GIMA
Geographical Information Management and Applications

A GIS-based analysis of offensive movement actions in football
Assessing the quality of off-the-ball, offensive positioning in football using a GIS platform

Master Thesis

Robin van de Wal
r.j.vandewal@students.uu.nl

Supervisor: ir. E. Verbree (TU-Delft)
Responsible professor: Prof.dr.ir. P.J.M. van Oosterom (Tu-Delft)

Date of submission: 13/08/2020
Data of defence: 11/09/2020

Universiteit Utrecht
TU Delft
UNIVERSITY OF TWENTE.
WAGENINGEN UR

Acknowledgements

I want to express my appreciation to SciSports for providing the necessary data, without it, this research project would have been impossible. I'd also like to thank ir. Edward Verbree and Prof.dr.ir. P.J.M. van Oosterom for their advice and patience.

Abstract

With the increased digitization of football, many researchers and professionals alike seek new ways to put the ever increasing amount of data to good use. On-the-ball actions are the main object of study in most of these projects, while a player spends the vast majority of time without the ball.

This research project proposes a spatial model to assess off-the-ball offensive movement actions in professional football. Through their presence, players control a certain amount of space by being the first player to be able to reach that space. This is the players *dominant region*.

The objective of an offensive movement action is to score a goal. To this end, the space in front off goal was assigned a value through the analysis of a large amounts of shots. This produced a continuous surface of *goal probability* around the goal and penalty area. When the *dominant regions* of an attacker controls a portion of this space, the attacker's current offensive movement action can be assigned a score based on the *goal probability*. This value is further modified by the chance an opponent intercepts the pass required to get the ball into that area.

The results of the analysis of several matches through the constructed spatial model point towards a positive relationship between high scoring offensive movement actions and better results (here meaning shots and goals). This is a promising result but more extensive testing needs to be done to determine the model's predictive power.

Table of contents

List of abbreviations and acronyms 6
1. Introduction 7
1.1 Statistics in football 8
1.1.1 General use of statistics in football 8
1.1.2 Statistical Analysis in football 9
1.2 Player Attributes 10
1.2.1 The multitude of player attributes 10
1.2.2 Offensive movement actions 11
1.3 Geodata in football 14
1.4 Spatial Analysis in football 14
2. Research Goals 16
2.1 Research objectives 16
2.2 Research questions 16
2.3 Scope 16
2.4 Relevance 17
3. Theoretical Framework 19
3.1 Introduction 19
3.2 Terminology 19
3.2.1 Position, location, area and place 19
3.2.2 Positional Roles 21
3.3 Football 22
3.3.1 Football: A spatiotemporal framework 22
3.3.2 Football analytics 24
3.4 Spatial models 24
3.5 Movement 24
3.5.1 Moving points objects 25
3.5.2 TRIAD Framework 26
3.5.3 Trajectory analysis 27
4. Methodology 28
4.1 Introduction 28
4.2 Offensive movement actions 28
4.2.1 Possession 29
4.2.2 Attacking phase 29
4.2.3 Formal definition 30

4.3 Practical elements 31
4.3.1 Data 31
4.3.2 Software 32
4.3.3 Assumptions 33
4.4 Space occupation 33
4.4.1 Possible and Dominant regions 33
4.4.2 Physical limitations 36
4.4.3 Conclusion 36
4.5 Spatial Value 37
4.6 Interception model 45
4.6.1 Flight time 45
4.6.2 Curve 47
4.6.3 Time to intercept 47
4.7 Conclusion and implementation 47
5. Results & discussion 56
5.1 Results 56
5.1.1 Dominant regions 56
5.1.2 Offensive movement actions 57
5.2 Discussion 58
5.2.1 Model accuracy 58
5.2.2 Applications 59
6. Conclusions and recommendations 62
6.1 Research questions 62
6.2 Further Research 63
Literature 66

List of abbreviations and acronyms

FIFA	Fédération Internationale de Football Association (International Federation of Association Football)
GIS	Geographical Information System
MPO	Moving Point Objects
UEFA	Union of European Football Associations

1. Introduction

This research project focusses on the sport of football, also known as soccer or association football. Football is the most popular sport in the world (McGarry & Franks, 2003). The sport attracts millions of spectators to the stadiums every weekend (Kotzbek & Kainz, 2014), many millions of players participate worldwide and it is the only individual sport which rivals the Olympic games in terms of viewership on television. The last World Cup in Russia reached an incredible 3.572 billion individual viewers, making it the third most viewed televised event of all time, only behind the summer Olympic Games of 2012 and 2016 (FIFA, 2018). Millions of players participate in the sport on a weekly basis.

But in essence, football can simply be described as a sport in which two teams of eleven players each try to score at least one more goal than the opposition, by kicking the ball in the opposition's net. From an aerial view, this looks like 22 dots divided into two colours: the players, chasing a white dot: the ball (Figure 1.1). From this perspective, it is only a small step to see a football match as a study of Moving Point Objects (MPO's) (Laube, Imfeld, & Weibel, 2005). These MPO's move around for a duration of 90 minutes in a constant attempt to get themselves in an advantageous situation. This shows how football can accurately be described as a union of space and time (Kotzbek & Kainz, 2014). The spatiotemporal aspects inherent to the sport, are often disregarded by professional analysts, as will be demonstrated in this chapter. This research project takes this MPO approach to the sport of football in order to perform spatial analysis in a GIS environment, in an attempt to properly address the spatiotemporal nature of football.

This chapter will introduce the use and analysis of statistics in football and highlight an important aspect of football where the use of statistics seems to be severely underdeveloped: off-the-ball movement. Besides that, the different attributes that footballer players can excel at and the use of (geo)data in football match analysis to gauge these player attributes will be discussed.

Figure 1.1 – An aerial view of a football match

1.1 Statistics in football

1.1.1 General use of statistics in football

The increased digitalisation in football of the past two decades (Kotzbek & Kainz, 2017a) has brought about a large amount of data. This data is used for match analysis, scouting, the assessment of individual and team performances and more (Kotzbek & Kainz, 2014). Statistics constructed of this data are presented by media outlets to provide the public with more information. The majority of these statistics (Figure 1.2, 1.3) are basic numbers like the distance travelled by each player, number of goals or assists, number of goal attempts or number of completed passes (Bransen & Van Haaren, 2018).

Notable in this list, is that only one metric concerns actions taken off-the-ball, without possession that is: distance covered. While Johan Cruyff concluded that *"the ball is an essential element of the game"*, the emphasis on possession practiced in these match statistics seems exorbitant. This is especially true if one considers the fact that at any given time, only one player can have possession of the ball. This results in the actions and movement of 21 player who are not in possession of the ball being all but disregarded.

Figure 1.2 and 1.3 – Typical football match statistics as presented in the media

Another noteworthy fact is that most of these statistics lack a spatial dimension. Apart from arguable distance covered, the statistics presented in both Figure 1.2 and 1.3 only tell *what* happened, not *where*. This is strange when one considers how crucial the space in which these actions took place is, in determining how important the action itself was. A shot from outside of the box is far less valuable than a shot from within 10 meters of the goal (Kotzbek & Kainz, 2016), but in these statistics, they are both counted exactly the same. The only way they are possibly separated is if one shot happens to translate into a goal. If football can be accurately described as a union of space and time (Kotzbek & Kainz, 2014), match analysis disregarding both the spatial and temporal aspect of football can't possibly be very descriptive of what actually happened on the field (Kotzbek & Kainz, 2015). Figure 1.2 and 1.3 speak to the fact that the spatiotemporal nature of football is regularly disregarded in match analysis (Kim, Kwon, & Li, 2011). This problem does not only exist in football, but also in other sports,

statistics only track how often certain action take place, providing an insufficient basis for analysis (Kotzbek & Kainz, 2015).

1.1.2 Statistical Analysis in football

The numbers presented in Figure 1.2 and 1.3 mostly constitute raw amounts of recorded events. These numbers themselves don't tell a complete story, and it is common knowledge both in football (Moruzzi, 2014) and geography (Clifford, French, & Valentine, 2010) that statistics presented on their own can be very misleading. An often cited quote surrounding statistics exemplifies their persuasive and possibly misleading nature best: *"There are three types of lies: lies, damned lies, and statistics".* More interesting and hopefully less misleading, is the analysis of raw data, refining it into meaningful information. The information derived from match data analysis is used in various ways in professional football (McGarry & Franks, 2003): (1) immediate feedback, (2) the development of databases, (3) the indication of areas of improvement, (4) evaluation and (5) as a mechanism for selective searching through video footage.

There are several characteristics of football which make this match data analysis difficult.

Firstly, the low scoring nature of football can make it difficult to assess performances through match statistics (Bransen & Van Haaren, 2018; Decroos, Bransen, Van Haaren, & Davis, 2019). While a football game has a duration of 90 minutes plus stoppage time, 7,96% of games have no goals, 26,30% of games have 1 goal or less, and 49,85% of games have 2 goals or less (Windrawwin.com, 2018). This means, by means of simple math, that most actions on the field do not contribute to a goal being scored, directly or indirectly. This does not mean however, that these actions might as well not have taken place, they still impact the game, just in ways that aren't directly reflected on the scoreboard. The rarity of goals in football has pushed analysts to using different statistics as a proxy for a team's or player's performance (Spearman, 2018). The reporting of total number of shots or shots on goal is the most typical example of this and can be observed in both Figure 1.2 and 1.3.

Secondly is football's dynamic and continuous nature. The most famous example of a sports team achieving success through a statistical approach to their sport, is undoubtably Billy Beane's Oakland A's. Their story was popularized by the book *Moneyball* (Lewis, 2003) and its subsequent adaption for cinema by the same name. Billy Beane was able to build a competitive baseball team on a tight budget, by identifying undervalued players in regard to their match statistics: their match statistics suggested they were better players than their market value reflected. The relative static nature of the sport of baseball makes it a prime subject for statistical analysis, it mainly consists of one-to-one contests with set outcomes (van Hoeve, 2017). In football, this is much less the case (Van Haaren, Dzyuba, Hannosset, & Davis, 2015), players can move freely over the pitch and while static situations exist (e.g. corner kicks, free kicks penalties) the majority of the match consists of continuous, dynamic play. It is not coincidental that sports analysis using probabilistic and computational processes first focussed on these more structured sports, like baseball and cricket, while only relatively recently branching out to more dynamic sports (Chawla, Estephan, Gudmundsson, & Horton, 2014).

Thirdly, is the fact that football is a team sport. It might be difficult sometimes, to dissect the individual performance from the team performance. If a player creates a great chance, for instance, by crossing the ball in front of goal, this is only counted as an assist if his teammate finishes the chance. However good this cross is, whether the player gets an assist to his name is determined by forces outside of his control: the finishing ability of his teammate. The other side of this equation is even harder to quantify: what if an attacking player makes a good run, getting him in a position where he can receive the ball and have a great chance at a goal, but the required pass for this never arrives?

Since simple statistics fail to capture football's complexity, several new statistics have been introduced, some even based on positional data (Van Haaren et al., 2015). These analytics which interpret existing data in useful ways are commonly known in sports as advanced metrics. The expected goals statistic (Green, 2012; Optasports, 2016) provides the likelihood of each shot becoming a goal, based on its position. This statistic allows comparison between how many goals a player or team *should* make versus how much they actually made or shot quality versus chance quality. Before a shot can be taken though, the ball has to end up with a player in an offensive area. In most cases, this means a pass has to be given to that player. But what if the quality of the chance is good, but the subsequent shot does not do it justice? To assess the quality of the pass which creates the chance independently, the expected assist model was introduced, which can be defined as: "the likelihood that a given pass will become a goal assist" (Optasports, 2016). Taken one more step back though, the pass can only be given if a teammate can get into the position where the pass is destined to end, before an opponent can. A metric to assess a player's ability to get into the right positions at the right time, to receive a pass, in order to score a goal, is currently not in existence.

1.2 Player Attributes

1.2.1 The multitude of player attributes

Different specific skills are required for different actions in football (Andrienko et al., 2019). While a rough consensus exists for Lionel Messi being the best player in the world right now, most would agree there are aspects of the game which others are more proficient at. Messi is for instance not renowned for his heading ability, nor his defensive prowess. A comprehensive picture of the different player qualities needed to play football, can be painted through two of the most prominent football video game series: FIFA and Football Manager. Figure 1.4 and 1.5 show the attributes player are rated on in these two videogames.

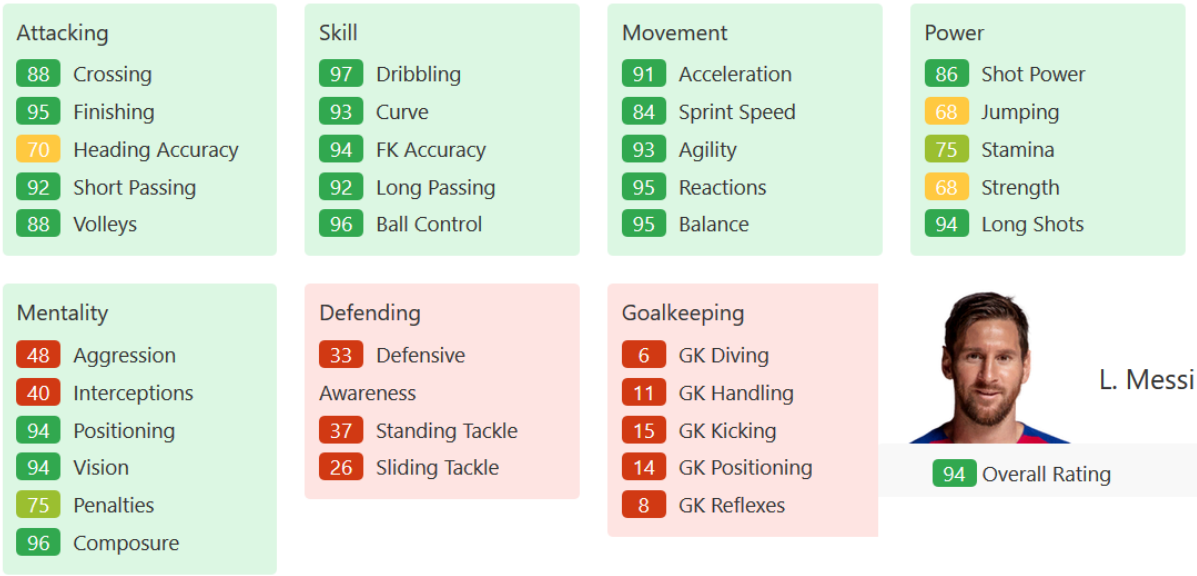


Figure 1.4 – Player attributes in the most popular football videogame, FIFA 20

ATTRIBUTES

Technical		Mental		Physical	
Corners	15	Aggression	7	Acceleration	18
Crossing	15	Anticipation	19	Agility	20
Dribbling	20	Bravery	10	Balance	20
Finishing	20	Composure	18	Jumping Reach	6
First Touch	19	Concentration	13	Natural Fitness	14
Free Kick Taking	19	Decisions	20	Pace	15
Heading	10	Determination	20	Stamina	13
Long Shots	17	Flair	20	Strength	9
Long Throws	4	Leadership	14	Height	170 cm
Marking	4	Off The Ball	16	Weight	72 kg
Passing	20	Positioning	5	Overall Physical Condition	97%
Penalty Taking	17	Teamwork	14	Match Sharpness	35%
Tackling	7	Vision	20	Morale	Superb
Technique	20	Work Rate	7		

Figure 1.5 – Player attributes (Lionel Messi's) in another popular football videogame, Football Manager 19

What's interesting about these numbers is that very ephemeral and elusive attributes like vision, flair and off the ball movement, are given a numerical score (from 1 to 99 for FIFA and 1 to 20 for FM). This is essentially the goal of the earlier discussed advanced metrics to take an array of complicated actions and reduce it down to a quick and easily interpretable number.

Besides the grading of these attributes, these numbers dictate particular behaviour of the virtual players in the game. In the case of *off the ball* movement, this means that hidden beneath these numerical scores is a complex algorithm which has a concrete framework laid out for what *good off the ball* movement is.

"I find it terrible when talents are rejected based on computer stats. Based on the criteria at Ajax now I would have been rejected. When I was 15, I couldn't kick a ball 15 meters with my left and maybe 20 with my right. My qualities technique and vision, are not detectable by a computer" – Johan Cruyff

Besides criticizing the use of statistics in football, Cruyff here identified how some player attributes are more difficult to assess through data analysis than others. While a player's skill in finishing or crossing can perhaps be determined by the number of goals or assists he collected over a season, more ephemeral attributes like vision and positioning require a more complicated approach which currently doesn't seem to exist.

1.2.2 Offensive movement actions

In the FIFA videogame, the only skill which determines offensive positioning is the *positioning* attribute (Figure 1.4). As a contrast, at least 9 different attributes are dedicated to essentially kicking the ball in different manners or situations (crossing, finishing, short passing, volleys, curve, free kick accuracy, long passing, shot power and long shots). In Football Manager, the formula is a little bit more complicated (Figure 1.5). The primary attribute which determines offensive positioning is *Off The Ball*, but other attributes factor in as well (anticipation, concentration, decisions, determination and work rate). What is discussed here are the mental aspects of positioning, knowing when to be where. The physical attributes necessary to actually carry out the movement are another matter.

Fernandez & Bornn (2018) identified that the vast majority of football analytics research focusses on the analysis of on-the-ball actions while much of the game's complexity resides in off-the-ball events.

It can be said that in many ways, the latter precedes the former: a pass can only be given if the receiving player has moved into an area where he can receive it. Furthermore, since only one player can have possession of the ball at any given time, there are 21 players solely occupied with their positioning and movement in regard to the ball, their position on the field, their teammates and their opponents, at all times. Off-the-ball movement therefore seems to be a massively understudied subject in football analytics research.

Passing was previously identified as an understudied area of football analytics research (Bransen & Van Haaren, 2018). While the majority of on-the-ball actions are passes (75%) most research was focussed on shots, which only comprise 2% of total on-the-ball actions. The earlier discussed disparity between on-the-ball and off-the-ball research is a continuation of this same logic. While a player averages about 3 minutes of possession per 90 minutes (Fernandez & Bornn, 2018), the vast majority of research is focussed on these 3 minutes, leaving the other 87 minutes a major opportunity for research, which is currently almost untapped.

In professional football scouting, the off-the-ball aspect of the game seems similarly understudied. The aim of scouting in professional football is to identify the skills that a player possesses (Zambom-Ferraresi, Rios, & Lera-López, 2018). While a typical scouting report might have a small section dedicated to a player's positioning, titled territorial coverage in Figure 1.6, the vast majority of the report is focussed on on-the-ball actions. Without the existence of a metric to express the quality of offensive movement, it is difficult to go beyond territorial coverage. While each of the other tabs (finishing, playmaking, assistance, individual, aerial and defensive) have added pages of details, the territorial coverage has no such added information. Most likely this is the case because there is not much more to add to it, the correct metrics currently just do not exist. Indeed, while many professional football clubs have access to large amounts of performance data, the valuable information that is hidden in this data is only used to a limited extent in their decision-making process (Van Haaren et al., 2015). What seems to be missing are computational methods to analyse this data in greater depth (Bialkowski et al., 2015). Raw data has no value in itself, only the information extracted from it does (Keim, Kohlhammer, Ellis, & Mansmann, 2010). Information can be defined as data enriched with meaning and context (Heywood, Cornelius, & Carver, 2011). Currently, companies like SciSports (SciSports, 2019) and Metrica Sports (Metrica Sports, 2019) operate in this existing gap between data and information (van Hoeve, 2017).

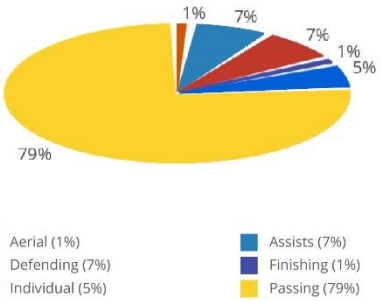
Cruyff once stated *"the most important thing is: what do you do during those 87 minutes when you do not have the ball? That is what determines whether you are a good player or not"* (from Fernandez & Bornn (2018)). It seems that in football reporting, football analytics research and, perhaps most surprisingly, professional football scouting, this vision is not shared. Almost all metrics used to express a player's quality or contribution to the game are centred around on-the-ball actions. What is good positioning? How can the skill of getting yourself into an advantageous position in football best be expressed in numbers? These are clearly complex spatiotemporal questions with a multitude of possible answers. To quantify the quality of of-the-ball actions, new metrics will have to be introduced.

wyscout

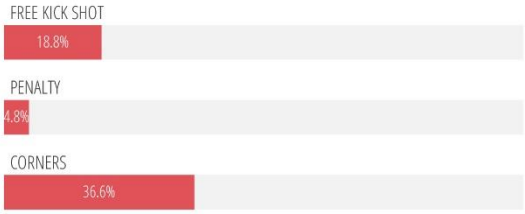
Player Scout Report

STATISTICAL SUMMARY

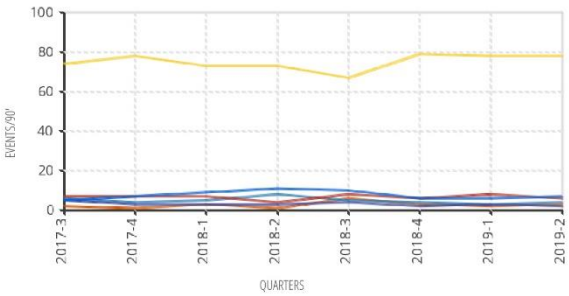
EVENTS DISTRIBUTION FOR TYPE



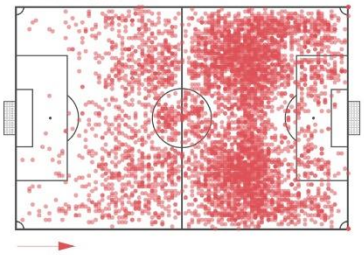
FREE KICKS



Player percentage contribution to the team considering free kicks, penalties and corners.



TERRITORIAL COVERAGE



FINISHING	EVENTS/90'	ON TARGET	OTHER	GOALS
Opportunity	0.8	50.9%	49.1%	22.8%
Head shots	0	100%	0%	100%
Shots	0.9	39.4%	60.6%	18.2%

PLAYMAKING	EVENTS/90'	ACCURATE	ASSIST	KEY PASS
Simple pass	60.9	93.4%	0.2%	0.6%
Smart pass	2.6	48.9%	2.2%	9.2%
Long pass	1.2	93%	0%	3.5%

ASSISTANCE	EVENTS/90'	ACCURATE	ASSIST	KEY PASS
Crosses	1.9	42.1%	2.3%	0%
Key pass	1.6	95.7%	15.7%	-

INDIVIDUAL	EVENTS/90'	WON	NEUT.	LOST
1 vs 1 Dribbling	2.6	70.3%	11%	18.7%
On the ball acceleration	0	-	-	-

AERIAL	EVENTS/90'	WON	NEUT.	LOST
Aerial duels	0.6	11.1%	4.4%	84.4%

DEFENSIVE	EVENTS/90'	WON	NEUT.	LOST
1 vs 1 defending	1.1	12.5%	12.5%	75%
Defensive duels	3.2	15.6%	27.1%	57.3%

Figure 1.6 – Player scouting report (kindly provided by Wyscout)

1.3 Geodata in football

The unavailability of the correct data lead to the focus on shots rather than passing (Bransen & Van Haaren, 2018). Indeed it can be observed that in recent year, this particular research imbalance seems to be fading, as more passing data has become available (Bransen, Van Haaren, & Van De Velden, 2019; Clemente, Martins, Wong, Kalamaras, & Mendes, 2015; Gyarmati & Stanojevic, 2016; Kotzbek & Kainz, 2017a; Link, Lang, & Seidenschwarz, 2016; Power, Ruiz, Wei, & Lucey, 2017; Rein, Raabe, & Memmert, 2017). The difference between on-the-ball and off-the-ball research, seems to be rooted in the very same issue: the unavailability of the correct data.

There are three main formats of football match data. The first, and most simple form, can be labelled as match statistics. These statistics simply record total numbers of events (e.g. passes, shots, tackles, etc.). These statistics are the simplest to produce and most often presented by media outlets (as exemplified by Figure 1.2 & 1.3), but lacks a lot of crucial information, most notably: the event's position.

The second form is event stream data, also called event logs (Gudmundsson & Horton, 2017). This datatype denotes the time and place of ball specific events (Kotzbek & Kainz, 2015), player events and technical events such as fouls, the start or end of a half or offside calls (Gudmundsson & Horton, 2017). accompanied by additional information depending on the specific type of event. Shots for instance might have the additional information of the precise placement in the goal and whether it resulted in a goal or not. This type of data can be seen as an expansion of the first category, with added variables and crucially: this data has a spatial component.

The third type of data is tracking data, also called object trajectory data (Gudmundsson & Horton, 2017). This type of data gives the exact position of all 22 players and the ball, per time unit (usually several time units per second) for the duration of a match. In some cases, the referee and linesman are also tracked, but this is not common as their position are not crucial to understanding the gameplay of a match. This last datatype is much more expensive to produce and therefore less widely available than the first two (Decroos et al., 2019), although the availability of both tracking and event stream data are constantly increasing (Van Haaren et al., 2015). For the analysis of off-the-ball actions, tracking data is an absolute necessity, its high cost and relative scarcity, might very well explain the lack of focus on off-the-ball actions in football analytics research.

