Collective behaviour monitoring in football using spatial temporal and network analysis: application and evaluations.

CORSIE, M.R.

2022

Collective Behaviour Monitoring in Football using Spatial Temporal and Network Analysis: Application and Evaluations

Martin Corsie

A thesis submitted in partial fulfilment of the requirements of the Robert Gordon University for the degree of Doctor of Philosophy

September 2022

Collective Behaviour Monitoring in Football using Spatial Temporal and Network Analysis: Application and Evaluations

PhD Candidate

Martin Richard Corsie

Supervisors

Paul Alan Swinton

Eyad Elyan

School of Health Sciences, Ishbel Gordon Building, Robert Gordon University, Garthdee Road, Aberdeen, AB10 7GJ, United Kingdom.

Contents

Tables and Figures

Acknowledgements

This taxing journey has been made energising by the wonderful people I have had the opportunity to share it with. It is unquestionable that without my primary supervisor Dr Paul Swinton, I would not have been able to accomplish this project and so I must offer endless gratitude to his guidance and support.

I would also like to give thanks to the many people who have afforded me opportunities through my PhD including Graeme Jones, and all of the marvellous people at the Scottish FA who it has been an honour to work alongside. Moreover, I have a great appreciation to the coaches who afforded me their time and whose interviews have been invaluable to my research.

Perhaps the best part of my project has been the friends I have made along the way. I have enjoyed coming in to work to see my comrades as we progressed through our degrees, a motivation that has been critical to my progress. There are too many people to mention individually, but it would be rude not to mention Teodora Stoyanova, a girl who in many ways I perceive to be the direct antithesis of myself, but someone I know I will remain friends with long into the future. This dynamic and sometimes chaotic group of friends I have made in this journey has given me a feeling of belonging, something I am not entirely sure I have ever truly felt before entering the fray of postgraduate research.

Last but far from least, my family and friends. To my Parents, Ian, Marian, and my sister Rachel, you have been ever-present in your tolerance and support through this 3-and-a-half-year voyage, I could not hope for a more supporting family. A PhD is in some ways a solitary obsession over a single question, and it is important to be able to escape from this every so often. To my friends Hamzah Khan and Richard Hyde, who achieved this with their wit and mischief I am forever grateful. And finally, to my Grandma Edna, who saw me start, but did not see me finish, your everlasting encouragement and energy remains to this today a well for motivation and inspiration.

Abstract

Analysis is an important part of understanding and exploiting performance of football teams. Traditional approaches of analysis have centred around events that may not fully incorporate the highly dynamic nature of matches. To circumvent this weakness, applications of collective behaviour metrics applying spatial temporal and social network analyses to data in football have been trending over the last 10 years. The aims of this PhD were to: 1) establish the strengths and limitations of current research investigating collective behaviour in football applying novel analytical procedures; 2) investigate the credibility of present methods informing coaching practice; and 3) provide guidance for practitioners in implementing complex analytical procedures with current data collection methods. These aims were achieved through the completion of five interlinked studies. The first two studies comprised systematic reviews evaluating the quality of previous research investigating collective behaviours. The first systematic review focussed on spatial temporal metrics and the second systematic review focussed on social network analysis metrics. In addition to standard review procedures, both systematic reviews included analyses of author quotes regarding the metrics used within each study. These included description and conceptualisation of each metric, along with practical applications and measurements of reliability. The first systematic review identified several limitations in the current literature base of spatial temporal metrics investigating collective behaviour in football. These included a lack of conceptualisation of the metrics used, assumptions of metric reliability, frequent use of broad and non-actionable practical recommendations, failure to justify sample sizes and a bias towards including males. Similar findings were found in the social network analysis systematic review where authors also seldom conceptualised metrics, provided vague practical applications and often failed to justify sample size. Literature including social network analysis were also inconsistent with the metric calculations and nearly all studies investigated elite male matches. The third study in this PhD attempted to quantify the reliability of spatial temporal metrics by simulating expected error values on top of real-world data. Through fitting linear mixed effects models on signal to noise ratios, metrics were established to be reliable where positioning systems are accurate to 0.5 m or less. In situations where positioning systems errors were approached 2 m, only some were considered to produce reliable values, (e.g. team centroid), whereas metrics using distances and numerical relations were considered to produce unreliable values. After assessing the literature and reliability, the PhD focussed on implementation of common and reliable metrics, leading into the final study of the PhD which employed an iterative design comprising multiple interviews to investigate coach perceptions of collective behaviour metrics. A thematic analysis identified themes that closely resembled the 10 traditional principles of play in football, further establishing their validity. Moreover, coaches reacted positively to presented measurements, most notable network intensity, distance between defenders, triads, team length, and team depth. Coaches stated they trained players with the concepts these measurements represent as a central focus. The PhD work was concluded with a final chapter set as pedagogical support for practitioners wishing to implement these techniques providing a guide to measuring the tactical concepts discussed within this thesis. Collectively, this PhD highlights that novel collective behaviour metrics have a place in current performance analysis systems in football. Additionally, a methodology is presented for practitioners to apply to their own teams and generate specific metrics relevant to the teams own tactical principles.

