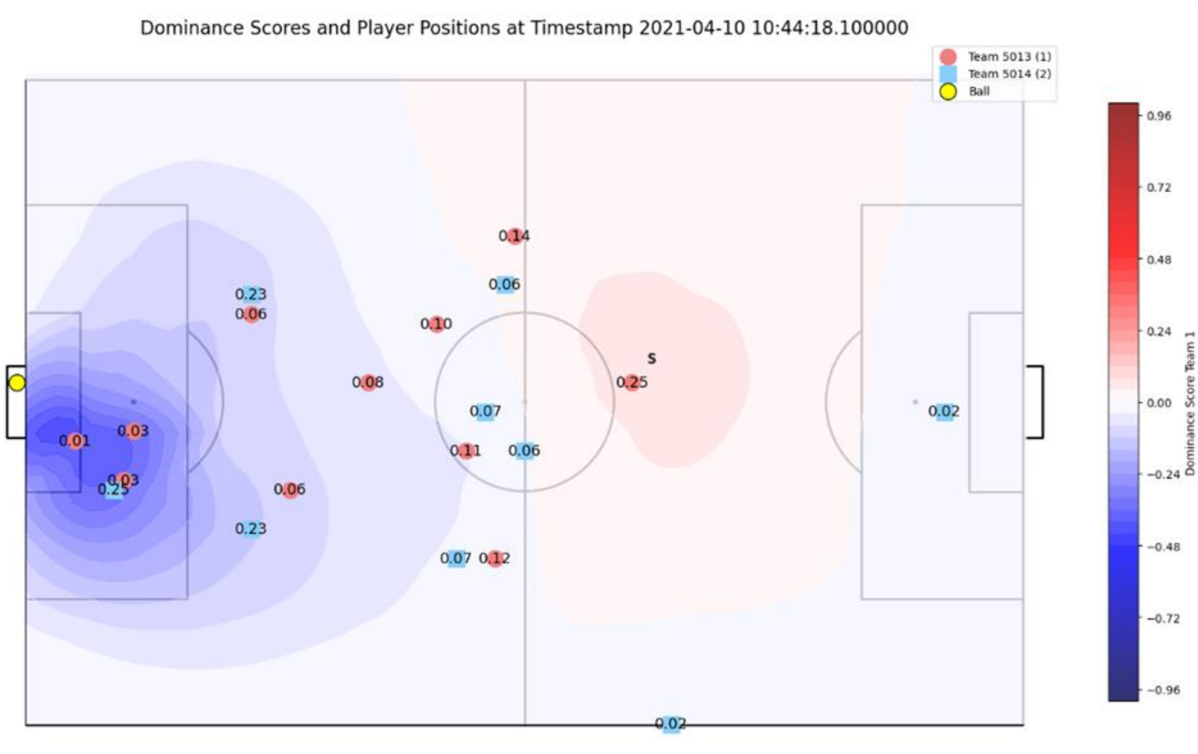

Figure 7: The fractional dominance scores at the final timestamp

Individual dominance combined with interpersonal distance

We will now compare the results in the same timestamps as before, but instead of the fraction of dominance, we plot the combined score of both space dominance and interpersonal distance to see how the individual contributions change.

In figure **8**, there is a noticeable change in individual dominance scores for the blue team, the main difference between the combined score and the fraction of dominance exhibited by the player is found near the sides of the pitch; the players that play on the wing or as fullback have higher scores. These players also have a lot of space to work with, as most of the open space is located at the sides of the pitch, rather than in the center. We see that the left winger (**LW**) of the blue team is one of the only wide players that does not make use of the area and therefore their score is relatively low compared to the other players. Compared to the fractional score, we see that players of the red team have relatively higher values, especially in the defensive and midfield areas, as these players get rewarded for a compact spatial distribution, even though their scores are still relatively low as their positioning could be improved, especially in the defensive line as there is a gap behind the defensive line that will be utilized by the blue striker (**S**).