Another difficulty presented by the analysis of tracking data, is its size and lack of categorization. While event data presents a table of carefully separated categories, with distinct variables, tracking data consists of a continuous stream of x, y and z coordinates, uncategorized and without any added variables. Besides that, just in an XML format, tracking data consists of roughly 500 times as many lines as event data. The text file only consisting of tracking data is also roughly 50 times larger than the text file only consisting of event data. While this is not an issue in text files, since they are still quite manageable, the ratio gives an indication of the disparity in file sizes and processing times when constructing different datatypes from the XML files. While tracking data can be used for many useful analyses, the raw data seems to induce an information overload (Keim et al., 2010) in many cases. It's clear that the availability of vast amounts of positional data presents new challenges for the way in which this data is analysed and interpreted (Link, 2018).

1.4 Spatial Analysis in football

Because of their spatial component, the last two datatypes described in the previous section: tracking and event stream data, can accurately be classified as *football specific geo data*: spatiotemporal

geodata which directly represents the gameplay of a football game (Kotzbek & Kainz, 2015). Since these datatypes feature spatial information within a local coordinate system, they are fit for analysis in a GIS environment (Longley, Goodchild, Maguire, & Rhind, 2005).

GIS has particular value when addressing spatial questions concerning location, patterns, trends and conditions (Heywood et al., 2011). The construction of a spatial model to express the quality of offensive movement actions as an advanced metric is a spatial question associated with all four of these conditions and therefore very fitting for a GIS environment.

Kotzbek and Kainz (2014) demonstrated how GIS can be applied in the field of football game analysis. In their work they developed a guide to football specific geodata (Kotzbek & Kainz, 2015), a shot quality assessment (Kotzbek & Kainz, 2016), a passing quality assessment (Kotzbek & Kainz, 2017a) and more (Kotzbek & Kainz, 2017c, 2017b), in a GIS environment. This research built the foundation for a very logical application of GIS, the analysis of the ever-increasing amount of football specific geodata.

2. Research Goals

2.1 Research objectives

The ability of a player to position himself in an advantageous position is a very important attribute in football which is currently an underdeveloped topic in football match analysis. This research project will use a GIS approach to build a spatial model to address the research imbalance between of-the-ball and on-the-ball actions.

This research has one main research objective:

To develop an advanced metric which assesses the quality of offensive movement actions on the basis of a spatial model in a GIS.

To achieve this main research objective, several sub research objectives will have to be completed first.

1. Establishing a formal definition of offensive movement actions.
2. Establishing how the quality of an offensive movement action can be defined and determining the elements contributing to this quality.
3. Determining how these elements can be represented through data in a GIS.

Additional sub research objectives have been set to determine the worth of the developed model.

4. Determining to what extent the constructed spatial model predicts the chance quality and quantity following the offensive movement actions.
5. Determining what other applications the constructed spatial model has in football match analysis.

2.2 Research questions

Research questions accompany these research objectives. The main research question is:

To what extent can the quality of an offensive movement action be determined and assessed through a spatial model?

The five sub research question associated with the five sub research objectives are:

1. What is an offensive movement action?
2. How is the quality of and offensive movement action defined and what are the elements contributing to this quality?
3. To what extent can the elements contributing to the quality of an offensive movement action be spatially modelled in a GIS?
4. To what extent can the quality of an attacking player's offensive movement actions be judged through the constructed spatial model?
5. What other applications does the constructed spatial model have in football match analysis?

2.3 Scope

The previous two sections explained what the research project was about, this section explains what it is not about. The research gap between off-the-ball and on-the-ball actions has been discussed extensively. This research project focusses on movement without the ball, dribbling does not factor into the equation. Defensive positioning is also an important and understudied aspect of football but

is also not a focus of this research project. Finally, offensive positioning is looked at in terms of trying to get oneself into a position in order to score a goal. Other research has looked at offensive positioning in terms of value offered to the team in retaining possession (Fernandez & Bornn, 2018) but this research project will purely look to how likely it is to score from certain positions.

2.4 Relevance

Besides the concrete research goals stated in section 2.1, a more abstract goal of this research project is to examine and demonstrate how GIS can be an appropriate platform for football match analysis, both to the scientific community and professional football analysts. This is not stated as a main or sub objective because its success can't be tested or demonstrated. There is a lot of value in constructing bridges between scientific disciplines, as old problems can be considered from new perspectives, prompting new approaches and tools, leading to new solutions and conclusions.

The construction of a new advanced metric can essentially be boiled down into expressing the quality of a complicated action, set of action or attribute into one number. This is valuable because numbers allow for easy comparison. A player can be compared to other players by means of this advanced metric in order to find who is better at a particular attribute. Similarly, a player can be evaluated against a whole host of other player, to establish a benchmark of sorts, in order to identify weak or strong aspects of his game. Finally, a recent match or set of matches can be compared to a previous match or set of matches to evaluate form. All these examples are only possible if the actions or attributes evaluated can be expressed numerically, which is the goal of this research project precisely. Additionally, performance analysis is always dependent upon how the *performance indicators* are defined (McGarry & Franks, 2003). Football clubs, media and fans alike benefit from the introduction of new performance metric, as it gives them more tools to gauge the performances of player and teams.

Three previous works are of particular inspiration to this research project and will therefore be highlighted in this section. Since their subject matter is very close to the subject of this research project this section will also point out how they differ, in order to demonstrate what is added to the scientific literature by this research project.

Van Hoeve (2017) is, to the author's knowledge, the only previous example of a GIMA Master thesis focussing on football. His work has helped in bridging the gap between Geographical Information Science and football. Additionally, the topic of concerted movements of MPO's is directly tackled in van Hoeve's (2017) work. Where this research project distinguishes itself is in the addressing of a specific football question: determining the quality of offensive movement actions. Van Hoeve (2017) studied movement in football from the perspective of a developer with the end goal of creating a conceptual analytical tool to help football professionals visualize and analyse football specific geodata. In this research project, the study of movement in football is done to provide a basis for spatial analysis.

Fernandez & Bornn (2018) is one of the few research papers explicitly focussing on off-the-ball movement. Their work looks at a very similar subject matter and some aspects will most likely heavily resemble some sections of this research project, most notably the movement model. This research project is different by specifically looking at attacking movement actions focussed on scoring goals, where Fernandez & Bornn (2018) focus more on possession and pitch control. The largest difference between the two models will therefore be the spatial value model, for which they use a dynamic variant, based on ball position, but this research project will use something more closely resembling an expected goals model.

Spearman (2018) is the other major influential football analytics researcher who focussed explicitly on off-the-ball movement. Spearman did specifically focus on movement actions aimed at creating chances at goal, and his research is the most similar to this research project in most aspects. While some aspects of their model are very in depth, some simplifications are made to speed up the process, most notably in the spatial value model and the interception model. These two aspects of the model will be addressed differently in this research project. The availability of tracking data makes it possible to model interceptions much closer to reality than done in Spearman (2018), even if that is more time consuming and resource intensive. As a result, the spatial value model employed in this research project will more closely resemble an expected goals model based on from *where* the shot was taken instead of simply how far from the goal the shot was taken.

Finally, while both Spearman (2018) and Fernandez & Bornn (2018) focus on a similar subject matter, they do so from a different background, likely resulting in different assumptions, methods and conclusions. Spearman's background is in physics while Fernandez & Bornn come from artificial intelligence and statistics. Tackling a similar problem from a different angle, geographic information science in this case, is therefore a worthwhile addition. Besides the worthwhile venture of building bridges between scientific disciplines, different approaches to similar problems will likely produce different theories and results.

3. Theoretical Framework

3.1 Introduction

This section will introduce and briefly explain the objective and content of the remaining sections of the theoretical framework.

Section 3.2 deals with the terminology used in this research project. Important concepts of both geography and football are defined and explained and possible overlaps are eliminated.

Section 3.3 will explain some football specific concepts in the literature. This section will start with the explanation of the spatiotemporal framework of football as interpreted in this research project. This section will conclude with a brief explanation of the term football analytics.

Section 3.4 will elaborate the use of spatial models. The design of an advanced metric to quantify offensive movement quality is the primary objective of this research project, this advanced metric is a numerical end product of an underlying system which can be described as a spatial model.

Section 3.5 will deal with the fundamental concepts of movement used in this research project.

3.2 Terminology

3.2.1 Position, location, area and place

Definitions are the formal statements of the meaning of words or phrases, as found in dictionaries. Different dictionaries might have slight variations in the meaning of the same word, and different words might mean several different things. Similarly, the same word might have different denotations in different (scientific) fields. Even within the same field, one word might mean several different things. Within football a goal both refers to the scoring of a point by one team, and the physical goalposts and net in which the ball ends up. In general parlance, the context in which these words are used often enough sufficiently clarifies the definition used. Words like location, position and place will be used almost interchangeably. In scientific writing however, it is important to be more precise and consistent in the use of language. Because of this, the definitions which will be used in this research project will be outlined in this section. Several important geographic, scientific and football-specific concepts will be defined.

Position is a term which is used often in both geography and football, but it carries a very different meaning in both fields. In geography, an object's position is defined as a quantitative expression of where the object is, for example in a coordinate system (Groves, 2013; Sithole & Zlatanova, 2016). This corresponds to the concept of **geometry** in a GIS setting, an object's geometry is its absolute or relative coordinates within a coordinate system (Gold, 2016a). In football, a position is somewhat similar to a player's role within the team, expressed through the general area of the pitch the player is expected to operate from. A left winger for instance is understood to be an attacking minded player operating on the left side of the field.

To distinguish these two concepts, the definition of position used in football will not be used in this paper from this point onward. Instead the words *positional role* or just *role* will be used. As explained, this concept is similar to the definition used for *position* in football *and* will suffice as a replacement outside of in-depth tactical discussion.

When the word *position* is used in this research project, this concept is understood as it's exact location on the football field, expressed in a coordinate system.

A **location** can be defined as a general placement relative to well defined physical space (Sithole & Zlatanova, 2016). It is often understood as a qualitative expression of the physical world e.g. a building or a room (Groves, 2013). An **area** is understood to be a distinct part of a wider space e.g. a floor in a building and can be thought to contain several locations (Sithole & Zlatanova, 2016). The field itself is divided up into different sections. Sometimes these sections are demarked by lines, like the two separate halves of the field or the tow penalty boxes. These sections most resemble the concepts of locations as defined by Sithole & Zlatanova (2016) because of the clear physical boundaries in which they are contained. Others are not demarked by lines, and are thus invisible to the human eye, like the final third. Where the final third exactly begins is thus not clear for everyone at all times, but the concept is still a useful one, to for instance mark when a key pass was made. In Sithole & Zlatanova's (2016) terms these sections would most closely align with the concept of areas except these areas do not necessarily contain multiple locations.

The distinction of location and area does not fit well in the context of a football pitch. The defined areas don't necessarily contain several locations and besides the small difference of the physical lines indicating the location's border's the two concepts really aren't any different. Using different terminology for sections of the pitch which are outlined and which aren't, doesn't help making this research project more clear. For this reason, one term will be used to refer to distinct areas of the pitch in this research project: an area. The word location will no longer be used in this research project in an attempt to make the text more understandable.

A different concept entirely are **places**. A place is the functional and fuzzy space directly surrounding a given object. Examples of this on the football field would be the goals or penalty spots. This term is distinct from the concept of area and will therefore be used in certain cases, when necessary.

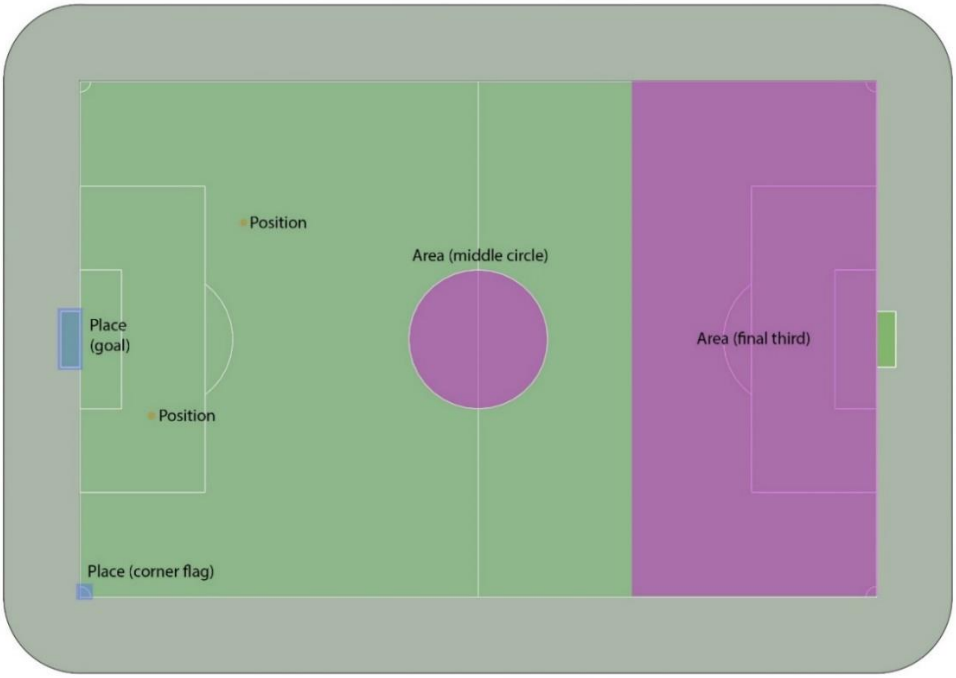


Figure 3.1 - Visual aid for the differentiation between the terms place, position and area, as used in this research project

3.2.2 Positional Roles

With these concepts separated, it is important to establish what positional roles existing in football. Since many research projects categorize player roles differently (Sarmento et al., 2018, 2014) the following section will outline what positional roles will be used and how they are understood in this research project.

Figure 3.2 – Positional roles

Table 3.1: Roles in football (from bottom to top, left to right)	
GK	Goalkeeper
LB	Left Back
CB	Centre Back
RB	Right Back
CDM	Central Defensive Midfielder
LM	Left Midfielder
CM	Central Midfielder
RM	Right Midfielder
CAM	Central Attacking Midfielder
LW	Left Winger
RW	Right Winger
ST	Striker

These are the general positional roles in football. More roles exist but they are just slightly altered variations of positional roles represented in Figure 3.2. It is important to mention that not all of these roles will be used in a formation and some roles can be occupied by several different players of the same team. A standard 4-3-3 with a flat midfield for instance has 1 left back, 2 centre backs, 1 right back, 3 central midfielders, 1 left winger, 1 right winger and 1 striker. Note that the CDM, LM, RM and CAM role are all missing from this formation while some roles have 2 or even three players assigned to them. Only the positional roles in the centre of the field can have multiple players occupying them simultaneously, on the left and right wing this does not occur. When multiple players occupy the same positional role, their placement on the field is still slightly different, this is indicated by adding “right” or “left” before their positional role. In the 4-3-3 example, the two centre backs would be broken down into the LCB (Left Centre Back) and RCB (Right Centre Back) and the midfield trio would be broken down into LCM (Left Central Midfielder) CM (Central Midfielder) and RCM (Right Central Midfielder).

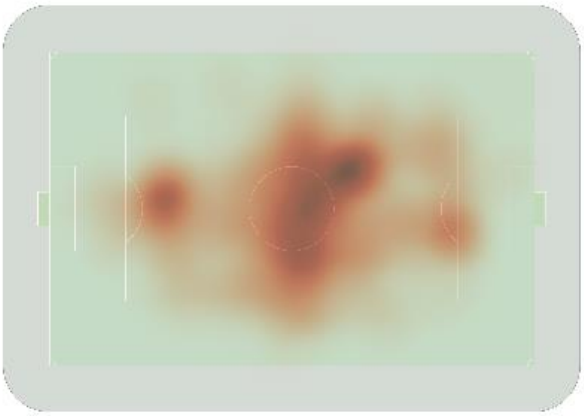


Figure 3.3 – Heatmap of a right defensive midfielder

While these positional roles do nominally correspond to a position on the field in name, outfield players don’t just move around in a 10-meter radius for the duration of the match. Figure 3.3 illustrates this by showing the movement in just one half by a Right Defensive Midfielder. In essence, the positional roles just hint at the placement of the player within the team shape, other factors contribute greatly to where the player can actually be found on the pitch throughout the match.

3.3 Football

3.3.1 Football: A spatiotemporal framework

From a geographic perspective, football can be observed as a spatiotemporal framework (Kotzbek & Kainz, 2014) in which objects act and interact in space over time. These objects can be separated into two categories: dynamic and static objects. The players and ball can be considered dynamic objects (Kim et al., 2011) and will mostly be represented as MPO's in this research project. The referee is technically also an MPO but should not affect the game as an MPO and will therefore not be counted as such. The inanimate objects are the field itself and both goals.

In general terms, the aim of both teams is to score at least one more goal than their opposition. A goal is scored when the entire ball passes the goal line, between the two goalposts and below the crossbar. The two goals are located on opposite sides of the field.

The space in which a football match is played is the football field, also named the pitch. The dimensions of a football field can be seen in Figure 3.4. It is important to note that these dimensions variate, but this is the official pitch size according to UEFA (UEFA, 2018). At the start of the game, each team occupies one half of the field, separated by the halfway line. Their objective is to defend the goal on their side of the field, and score in the goal opposite of them. After the halftime break, the teams switch sides.

The time discussed in terms of a football match, usually refers to the time elapsed since the beginning of the match. The clock is reset to 45 minutes at the beginning of the second half, regardless of any added injury time in the first half. This means that while the second half will start at 45 minutes and 0 seconds, it is often the case that more time than that has been taken in the first half. Note that this time does not refer to time the ball is in play, like in some other sports (e.g. futsal). In reality, a large percentage of time in a football match is dead time, meaning the ball is currently out of play. Exceptions to this duration of 90 minutes plus injury time exist (overtime) but will be disregarded in this research project, since they do not occur in league games.

Figure 3.4 – Dimensions of an officially UEFA licensed football field

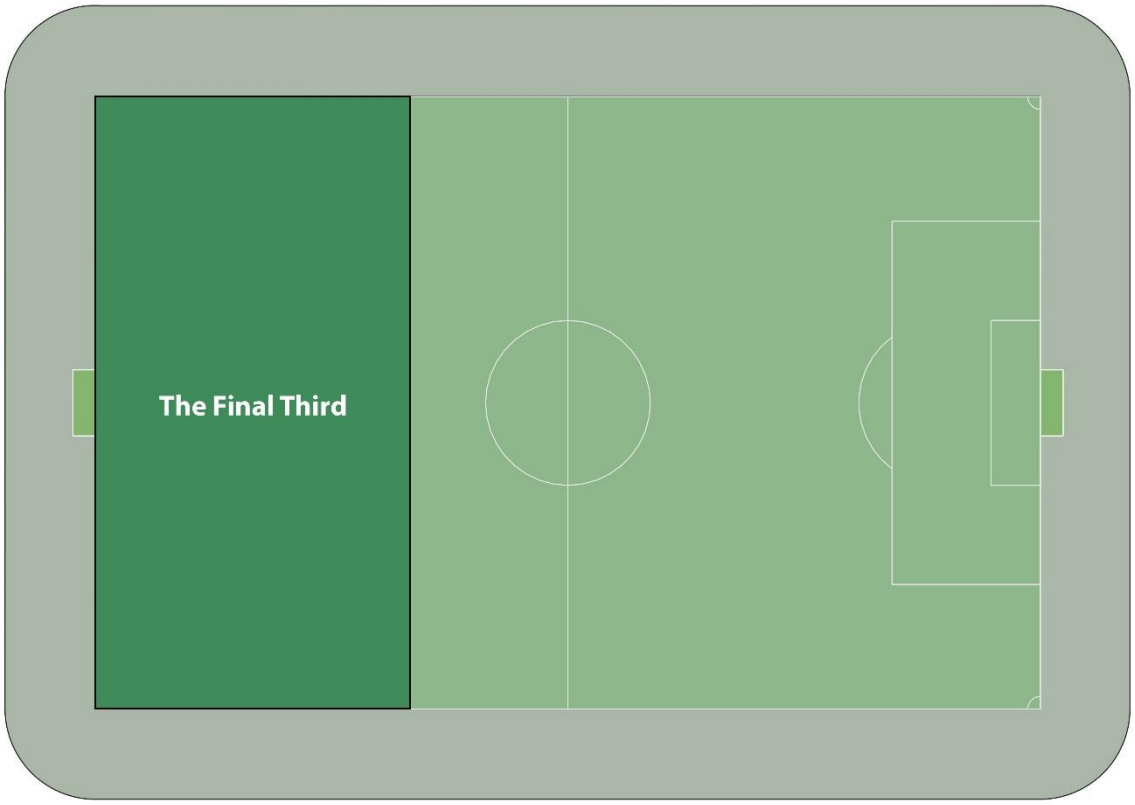


Figure 3.5 – The final third for a team attacking the goal on the left.

3.3.2 Football analytics

A majority of the literature cited in this research project can be considered football analytics research. *Sports analytics* is a term which was first used in American professional sports, it refers to a wide array of performance analyses which blur the line between scientific approaches and the perspectives of club management, media, and business (Link, 2018). Football analytics is simply the continuation of this new way of performance analysis in professional sports applied to the sport of football.

Gudmundson & Horton (2017) reviewed the trends in football analytics research and found that it predominantly focussed on (1) playing area subdivision, (2) network techniques for team performance analysis, (3) specific performance metrics, (4) application of data mining methods for labelling events, (5) predicting future event types and positions, (6) identifying team formations, (7) tactical group movement and (8) temporally segmenting the game. The fact that five of these trends are distinctly spatial (1, 2, 5, 6 & 7) and the other three easily have a spatial component introduced, once again emphasizes the importance of spatial research in football analytics research.

Out of these eight identified trends, the topic of this research project most comfortably fits in the construction of specific performance metrics; the construction of an advanced metric is the primary objective of this research project. Furthermore, the topic is at least tangentially related to five of the other identified trends (playing area subdivision, applications of data mining for labelling events, predicting future event types and positions, identifying team formations and tactical group movement).

3.4 Spatial models

In science, a (scientific) model refers to the abstraction of the real world into concepts, data or simple systems. This is done in order to test or simulate problems without having to build a simulation which mimics the real world exactly. Models are a simplification of reality in this way, gathering the relevant concepts and variables while cutting out all surrounding white noise, in an attempt to get a clearer picture of the examined problem.

In geographical information science, creating a model to simplify the world in order to represent it in a GIS is key stage in any project (Heywood et al., 2011). Such a model, specifically used for spatial analysis, is often referred to as a spatial model. The infinite complexity of reality is reduced to a carefully curated sample of data, specifically chosen to represent a certain measurable attribute (Clifford et al., 2010). It can be said that the use of such a model is the only way of evaluating our understanding of the complex behaviour of spatial systems (Heywood et al., 2011).

In this research project, the spatial model which is to be constructed seeks to evaluate the quality of offensive movement actions in football. Different aspects of the game of football will be abstracted and represented through the available spatial data (tracking and even stream data). The goal of all this abstraction is to analyse a real life phenomenon: an offensive players ability to find space, in a model test setting so that it and its components can be further examined without the endless complexity of a real world football match surrounding it.

3.5 Movement

Movement is the change of an object's or individual's spatial position in time, it is essential to almost all organisms and spatiotemporal processes (Dodge, 2015). For this reason, movement is a crucial object of study in a multitude of fields, such as but not limited to, behavioural studies, ecology,

environmental studies, epidemiology, transportation, mobility, geographical information science and of course sport science (Andrienko et al., 2017; Dodge, 2015).

3.5.1 Moving points objects

Through the lens of Geographical Information Science, a team of football players can be represented as eleven Moving Point Objects (MPO) moving through space: the pitch, in a well defined time period: the duration of the match (Laube et al., 2005). The match then becomes two sets of eleven MPOs, competing over another MPO: the ball. Laube et al. (2005) discussed how from a non-scientific perspective, the movement of these MPO's is probably the most intensively observed and most competently discussed motion of individuals ever created by human culture. As has been established though, the scientific community has mainly focussed its research on on-the-ball actions rather than movement. This while the study of movement is the key to understanding underlying mechanisms of dynamic processes (Dodge, 2015).

The treatment of football players as MPO's is not foreign outside of the scientific literature. The videogame series Football Manager has used MPO's to represent the players, ball and referee during matches for a long time now (Figure 3.6).