Key Words: Football; soccer; collective behaviour; performance analysis; position tracking; network analysis; spatial temporal; data visualisation; reliability; dynamic systems theory;

Publications and Presentations

Published Peer Reviewed Articles

Corsie, M. Craig, T. Swinton, PA. Buchanan, N. (2021) Spatial-Temporal Metrics to Assess Collective Behavior in Football: A Systematic Review and Assessment of Research Quality and Applicability. *Journal of athletic enhancement,* 10(5).

Towlson, C. Abt, G. Barrett, S. Cumming, S. Hunter, F., et al. (2021) The effect of bio-banding on academy soccer player passing networks: Implications of relative pitch size. *PLOS ONE* 16(12): e0260867.

Peer-reviewed articles currently under review

Corsie, M. Swinton. P. A. (2022) Reliability of spatial-temporal metrics used to assess collective behaviours in football: An in-silico experiment. Science and medicine in football – in second review.

Conference Proceedings

Corsie, M. Craig, T. Swinton, PA. Buchanan, N. (2019) Spatial-temporal metrics to assess collective behaviour in football: a systematic review. *8th International Workshop and Conference of the International Society of Performance Analysis of Sport.*

Corsie, M. Craig, T. Swinton, PA. Buchanan, N. (2019) Network Metrics to Assess Collective Behaviour in Football: A Systematic Review (SOCCER). *8th International Workshop and Conference of the International Society of Performance Analysis of Sport.*

Abbreviations

ApEn	Approximate Entropy
BV	Between Variance
DC	Distance between Centroids
EASA	Estimated Adult Stature Attainment
FIFA	Federation Internationale de Football Association
GPS	Global Positioning System
GTSC	Game Technical Scoring Chart
ICC	Intraclass Correlation Coefficients
IMU	Inertial Measurement Unit
LPM	Local Positioning Measurement
LPW	Length per Width
ODC	Out-Degree Centrality
SampEn	Sample Entropy
SSG	Small Sided Game
UEFA	Union of European Football Associations
WV	Within Variance

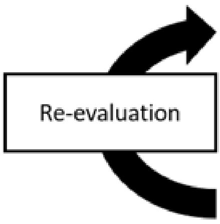
CHAPTER 1

Introduction

Performance analysis has become a pillar of practice in football with widescale adoption of strategies informed by performance analysis in professional teams. This comes as part of a sport science wave that has influenced how coaches and support staff are able to prepare their teams to perform at their highest level and takes the form of evidence-based practice, where decisions are made at least in part by scientific investigation and research along with other considerations such as coaching knowledge and player values (Coutts et al, 2017). For the last 30 years, performance analysis has predominantly focused on descriptive actions and events. (Lord et al 2020; Mclean et al, 2017a). In football, this often relates to passes, shots, dribbles, crosses and tackles with additional constructs such as team possession also considered. Coaches have highlighted the importance of these aspects, however, when considered in isolation, these data miss out on critical information of game states. This was highlighted by Mackenzie and Cushion (2013) in their critical review where match analyses were often identified as not including appropriate information pertaining to situational variables such as match status, match location, stage of competition, type of competition and opponent, as well as match context including position on the pitch and match score.