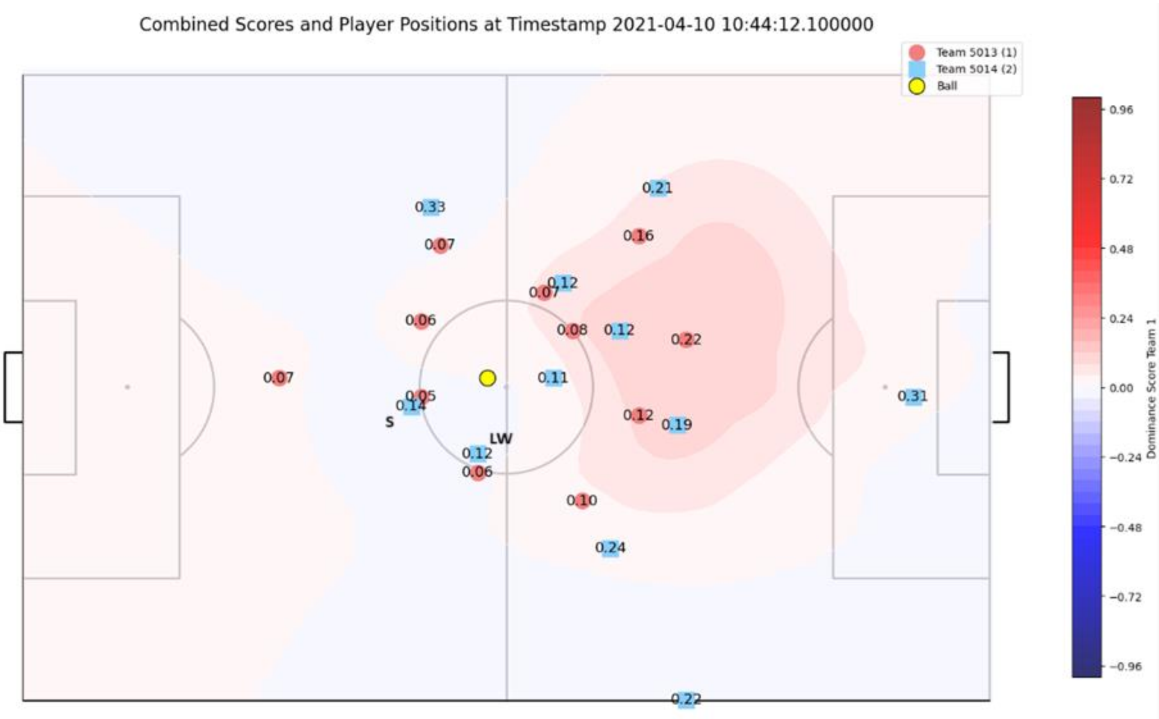

Figure 8: The combined dominance scores at the first timestamp

The following figure (**9**) shows the situation 4 seconds later, where we see the blue striker (**S**) getting to a threatening position. The dominance values are slightly higher for the attacking players, which is caused by them being in potentially valuable positions. Something that is noticeable, however, is the relatively high space dominance score for the blue goalkeeper (**GK**). The blue goalkeeper has such a high score because both their influence area is big, as there are not many other players around, as well as their high interpersonal distance score. The goalkeeper is not in a threatening position for the opposing team, so it may seem counterintuitive that they would have such a high score. However, it should be noted that the goalkeeper is a very important player in a football team that can also significantly change the state of the game by utilizing their positioning. If the red team is able to take the ball and make a quick turnaround with a long ball, the blue team can be very dependent on the spatial dominance of the goalkeeper to ensure they do not get scored on. Another thing to notice is that the red team's striker (**SR**) has a significantly lower combined score than their fractional score in the same situation. This is due to the fact that the red striker is in the defending team and is relatively far from their teammates and the ball. This player is punished for being too far from active play, therefore not contributing a lot to spatial dominance.