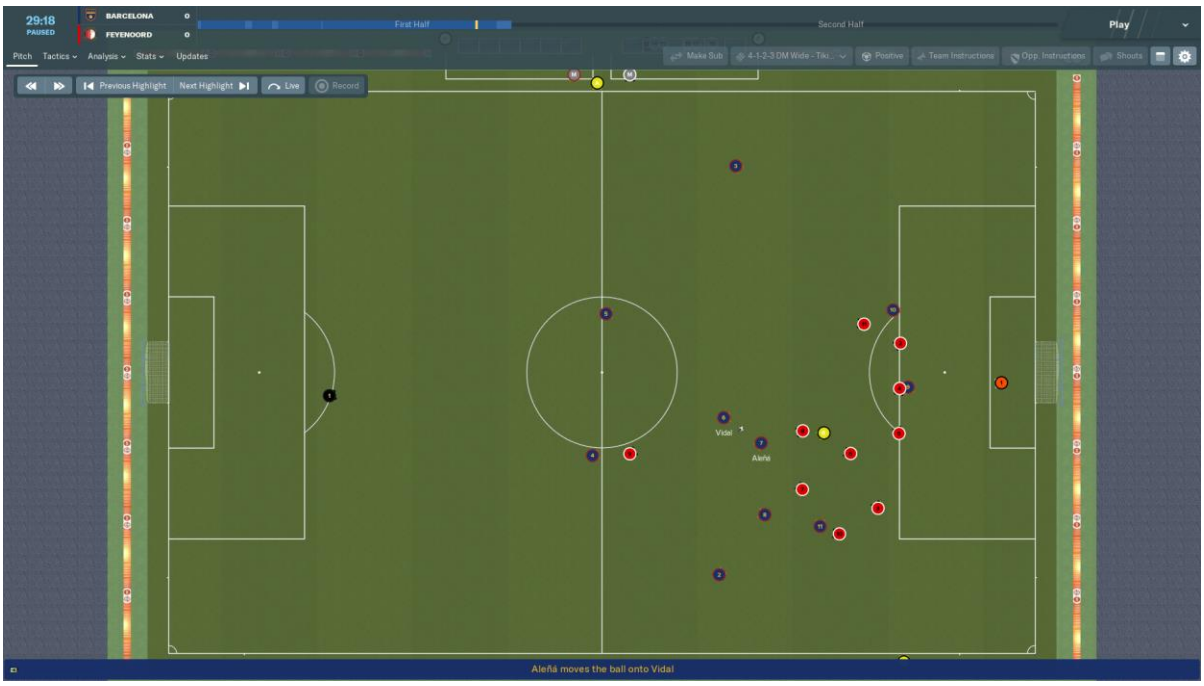


Figure 3.6 – A football match represented through MPO's in Football Manager 19

The MPO's representing the 22 players and the ball move for the full duration of a match. They do not move about randomly, hopefully, they move in a coordinated manner (Andrienko et al., 2019). Though this is not always the case, as Cruyff famously complained: *"I hate someone who is moving but doesn't know where he's going.",* we can safely assume that most of the time, some form of tactical or strategic thought guides the movement of players. This thought can be in relation to the opposition's goal, the ball, opposition players or teammates.

Since the last three objects on that list are also MPO's, the study of these particular kinds of movements pose a truly complex question of multiple different MPOs moving in relation to one another, acting and reacting continuously in space over time. This coordinated movement of multiple objects in relation to each other and their spatiotemporal context is called a concerted movement (van

Hoeve, 2017). It is the study of these concerted movements which can give answers to what Fernandez & Bornn (2018) called "one of the most important questions to improve in football": what should I do, when my teammate has the ball?

3.5.2 TRIAD Framework

Sinton (1977) defined the three basic components of geographic information as; Attributes, Time and Space. These are the basic three components that form the TRIAD framework (Figure 3.7) (Peuquet, 1988; Siabato & Manso-Callejo, 2011). Attributes are properties attached to objects, space represents the geometry of an object and time dictates at what time a particular object is relevant (Gold, 2016a).

MPO's can be represented in this framework at a given time, since at each timestep they have a given position. *Dynamic* aspects of the MPO's like velocity and direction of movement can be translated in to variables.

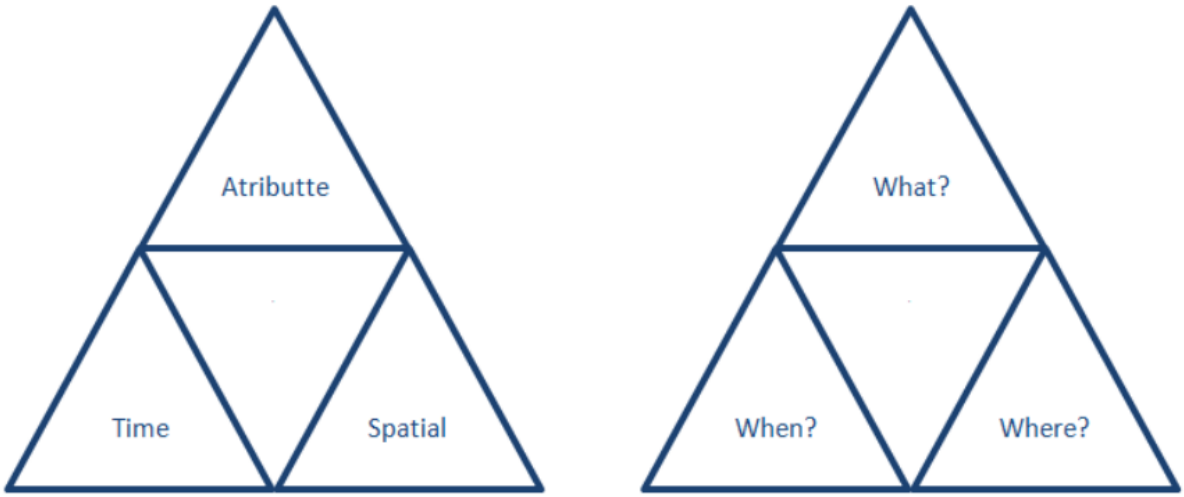


Figure 3.7 - TRIAD framework (Siabato & Manso-Callejo, 2011)

Time, space and attributes are three of the six types of information a spatial model can have according to Gold (2016a). The other three, topology, semantics and functions are not as relevant to MPO's and therefore not the main focus of this research project.

MPO's will most often be represented through the TRIAD framework in this research project. The table below shows the variables of a player MPO in this researched project for each component within the TRIAD framework. The ball will of course also be represented as an MPO, it will have very similar variables as the ones listed below, excluding the *team* attribute.

Table 3.2: Variables of a player MPO, divided into the TRIAD components

Component	Variables	Unit
Time	T	10th of a second
Space	X coordinate	Meters
	Y coordinate	Meters
Attribute	Velocity	Meters/Second
	Direction	Radians
	Team	A or B
	ID	Unique integer for each player

3.5.3 Trajectory analysis

Research surrounding concerted movement should discuss the key concepts of movement trajectories, movement parameters and movement patterns (van Hoeve, 2017). Movement trajectories can be defined as paths through space over time. Movement parameters are derivatives of movement trajectories, like velocity and acceleration. Movement patterns are frequent spatiotemporal behaviours of MPOs, derived from a combination of movement trajectories and parameters.

Andrienko et al., (2019) examine the 10 second 'episode' leading up to a shot. A shot is a clear border between one episode and the next, but for this research project it will not suffice since many offensive movement actions do not lead to shots. The point of examining offensive movement actions is isolating it from other factors. If an offensive movement action does not lead to a shot, that does not necessarily mean it was a bad attempt, perhaps the pass was not correct or the attacker's movement action was overlooked by his teammates. But the end objective of an offensive movement action is a shot on goal. Because of this, a 10 second timeframe preceding the potential shot can be used, similar to Andrienko et al (2019). Crucial here is the word potential, at the moment an offensive movement action is identified, the shot has not occurred yet, and may not occur at all. The offensive movement action takes place within the ten seconds preceding this potential shot. The time at which a potential movement action is identified will serve as the midpoint of the 10 second timeframe, meaning the 5 seconds after will be used to see its success (did it lead to a chance on goal?) while the 5 seconds before will be used to complete its trajectory.

4. Methodology

4.1 Introduction

This section outlines the methodology used to construct a spatial model which analyses the quality of offensive movement actions. In this section, the relevant variables contributing to offensive movement quality are listed, which are then further elaborated upon in the following sections.

Section 4.2 is about identifying offensive movement actions. The main objective of this section is to identify when such an offensive movement action is taking place. It is not reasonable to perform the analysis proposed in this research project for all 23 points in play (22 players and the ball), on every frame (roughly 57.000 per game) for several games, therefore, it has to be determined when an attack is taking place based on existing literature (Ueda, Masaaki, & Hiroyuki, 2014; Van Haaren et al., 2015).

Section 4.3 will explain some practical elements of the methodology. The data is described and any data processing which occurred will be explained. The software used for these processes is also mentioned. Finally, any assumptions under which this research project will operate are mentioned and explained.

Section 4.4 will be dedicated to the study of space occupation. This section seeks to answer the question: who controls what space, when? A methodology which addresses this question is chosen from the literature and this choice is explained in detail.

Section 4.5 explains the study of spatial value in sports analytics research. The space occupied by a player at any given time has a certain value associated with it, this section will explore this value and determine the best method of quantifying it in a GIS, based on existing literature (Fernandez & Bornn, 2018; Fernández, Bornn, & Cervone, 2019; Kotzbek & Kainz, 2016; Mackay, 2017; Spearman, 2018).

Section 4.6 will examine the question of ball trajectories and interceptions. The space occupied by an offensive player at any given time is only worth something if the ball can actually reach it without being intercepted by an opponent. This section will present a model for ball trajectories based on existing literature. Then, the previously constructed model for space occupation will be used to simulate a model for interceptions.

Finally, section 4.7 concludes the methodology chapter, aggregating all previous section into one spatial model to assess the quality of offensive movement actions. This section will give a direct answer to sub question 3: *To what extent can the elements contributing to the quality of an offensive movement action be spatially modelled in a GIS?* The remaining two sub questions and main research questions will be answered by actually using the constructed model to analyse match data.

4.2 Offensive movement actions

Establishing a formal definition for offensive movement actions is a crucial step in the building of a spatial model which assesses their quality. This section will outline such a formal definition and thereby demarcate what is and what isn't an offensive movement action. The elements needed to be present for an offensive movement action to take place will be listed, and then these will all be condensed into one definition.

To more concisely discuss hypothetical situations in a football game, the player making the offensive run will be referred to as: "the attacker". When advantageous, the opposition players trying to stop the attacker's run are referred to as: "the defenders".

Rein et al. (2017) found that there are two main attacking strategies. The first relies on the attacking team trying to establish a majority situation in certain areas of the pitch in order to outplay the opposition (Silva et al., 2014). This strategy is more methodical and typically indicative of a stronger team with more possession of the ball, trying to break down a weaker team's defence. The second strategy aims to get the ball into critical spaces in front of goal, leading to scoring opportunities (Duarte et al., 2012; Rein et al., 2017). In general terms, this strategy is more opportunistic, relying on low-percentage chance passes to land in areas crowded with defenders. Because of the high frequency of risky passes employed, this strategy usually does not rely on keeping possession and is more often adopted by weaker teams trying to catch their stronger opponents off guard in a counter attack (Gyarmati & Stanojevic, 2016).

But in essence, the end goal of establishing a majority situation, is still to set up an attacker for a shot on goal in a favourable area, since, in the words of Cruyff: *"You have to shoot, otherwise you can't score".* In other words, the objective of strategy one (creating a majority situation) is creating a situation from which it is easier to attempt strategy two (getting the ball in a critical space in front of the goal). It seems that in football, attacks can be boiled down to the effort of an entire team to create the most favourable situation for an attacker from which to take a shot on goal (Andrienko et al., 2019). The movement of the attacker leading up to this situation is what is examined in this research project: the Offensive Movement Action.

4.2.1 Possession

Team ball possession is the most commonly investigated performance indicator (Link, 2018). This makes sense, because without possession of the ball, any offensive move towards the opponent's goal is impossible. An attacker can therefore only make an offensive movement action if one of his teammates is in possession of the ball. An offensive run can only ever be capitalized upon if the attacker receives a correct pass in the advantageous area he has gotten himself into. It can safely be assumed that the opposition team will not actively try to provide this correct. Even if in some rare occurrences, the attacker might intercept a pass given from one defender to the other, the positional awareness required to do this is more of a defensive skill than an offensive one and therefore not part of an offensive movement actions. Thus, an attacker can only ever realistically count on a pass being given into the area he is running towards, if one of his teammates has possession of the ball.

While it is essential that one of the attacker's teammate is in possession of the ball in order for him to make an offensive movement action, the player with the ball can not himself make an offensive movement action. An offensive run with the ball is usually referred to as "a dribble" and requires a completely different skillset than an offensive movement action without the ball. This research project specifically aims to quantify the ability of an attacker to find space for himself without the ball, not the ability to successfully dribble past opponents.

4.2.2 Attacking phase

This research project looks at an attacker's ability to find space for himself in order to receive the ball and subsequently score a goal. The likelihood a goal will be scored is different from every position on the field (Kotzbek & Kainz, 2016; Link et al., 2016). If the attacker can not reach any position from which a shot has a realistic chance to end up in the back of the net, he is currently not in an offensive position and therefore he is not performing an offensive movement action. A more concise way of stating this is: an offensive movement action can only take place during an attack. There are different methods of determining when an attack is taking place.

Clemente et al., (2015) pinpoint the start of an attacking sequence after one successful pass was made, anywhere on the field, and determined it ends after the team in question loses possession of the ball. For their research, which focusses on the pivotal role of midfielders in building attacks, this is a satisfactory framework, but when the research subject specifically focusses on attackers, like in this research project, this definition includes many irrelevant sequences of play in which the ball never gets anywhere close to the goal and the attackers are not involved. An easy method of selecting the situations which have a chance of developing into actual attacks, is only looking at moments where a team has possession in their opponent's half. The chances that a direct opportunity is missed with this methodology is quite slim, as in an analysis of the 2011/12 Premier league season, less than 2,5% of assists were given from the own half (Davies, 2012), indicating that passes from one's own half rarely seek to capitalize on an offensive movement action.

Both Link et al. (2016) and Kotzbek & Kainz (2016) determined from which areas of the field, goals are most often scored (Figure 4.6 and 4.7). It can be observed that the chance of a shot actually resulting in a goal drops significantly when the shot is attempted from outside the penalty area. One could argue that if an attacker can not reach any area from which a shot is at all likely to succeed within a given timeframe there is no need to further analyse the situation, since an outcome can't possibly indicate a successful offensive movement action. The timeframe which will be used for this, are 10 second long offensive "episodes", defined by Andrienko et al. (2019). To this end, a situation will only be counted as an attack when at least one attacking player is located in the final third of the pitch. The final third serves as a clear and often used area from which an attacker can reach a dangerous position within a matter of seconds. Players who are not in the final third are simply too far away from the areas from which shots on goal should be taken to end their movement actions in these zones. This is not to say that scoring from outside the box is impossible and one should never attempt it, only that when such a goal is scored it is very rare that the goal scorer's movement prior to the goal is seen as a major contribution to the goal. These goals are usually scored because of the excellent shooting technique of the goal scorer, other attributes do not really factor into the equation as much.

Important to not is that this research project specifically seeks to address the ability to create chances from open, continuous play. This means that situations coming from set pieces (free kicks, goal kicks, throw-ins, penalties and corner kicks) will not be looked at.

Each frame of a match (roughly 57.000 per game) will be tested to see if the above-mentioned parameters are met. If all factors are present, then the player or players in the final third is or are performing an offensive movement action and steps will be taken to assess its or their quality.

4.2.3 Formal definition

An offensive movement action refers to the movement of an attacker, aimed at creating the most favourable situation for himself in order to score a goal. This means that the attacker seeks to get himself into an area where (1) he can potentially receive the ball and (2) attempt a shot from as favourable apposition as possible. The space the attacker occupies is quantified by these two factors in order to determine his offensive movement action's quality.

An offensive movement action can only exist in certain situation.

1. The attacker can only make an offensive movement action if one of his teammates has possession of the ball. It is essential that this possession is in open play, waiting for a free kick to be taken does not count as an offensive movement action. The team which has possession

of the ball is listed in the event stream data. The event stream data also provides information about the phase of play, be it open or not.

2. An offensive movement action can only be made during an attacking episode of play, meaning that the ball has to be in the opponent's half. The position of the ball at any point in time can be gathered from the tracking data.
3. An offensive movement action can only be made from an offensive position, meaning starting from within the final third. It is also essential that this attacker is not in an offside position. The exact position of all attackers and defenders (for the offside decision) can be gathered from the tracking data. The dimensions of the field are listed with this tracking data as well, making the construction if a final third an easy assignment.

Frames will be selected based on these three rules, these will be known as "key frames". The trajectories which will be analysed will consist of at least 100 points per player, starting 5 seconds before the key frame and ending 5 seconds after. It is very likely that, once the conditions are met, an attacking phase will have multiple consecutive keyframes. The ending of the offensive movement action will be set when the last of this array of keyframes has occurred. This means that an offensive movement action will start once a somewhat dangerous attack begins, and ends when (1) the attacking team loses possession of the ball, (2) the attacking team plays the ball back to their own half or (3) all attackers leave the opponent's final third.

4.3 Practical elements

4.3.1 Data

Spatial models are commonly constructed in a GIS using spatial data as raw materials (Heywood et al., 2011). The data used in this research project was kindly provided by SciSports (SciSports, 2019). It consists of both tracking and event data of 14 matches played some years ago in competitive professional football. All matches are played by one particular team versus various opponents which creates the opportunity to follow the players of that particular team for several games to hopefully discover offensive patterns. SciSports has requested that no names should be explicitly mentioned so neither the club, nor the players in questions will be named in this research project. Instead, each player will be given an anonymous ID.

The format is an XML file. In the event data, all events are categorized, giving a position on the pitch in X and Y coordinates and given a *t* in milliseconds to demark the exact time in the match the event took place. Additional information is added depending on the type of event, a shot for instance will have its result noted (e.g. miss, block, save or goal). The tracking data has a temporal resolution of 10 Hz, and all 22 players currently in play and the ball are tracked in the X and Y dimension. These X and Y dimensions are tracked. Noteworthy is that the ball does not have a Z dimension, as is the case for some forms of tracking data, and the positions of the arbitral team is not tracked. This tracking data can be characterized as 3D spatiotemporal points (X, Y and t) but could easily be translated into a 3D polyline because of its temporal consistency and lack of any other attributes (Baars, Oosterom, Verbree, & Gorte, 2004).

These X and Y dimensions are given in a coordinate system which has the centre spot as the 0,0 coordinate. The X and Y dimensions stretch just beyond the outer limits of the field, meaning that it is possible to see the entire field and when an object has crossed the line and is out of play.

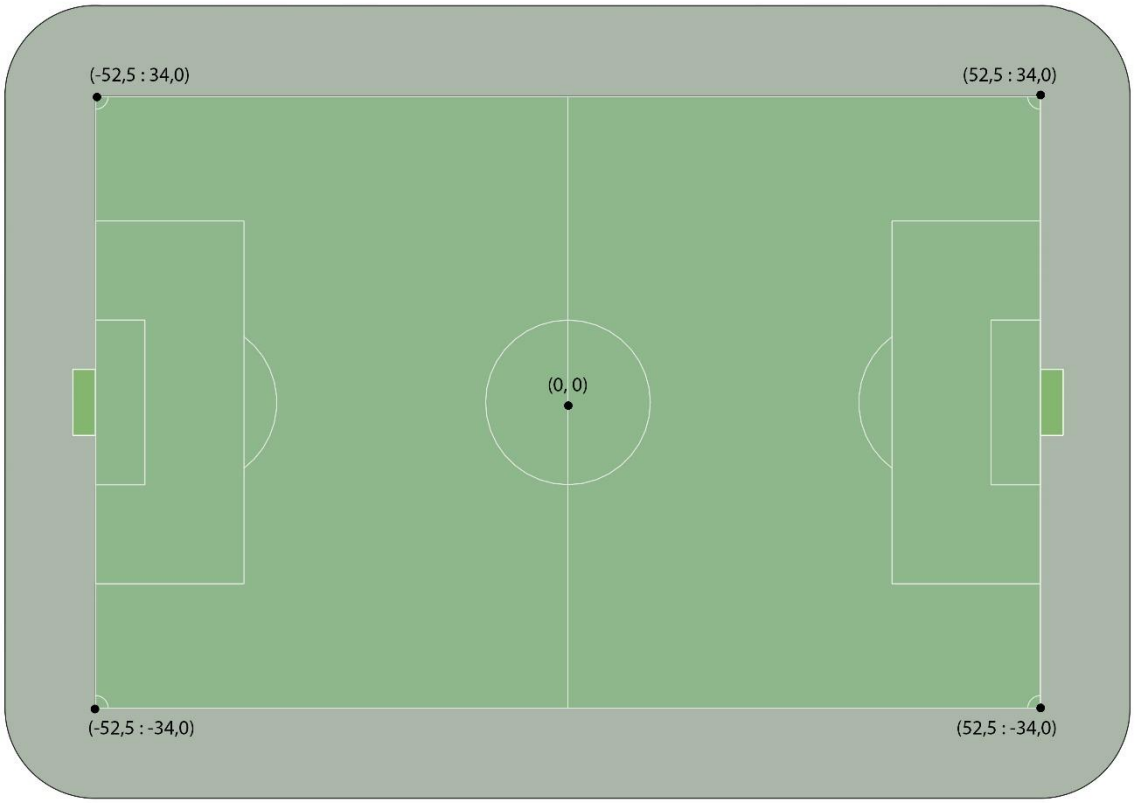


Figure 4.1 – Coordinates on the football field

This data does not contain any added information beyond what is described here. Derivatives such as speed or acceleration, have to be calculated by the user.

Besides the tracking data kindly provided by SciSports, they also provided around 50 games worth of event data. This data will be used to construct a spatial value model. The roughly 65 games worth of shots data proved not extensive enough for this purpose, so more data was brought in to supplement it. This data came from 2 large football data gatherers who provided some of their data for public use: StatsBomb (StatsBomb, 2020) and Wyscout (Pappalardo et al., 2019). Since these are different data gatherers, they use different standards, most notably a different coordinate system. The coordinates were translated to the system explained above. The end result was comparable event data for more than 1000 matches.

4.3.2 Software

Kotzbek and Kainz (2014, 2015, 2016, 2017a, 2017b, 2017c) showed how appropriate the platform of GIS can be for football match analysis and urged fellow GI experts to do further research in this field. Their call to actions will be answered in this research project and a GIS platform will be used to analyse the acquired football specific geodata. Specifically, QGIS will be used.

Besides the GIS platform, various other software programs will be used in this research project. Microsoft Word will be used to write the reports. Microsoft Excel has and will be used to bring the data from the XML file into QGIS and making graphs and tables. It is likely that in the future of this research project, other programs will be added to this list.

The coding language Python3 will be used to write any necessary scripts. Python3 works well in tandem with QGIS.

4.3.3 Assumptions

Spatial models are simplifications of the real world (Heywood et al., 2011). One simplification method is to assume that very rare occurrences do not occur. The following list will explain such assumptions made in this research projects.

1. Perfect decision-making of the arbitral team. While fans love to discuss mistakes which are made by arbitral teams in real life football, this research project will act as if they do not exist, the arbitral team will always make the correct decision. A player who is in an offside position therefore does not have a percentage chance that he can still receive the ball and continue on with the attack, even though this might be true in real life.
2. The referee is incorporeal. This assumption also focusses on the arbitral team, specifically on the head referee since he or she is the only one on the pitch during active play. The assumption is made that the ball will ever hit the referee, even though there are instances of this occurring in real life. This assumption is made because the dataset provided does not have the necessary tracking data of the referee.
3. The end objective of an attack is always to score a goal. Andrienko et al., (2019) point out how in some rare occurrences the team in possession might not have the objective of scoring a goal, instead opting to keep possession to run out the clock in order to protect a narrow lead. This situation will be disregarded as a possibility in this research project, it will be assumed that attacker will always want to score more goals.
4. The personal objective of the attackers is always to score a goal. The previous assumption refers to the objective of the team, this assumption refers to the objectives of single actors within the team. It is assumed that the movement of attackers is aimed at getting themselves in the best position from which to score a goal. This assumption is not always true, this might even be common, Fernandez et al. (2018) for example, specifically researched player's ability to create space for their teammates by their movement actions. In this research project though, the end goal of an offensive movement action will always be assumed a goal scored by the attacker and their quality will be judged accordingly. This means that an offensive movement action judged as "bad" in this research project might in fact be very useful for the team. For this reason, "bad" offensive movement actions are not paid much regard, good offensive movement actions are much more interesting discussion topic.

4.4 Space occupation

4.4.1 Possible and Dominant regions

A player can be said to control an area, if he is the first of all player on the pitch who is able to reach that area. The total area which is controlled in this manner by a player, is the player's dominant region (Fujimura & Sugihara, 2005; Taki & Hasegawa, 2000). Of course, a player's position is crucial in the determination of his dominant region, but so is his velocity and the direction in which he is moving. There are different methods of assigning dominant regions to a set of points.

The most common and simple approach is the use of Voronoi cells, also named Thiessen polygons, which simply assign an area to the point closest to it, based on Euclidian distance (Cervone, Bornn, & Goldsberry, 2016; Gold, 2016b; Rein et al., 2017; Rein, Raabe, Perl, & Memmert, 2016). This approach

ignores all movement parameters and simply bases the dominant regions on the positions of the players.

The inclusion of movement parameters is essential for this research project, so the use of simple Voronoi cells will not suffice. A model should be used which uses the *possible region* concept (Kang, Hwang, & Li, 2006) at multiple timesteps. A possible region is defined as the region where moving object *m* could be possibly located at future timesteps (e.g. *t)*, expected at *t0*. The size and shape of these regions can be influenced by velocity, orientation, acceleration and intention.