More recent approaches in performance analysis have sought to conceptualise a football match as a complex system (McLean et al, 2017a). In such a system, events and situational variables are still important in the explanation of the collective behaviour, however, metrics that evaluate subgroup, team, and match dynamics offer potential to comprehensively understand performance through framing the match as a complex system instead of reducing it into its constituent parts (McLean et al, 2017a). Consequently, changing an aspect can have far reaching impact on other parts within the system. This aligns with coaches manipulating ecological constraints to develop desired behaviours and skills of individuals and teams that has previously been linked to developing expertise of elite footballers (Araujo et al, 2010). In parallel with a changing view of football as a complex system, there has been an increase in research investigating how manipulations of constraints including field dimension, goal, formation, player number, pitch location and spatial references influence match dynamics. The growth of data and performance analysis over the last twenty years (Lord et al, 2020) has afforded practitioners the opportunity to apply metrics that observe match dynamics through spatial-temporal and network analysis.

With this branch of performance analysis still being in its infancy, a clear understanding of how such approaches can be applied is important for the research evaluating dynamic systems in football through spatial temporal and network metrics. For this to be realised, harmony between academics and practitioners must be achieved and several barriers must be overcome. For coaches, many publications lack clear practical applications that are relevant to them. As such, researchers often struggle to achieve buy in from coaches and are unable to dedicate the time necessary for investigations centred on the needs of the coach due to pressures of publication and other constraints that scientists face (Fullagar et al, 2019). The aims of this PhD were focussed on further development of football research and in particular exploring novel spatial temporal and network analysis approaches to be more practical and valuable to coaches and other practitioners. The model presented by Fullagar et al, (2019) reflects the processes followed to complete this research and an adapted version is shown in figure 1.1 along with a list of chapters that highlight the sections of the thesis that constitutes each stage.



Stage	Location in Thesis
1. Identify relevant research questions	Chapter 1, 2, 3 & 4
2. Evaluate validity, reliability and impact	Chapter 3, 4, 5 & 6
3. Develop and integrate strategies into practice	Chapter 6 & 7
4. Assessment of new practices	

Figure 1.1 Applying research and evidence-based practice in sports, adapted from Fullagar et al, 2019.

The research presented in this thesis is directed through the need to provide guidance and understanding of how coaches and practitioners can apply data analysis techniques to inform their decision making and support athlete performance and development. The first stage of this requires an understanding of what the relevant research questions are relating to the practicality of this area. To do this, a review of literature was conducted, with the intention of scoping out how current theory and practice is represented in academia. Through this process, a need for systematic review of collective behaviour metrics in football was evident due to limited conceptual and practical information. Two systematic reviews were performed to evaluate spatial temporal (chapter three) and social network analysis metrics (chapter four), with specific interest in the value of the methods presented to practitioners. This appraisal of the literature formed the basis for shaping the direction of the research, identifying relevant questions, allowing investigations that focus on the potential impact of a range of approaches presented in the literature. A critical feature, that was identified through the systematic reviews and will be discussed in depth in the relevant chapters was a scarcity of research supporting reliability of metrics used. In turn, an in-silico simulation of how measurement error impacts metric value was conducted (chapter five). Chapter six uses interviews to help translate the metrics into concepts that are relevant to the game through the perceptions of coaches. These steps offer the creation of systems that monitor collective behaviour dynamics through passing and positional data and can be used to inform practice design. An iterative process was used to allow adaptations and refinements to be made to the metrics. Based on the work exploring coach perceptions of these metrics, chapter seven provides pedagogical support for practitioners wishing to incorporate the techniques used in this thesis within their own setting. This is done through a step-by-step process, using R studio to manipulate data and calculate metrics. The final chapter provides an overarching summary of the thesis, including conclusions, limitations, and areas for future research.