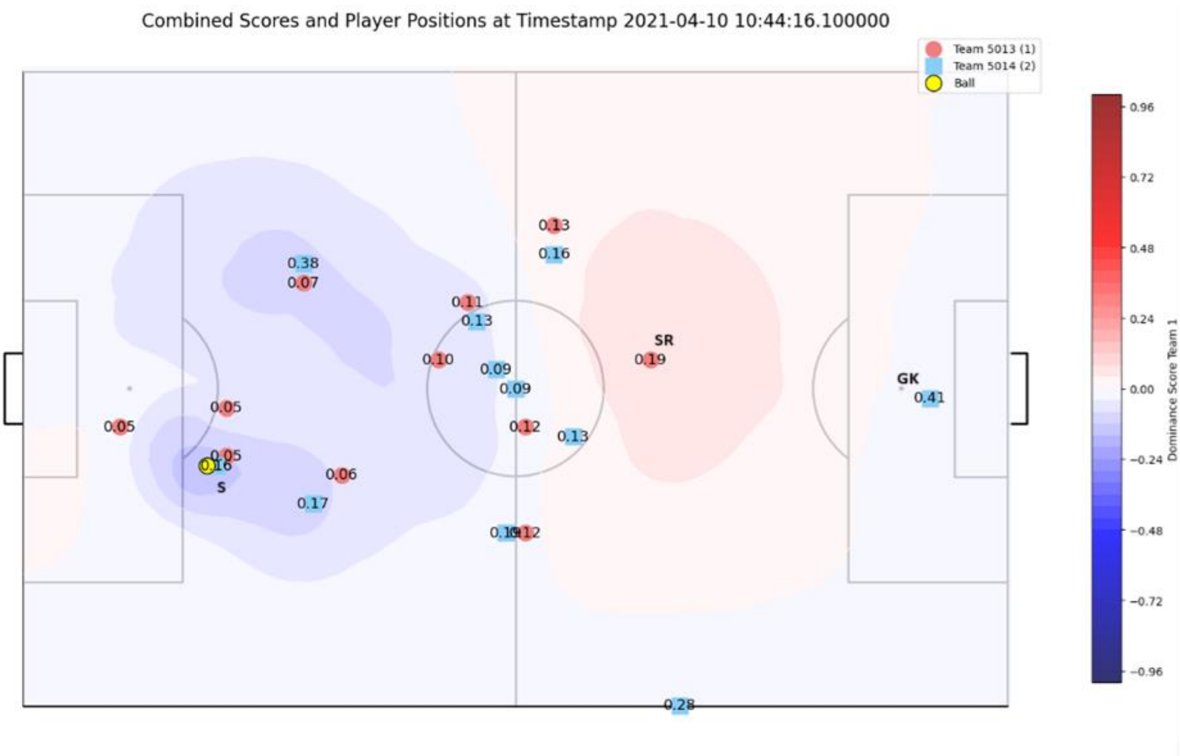

Figure 9: The combined dominance scores at the second timestamp

The final figure (**10**) once again shows the time at which the blue team scores a goal. What is interesting about this particular timestamp is that the dominance scores for the red team have significantly increased, while those of the blue team have decreased. This is due to the fact that the 'attacking team', which is the team in possession, has changed from blue to red, as the red players are now closer to the ball. While the team that is closest to the ball is usually the team in possession, this is not the case at all timestamps as you can imagine a player passing the ball over an opponent. For a brief amount of time this opponent will be closer to the ball and therefore in the attacking team.