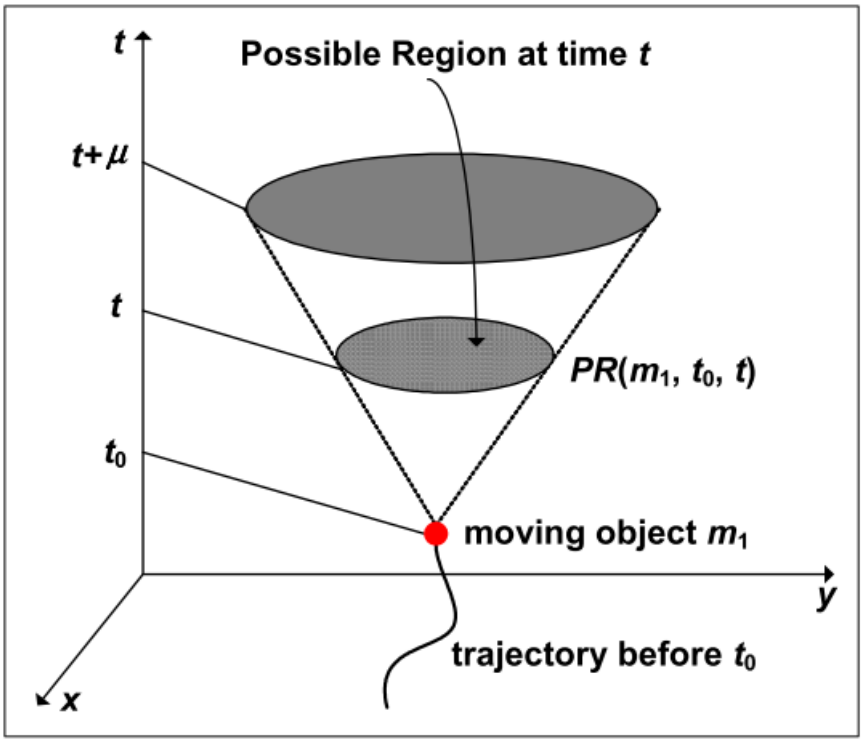


Figure 4.2 - Possible regions (Kang et al., 2006)

Taki & Hasegawa (2000) used a more complex model, taking into account velocity of the players at the time of calculation and using a constant acceleration and maximum speed, to draw a set of time contours per point. The time contours represent the outlines of the possible region at different timesteps (different *t´*s). Based on the points of intersection between these time contours, dominant regions can be drawn. It is essential to limit the amount of timesteps to a reasonable maximum amount (Kang et al., 2006). This is to prevent the entire field being covered by every player's dominant region.

The time contours themselves provide the option to extract additional information. It is possible for instance to determine how much time it takes for a player to get to a certain position and subsequently calculate the amount of time this player has before an opponent can also reach that position. This creates the possibility to establish how much time a pass has to traverse a passing lane before it is blocked or how much time an attacker receiving the ball has to control and shoot the ball. Another benefit of these time contours is that a *contested region* can be created, based on the amount of time a player has before an opponent arrives, if this time is below a set threshold, the region is contested. These applications are very suitable for this research project and coupled with the fact that velocity is included in the equation, Taki and Hasegawa's model is highly preferable to the use of Voronoi cells.

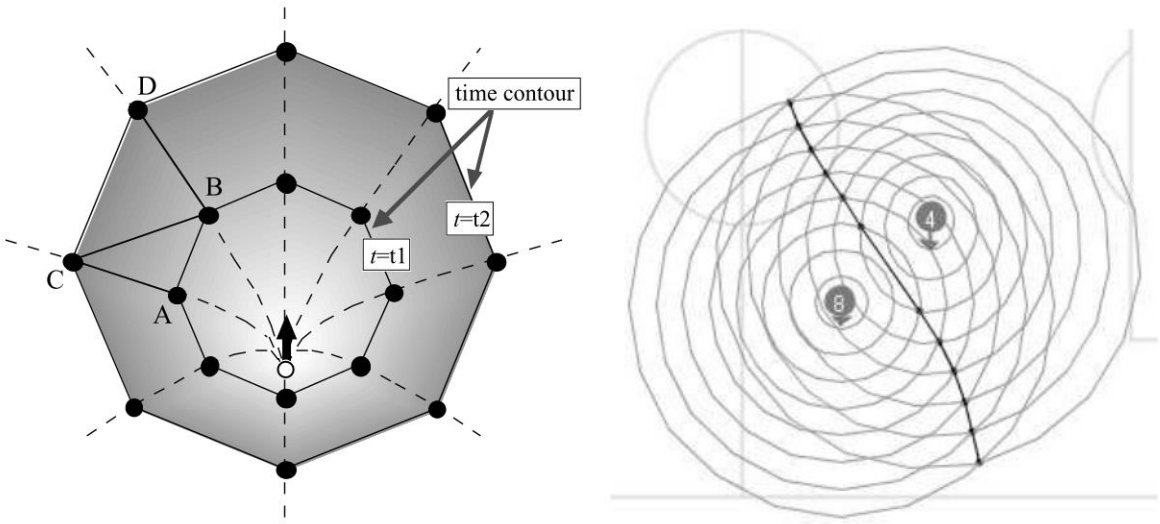


Figure 4.3 and 4.4 - Taki and Hasegawa's (2000) shortest time contours (left) and border between dominant regions based on the resulting time contours (Chawla et al., 2014) (right).

The constant acceleration used in Taki and Hasegawa's (2000) model eliminates a problem existing in the data used in this research project. While the tracking data determines the positions of all players and the ball it does not provide any information on which way the players are facing. While this is possible to discern when players are moving at higher speeds, since sprinting isn't possible in any direction but forwards, the majority of the time it is uncertain which direction the players are facing.

Taki and Hasegawa's (2000) model theoretically allows for infinite speed given infinite time, Fujimura and Sugawara (2005) eliminated this problem by introducing an equation for possible regions which includes a maximum speed. Their method consists of two formulas, the first determining the middle point of the circular possible region and the second determining its radius.

Formula 1: Centre point of possible region

$$X_0 + \frac{1 - e^{-\propto t}}{\propto} * V_0$$

With X_0 being the starting position (position at t = 0), $\propto$ being the acceleration modifier, t being the time and V_0 being the starting velocity (velocity at t = 0).

Formula 2: Radius of possible region

$$Vmax(t - \frac{1 - e^{-\propto t}}{\propto})$$

With Vmax being the maximum possible velocity.

This model has all the benefits Taki and Hasegawa's model has, with the added benefits of having a clearly defined maximum speed and a guide of implementation, described by Ueda et al (2014).

Fernandez & Bornn (2018) proposed a pitch control method accounting for the position, velocity and distance to the ball. A player who is further away from the ball is understood to have a wider area of influence, since it would take longer for the ball to reach him, thus giving him more time to reach the ball. More time means that the player is able to traverse more distance, resulting in a larger area of influence.

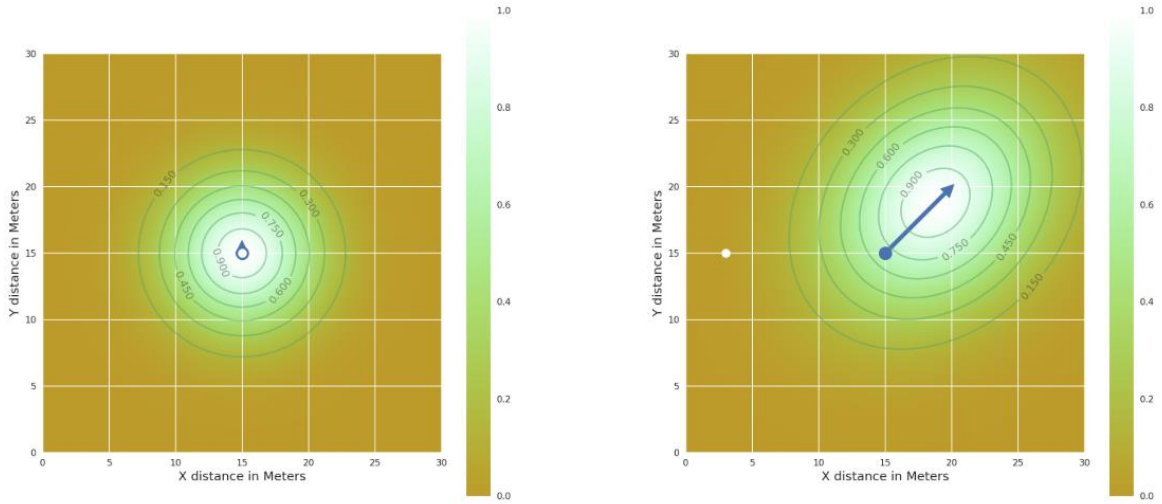


Figure 4.5 - Fernandez & Bornn's (2018) areas of influence

It can be said that Fernandez & Bornn's (2018) model goes beyond a movement model with its incorporation of position, velocity and distance to the ball. In some ways, this inclusion tries to tackle issues which are already addressed with the use of time-contours. Distance from the ball does not give a player increased speed, just more time to get to more places.

A note on players being offside. Players who are offside should exert no dominant region at all, since whether he is somewhere first or not does not matter, since he can not receive the ball in any area of the pitch, until he is no longer offside.

4.4.2 Physical limitations

The model chosen relies on creating time contours (Taki & Hasegawa, 2000) or possible regions (Kang et al., 2006) for different timesteps. The extent of these possible regions relies on two parameters: maximum speed and acceleration. It would be possible to assign each individual player a maximum speed and acceleration based on test data, gathered on training for instance. Such data is not available to this research project and can another method of determining maximum speed and acceleration has to be chosen.

Another option would be to gauge individual player's maximum speed and/or acceleration from the match data itself. It is of course always possible that a player does not need to sprint at his full speed at any moment in a match. This player would then be assigned a significantly lower maximum speed than he deserves. In the case where data for some players is limited to one match, or even only part of one match, the likelihood of them not reaching their maximum speed in this timeframe rises. It does not seem appropriate to determine maximum speed based on these snapshots.

It is likely that for these reasons, some researchers choose to simply assign a uniform maximum speed and acceleration to all players. Fujimura et al. (2005) determined their maximum speed at 7.8 m/s and this value was later used by different studies (Ueda et al., 2014). Spearman (2018) chose to approximate an average maximum speed sustainable throughout the entire match, and so landed on a much lower maximum speed of 5 m/s. The acceleration Spearman (2018) used was 7 m/s/s.

4.4.3 Conclusion

The model designed by Fujimura and Sugihara (2005) seems the best fit for this research project. There are several reasons for this. (1) The use of time contours fits very well with other aspects of the model.

It allows for a very easy integration of the interception model (more on that in section 4.6) and allows easy calculation of the time an attacker has at any given position they control. (2) The model is a good fit for a GIS platform. All aspects of it are easily translatable to GIS concepts and Ueda et al (2014) even provided a guide which could easily be followed within a program like QGIS.

The model uses two formulas (Formula 1 & 2) which calculate the middle point of the circular possible region (1) and its radius (2). These concepts can directly be incorporated in QGIS using the buffer tool.

Since Fujimura and Sugihara's (2005) model is used, it seems sensible to also use their prescribed acceleration and maximum speed.

4.5 Spatial Value

The end objective of an offensive movement action is a shot on goal. The specifics of the attack determine the situation from which this shot is taken. Space is crucial is in this matter, a shot from within the penalty box has a much higher chance to convert to a goal than a shot from 30 meters out (Kotzbek & Kainz, 2016; Link et al., 2016). For this reason, an attacker performing an offensive movement action seeks to control crucial space in front of the goal, from which a shot is most likely to succeed (Andrienko et al., 2019). The *control* of space has been explored in the previous section through the dominant region concept, this section seeks to establish what space is *crucial*. The objective of this section is to formalize a framework from which to assess the spatial component of chance quality. In other words: assign a value to space, based on potential shot success probability.

When opposition, chance type, assist type and time to control the ball are all discarded, the only factor that is left is the position from which the shot is attempted itself. This position itself becomes crucial when translated into a position relative to the position of the opponent's goal. This truly is the spatial component of chance quality, since pitch sizes can differ. It can be broken down further into two distinct factors which influence the potential of a shot: distance to goal and angle to goal.

Different methods can be used to establish a spatial value model. The first is by an equation, based on distance to goal and angle to goal. This approach would calculate the spatial value precisely for a given position. Another approach is to separate the area in front of goal in several zones and assign a spatial value to each zone. This value could be based on an equation, essentially combining the equation and zone approach, or based on empirical data: e.g. 23% of all shots fired from zone A resulted in a goal so the assigned expected goal value of zone A is 0,23. These zones could be square, creating something akin to a raster grid pattern covering the entire relevant area, or have different shapes, dividing the pitch up into a patchwork of polygons. It could be said that this *zone approach* is slightly less precise, since two points which are close together but not exactly occupy the same position, would get the same value assigned to them, while they would have different values in the *equation approach*. This is something to consider but could me mitigated depending on the size of the areas; smaller areas would mean this is not so much of an issue.

Link et al. (2016; 2018) dubbed the spatial component of their *Dangerousity* model; "the zone". This referred to the 2 by 2 meter square from which the shot was taken, within their 17 by 34 square grid (Figure 4.6). Each cell has a value associated with it, ranging from 0,04 to 1,00. This model does not have any underlying equation, instead it was constructed from 5 assumptions. (1) Distance to goal decreases and centrality increases danger. (2) Cells inside the penalty area more dangerous, because of the chance of attaining a penalty kick. (3) There is a homogenous area in front of goal in which the danger can't increase further. (4) An acute angle to goal reduces danger. (5) Areas with an acute angle to goal, but close to the backline have added danger because of the possibility to cross with a

decreased chance of being offside. It is important to mention that this dangerousity model does not only determine spatial value by the chance that a shot taken from there becomes a goal, like an expected goal model would. Other factors are included as well, like the chance of receiving a penalty kick or the possibility of crossing, as is evidenced by assumptions 2 and 5 respectively. This means that Link et al's (2016; 2018) model can not directly be implemented in this research project, since spatial value is expressed differently.

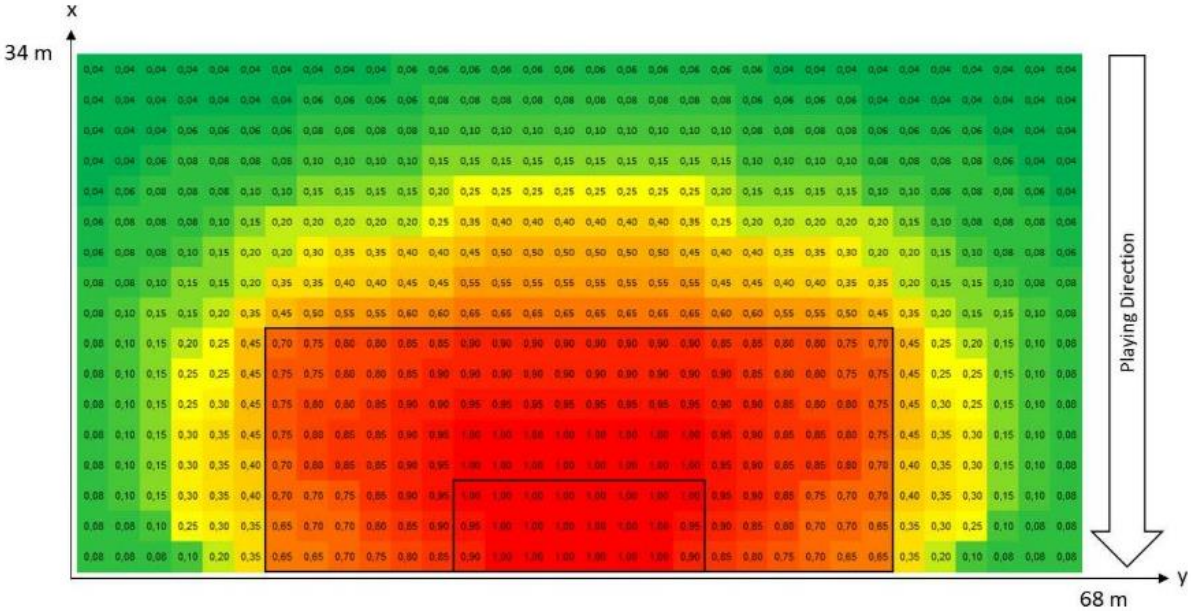


Figure 4.6 - Link et al's (2016; 2018) dangerousity zones

While the 5 assumptions on which this model is built are sensible, it is difficult to see how a direct numerical value is calculated through them. Some factors seem to get an enormous amount of weight, without a clear explanation as to why. The best example of this is the huge discrepancy between danger within and outside of the penalty area. While penalty kicks are very dangerous, since over 75% of them convert into goals (InStat, 2020), they are not a very common occurrence. In the 2018/2019 premier league season for instance, only 0,27 penalties were given per game (WhoScored, 2020). Keeping this fact in mind, the chance to attain a penalty kick seems too influential in Link et al's (2016; 2018) model, and without any supporting data, it is difficult to discern the validity of this particular aspect of the model.

Another method of assessing the spatial aspects of scoring attempts was developed by Kotzbek & Kainz (2016). Their *Spatial Scoring Probability Index* (SSPI) establishes a semi-circle in front of the goal, divided into different zones. Each of these zones has an expected goal value associated with it ranging from 0 to 99. An equation was used to calculate these expected goal values:

Formula 3: SSPI (Kotzbek & Kainz, 2016)

$$x_d = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$
$$x_a = \{0, 11.25, 33.5, 56.25, 78.75\}$$
$$SSPI = (x_d * 10) * 0{,}33 + (\cos x_a * 100) * \ 0{,}66$$

In which Xd is the distance in meters from the goal and Xa is the angle in degrees. This equation makes it possible to use a continuous model, calculating the SSPI from a given position instead of assigning it to a zone.

Compared to Link et al's (2016; 2018) *zones*, Kotzbek & Kainz's (2016) SSPI seems to assign much more weight to the angle to the goal. Even shots attempted from within one-meter distance from the goal have a relatively low success chance if taken from an unfavourable angle, while shots from outside of the box still have a decent chance to convert to a goal when attempted from a favourable angle.

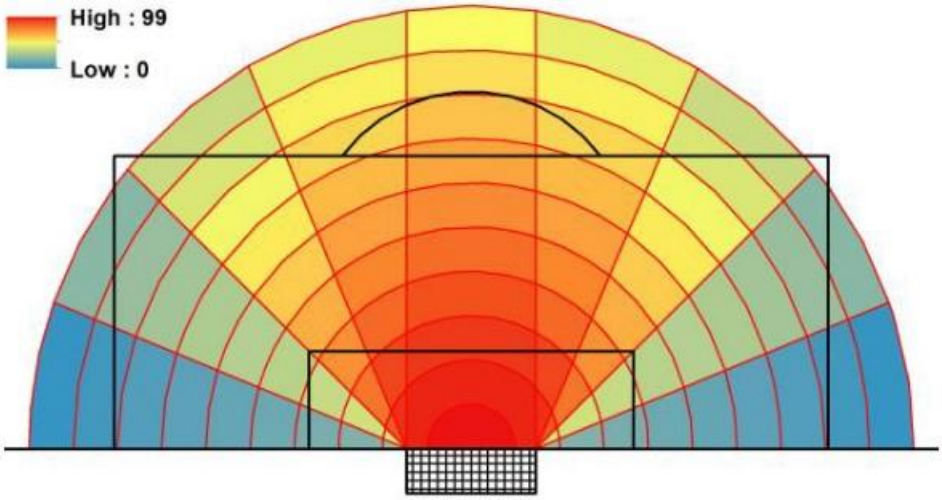


Figure 4.7 - Kotzbek & Kainz's (2016) SSPI

Other models simplify the spatial value model to a simple equation based solely on distance. One such model is Spearman's (2018) goal model. Based on the analysis of event data, the probability of scoring a goal was computed, conditioned on distance to goal. The goal probability approaches 1 at 0 meters and approaches 0 at roughly 35 meters.

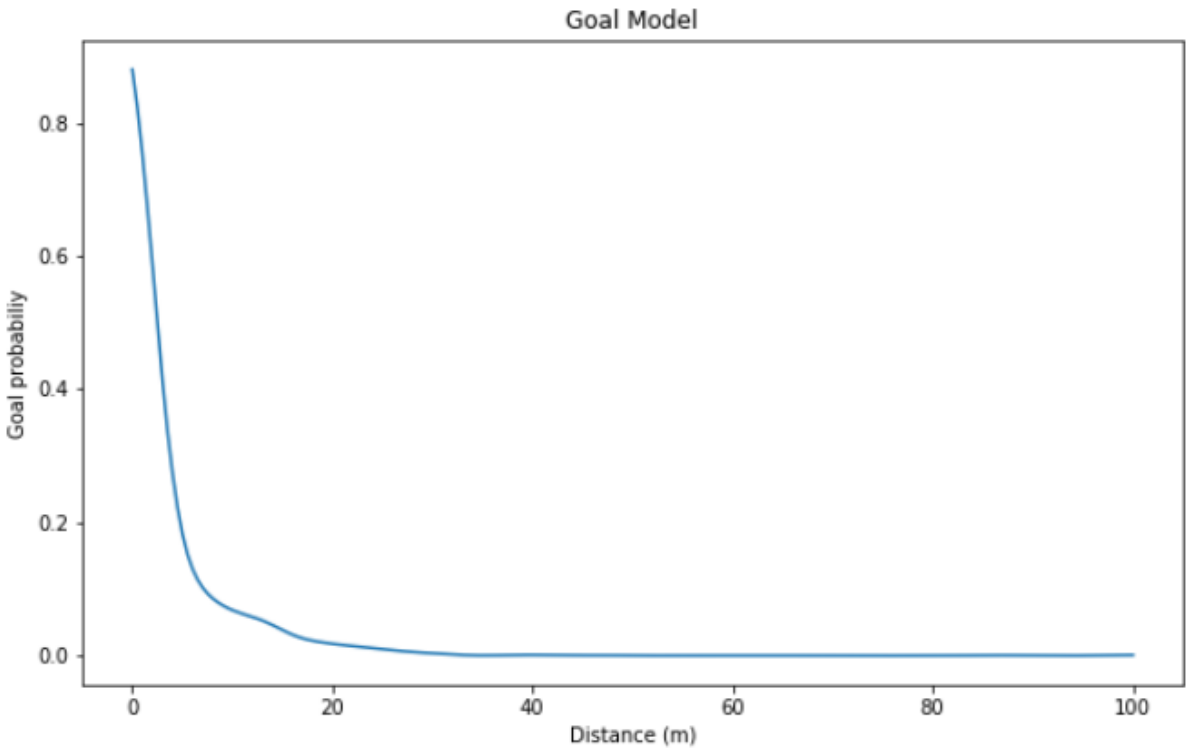


Figure 4.8 - Spearman's (2018) goal model

This approach simplifies and therefore streamlines the calculation process but ignores certain complexities, most notably the angle to goal.

Fernandez & Bornn (2018) developed a more dynamic approach, accounting the positions of the players and the ball when calculating spatial value. This approach then does not only look at the potential certain spaces have in terms of attempting shots, but also accounts for possible further passing combinations. It can be said that this approach is more holistic, and more closely approaches the *true* value of space in football. This research project however, specifically focusses on offensive movement actions with the objective of scoring a goal, other options are not taken into consideration. This is a slightly different skill than the one examined in Fernandez & Bornn's (2018) model, one more crucial for the role of striker.

In terms of the purely spatial aspects of the game, a dynamic model such as developed by Fernandez & Bornn (2018) does not make much sense, since the distance and angle to goal are static parameters. The dynamic influences like opposition players, time to control the ball and type of pass, will be introduced in different stages.

Another approach, originally developed for the sport of basketball, is based on economic reasoning (Cervone et al., 2016), viewing passes as transactions. These are not *fair* transactions, which would imply that both sides of the exchange walk away with equal value. Players within the same team are cooperating instead of competing and are thus looking to create value through *unfair* transactions, by *investing* the ball from a less to a more valuable stretch of *real estate*. Assuming that all players in the team are rational actors, it can be reasoned that in a transaction, the receiver of the pass occupies more valuable space than the passer, otherwise the transaction would not have taken place. It is important to note that this transaction is judged by the values of the entire zones of control of the two players, not just their respective position. With many of these transactions taking place each match, and with many matches to gather data from, general, team and player specific spatial value models can be established.

A deep learning approach to assess scoring probability from game state was developed by Decroos et al (2019). This model estimates the chance that any given game state translates into a goal for either side. The actual input of their probabilistic classifier is event stream data, meaning that what is tested here are on-the-ball actions and their contribution to the chance of a goal being scored on either side. This means that Decroos et al's (2019) model is not a correct fit for this research project, since spatial value is calculated entirely different. A pass made from a given position contributes entirely different to the probability of a goal being scored than a shot, or a cross. Their model does not assign value to space in the same way that is required for this research project.

Many of these approaches are data driven in nature (Cervone et al., 2016; Decroos et al., 2019; Spearman, 2018). Such a data driven approach was attempted in this research project in order to test the validity of different models. To this purpose, event data from over a 1000 matches was collected from various sources (Pappalardo et al., 2019; SciSports, 2019; StatsBomb, 2020). The games were gathered from different professional leagues and competitions. This resulted in the gathering of over 65.000 shots, roughly 720 of which resulted in goals.