CHAPTER 2

Background & Literature Review

2.1 Developments in Sports Performance Analysis

2.1.1 Traditional Performance Analysis

Before considering the many varying complexities of performance analysis, it is prudent to first define what the phrase "performance analysis" commonly refers to in the context of sport. As a branch of sport science, performance analysis can offer coaches objective information that can be used to generate insight into and improve both the understanding and quality of performance (Lord et al, 2020). However, it is noted by Sampaio et al (2013) that providing a single agreed upon definition of performance analysis is ultimately challenging due to the interdisciplinary nature of the field. Nevertheless, Sampaio and colleagues agree with Lord et al (2020) by highlighting two primary aims of performance analysis: 1) to improve the scientific understanding and 2) to support the coaching process through augmented information. Some have attributed the beginning of sports performance analysis in conjunction with the arrival of the personal computer approximately 40 years ago (Nevill et al, 2008). Conversely, Eaves (2015) argues that performance analysis has been used for well over a century with notation analysis being performed on sports such as baseball and tennis in the late 1800's. These early analyses would often feature within sports reports in newspapers and periodicals providing simple summaries of frequency data and even occasionally being used as an analytical tool to help inform practice.

The first scientific study applying notational analysis is considered to be a statistical paper by Reep and Benjamin (1968) that examined passing in association football. However, the field of performance analysis only gained traction over a decade later when a slow but steady stream of articles was produced relating to performance in sport (Nevill et al, 2008). Traditionally, researchers measured and quantified events in a form of notation to gather data. For example, the number of aces performed in a tennis match, or the number of shots taken by a football team (Lord et al, 2020). Currently, this level of data is often displayed in match reports that fans can access easily. Scientific research investigating game actions would often include additional dimensions to the data. Analysis of shots in team sports such as handball could be analysed alongside variables such as player position, match status, type of attack, match outcome, quality of opposition and individual characteristics to name some of the additional dimensions that game actions are analysed with (Ferrari et al, 2019). Despite a range of alternative and innovative approaches to performance analysis in sport, investigations of game actions are still the most common approach as identified in a recent systematic review by Lord et al (2020).

In contrast to traditional analyses that served to provide simple summaries across a sporting event, it is now recommended that performance analysts gather game action data within a dynamical context to adequately capture the complexity within team sports (Lord et al, 2020) This means identifying components of space, time, speed, or interactions with opponents. It is argued that these additional dimensions can provide deeper insights when analysing the data. The most common approach to analyse performance data in sport is the use of inferential statistics. Often, inferential methods attempt to identify key performance indicators through relating the frequency or percentage of game actions to team success. Predictive models take this one step further by attempting to forecast outcomes from patterns of game actions or other measurements. Inferential and predictive models have the potential to help shape practice by identifying specific characteristics possessed by successful individuals or teams (Lord et al, 2020). So far, these approaches appear to have limited success in stablishing validity. Alternatively, descriptive statistics that are still used consistently within the research base only provide frequencies or percentages of data which can still be useful, if only superficially. To date, predictive and inferential statistical methods have only had limited success, especially in complex sports such as football. Consequently, non-linear methods such as approximate entropy, Lyupanov exponent and artificial neural networks have been suggested as means to analyse sports performance (Dutt-mazunder et al, 2011). Neural networks have been used sporadically in sports analysis over the last 15 years investigating game events, collective behaviour and movement patterns (Passos et al, 2006; Lord et al 2020).

2.1.2 Movement Analysis

Understanding the movement demands and patterns in sports has become a critical feature in the performance analysis discipline. The first such attempt to collect data and perform analyses that could enhance understanding to this context was performed by Reilly and Thomas (1976). Research into player movement was rare due to the technological limitations at the time requiring analysts to notate movements visually (Nevill et al, 2008). However, with the introduction of global position systems (GPS), measurements could be automated. Originally used in military settings, the first attempt at applying GPS to measure athlete movement was performed in 1997 (Schutz and Chambaz 1997). Since then, numerous developments have been made to improve the accuracy of measuring player movements. One such innovation has been an integration of triaxial accelerometers into GPS units that assist in measuring accelerations as well as being able to calculate the impact force from physical collisions that athletes may receive in sports such as rugby (Cummins et al, 2013). These measurements can help calculate body load during training and matches which have been suggested to predict injury likelihood (Ehrmann et al, 2016). The sampling frequency has also increased in the years since the introduction of GPS in sport performance analysis where units recording at 20Hz have become regular in recent years whereas previously frequencies as low as 1Hz were used (Zlojutro et al, 2020; Cummins et al, 2013). Increases in sampling rate have generally resulted in an improvement in device validity and reliability. Although this is not always the case with some studies reporting that 10Hz GPS units are just as accurate if not more so when compared to 15Hz units (Scott et al, 2015; Johnston et al, 2014). Other considerations when using GPS should also be considered such as differences in interunit reliability, applications of different filtering techniques to smooth the data and where to position the GPS unit on the body (Malone et al, 2017).