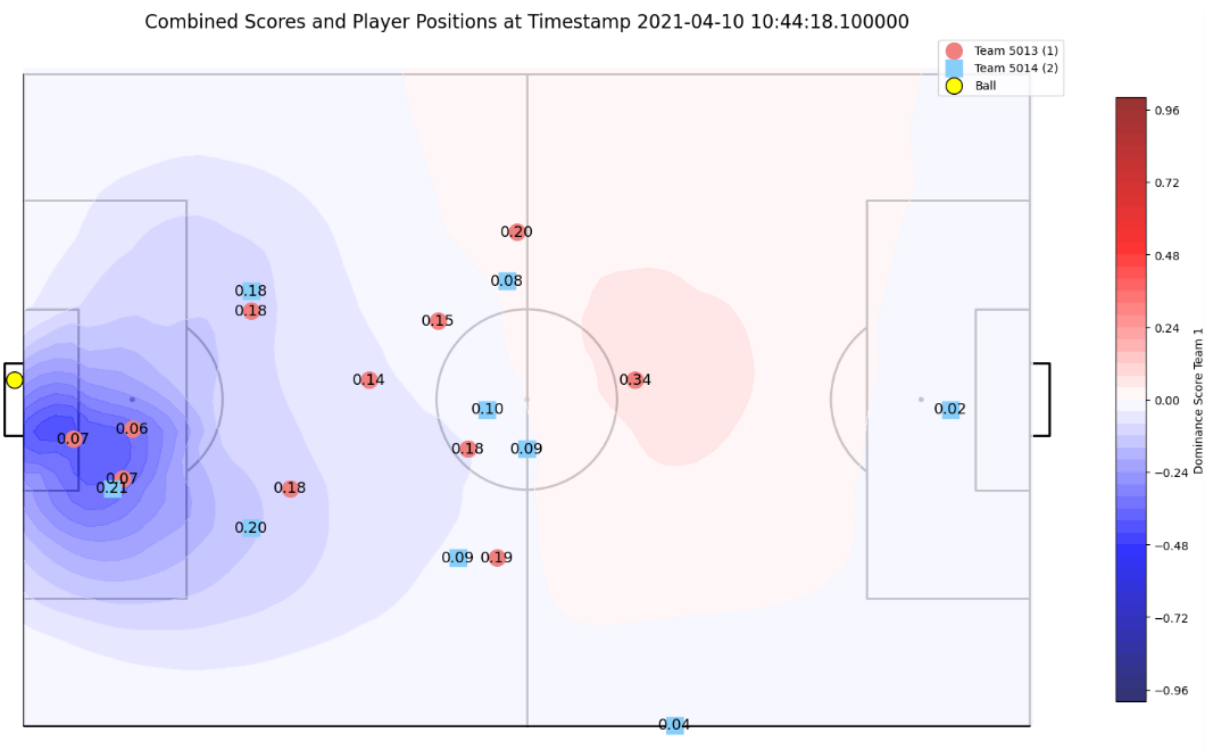

Figure 10: The combined dominance scores at the final timestamp

It seems clear that the fraction method produces more logical results in this particular timestamp. This also shows that even though the player dominance scores make sense in attacking play, the switch to defending play overturns the dominance scores very quickly. This makes sense in a way as the attacking team is usually more spatially dominant in terms of deciding where the ball is going, whereas the defending team will be forced to react to this. However, it should be noted that these abrupt changes in dominance scores might be too excessive.

Key Findings

During the visual investigation of the results, several discoveries were made, leading to the following key implications:

The fraction method provides the most basic, accurate interpretation of the contribution of a player to the team dominance scores. The scores are easier to interpret due to the fact that all values are percentages that sum to one, whereas in the combined score this is not necessarily the case.

The combined score integrates interpersonal distance as an additional metric that is able to reward players for more tactical spatial distribution like keeping the pitch wide in attacking play. It also reduces the spatial dominance for players of the defending team, which alleviates the problem of attackers from the defending team lingering up front and having a relatively high space dominance score, while they are far from active play. When a team switches from attacking to defending, however, their space dominance scores plummet as attacking play is prioritized more in this combined score, as well as in the concept of space dominance.

Conclusion & Discussion

Summary

This thesis set out to answer the question: “How can we quantify and assess the amount of space dominance exhibited by individual players at different timestamps during a match?”. After introducing the necessary literature on key performance indicators in football, space dominance on the team level and interpersonal distance, a methodology was developed to quantify individual space dominance. This methodology uses two different metrics: a fractional score and a combined score that integrates interpersonal distance as a factor. Both metrics quantify individual space dominance as a percentage of contribution to space dominance on the team level, where the combined score is an extension to this that adds an interpersonal distance score as well.

Key findings

Both the fractional score and combined score offer valuable insight into individual contribution to space dominance, with the fractional score providing more simplicity in terms of interpretation and accuracy, while the combined scores offers a more nuanced approach, leading to a richer analysis compared to the fractional score.