The factors distance and angle to goal were isolated to see how they impacted the chance a shot translated into a goal.

Distance to centre of goal

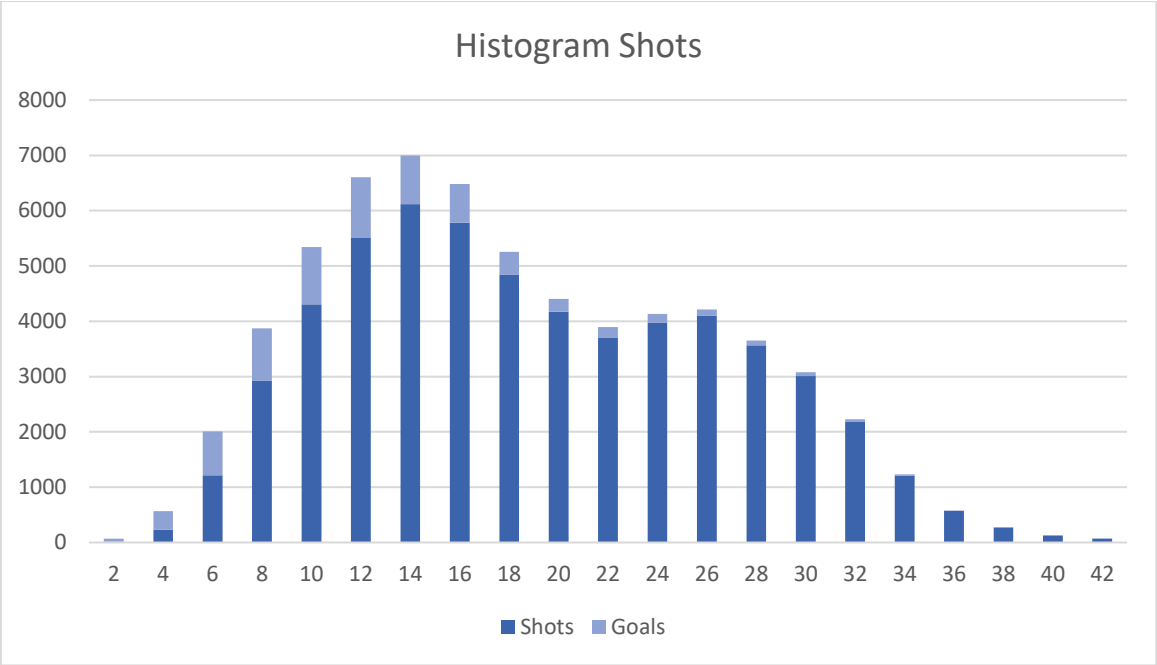


Figure 4.9 - Shot and goal histogram

It is important to note that the labels on the X axis indicate the upper bound of the category (e.g. category 2 is 0-2 meters, 4 is 2-4 meters, etc.). The Y axis gives the total amount of shots attempted.

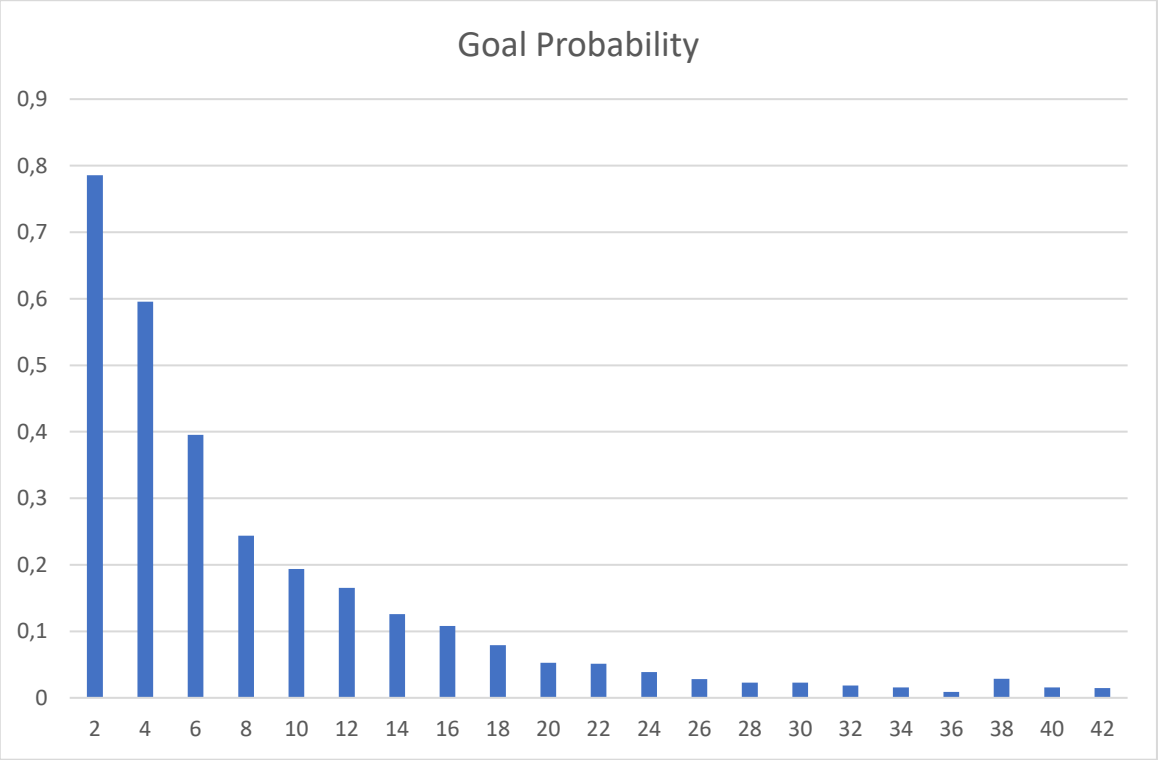


Figure 4.10 - Goal probability over distance

The goal probability is the proportion of goals of the total amount of shots.

Beyond 42 meters out from goal, too few shots are taken to eliminate statistical anomalies sufficiently. The amount of shots taken from within 42 meters distance from the centre of the goal accounts for

99,7% of the total amount of shot-data collected. It might even be said that this is the case, to a lesser extent, for all distances beyond 34 meters out, since single goals weigh quite heavily due to the relatively low amount of shots taken. This seems to account for the noticeable difference between the category 34-36 and 36-38.

Angle to centre of goal

The angle to goal is calculated from the centre of the goal. A symmetrical spatial value model is assumed, so both sides of the pitch are counted in the same angle categories. The angles are calculated in degrees.

Similar to the distance factor, the amount of shots attempted from a steeper angle than 50 degrees is too low to accurately represent a wider population. 99,7% of all shot-data collected is represented in the graph.

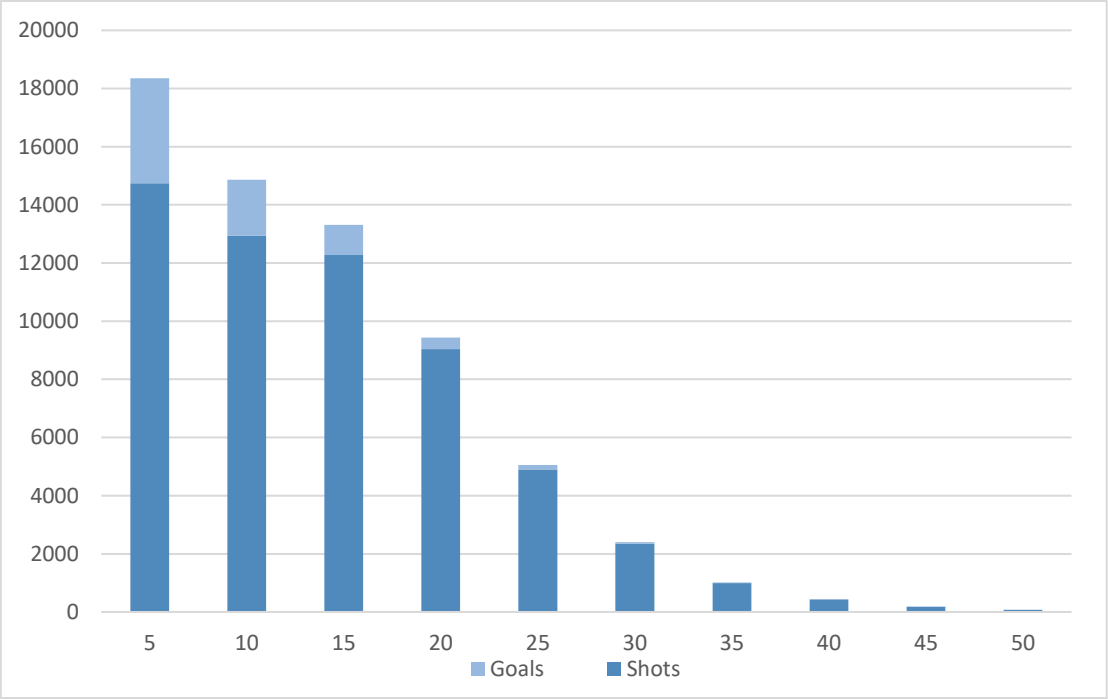


Figure 4.11 - Amount of shots and goals over angle to goal

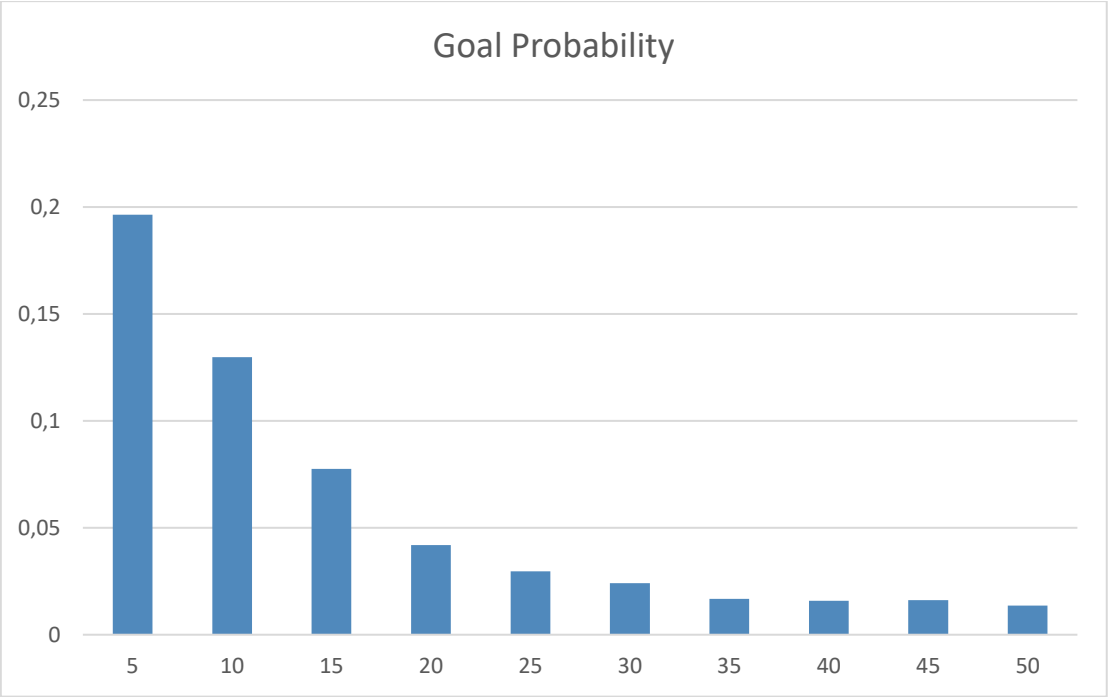


Figure 4.12 - Goal probability over angle in degrees

It can clearly be observed from these Figures (4.9 - 4.12), that the best predictor of the success of a shot, is a close proximity to the goal. While angle to goal is also an important factor, the difference between having a very favourable angle and having a very unfavourable angle, is nowhere near as impactful as the difference between being very close to goal and being very far away from goal.

It seems that, compared to the data available to this research project, Kotzbek & Kainz (2016) indeed put too much emphasis on angle to goal and severely underestimated the impact of distance to goal. Besides that, Kotzbek & Kainz (2016) identified a linear formula for the distance factor in shot susses rate, while it can be observed that an asymptotic formula would be more accurate, creating a sharp downward slope, before mellowing out in a long tail, similar to the figure used by Spearman (2018) (Figure 4.8).

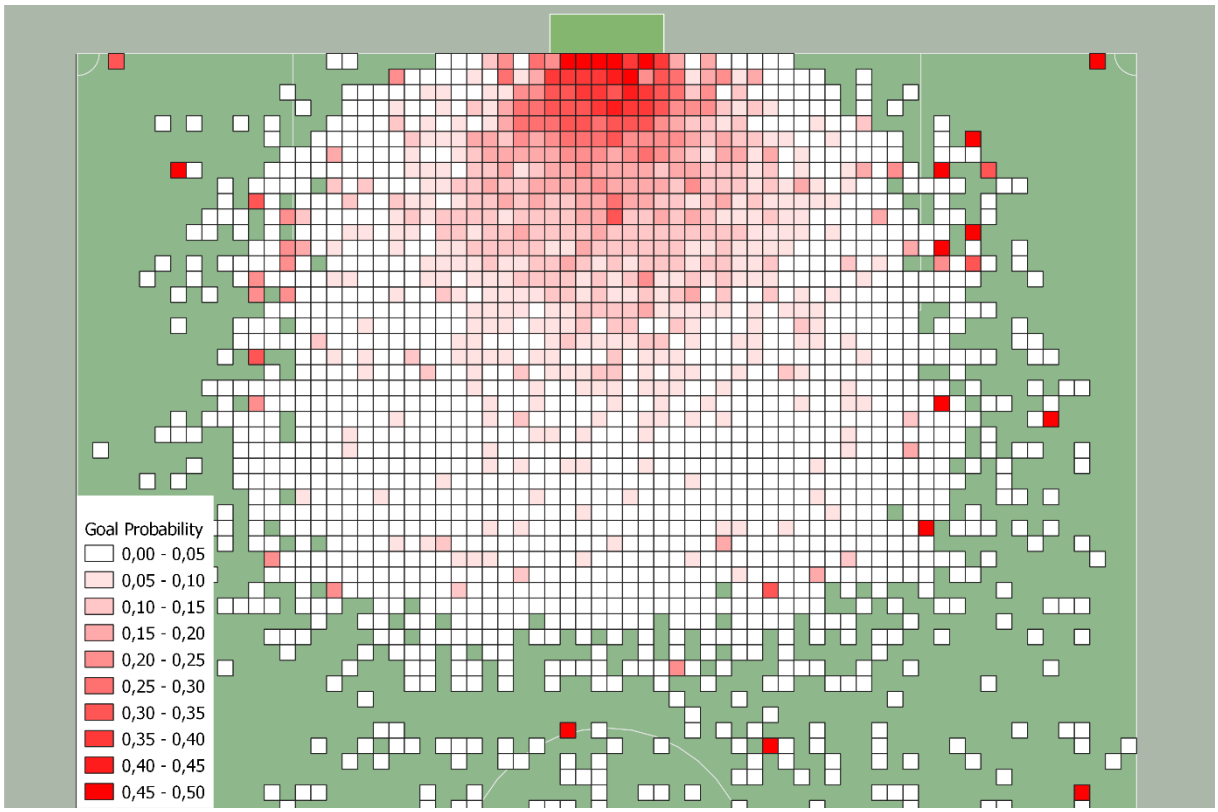


Figure 4.13 - 1 by 1 meter grid of goal probabilities

Figure 4.13 clearly shows the best areas to shoot from. What also stands out is the validity of Link et al's (2016) third assumption; the existence of an area just in front of the goal where the spatial value is almost uniform.

What can also be observed are slight weaknesses in some parts of the data. The data does not differentiate between shots from open play and shots from standard situations, like penalty kicks. The area from which penalty kicks are taken clearly stands out compared to its neighbours. These two particular cells having a higher goal probability than might be accurate shouldn't be enough to discount the entire dataset, but it's something to consider.

Similarly, it can be observed how cells with higher distance and/or steeper angles to goal are prone to statistical inaccuracies. Since they have a relatively low amount of total shots, 1 or 2 lucky goals can severely impact the goal probability in these cells. Similar to the penalty kick issue, this doesn't discount the entire dataset but is important to mention and keep in mind.

Because the intended outcome of the spatial value section is a continuous and symmetrical spatial value surface, a three-dimensional polynomial surface fit was generated using MATLAB (Figure 4.14). This surface can be used to provide any relevant position (meaning at least within 36 meters of the goal) on the pitch with a goal probability.

In an attempt to eliminate statistical anomalies, 100.000 points were randomly placed within a 40-meter radius of the centre of the goal. A circle with a 1-meter diameter was places around each circle. Circles with at least 10 shots within them were assigned a goal probability score by simply dividing the amount of shots that resulted in a goal by the total amount of shots. Each of these circles served as a datapoint to construct the surface fit shown in Figure 4.14. This goal probability model will determine spatial value in this research project.

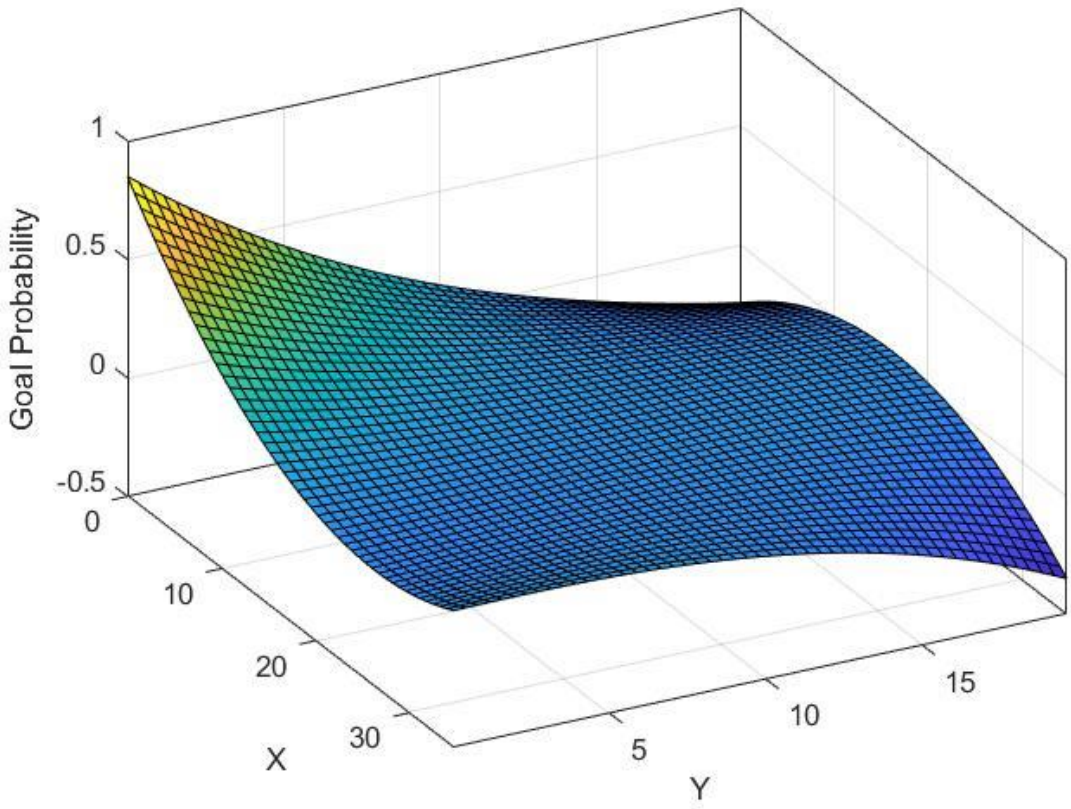


Figure 4.14 – Continuous three-dimensional surface showing the interaction of distance to goal in two dimensions and goal probability

The goal of the spatial model is to express the quality of an offensive movement action in numbers. To this end, all spatial value of all the space controlled in an attacker's dominant region needs to be integrated into a single number. The spatial value, expressed in figure 4.14, is multiplied by the total area it is tied to. All these values within the attacker's dominant region are then aggregated into a single number: the attacker's spatial value score.

4.6 Interception model

The space occupied by attackers is only worth its spatial value if the ball actually has a reasonable chance to end up in that space. In order to determine this chance, the trajectory between the space in question and the current position of the ball is examined. The obstacles in this trajectory are of course the opposition players, or to be more precise, the possible regions they project as determined in section 4.4. Depending on how long it takes before these possible regions overlap with the trajectory the spatial value of the space in question becomes worth more or less.

4.6.1 Flight time

In football the ball leaving the ground is no uncommon sight. Naturally, when the ball is too high for any opposition player to reach it, it can not be intercepted. The duration the ball is in the air depends on the angle and power with which it is kicked. Figure 4.14 demonstrates how passes with the same beginning and end point can have a drastically different trajectory and therefore flight time.

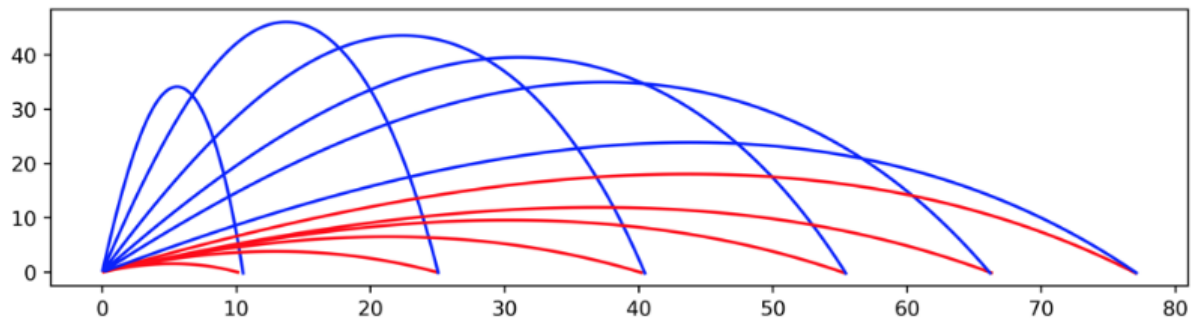


Figure 4.15 - Ball trajectories (Spearman, 2018)

A height has to be determined above which the ball is no longer reachable for opposition players. This height is, somewhat arbitrarily, set at 2 meters for outfield players and 2,5 meters for goalkeepers. While some player might be able to reach balls traveling at a higher altitude than 2 meters, most will not be able to do so consistently at full speed.

Depending on the trajectory, the ball can take significantly longer to reach its destination. Balls which are kicked higher will take more time to reach their end destination than balls which are kicked directly at their intended target. This will impact the amount of time the receiver of the pass has to do something with the ball. So while higher passes are able to ignore parts of the opposition's possible regions, they are not necessarily preferred.

Assuming the ball's trajectory takes the shape of an arc, the ball spends as much time going up as it does coming down. This means that arc's zenith is the exact middle point between the start and end of the pass, both in time and distance.

The force exerted by kicking a ball can be separated into two different forces: the horizontal force, and the vertical force. The horizontal force will determine how far the ball travels in the X dimension, as seen in Figure 4.15 and the vertical force will determine how far the ball travels in the Y dimension, before coming back down. These forces directly translate into velocities, so we have a horizontal and a vertical velocity at the point of departure. For simplicity's sake, all forms of friction are ignored, meaning that the horizontal velocity will remain constant until the ball reaches its intended destination. The horizontal velocity will decline because of gravity with a downward acceleration of 9,81 m/s/s. So, the time of flight is the trajectory where the ball is unreachable for outfield players, between point A where the ball ascends above 2 meters and point B, where the ball descends below 2 meters again.

The vertical height is described by Formula 4:

$$Y = V * t - 9{,}81 * t^2$$

With Y being the vertical height in meter, V being the vertical velocity at departure in m/s and t being time in seconds.

By calculating Y = 0, the t for which the ball reaches the ground again can be calculated for each given V.

Figure 4.15 demonstrates how two different passes, both originating from the same position (point A) and both traveling to the same destination (point B) can still have very different trajectories. Instead of calculating the optimal trajectory for each possible destination, this research project will use the optimal trajectory, illustrated by the red lines in Figure 4.15.

This optimal trajectory will be based on how long it takes for the attacker to reach the pass destination, derived from the attacker's time contours. The max pass velocity is 100 km/h, if a pass needs to travel faster than that in order to reach its intended destination at the same time as the attacker, it will simply have to arrive at a later time.

4.6.2 Curve

In football, good players have the possibility to add curve to the ball, altering their trajectory from straight lines to arcs. These arcs can have a significant impact on the trajectory and therefore end position of the ball. Despite this, the ball trajectory model in this research project will not use curved trajectories, for several reasons.

Curving a pass is a very difficult skill, only consistently successfully accomplished by the greatest of players. It is no wonder that when discussing this topic McGarry & Franks (2013) invoke the name of Roberto Carlos since he was well know for his mastery of it, but he is also widely considered one of the greatest ever left backs to ever play the game. The difficulty of this skill is such that in normal situations, one can't count on a teammate executing it perfectly. In other words, if a pass can only safely reach a destination by being curved perfectly, then there is a very slim chance that this destination is reached indeed. For this reason, it is assumed that passes travel in a straight line from their starting point to their end destination.