2.1.3 Recent trends in analysis

Other methods for tracking athlete movement have also been developed including semi-automatic video tracking and local positioning measurement

(LPM) systems (Linke & Lames 2018). Indeed, LPM systems have been reported to be highly accurate when measuring athlete position (Linke & Lames 2018; Frencken et al, 2010; Ogris et al, 2012; Siegle et al, 2013). On the other hand, when analysing team sports such as football, video tracking allows the unobtrusive measurement of both teams and the ball. Tracking systems have been used to profile the physical exertions of players, allowing for sport scientists and coaches to tailor training programmes specific for individuals to perform at their best based on expected match demands (Palucci et al, 2019). However, this data can also be used to measure tactical movements and how players and teams organise themselves to perform as a unit. This application of spatial-temporal data has gained traction in team sports performance analysis over the last decade (Low et al, 2020) and goes beyond the traditional applications of motion analysis that look at physical profiles of individuals, and instead investigates the interactions of multiple players. At a dyadic level this may simply involve calculating a relationship between two players such as the distance (Duarte et al, 2012a). Other mathematical models have been applied to calculate collective team behaviours such as the convex hull which observes how the space a team occupies contracts and expands over the course of a game (Frencken et al, 2013). Some approaches even observe the match dynamics. For example, Voronoi cells are a method of measuring the areas on a pitch closest to each player at any given point in time. This model has been suggested to quantify passing behaviour between opponents in team sports such as football (Rein et al, 2016). By analysing data beyond the individual, deeper insights into the underlying mechanics of performance may be found and could particularly benefit team sports.

In recent years, social network analysis has been increasingly applied to game action data, most commonly passes in football (Ribeiro et al, 2017). By framing a sports team as a social network, interactions between players through passing actions can be applied to understand interactive processes used by teams to overcome their opponents (Mclean et al, 2018). Such analysis can identify players who are critical in linking play between two members of the team using betweenness centrality, or how subgroups of players cluster within a team by measuring the clustering coefficient (Aquino et al, 2018). More and more data are being gathered in a wide variety of sports and innovative approaches are being used to investigate how this data can lead to an improvement in performance in what is often a highly dynamic and unpredictable environment.

2.2 Performance Analysis in Football

Association football is one of the most prominent sports worldwide and this popularity is reflected in the number of articles examining performance analysis methods within the sport over recent years (Sarmento et al, 2014; Aquino et al, 2017a; Sarmento et al, 2018a; Palucci Viera et al, 2019; Low et al, 2019). Football is a team sport where 2 teams of 10 outfield players and a goalkeeper attempt to score in their opponents' goal. To achieve this, football players perform actions including passing, shooting, crossing, dribbling, tackling and intercepting whilst their teammates manoeuvre themselves to facilitate the collective team objectives of scoring goals whilst limiting their opponents' opportunities to do the same. Traditional means of performance analysis gathers event data relating to these actions with a large section of literature focussing on elite players performing at national or international level (Aquino et al, 2017a). Through understanding the actions of elite players and teams, coaches can replicate these behaviours through training practices with their own team. yet, this descriptive approach only offers superficial information to develop team performance. Consequently, contextual information is required beyond simple event data to explain the deep underlying mechanisms related to successful performance in a complex sport such as football.