The results show that both scores have good potential to quantify individual space dominance. As the fractional score adds up to one, it is very easy to see which players contribute a lot to space dominance and which players could increase their individual space dominance scores by improving their positioning on the pitch. The simplicity of this approach is valuable as well as this does not only help interpretation, but also keeps true to the nature of how spatial dominance is defined on the team level. The combined score, on the other hand, also shows good potential as it correctly assigns high values to players that contribute highly to space dominance. It does so while also rewarding a compact or wide formation, depending on which team is in possession, therefore rewarding a more tactical spatial distribution.

Limitations

Both scores also come with their limitations: Quantifying individual space dominance by using the fraction of the player’s contribution to team dominance does not necessarily relate to the player’s actual contribution to space dominance. A player might have a relatively large proportion in the teams’ space dominance score, but because the player is surrounded by a few other players of the opposite team, their space dominance scores cancel out the space dominance score of the first player. This happens because the team scores get subtracted with each other to find which areas belong to which team. Therefore, the player does not necessarily contribute a lot to the team’s space dominance, even though it seems like they do because they have a rather large contribution to the score.

The dynamics of the interpersonal distance score can also have a negative impact on the individual scores: Defining the attacking team based on proximity to the ball leads to inconsistencies, especially in moments where the ball closely passes a player that is not on the attacking team. This can inaccurately lead to that player’s team being recognized as the attacking team. This will, severely affect the combined individual dominance scores the players exhibit for a brief amount of time.

Next to this, complexity of team dynamics can be a disturbing factor in individual dominance scores, as football is a dynamic sport where roles and positions continuously change based on game phases, tactics or other events like red cards. The problem with this is that in the combined score, it is assumed that the attacking team will try to play as wide or expansive as possible. In some practical cases, however, coaches or players can deploy different tactics, which interferes with the combined score.

Future research

Future research can build upon the framework that was set out in this thesis.

Methodological improvements can be made in the area of the interpersonal distance score. Defining a more refined metric to quantify the attacking or defending team, as well as more in depth analysis on the defending team's interpersonal distance score can further improve the combined score metric.

Next to this, the generation of space through making runs that create more spatial freedom for other players is an important concept that is neglected in this thesis. This is a crucial concept if we think about a team's ability to create free space though.

Other areas that can be explored are the definition of expected threat, which could be altered for defending teams. The expected threat could also be defined per team, as some teams might have different threatening positions due to their unique playstyle and squad.

Future research can also investigate the impact of player movement patterns on space dominance. By analyzing player trajectories and movement behaviour, researchers can gain insight into how specific player movement influences space dominance both positively and negatively. These insights can be used to tweak formational tactics or refine the individual player dominance score.

Ultimately, the findings of this study will not only contribute to the understanding and development of individual space dominance, but also to future research in football dynamics and off-ball qualities of individual players. By addressing the need for quantification and formulating a new additional metric for individual space dominance, this work lays foundation for future research in space dominance in football matches.

Appendix

Appendix A: Linear interpolation

To estimate a the coordinates (x, y) at timestamp *t* using linear interpolation, given the values (x_{t-1}, y_{t-1}) and (x_{t+1}, y_{t+1}), we can use the following formulas:

$$x_t = \frac{x_{t-1} + x_{t+1}}{2}$$

$$y_t = \frac{y_{t-1} + y_{t+1}}{2}$$

These formulas calculate the average of the known values at the previous and future timestamps, assuming linear change between the coordinate points. The following example shows how this is done in practice.

Given (x_{t-1}, y_{t-1}) = (1, 4) and (x_{t+1}, y_{t+1}) = (3, 8), we can estimate (x, y) at timestamp *t* by plugging in the numbers in the formula.

$$x_t = \frac{1+3}{2} = 2$$

$$y_t = \frac{4+8}{2} = 6$$

By doing this, we get the estimated values for (x_t, y_t) = (2, 6)

References

Bloomfield, J. R., Polman, R. C. J., and O'Donoghue, P. G. (2005). Effects of score-line on intensity of play in midfield and forward players in the FA Premier League. J. Sport Sci. 23, 191–192.
Casal, C. A., Maneiro, R., Ardá, T., Marí, F. J., & Losada, J. L. (2017). Possession zone as a performance indicator in football. the game of the best teams. Frontiers in Psychology, 8. https://doi.org/10.3389/fpsyg.2017.01176

Fernandez, J., & Bornn, L. (2018). Wide Open Spaces: A statistical technique for measuring space creation in professional soccer. In Sloan sports analytics conference (Vol. 2018).