Besides this, curve is not adopted into the model for practical reasons. Several different speeds at which the ball can be played already have to be accounted for, if all of these passes with all a different speed would all get additional options for curve, the processing time would rise exponentially.

While it would be interesting to examine the possible curve of a football's trajectory in more detail, the subject is simply too far removed from the subject matter, offensive movement actions, to warrant an in-depth look. A research project focussed on the physics of the kick of a football or at least specifically focussed on passes or shots would undoubtably need to spend more attention to it. This research project however is specifically focussed on off-the-ball movement and therefore, the curve of a football's trajectory is just out of its scope.

4.6.3 Time to intercept

Depending on all these factors, the length of the relevant sections of the trajectory which overlaps with time contours projected by opposition players can be determined. This gives the amount of time such an opposition player has to actually reach out to the ball in order to make an interception. Naturally, more time means the player has a better chance to perform a successful interception (Spearman, 2018; Spearman, Basye, Dick, Hotovy, & Pop, 2017).

In order to simplify the model, all players are considered to have the same chance to intercept the ball given the same amount of time. Each quarter of a second adds twenty-five to the percentage chance that the ball is intercepted. This means that if a defender can occupy a certain space for a full second before the ball traverses it, he will certainly intercept the ball.

4.7 Conclusion and implementation

With all the components from the previous chapters, the quality of an offensive movement action can be determined. This section will briefly summarize the steps taken in this methodology chapter and illustrate the interaction between the different models. A visual representation of the spatial model, its components and their interaction can be seen in the methodological model (Figure 4.16).

Furthermore, the implementation of the spatial model into QGIS is explained in detail, since one of the objectives of this research project is to show how GIS can be used in football match analysis.

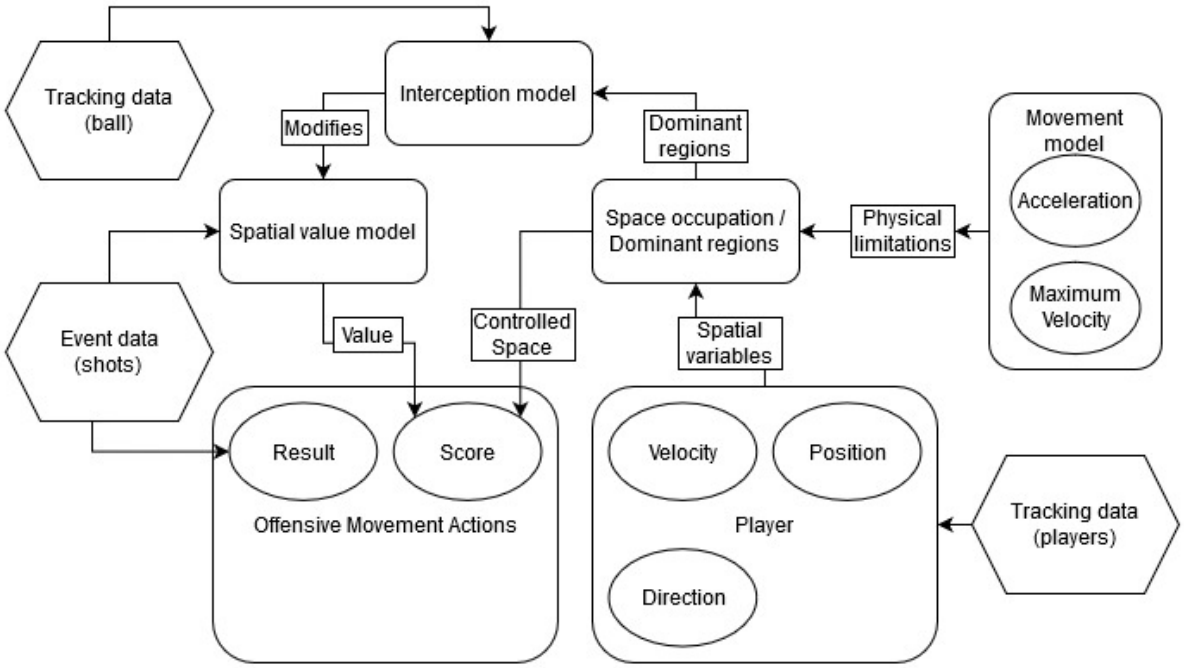


Figure 4.16 – Methodological model

Figure 4.16 visualizes the spatial model in a diagram format. Data is represented through hexagonal shapes, while concepts and models are represented by rounded rectangles. Circles found inside the rounded rectangles are aspects, derivatives or variables of the concept or model.

First it was established when an offensive movement actions is possibly taking place. An attacker can be considered to make an offensive movement action if (1) one of his teammate's is in possession of the ball (2) on the opponent's half and (3) the attacker himself is in the final third. These three rules were simple implemented in a Python script to select which frames are going to be analysed in the next steps.

The steps of the process described in the next paragraphs will be supported by an example from the data. In this example, the attackers (the red points) have managed to reach the backline with the ball (the white point). The attacker with the ball is not in a good position to shoot, which will most likely mean he is looking for a teammate. Thankfully, three of his teammates have entered the opposition box and one is lurking on the edge. These attackers are numbered from left to right in order to make further discussion easier.

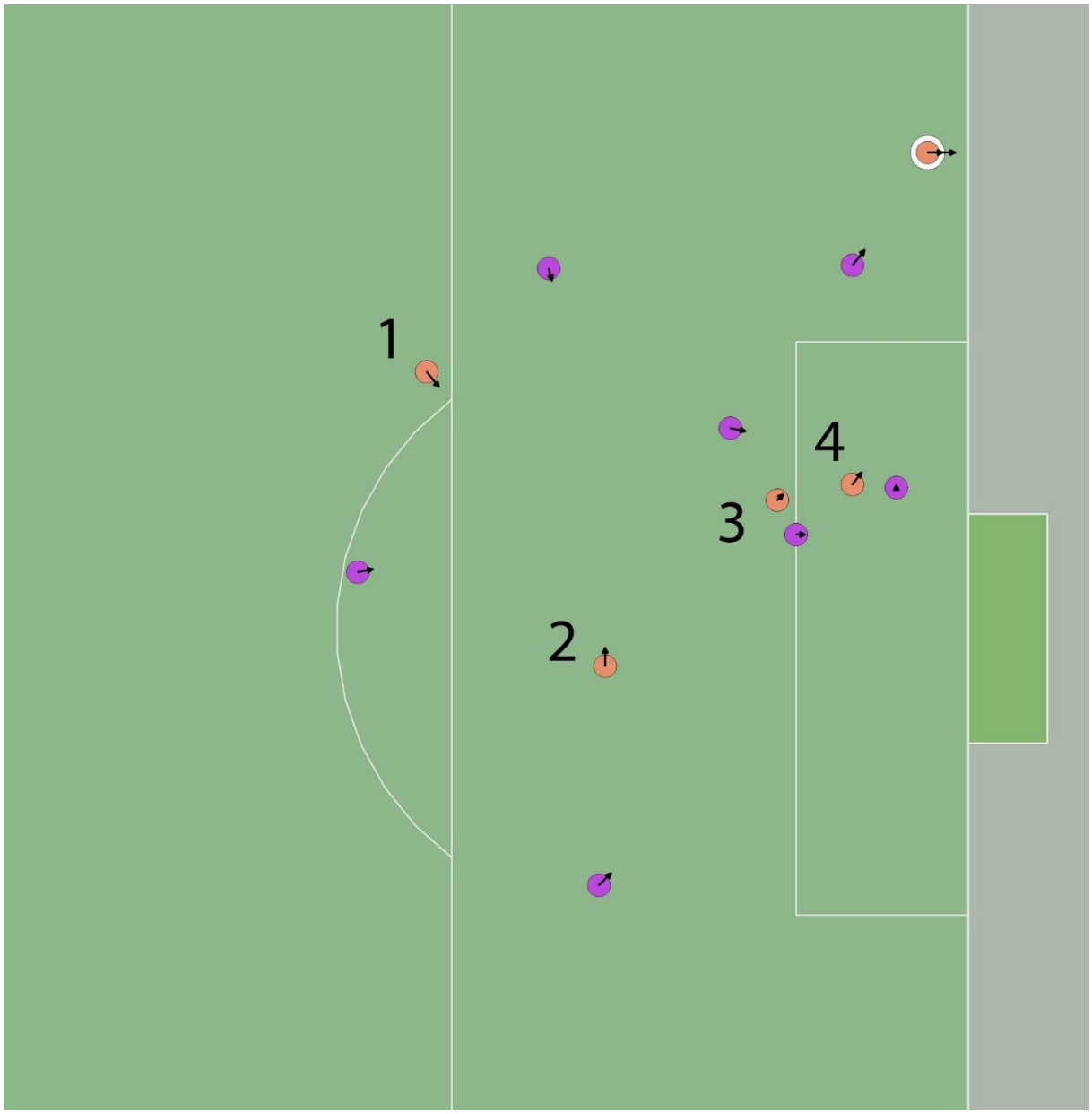


Figure 4.16 – Example attack situation

Then it was determined what player controlled what space through dominant regions. This model required several steps to implement in QGIS. The first was the drawing of possible regions through a point and buffer method. The position of the points and extent of the buffers were calculated through formula 1 and 2 respectively. This process was repeated for several timesteps, for each player in each relevant frame. Then, the points at which these possible regions of the same timestep intersected were identified. These points served as the vertices of the borders between the dominant regions.

To keep the example image clear, only the possible regions of two players are shown in the example image. These players are from opposite teams and are located in the middle of the image below.

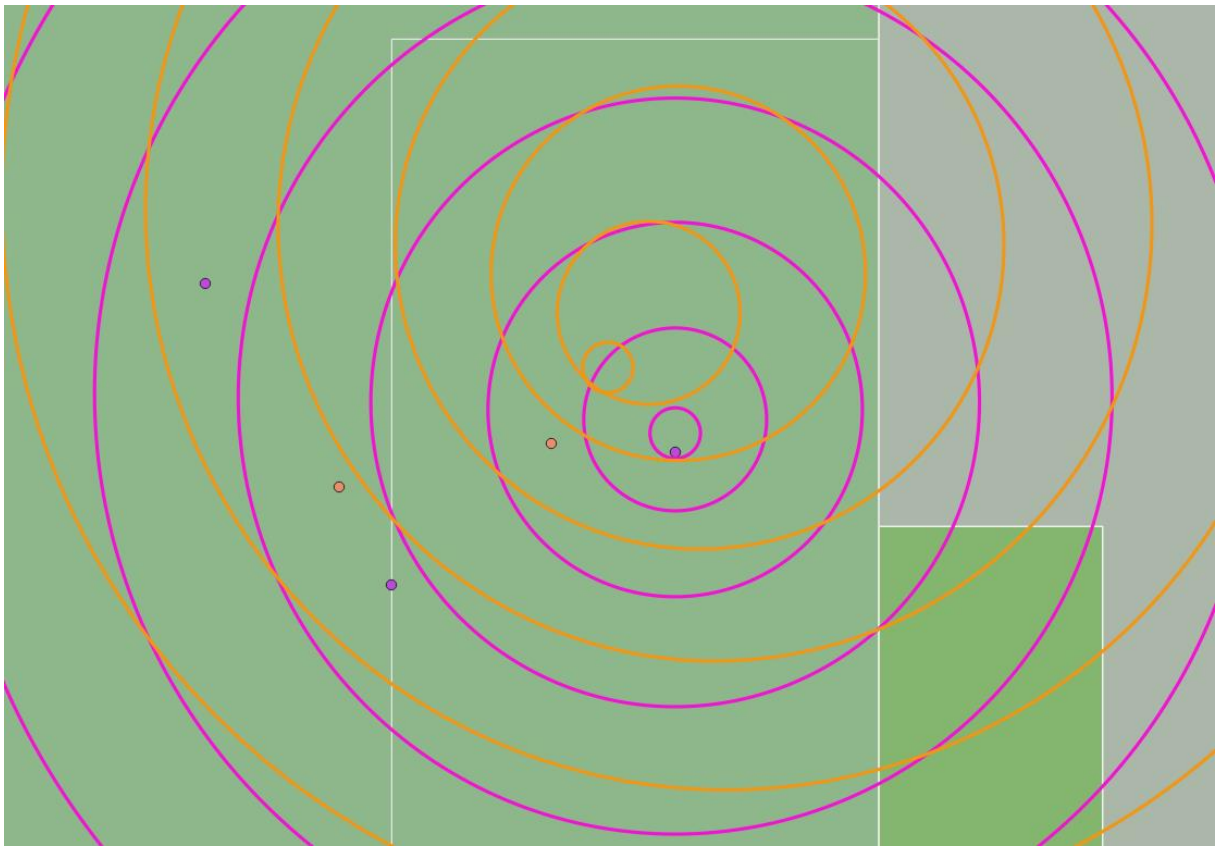


Figure 4.17 – Dominant regions of an attacker (orange) and a defender (purple)

It can be seen by these possible regions, that the attacker is moving at a higher velocity than the defender, or in this case, the goalkeeper of the purple team. As a result, the attacker can be seen to reach the area in front of the goalkeeper before the goalkeeper can.

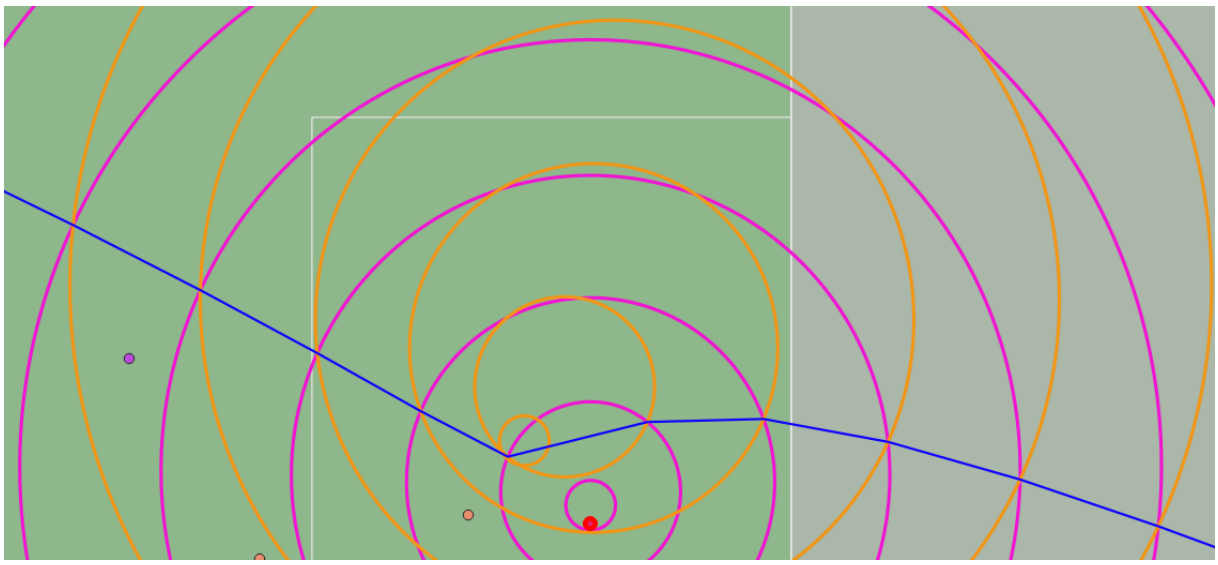


Figure 4.18 – Border between attacker's and defender's dominant region

The intersections between the possible regions with the same timestep can then be used to draw this line. This line is the border between the dominant regions of both players, every position on one side of the border can be reached first by the attacker and every position on the other side can be reached first by the goalkeeper. Repeating this process for every player makes it possible to draw dominant

regions covering the entire pitch. Polygons are a good representation for dominant regions as they really separate space into distinct regions, which is the objective of the dominant region model.

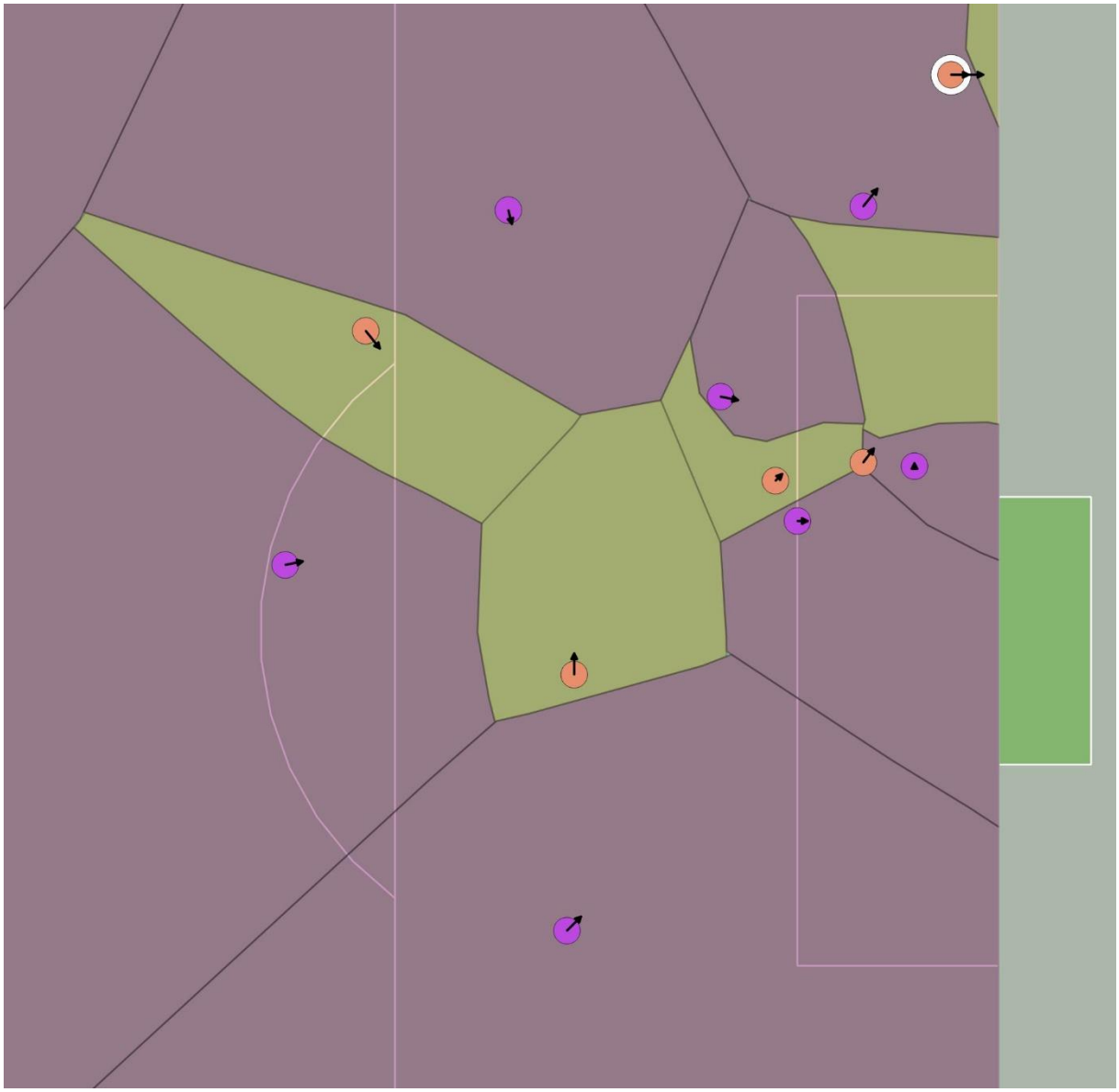


Figure 4.19 – Dominant regions in example situation, these regions are calculated starting at 0,25 seconds to account for the minimum time the ball needs to travel

The dominant regions drawn in this manner, can be seen to act quite different than Voronoi cells based on Euclidean distance. In some instances it can be observed, when players are close to other players and traveling at high velocities their current position is outside of their dominant region, because they can't decelerate fast enough to reach the position they are currently in, before another player can reach it. The base for the attacking regions is not the current position of the player, but the possible regions for several timesteps, starting at 0,25 seconds after the current situation. Attacker 4 exemplifies this in Figure 4.19.

Then, this space was given a quantifiable value, expressed as a goal probability, based on the controlled space and relative position of the ball. The data used to determine goal probability from different positions on the pitch was condensed into a single three-dimensional fit curve, in order to create a symmetrical spatial value model (Figure 4.14). The data from this curve was exported back into QGIS

in the form of points with X, Y coordinates based on their placement on the fit curve and an associated goal probability value. A raster based approach would also have sufficed here, but points proved a better fit with both the dominant region model (points have a precise position and QGIS can easily calculate whether that position is within the polygon) and the interception model (lines can simply be drawn from the position of the ball to each end position, represented by these points).

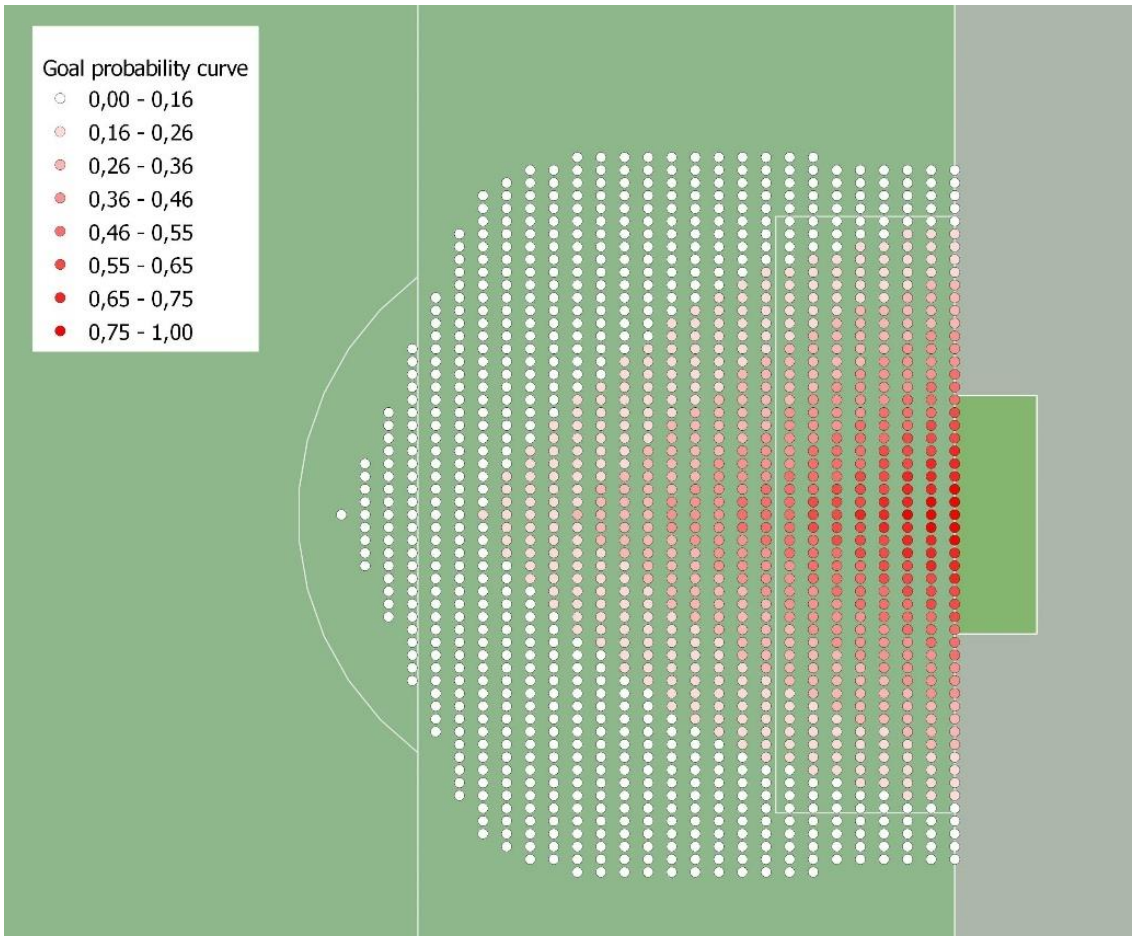


Figure 4.20 – Goal probability curve integrated in QGIS

Finally, it was determined how difficult it would be for the ball to reach the controlled space through the interception model. A line trajectory was drawn between the current position of the ball and goal probability points, as described in the previous paragraph, which are located in an attacker's dominant region. A line is the most accurate way to represent a trajectory in QGIS.