2.2.1 Considerations of Performance Analysis in Football

When performance analysts record data gathered from football matches, it is important to include additional situational variables. Aquino et al (2017b) highlights the importance of factors such as match location, match status and opponent strength. Home advantage is a commonly accepted factor in performance of football and several attempts have been made to model the effects of it (Goumas 2017; Pollard and Gomez 2009; Liu et al, 2019). It has also been demonstrated that match location has a significant ($p<0.001$) impact on actions including shots, shots on target, crosses, successful passes and ball possession. Match score can also impact the dynamics of a football match. A team that is winning may become less aggressive and attempt to defend their lead. This effect was demonstrated by Almeida et al (2014) who also identified effects of opposition quality on team behaviour. Opposition quality can be slightly more challenging to accurately categorise compared with the binary match location (home, away) and the ternary match status (winning, drawing, losing). Various approaches have been used to separate teams depending on their strength with a k-means cluster approach providing satisfactory results (Aquino et al, 2017a; Aquino et al, 2017b). These factors cause football to become a dynamic and challenging sport to quantify with situational variables intermingling with each other increasing the complexities of performance. However, most research articles in football do not account for situational variables such as match location, opponent strength and match status (Mackenzie and Cushion 2013; Aquino et al, 2017b).

Another consideration within the recording of event data through notation is the reliability of the process. To achieve appropriate reliability, clear definitions must be outlined for each in game action, along with descriptions relating to what is a successful and unsuccessful action. This is demonstrated by Liu et al (2013) who found that experienced operators agreed on nearly all actions in a match (ICC ≥ 0.96). Such high scores can be attributed in part to clear definitions. However, much like situational variables, many articles do not provide sufficient information in the processes used to classify actions and many do not contain clear descriptions of actions, risking contradictory interpretations (Mackenzie and Cushion 2013; Aquino et al, 2017a). Consequently, it becomes challenging to perform meta-analyses on the literature base as data collection methods may categorise match events such as successful passes inconsistently due to varying definitions compounded by analyst's different interpretation of such definitions. Due to the complexity of football, large data sets and rigorous procedures are required to provide actionable insights, however, Mackenzie and Cushion (2013) suggest that many studies in the field fall short of this description. Acquiring a large and reliable data set is a key factor in high quality research in all disciplines. That being said, there is no clarity on what an appropriate sample size would be for match analysis as power calculations are not used to establish this within the literature (Aquino et al, 2017a).

2.2.2 Current Practices of Performance Analysis in Football

With the rapid increase in performance analysis research in football over the last 30 years, many football clubs will gather and record their own performance data. However, teams have often been secretive with their own information and have rarely shared it with researchers or the general public. Although, in recent years large data sets have become more widely available offering scope for more extensive research into performance analysis (Pappalardo et al, 2019). A number of football fans have taken to amateur performance analysis in recent years creating blogs and presenting findings identified after analysing arcane data sets. For example, a recent and common phrase within football terminology is 'expected goals'. However, there appears to be very little academic research expounding this process (Rathke, 2017). Expected goals is a model that examines the shots attempted by a team and categorises the danger of each shot based on factors such as distance and angle to goal, shot power and part of the body striking the ball. This helps identify that creating shooting opportunities closer to the opponent's goal yields better results. However, this finding may offer no true practical application as coaches understand this concept already through intuition and common sense. Nevertheless, expected goals models offer an alternative means of calculating team performance beyond regular means that focus on outcome measure and might miss complexities in team performance.

A common approach to simplify the intricate nature of team performance in football is to scrutinise the different roles within a football team. Differences have been identified between technical actions performed by players across different positions (Dellal et al, 2011). Yet, more emphasis appears to be placed upon positional differences in physical fitness. Both anthropometric and physiological tests are frequently used by sports scientists to measure the physical ability of each player. Anthropometric and physiological differences have been identified in elite male footballers (Marcos et al, 2018). This was identified using laboratory-based tests, however, with the increasing use of tracking technologies such as GPS, a similar analysis can be applied to identify differences between positions during match play. Dalen et al (2016) found clear positional differences in running profiles by observing GPS data of players in 45 domestic matches. This study identified that full backs and wide midfielders perform more accelerations across a match compared with central defenders, central midfielders and attackers. This data also demonstrated clear positional differences between distance travelled when walking, jogging, running, high-speed running and sprinting. Understanding specific physiological needs of players can help sports scientists and coaches develop training programmes tailored to developing fitness requirements for each position (Palluci Viera et al, 2019).