Fonseca, S., Milho, J., Travassos, B., & Araújo, D. (2012). Spatial dynamics of team sports exposed by Voronoi diagrams. Human Movement Science, 31(6), 1652–1659. https://doi.org/10.1016/j.humov.2012.04.006

Fujimura, A., & Sugihara, K. (2005). Geometric analysis and quantitative evaluation of sport teamwork. Systems and Computers in Japan, 36(6), 49–58. https://doi.org/10.1002/scj.20254

Grehaigne, J., Bouthier, D., & David, B. (1997). Dynamic-system analysis of opponent relationships in collective actions in soccer. Journal of Sports Sciences, 15(2), 137–149. https://doi.org/10.1080/026404197367416

Gu, C., De Silva, V., & Caine, M. (2024). A machine learning framework for quantifying in-game space-control efficiency in football. Knowledge-based Systems, 283, 111123. https://doi.org/10.1016/j.knosys.2023.111123

Hook, C., and Hughes, M. (2001). "Patterns of play leading to shots in 'euro 2000," in Center for Performance Analysis, ed. Pass.com (Cardif: UWIC).

Hughes, M., Caudrelier, T., James, N., Donnelly, I., Kirkbride, A., & Duschesne, C. (2012). Moneyball and soccer - an analysis of the key performance indicators of elite male soccer players by position. Journal of Human Sport and Exercise, 7(2), 402–412. https://doi.org/10.4100/jhse.2012.72.06

Jamil, M., Liu, H., Phatak, A., & Memmert, D. (2021). An investigation identifying which key performance indicators influence the chances of promotion to the elite leagues in professional European football. International Journal of Performance Analysis in Sport, 21(4), 641–650. https://doi.org/10.1080/24748668.2021.1933845

Jones, P. D., James, N., & Mellalieu, S. D. (2004). Possession as a performance indicator in soccer. International Journal of Performance Analysis in Sport, 4(1), 98–102. https://doi.org/10.1080/24748668.2004.11868295

Martens, F., Dick, U., & Brefeld, U. (2021). Space and control in soccer. Frontiers in Sports and Active Living, 3. https://doi.org/10.3389/fspor.2021.676179

McGarry, T., Anderson, D. I., Wallace, S. A., Hughes, M. D., & Franks, I. M. (2002). Sport competition as a dynamical self-organizing system. Journal of Sports Sciences, 20(10), 771–781. https://doi.org/10.1080/026404102320675620

Michalak, J. (2022). Identifying football players who create and generate space. In Ola Larsson (Ed.), & Lars-Åke Nordén (Trans.), *Civilingenjörsprogrammet I Informationsteknologi* [Thesis]. https://www.diva-portal.org/smash/get/diva2:1687971/FULLTEXT01.pdf

Narizuka, T., Yamazaki, Y., & Takizawa, K. (2021). Space evaluation in football games via field weighting based on tracking data. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-84939-7

Nordahl, O. A., Kim, J., James, N., Gwon, U., Parmar, N., Nicholls, S. B., & Vuckovic, G. (2023). EVALUATING GOAL THREAT IN FOOTBALL USING PLAYER AND BALL LOCATIONS. *Kinesiologia Slovenica/Kinesiologia Slovenica*, *29*(3), 26–48. https://doi.org/10.52165/kinsi.29.3.26-48

Rein, R., Raabe, D., & Memmert, D. (2017). “Which pass is better?” Novel approaches to assess passing effectiveness in elite soccer. Human Movement Science, 55, 172–181. https://doi.org/10.1016/j.humov.2017.07.010

Singh, K. (2018). Introducing expected threat (XT). https://karun.in/blog/expected-threat.html

Taki, T., Hasegawa, J., & Fukumura, T. (1996). Development of motion analysis system for quantitative evaluation of teamwork in soccer games. In Proceedings of 3rd IEEE International Conference on Image Processing (Vol. 3, pp. 815-818). Lausanne, Switzerland. https://doi.org/10.1109/ICIP.1996.560865