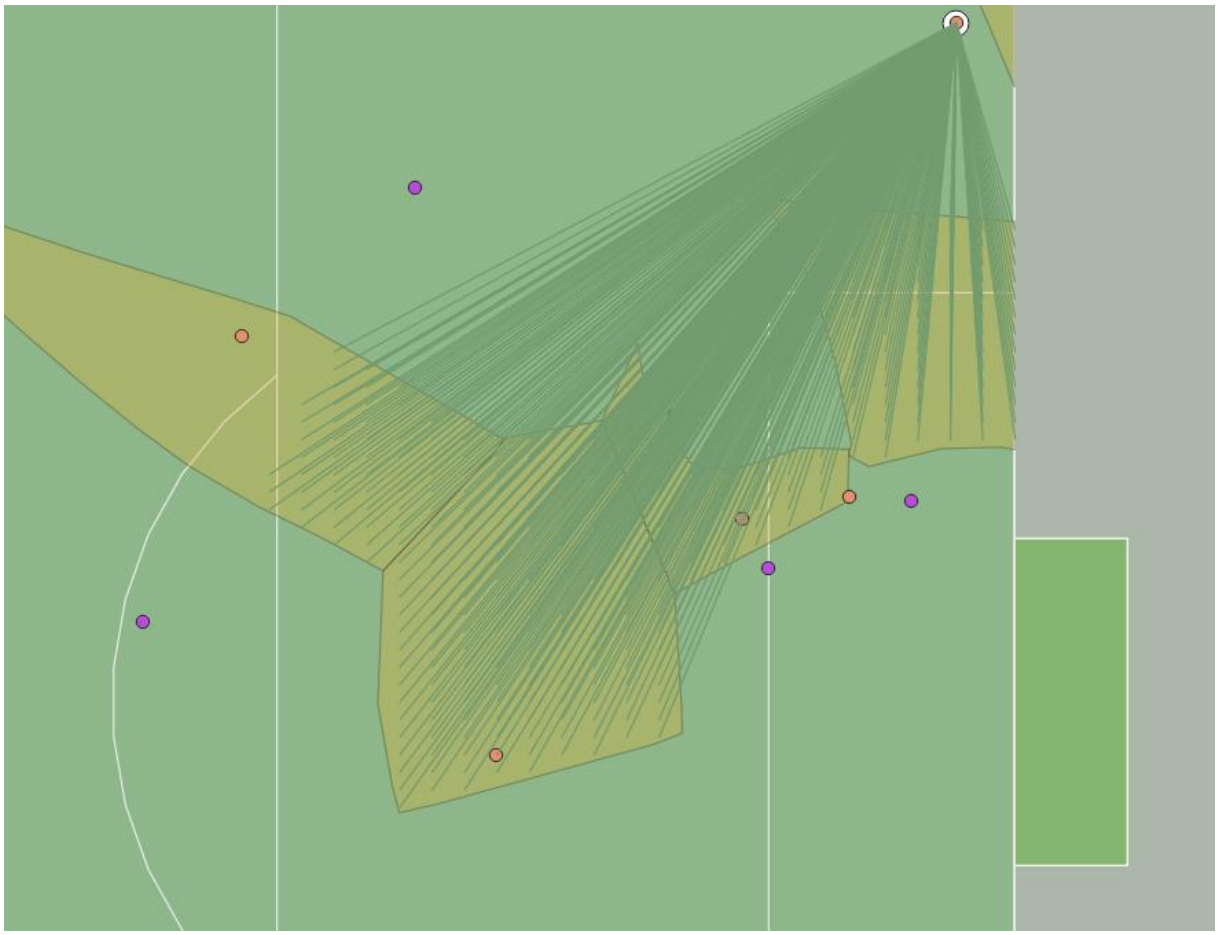


Figure 4.21 – All trajectories from ball to attacker-controlled goal probability points in example situation

The position of the ball is, of course, the trajectories starting point and the goal probability point is its destination. Then, it was determined how much time the ball had to reach its destination. This was determined based on how much time the attacker took to get there, derived from the attacker's possible regions at different timesteps. When this was established, the ball's intended trajectory was divided up into segments of a quarter of a second long, their length depended on the speed the ball needs to travel to reach its intended destination at the correct time. Each of these segments was given an average height, to determine whether or not it is subject to interceptions, a ball traveling over 2 meters high can not be intercepted. The defender's possible regions for the relevant timesteps were then compared to the trajectory segments, this made it possible to determine the amount of time the defenders had to perform an interception. When a defender can occupy a space amongst the ball's intended trajectory for a full second, the ball will certainly be intercepted, meaning that the goal probability value of the destination point, will not be counted towards the attacker's spatial value score. Likewise, a partial chance of interception means only a part of the goal probability value is counted.

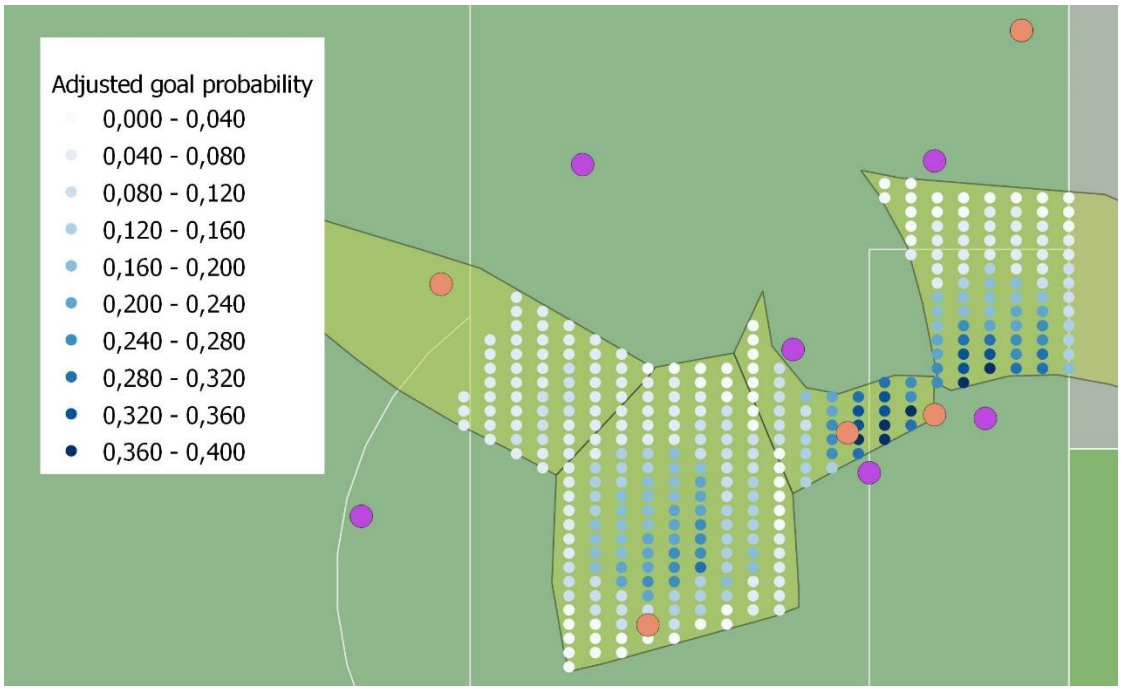


Figure 4.22 – Attacker controlled goal probability points, adjusted for interception chance

In the example situation (Figure 4.22) the impact of interception chance can clearly be observed. In theory, the area controlled by the attacker

All of this resulted in a set of goal probability values per relevant dominant region. The collection of points was translated into numbers (e.g. sum of all goal probabilities, total number of points) and additional statistics (e.g. average, variance and standard deviation) were generated in order to further explore the results.

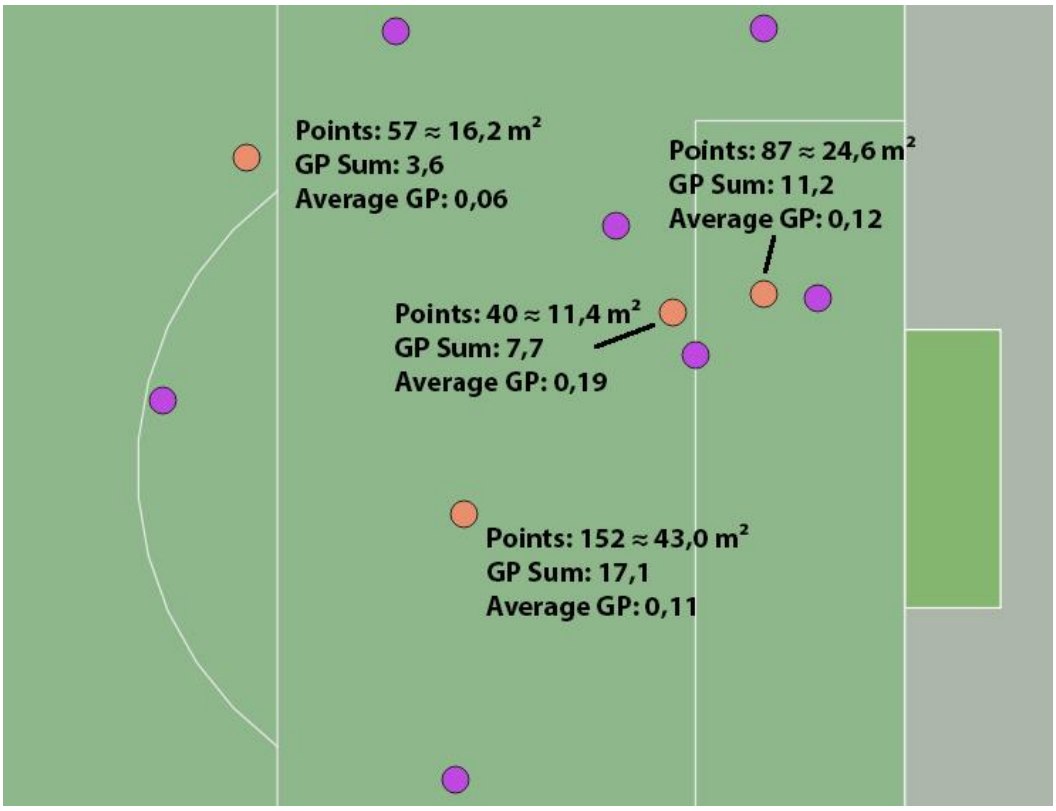


Figure 4.23 – Goal probability statistics per attacker controlled dominant region

The current approach to gauging the quality of an offensive movement action is by simply taking the sum of all goal probability points controlled by an attacker. This seems to favour attackers who control a large area, as opposed to a smaller area with higher goal probabilities, since every goal probability point with a positive value will count towards the sum. In one way, this makes sense, since the larger the area covered by the attacker, the more options the passer has to get the ball to him. A player who controls only a very small area just in front of goal will require a pass to be perfect, this is much less the case for a player who covers a wider area. But on the other hand, the end goal of an attack is to get the ball to an attacker in a critical space in front of goal (Duarte et al., 2012; Rein et al., 2017). It could be argued that the attacker controlling the space with the highest adjusted goal probability value, controls the most *critical* space. In the example situation, this would be attacker 3, in the middle, controlling roughly 11,4 square meters of space but with a much higher average goal probability per controlled point. As it stands, his goal probability sum is lower than two of his colleagues, simply because he controls less space.

The quality of the offensive movement action is calculated by taking the highest goal probability sum achieved during the offensive movement action. This rating is called the Offensive Movement Action score, sometimes simply referred to as the *score*.

5. Results & discussion

This chapter will present the findings of the analysis of the data through the constructed spatial model, described in the methodology chapter. Then these results will be discussed in more detail in order to answer the two remaining sub questions (4 and 5).

5.1 Results

5.1.1 Dominant regions

A player's dominant regions is the area in which the player can arrive at every position before any other player can. In total, 9372 dominant regions were found to occupy crucial space in front of goal, so that they could be assigned a score, through the spatial value model. This score ranged anywhere from 0 to 129. Most scores are on the lower end of this range; the mean score was 12,4.

The crucial space occupied in front of goal, is any space with an associated goal probability value (see Figure 4.20). The total amount of controlled crucial space was calculated for each dominant region. These are not the total controlled areas; only crucial space is counted towards this number.

In order to gauge some measure of success, it was determined whether these offensive movement actions resulted in shots or goals. This was done through identifying the time and position of all taken shots (on and off target) throughout all analysed games, and counting back 5 seconds in time to see if any offensive movement actions made by the same player who made the shot, directly preceded them. This makes it possible to see how many offensive movement actions resulted in shots, the success of these shots (on or off target, goal or no goal) and compare the groups of offensive movement actions preceding these different outcomes to each other. The ordinal variable resulting from this is dubbed the *outcome* of the offensive movement action. This also makes it possible to see how many of these shots were actually preceded by offensive movement actions. All these factors and their interaction are shown in Figure 5.1.

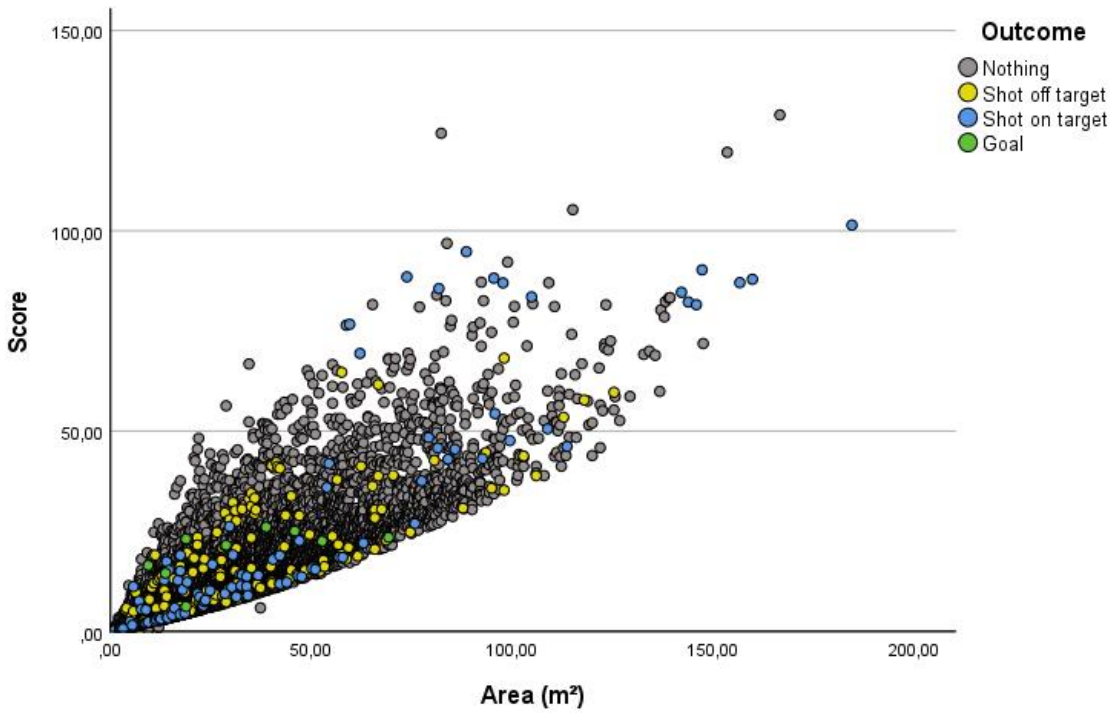


Figure 5.1 – Scatter plot of controlled *crucial* space and score, divided into categories based on outcome

5.1.2 Offensive movement actions

Since the objective is to assign a score to offensive movement actions, not just controlled zones at exact timepoints, the zones belonging to the same offensive movement actions (meaning within 5 seconds of one another) were grouped together. For each of these groups a trajectory of roughly 10 seconds was drawn. This trajectory inherits its score from the highest score achieved throughout all its dominant regions.

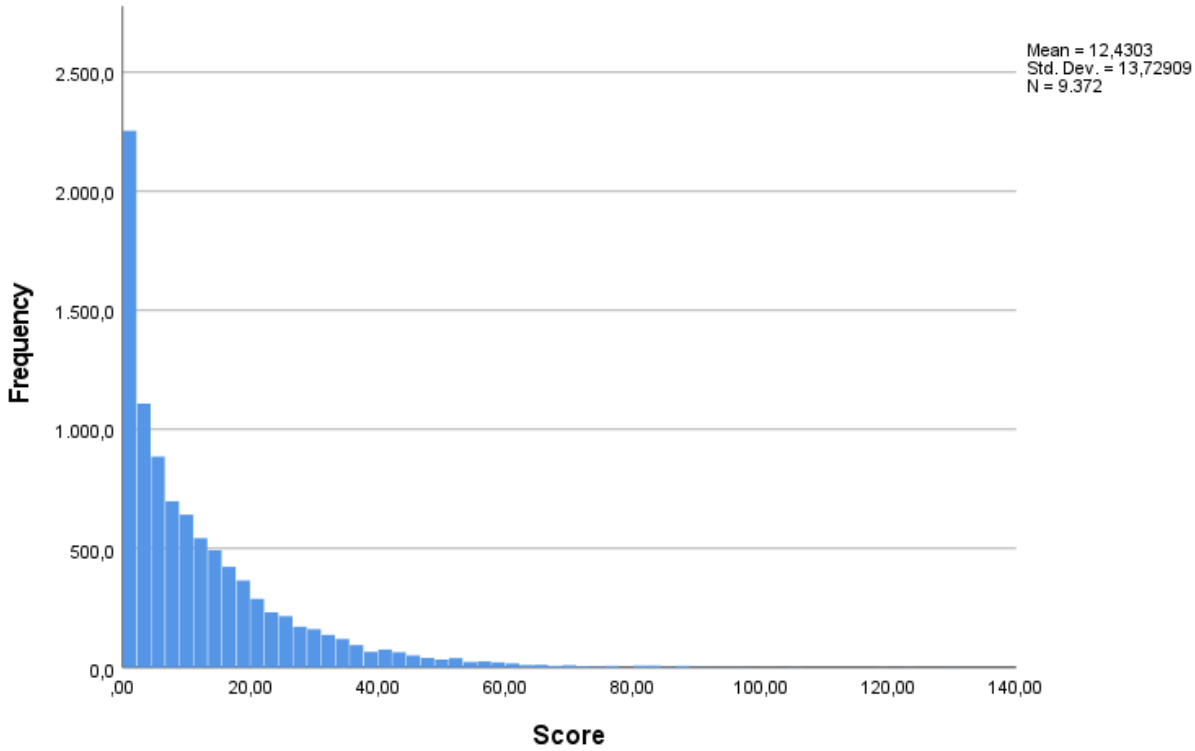


Figure 5.2 – Histogram of offensive movement action scores

This all resulted in a total of 1882 offensive movement actions. The vast majority of these, 1792, did not lead to any outcome. The scores of these offensive movement actions ranged from 0 to 129, skewing heavily to the right, with a mean score of 9,14. A total of 53 offensive movement actions lead to a shot off target, the mean score for this group was slightly higher, at 12,22. There were 32 offensive movement actions directly leading to a shot on target, this group had a mean score of 13,94. Finally there were 5 offensive movement actions leading to a shot on goal resulting in a goal, with a mean score of 18,75. The fact that the mean scores of these groups is positively correlated with the outcome is a positive sign.

5.2 Discussion

5.2.1 Model accuracy

This section will discuss the accuracy of the constructed spatial model and therefore directly answer sub question 4: “To what extent can the quality of an attacking player’s offensive movement actions be judged through the constructed spatial model?”.

Offensive movement actions

The initially set rules for determining when an offensive movement action was taking place, did not correctly identify all such occasions. Many shots, directly preceded by passes or crosses went unnoticed, because the data did not match the set criteria of: (1) one of the attacker’s teammate’s has the ball (2) in the opponent’s half while (3) the attacker himself is located in the final third.

This mismatch is likely due to poor implementation of these rules in the spatial model. The possession of teammates was checked through event as well as tracking data. The event data served to identify which team has possession during the phase of play. The tracking data was used to identify whether or not the ball could be played at that very moment, meaning that the ball had to be within one meter of the teammate. This one-meter limit was possibly too strict, eliminating too many viable situations.

In order to correct this, shots preceded by a pass or a cross where retroactively provided an offensive movement action, including a dominant region and a spatial value, through the same system as the other offensive movement actions. This was a necessary correction to make the most of the analysed match data.

Dominant regions and shot positions

To assess the validity of the dominant region model, the position of the shots resulting from the offensive movement actions can be compared to the extent of the dominant region. The expectation is that the position of the majority of shots will be found inside the dominant regions. It is of course not impossible for shots to have taken place from outside the dominant region, for various reasons. One explanation could be that the defender whose dominant region covers the position from which the shot was taken did not act quick enough. Another reason could be that the attacker can simply run faster than the defender, remember in this spatial model all players have equal physical capabilities.

For each shot taken directly following an offensive movement action (a total of 91 cases) 39,9 percent were taken from outside of the latest predicted dominant region and 60,4 percent from within that region. As established, some shots taken from outside of the dominant region are to be expected, but the fact that this is the case for such a large portion of all shots is a cause for concern. This could point to the movement model being too simple, for instance not taking into account a minimum human reaction time of 0,25 seconds before changing direction. Another explanation could be that dominant regions just do not hold much explanatory or predictive power in football match analysis, despite their prevalence in the literature. Perhaps a more fluid approach to spatial occupation (Fernández et al., 2019; Spearman, 2018) is a more truthful representation of the sport. This would mean that spatial model constructed in this research project is flawed because it heavily relies on the use of dominant regions.

Outcome and scores

If the spatial model accurately identifies dangerous offensive movement actions, then one would expect that over several matches, a pattern would emerge where high scoring offensive movement

actions would often be found in the lead up to goals and shots. In other words, one would expect to find a correlation between the *outcome* and the *score* of offensive movement actions.

One should not expect that every high scoring offensive movement action will lead to a shot or a goal. The whole point of the constructed spatial model is to isolate the offensive movement action from other factors like the final pass or the eventual shot quality. In other words, a good offensive movement action in itself does not necessarily lead to a goal, other factors are needed for that, but that does not diminish the quality of the offensive movement action. The expected relationship between high scoring offensive movement actions and shots is not one to one.

In order to perform a statistical test, the offensive movement actions were divided into two groups: (1) offensive movement actions that lead to a shot on or off target or a goal and (2) offensive movement actions that do not lead to anything. Because the scores were not normally distributed in either groups, a nonparametric statistical test was employed to compare the two groups. The medians of the scores of these groups were compared through an independent sample median test. This test showed a highly significant difference (significance = 0,000) between the medians of group 1 and group 2. This implies that there is a statistical difference in score between offensive movement actions that lead to a shot or goal and offensive movement actions without such an outcome.

As already mentioned in the result section, groups with more favourable outcomes (shots on/off target and goals) had a higher mean score. Again, a nonparametric statistical test was selected in order to further explore the differences *score* between these four groups. A Spearman's rho test on the variable *outcome* and the variable *score* produced a highly significant (significance = 0,000) correlation coefficient of 0,83. This suggests a statistically significant relationship between these two variables. This finding implies that high scoring offensive movement actions are a prediction of better outcomes, which speaks to the validity of the spatial model; what is considered a good offensive movement action does indeed seem to lead to more dangerous attacking situations.

While the highly significant relationships found in these two statistical tests are encouraging, it is important to keep in mind that there is only a very low number (5) of offensive movement actions directly leading to goals. It would therefore, based purely on this analysis, be inappropriate to state that high scoring offensive movement actions are a reliable predictor for goals, which in terms of football, are the only result that matters at the end of the day.

5.2.2 Applications

In order to answer sub question 5: "What other applications does the constructed spatial model have in football match analysis?", possible applications are explored in this section. These applications are listed under the assumption that the determined quality of offensive movement actions is somewhat accurate.

Player comparisons

The objective of the constructed spatial model is to express a certain player quality or attribute (offensive positioning in this case) in a number; an advanced metric. A major reason to expressing such an attribute in a number, is that numbers are extremely easy to compare with one another. The metric can be used to point out when and where a player makes his most dangerous offensive movement actions, but also to compare that with other players. This could be used in coaching for instance, to identify areas of improvement for attacking players, and perhaps even suggest adjustments in movement trajectories or timing.

The *offensive movement action score* currently grades one specific offensive movement action but could easily be expanded into a number which grades a player's performance throughout a match. One simple way to do this is by adding all offensive movement action scores achieved by one player over the duration of a match together for an *offensive movement action score per match* metric. Another metric, possibly fairer towards weaker teams with less attacks per match, could be to calculate *offensive movement action score per attack*. These metrics could then be used to compare player with each other on more than just one specific moment. This could for instance be of use in the scouting of players, by finding attackers with a keen sense for space. Likewise, different matches of the same player can be compared to each other, to identify when or against what types of opponents or formations this player is able to find the most space.

Opponent Marking

The spatial model can be used to assess the opponents of the subject as well. If an attacker consistently performs high scoring offensive movement actions, then that attacker could be very good at finding space, or perhaps the defenders are not properly marking him.

In general, there are two different ways teams choose to defend in football, although these approaches are somewhat fluid and very few teams strictly always use one of these approaches and never the other. There is zonal marking, meaning that each defender marks a certain zone, and man marking, meaning the defenders are responsible for marking a certain opponent. In either case, the constructed spatial model is an excellent tool to assess the defenders marking.

In the case of zonal marking, we can see how many offensive movement actions gathered their score in the zone for which a certain defender is responsible because the score is directly tied to areas of the pitch. In the case of man marking, one could simply take the *score per match* or *score per attack* of the attacker for which a defender is responsible as a measure of their success.

On-the-ball decision making

The spatial model can also be used to grade the decision-making of the player with the ball, although this is a little more complicated than previously mentioned applications. This application is best illustrated through a simple scenario. This scenario will be the same as the scenario used in the conclusion section of the methodology chapter (4.7) (Figure 5.3, 5.4 & 5.5).

The player with the ball has four teammates (attackers) which he can play the ball to. The attacker with the highest *score* is attacker 2, mostly because he covers a much larger area of *crucial space*. A simple analysis could therefore be that the player with the ball should try to pass the ball to attacker 2. Even the most skilled player can not kick the ball perfectly every time and attempting to pass to attacker 2 will leave the most margin for error. In other words, passing to attacker two has the highest chance of converting this attack into a goal.

But on the other hand, the most valuable space controlled by an attacker is occupied by attacker 3 and 4 (the dark blue areas). If the player with the ball manages to get a pass into one of these areas this would, according to the spatial value model, result in a better chance than any pass towards attacker 2 could. It could therefore be argued that the player should attempt this pass with perhaps slightly less margin for error, but a higher reward if he succeeds.

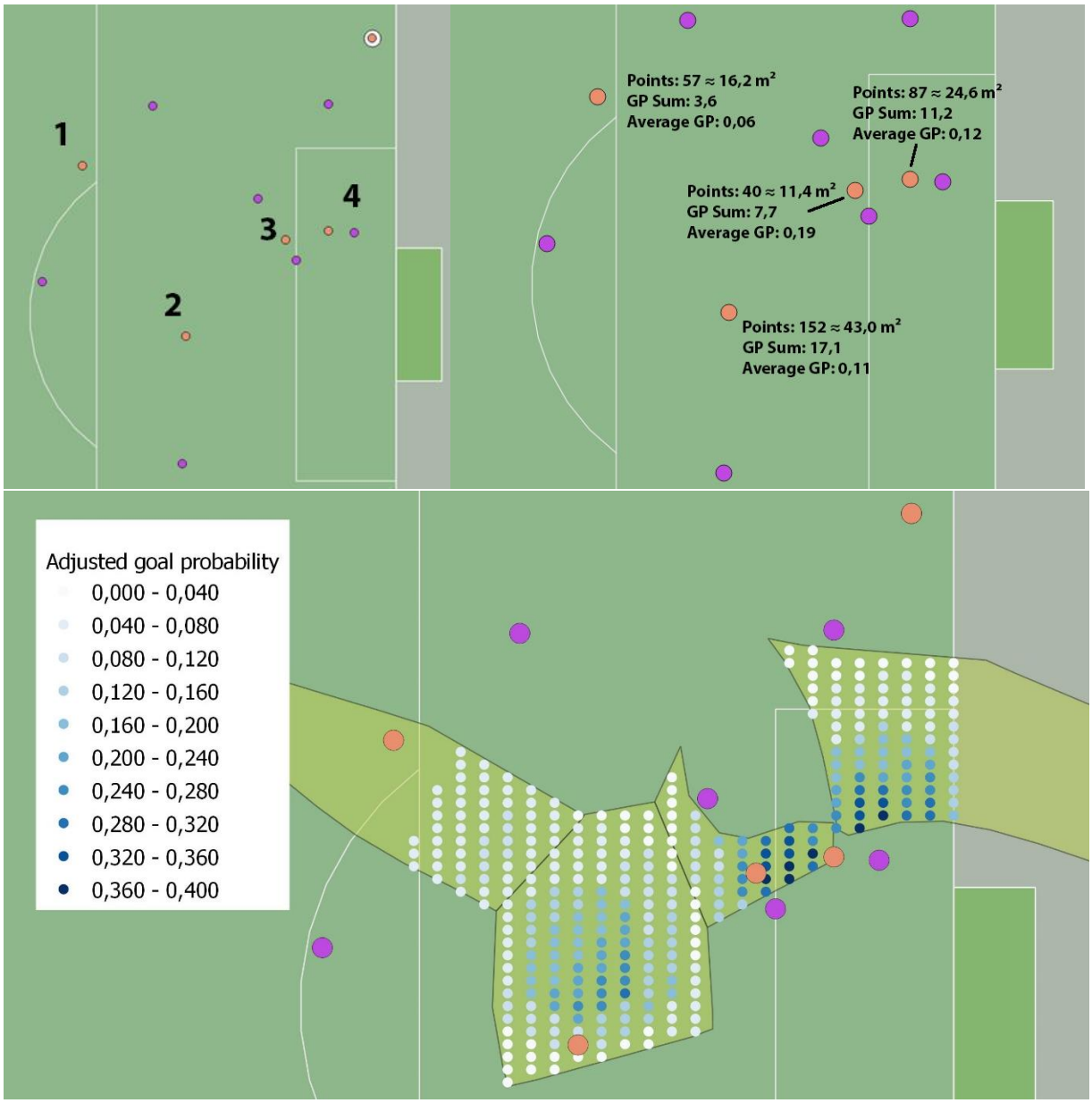


Figure 5.3, 5.4 & 5.5 – Scenario to illustrate player decision making

6. Conclusions and recommendations

6.1 Research questions

This research project set out to create an advanced metric to express the quality of offensive movement actions using a GIS platform. That goal was accomplished though it is currently unsure how useful the constructed spatial model is for actual football match analysis, since there was only time for limited testing. This limited testing did however point to high scoring offensive movement actions being predictive of better results, meaning more shots and goals.

The main research question was:

To what extent can the quality of an offensive movement action be determined and assessed through a spatial model?

The constructed spatial model incorporates all relevant elements to satisfaction and produces results which seem to corroborate its process. Some elements might have been oversimplified (e.g. the movement model) or based on assumptions which don't quite seem to speak to the essence of football (e.g. the dominant region model) but that is the nature of a scientific model; it's not meant to be an exact copy of reality, it's a simplification to examine one particular issue.

The five sub research question were answered accordingly:

1. What is an offensive movement action?

An offensive movement action refers to the movement of an attacker, aimed at creating the most favourable situation for himself in order to score a goal. An attacker can make an offensive movement action when he is located in the final third and one of his teammate's has the ball in the half of the opponent.

2. How is the quality of and offensive movement action defined and what are the elements contributing to this quality?

The quality of an offensive movement action is determined by the amount of space controlled by the attacker and the value of this space, at key moments where it is possible for a teammate to attempt a pass to him. The value of this space is determined by the likelihood that a shot from that position will result in a goal and is negatively affected by the difficulty of passing to that position from the current position of the ball.

3. To what extent can the elements contributing to the quality of an offensive movement action be spatially modelled in a GIS?

All required elements were modelled into a GIS to satisfaction. The representation of these elements as vector data is sufficiently close to reality and offers a basis for further calculations in a GIS. Whether or not some of these elements hold much predictive power is uncertain though. Spatial occupation for instance did not seem as rigid a concept as presented in this research project; a large portion of shots attempted after an offensive movement action were attempted from outside the last dominant region, indicating that these regions are not fully controlled by a single player.

4. To what extent can the quality of an attacking player's offensive movement actions be judged through the constructed spatial model?

A highly significant correlation was found between the outcome and the score of the offensive movement actions, meaning that offensive movement actions leading to shots or goals were deemed significantly more dangerous than those which did not. The dataset on which the spatial model was tested was small, but this is an encouraging result that does point towards high scoring offensive movement actions being an indicator for more shots and more goals, in other words, high scoring offensive movement actions truly do seem more dangerous than low scoring ones.

5. What other applications does the constructed spatial model have in football match analysis?

Besides simply judging a player's offensive positioning, the spatial model's main purpose, several possible applications. Firstly, being able to easily compare players offensive positioning to different players and to the same player in different matches, which is very difficult without any numerical expressions of quality. Secondly, a defender's man or zonal marking ability can be judged by the quality (or lack thereof) of their opponents. Thirdly, the player with the ball's decision-making skills could be analysed or at least discussed on the basis of this spatial model.

Finally, a more ephemeral research goal was to further explore GIS as a platform for football match analysis. A spatial model which can be used for football match analysis was fully developed on a GIS platform in this research project, proving that indeed GIS can be a very appropriate platform, but some clear limitations were also uncovered. The match analysis took an extremely long time, leaving many opportunities for research unexplored, simply because of time pressure. While it could be entirely possible that this was solely because of the poor programming skills of the author, it could also be that the version of QGIS used in this research project, simply did not have tools which were appropriate for such large-scale spatiotemporal analysis. It might prove that the latest version of QGIS 3.14, released during this research project, will be the solution to this problem with the inclusion of a temporal manager. All in all, of course GIS is an appropriate platform for any spatial analysis, including football match analysis, but the tools employed need to be chosen carefully.

6.2 Further Research

This section will put forth future research opportunities which can build on the findings of this research project.

An obvious avenue of improvement upon this research project would to test the constructed spatial model with more data in order to present more significant results. This research project had more data available to it than was ultimately used, but likely due to suboptimal implementation, data analysis proved an extremely time-consuming process. This also resulted in none of the set parameters (for instance maximum movement speed, acceleration and time to intercept a ball) being tested at all, they were simply taken from the literature and implemented without any testing because that would take too much time to do on any meaningful scale. A better implementation of the spatial model that doesn't need several days to analyse a game of football would be a definite improvement.

One implementation method which would speed up the analysis process is to make use of dynamic data structures. This would allow for the calculation of dynamic/kinetic Voronoi cells which can be calculated faster, using the knowledge that the Voronoi cell will differ only slightly from the Voronoi cell constructed in the last timestep. This approach can significantly speed up calculations and is a very good fit for the problem tackled in this research project. A researcher with a background in (dynamic) computational geometry would likely find a great research opportunity here.

The constructed spatial model is quite modular, meaning that each component (spatial occupation model, spatial value model & interception model) can easily be adapted without changing any of the

other components. These adaptations could help to make the spatial model more realistic, for instance by adding a minimum human response time to the movement model as mentioned earlier. Of course, the model can also simply be expanded by adding more components. One factor that is currently not calculated into the spatial value model for instance is the chance for an opponent to block an attempted shot. While this could prove slightly more difficult to implement than anything currently in the model, it would be a welcome addition in order to make the spatial model more realistic.

Another example is a model which further modifies the spatial value by accounting for the angle of pass reception to goal. A cross coming from the backline is usually considered much more dangerous than a pass coming from the centre of the field, because the former allows for an easy shot while the latter requires the attacker to shoot at an awkward angel or take more time to control the ball. A research project incorporating this would likely need a lot of data analysis in order to quantify this, as there doesn't seem to be a comprehensive model presented in the literature, but therefore provide a worthy contribution.

It has been mentioned earlier, but some sort of curve to the trajectories of the ball would further deepen the interception model. A pass over a long range will almost always have a visible curve to its trajectory, which is something professional footballers account for and even sometimes emphasize in order to get better results. It would be complicated to implement this, but definitely an interesting addition.

Both opponent marking and on-the-ball decision making were briefly discussed in section 5.2.2 but provide a major opportunity for future research. The spatial model constructed in this research project could easily be used for these ends, with only minor adaptations.

Marking or defensive coverage can basically be directly assessed through the spatial model presented in this research project. This can be done by determining what defender is responsible for which opponent player or which area of the pitch and then evaluating the offensive movement actions of that opponent player or made in that area of the pitch. Defensive coverage seems to be another understudied topic in football analytics research, so a research project focussing on this would likely be a welcome addition to the existing literature.

On-the-ball decision making is, as mentioned, slightly more difficult to assess directly through the constructed spatial model. Some additional research specifically focussed on on-the-ball actions is probably needed in order to approximate *correct* and *false* decisions, or more likely create a continuous grading system, but as mentioned, plenty of on-the-ball research exists, so this shouldn't present an issue.

Any in depth analysis of offensive movement action trajectories was beyond the scope of this research project but could undoubtably be an interesting research topic. The *score* could for instance be determined at different timesteps along an interval and be compared with movement parameters: do players achieve the highest score when they are simply running straight towards goal or when they are curving their run more thoughtfully? Do players achieve the highest score when sprinting or jogging? A framework could also be established to judge the shape of trajectories and devise categories.

This research project specifically focussed on the specific ability of an individual player to find space for himself in order to receive the ball in as favourable a position as possible to shoot on goal, but a research project which explores the team dynamics of such movement actions would undoubtably provide interesting results. More complicated concepts like making a run, in order to create space for teammates by drawing opponent defenders with the runner could be explored, as Fernandez & Bornn (2018) did. The spatial model presented in this research project is much too simple for these ends but

could provide a basis for more this more complicated goal. Such a research project would grapple more witch concepts like team tactics, strategies and concerted movements.

Literature

Andrienko, G., Andrienko, N., Anzer, G., Bauer, P., Budziak, G., Fuchs, G., … Wrobel, S. (2019). Constructing Spaces and Times for Tactical Analysis in Football. *IEEE Transactions on Visualization and Computer Graphics*, *PP*(8), 1. https://doi.org/10.1109/TVCG.2019.2952129

Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., von Landesberger, T., & Weber, H. (2017). *Visual analysis of pressure in football*. *Data Mining and Knowledge Discovery* (Vol. 31). https://doi.org/10.1007/s10618-017-0513-2

Baars, M., Oosterom, P. Van, Verbree, E., & Gorte, B. (2004). Structuring , indexing , querying and visualizing moving objects in a DBMS, 251–260.

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2015). Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. *Proceedings - IEEE International Conference on Data Mining, ICDM*, *2015-Janua*(January), 725–730. https://doi.org/10.1109/ICDM.2014.133

Bransen, L., & Van Haaren, J. (2018). Measuring football players' on-the-ball contributions from passes during games. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11330 LNAI*, 3–15. https://doi.org/10.1007/978-3-030-17274-9_1

Bransen, L., Van Haaren, J., & Van De Velden, M. (2019). Measuring soccer players' contributions to chance creation by valuing their passes. *Journal of Quantitative Analysis in Sports*, *15*(2), 97–116. https://doi.org/10.1515/jqas-2018-0020

Cervone, D., Bornn, L., & Goldsberry, K. (2016). NBA Court Realty Intro: the Basketball Court is a Real Estate Market. *MIT Sloan Sports Analytics Conference*, (1), 1–8.

Chawla, S., Estephan, J., Gudmundsson, J., & Horton, M. (2014). Classification of passes in football matches using spatiotemporal data. *ACM Transactions on Spatial Algorithms and Systems*, *3*(2). https://doi.org/10.1145/3105576

Clemente, F. M., Martins, F. M. L., Wong, D. P., Kalamaras, D., & Mendes, R. S. (2015). Midfielder as the prominent participant in the building attack: A network analysis of national teams in FIFA world cup 2014. *International Journal of Performance Analysis in Sport*, *15*(2), 704–722. https://doi.org/10.1080/24748668.2015.11868825

Clifford, N., French, S., & Valentine, G. (2010). *Key Methods in Geography*. *Geographical Research*. https://doi.org/10.1017/CBO9781107415324.004

Davies, J. (2012). An analysis of every Premier League goal & assist of 2011/12. Retrieved November 18, 2019, from https://eplindex.com/21057/analysis-premier-league-goal-assist-2011-12.html

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions Speak Louder than Goals, (1), 1851–1861. https://doi.org/10.1145/3292500.3330758

Dodge, S. (2015). From Observation to Prediction: The Trajectory of Movement Research in GIScience. *Advancing Geographic Information Science: The Past and Next Twenty Years*, (February), 123–136.

Duarte, R., Araújo, D., Freire, L., Folgado, H., Fernandes, O., & Davids, K. (2012). Intra- and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. *Human Movement Science*. https://doi.org/10.1016/j.humov.2012.03.001

Fernandez, J., & Bornn, L. (2018). Wide Open Spaces : A statistical technique for measuring space creation in professional soccer. *MIT Sloan Sports Analytics Conference*, 1–19.

Fernández, J., Bornn, L., & Cervone, D. (2019). Decomposing the Immeasurable Sport: A deep learning expected possession value framework for soccer. *MIT Sloan Sports Analytics Conference*, 1–18. Retrieved from https://lukebornn.com/sloan_epv_curve.mp4

FIFA. (2018). 2018 FIFA World Cup™ - News - More than half the world watched record-breaking 2018 World Cup - FIFA.com. Retrieved October 16, 2019, from https://www.fifa.com/worldcup/news/more-than-half-the-world-watched-record-breaking-2018-world-cup

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

Gold, C. (2016a). Tessellations in GIS: Part I—putting it all together. *Geo-Spatial Information Science*, *19*(1), 9–25. https://doi.org/10.1080/10095020.2016.1146440

Gold, C. (2016b). Tessellations in GIS: Part II–making changes. *Geo-Spatial Information Science*, *19*(2), 157–167. https://doi.org/10.1080/10095020.2016.1182807

Green, S. (2012). Assessing the performance of Premier League goalscorers. Retrieved September 11, 2019, from https://www.optasportspro.com/news-analysis/assessing-the-performance-of-premier-league-goalscorers/

Groves, P. D. (2013). Fundamental Concepts. In *Principles of GNSS* (pp. 1–22). Artech House.

Gudmundsson, J., & Horton, M. (2017). Spatio-temporal analysis of team sports. *ACM Computing Surveys*, *50*(2). https://doi.org/10.1145/3054132

Gyarmati, L., & Stanojevic, R. (2016). QPass: a Merit-based Evaluation of Soccer Passes, 0–3. Retrieved from http://arxiv.org/abs/1608.03532

Heywood, I., Cornelius, S., & Carver, S. (2011). *An Introduction to Geographical Information Systems* (4th ed.). Essex: Edinburgh Gate. https://doi.org/10.1007/978-0-585-31560-7_4

InStat. (2020). Penalty Research. Retrieved January 27, 2020, from https://instatsport.com/football/article/penalty_research

Kang, C. H., Hwang, J. R., & Li, K. J. (2006). Trajectory analysis for soccer players. *Proceedings - IEEE International Conference on Data Mining, ICDM*, 377–381.

Keim, D., Kohlhammer, J., Ellis, G., & Mansmann, F. (2010). *Mastering the Information Age*. *Los Angeles: Jeremy P. Tarcher*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Mastering+the+Information+Age#2

Kim, H. C., Kwon, O., & Li, K. J. (2011). Spatial and spatiotemporal analysis of soccer. In *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*. https://doi.org/10.1145/2093973.2094029

Kotzbek, G., & Kainz, W. (2014). Football Game Analysis: A New Application Area for Cartographers and GI-Scientists? *5th International Conference on Carthography and GIS*, 299–306. Retrieved from http://homepage.univie.ac.at/gilbert.kotzbek/publikationen/2014_Football_Game_Analysis_G_Kotzbek_W_Kainz.pdf

Kotzbek, G., & Kainz, W. (2015). GIS-Based Football Game Analysis – A Brief In- troduction to the

Applied Data Base and a Guideline on How to Utilise It. *International Cartographic Conference*.

Kotzbek, G., & Kainz, W. (2016). Towards Automated GIS-based Analysis of Scoring Attempt Patterns in Association Football, 3–7.

Kotzbek, G., & Kainz, W. (2017a). *Beyond the pass completion rate: Towards analysing the players' passing quality in association football*. Retrieved from http://homepage.univie.ac.at/gilbert.kotzbek

Kotzbek, G., & Kainz, W. (2017b). Dynamic Zoning in the Course of GIS-based Football Game Analysis University of Vienna ; Faculty of Earth Sciences , Geography and Astronomy ; Department of Geography and Regional Research ; Universitätsstraße 7 , 1010 Vienna , Austria. *GIS Ostrava - Dynamivs in GIscience*.

Kotzbek, G., & Kainz, W. (2017c). Über die Zweckmässigkeit GIS-Gesützter Fussbal-Spielanalysen. *Gis.Science*, *4*(4), 125–134.

Laube, P., Imfeld, S., & Weibel, R. (2005). Discovering relative motion patterns in groups of moving point objects. *International Journal of Geographical Information Science*, *19*(6), 639–668. https://doi.org/10.1080/13658810500105572

Lewis, M. (2003). *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton & Company.

Link, D. (2018). *Data Analytics in Professional Soccer*. *Data Analytics in Professional Soccer*. https://doi.org/10.1007/978-3-658-21177-6

Link, D., Lang, S., & Seidenschwarz, P. (2016). Real time quantification of dangerousity in football using spatiotemporal tracking data. *PLoS ONE*, *11*(12), 1–16. https://doi.org/10.1371/journal.pone.0168768

Longley, P. a, Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographical Information Systems and Science. 2nd Edition. John Wiley & Sons, Ltd*. https://doi.org/10.2307/215736

Mackay, N. (2017). *Predicting goal probabilities for possessions in football*. Vrije Universiteit Amsterdam. Retrieved from https://science.vu.nl/en/Images/werkstuk-mackay_tcm296-849981.pdf

McGarry, T., & Franks, I. M. (2003). *The science of match analysis*. *Science and Soccer: Second Edition*. https://doi.org/10.4324/9780203417553

Metrica Sports. (2019). Metrica Sports. Retrieved October 16, 2019, from https://metrica-sports.com/

Moruzzi, M. (2014). Does analysing football through statistics miss the point of the game? Retrieved October 25, 2019, from https://www.theguardian.com/football/when-saturday-comes-blog/2014/feb/03/statistics-football-analysis-miss-point-game

Optasports. (2016). Advanced metrics. Retrieved September 11, 2019, from https://www.optasports.com/services/analytics/advanced-metrics/

Pappalardo, L., Cintia, P., Rossi, A., Massucco, E., Ferragina, P., Pedreschi, D., & Giannotti, F. (2019). A public data set of spatio-temporal match events in soccer competitions. *Scientific Data*, *6*(1), 236. https://doi.org/10.1038/s41597-019-0247-7

Peuquet, D. J. (1988). Representations of Geographic Space: Toward a Conceptual Synthesis. *Annals of the Association of American Geographers*. https://doi.org/10.1111/j.1467-8306.1988.tb00214.x

Power, P., Ruiz, H., Wei, X., & Lucey, P. (2017). "Not all passes are created equal:" Objectively

measuring the risk and reward of passes in soccer from tracking data. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *Part F1296*, 1605–1613. https://doi.org/10.1145/3097983.3098051

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

Rein, R., Raabe, D., Perl, J., & Memmert, D. (2016). Evaluation of changes in space control due to passing behavior in elite soccer using Voronoi-cells. *Advances in Intelligent Systems and Computing*, *392*, 1–3.

Sarmento, H., Clemente, F. M., Araújo, D., Davids, K., McRobert, A., & Figueiredo, A. (2018). What Performance Analysts Need to Know About Research Trends in Association Football (2012–2016): A Systematic Review. *Sports Medicine*. https://doi.org/10.1007/s40279-017-0836-6

Sarmento, H., Marcelino, R., Anguera, M. T., CampaniÇo, J., Matos, N., & LeitÃo, J. C. (2014). Match analysis in football: a systematic review. *Journal of Sports Sciences*, *32*(20), 1831–1843. https://doi.org/10.1080/02640414.2014.898852

SciSports. (2019). SciSports. Retrieved October 16, 2019, from https://www.scisports.com/

Siabato, W., & Manso-Callejo, M. A. (2011). Integration of temporal and semantic components into the geographic information through mark-up languages. Part I: Definition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6782 LNCS*(PART 1), 394–409. https://doi.org/10.1007/978-3-642-21928-3_28

Silva, P., Travassos, B., Vilar, L., Aguiar, P., Davids, K., Araújo, D., & Garganta, J. (2014). Numerical relations and skill level constrain co-adaptive behaviors of agents in sports teams. *PLoS ONE*. https://doi.org/10.1371/journal.pone.0107112

Sinton, D. (1977). The Inherent Structure of Information as a Constraint to Analysis: Mapped Thematic Data as a Case Study. *First International Advanced Study Symposium on Topological Data Structures for Geographic Information Systems.*

Sithole, G., & Zlatanova, S. (2016). Position, Location, Place and Area: an Indoor Perspective. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, *III–4*(July), 89–96. https://doi.org/10.5194/isprsannals-iii-4-89-2016

Spearman, W. (2018). Beyond Expected Goals. *MIT Sloan Sports Analytics Conference*, 1–17.

Spearman, W., Basye, A., Dick, G., Hotovy, R., & Pop, P. (2017). Physics-Based Modeling of Pass Probabilities in Soccer. *Research Papers Competition*, 1–14.

StatsBomb. (2020). StatsBomb Open Data. Retrieved from https://github.com/statsbomb/open-data/tree/master/data/events

Taki, T., & Hasegawa, J. ichi. (2000). Visualization of dominant region in team games and its application to teamwork analysis. *Proceedings of Computer Graphics International Conference, CGI*, 227–235.

Ueda, F., Masaaki, H., & Hiroyuki, H. (2014). The casual relationship between dominant region and offense-defense performance. *Football Science*, *11*, 1–17.

UEFA. (2018). *UEFA Stadium Infrastructure Regulations Edition 2018*.

Van Haaren, J., Dzyuba, V., Hannosset, S., & Davis, J. (2015). Automatically discovering offensive

patterns in soccer match data. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9385*, 286–297. https://doi.org/10.1007/978-3-319-24465-5_25

van Hoeve, L. T. (2017). *Geovisual football analytics: towards the development of an interactive visual interface for football coaches, analysts and players*. Universiteit Utrecht, GIMA programme. Retrieved from https://dspace.library.uu.nl/handle/1874/356830

WhoScored. (2020). WhoScored Database. Retrieved October 10, 2020, from Whoscored.com

Windrawwin.com. (2018). What Are The Most Common Football Scores? - WinDrawWin.com. Retrieved October 1, 2019, from https://www.windrawwin.com/statistics/full-time-scores/

Zambom-Ferraresi, F., Rios, V., & Lera-López, F. (2018). Determinants of sport performance in European football: What can we learn from the data? *Decision Support Systems*, *114*(March), 18–28. https://doi.org/10.1016/j.dss.2018.08.006