Investigations of performance analysis methodologies do not just concern professional male footballers. Several studies have investigated elite female footballers as well. However, it is highlighted that the majority of studies focus predominantly on males (Low et al, 2019). This perhaps reflects the disparity in both participation and finances in male compared with female football. Some systematic reviews have even excluded performance analysis of female athletes entirely from their research (Sarmento et al 2018a; Sarmento et al 2018b; Sarmento et al 2014). No justification is provided for the decision to omit female data; however, this may be due to changes in the observed relationships when pooling male and female data. Considering the growth of the women's game, there may be greater interest in generating analyses and comparing game structures and pattern differences between elite male and female football such as the investigation by Tenga et al (2015). Alternatively, some performance analysis research in football explores differences between professional and elite youth cohorts. One such study investigated differences in male Brazilian professional football players compared with their youth counterparts at U20, U17, U15, U13 and U11 level (Palucci et al, 2018). This analysis observed differences in skill related, running and collective behaviour performance and found several trends which demonstrated clear improvements in performance from pre-pubescent players up to elite adults. However, few differences were identified in this analysis between U20 players and professionals highlighting the challenge to quantify the nuances which help players ascend to the elite of the game. Whilst literature has observed amateur or semi-professional football, most literature concentrates on youth players focussing on developing and nurturing talent, and professional players as a means to create a model for successful football performance.

The creation of models is an important aspect in sports performance analysis and links directly with the practical applications of the discipline. But models can often be descriptive, providing normative data to indicate target values for certain measurements. As an example, Palucci et al (2018) offer their research primarily as a source of reference values for male footballers at varying competition levels, providing information on physical, technical and tactical activities during training. This can be useful for practitioners when applying these tests to their own players, helping coaches understand how these values develop over age groups. Yet, the model fails to provide any deeper understanding of the mechanisms explaining successful performance. From a data analysis standpoint, this can be challenging to identify novel insights into the game when coaches already have a deep understanding of the game through their own intuition and study. As previously stated, an analyst may present a model such as expected goals, explaining that the closer you are to the goal, the more likely you are to score, a concept all coaches already understand clearly. Nevertheless, further investigation into expected goals can yield predictive insight into future performance and also be used for player recruitment (Rathke 2017). One such model suggested as a framework for talent identification is that proposed by Decroos et al (2019). This analysis model comprised machine learning algorithms to identify the values of individual actions and measured how each event increased or decreased the likelihoods of a goal being scored by either team. Such a framework could be a potentially effective approach for player recruitment strategies and further uses may even extend to explaining the decision making of players and their behaviours in competitions.

2.3 Theoretical frameworks in football

Traditional performance analysis in football has largely considered relationships between simplistic variables and attempted to identify associations between high level performance (Sarmento et al, 2014). Over the last decade, there has been more emphasis on understanding the collective behaviours with researchers applying theories from sociobiology to frame the sport of football (Duarte et al, 2012b). By framing a football match as a system or even a superorganism of many agents interacting and coordinating with each other, interactions and collective behaviour of the component parts becomes the focus of performance analysts. A dynamic system such as a football match can appear chaotic, however, like other complex systems such as the weather, patterns and order emerge from the laws which regulate the system (Davids et al, 2013). In football, each player is constantly integrating information and adjusting their position on the pitch accordingly. A team, especially one with experienced players, will create coordinated behaviour attempting to overcome and disrupt the opposition. These emergent patterns of play are based on the collective movement of players. Understanding and analysing these aspects can help develop a deeper understanding of the team as a collective and provide insight into the complexities inherent in a sport such as football which appears unpredictable.

Players within this system or 'superorganism' make decisions through their own self-organisation. This effect has been demonstrated in numerous dynamic systems including swarms of animals (Duarte et al, 2012b). In a swarm, the collective demonstrates what appears to be intelligent and coordinated behaviour, however, is found to be the product of individual agents making decisions based on the position and movement of their neighbours. This helps the dynamic self-organising system achieve what is impossible for the individual to (Chassy et al, 2018). Extrapolating this theory into sport helps identify the decision-making processes made by players. Elite football players will naturally make intelligent movements guided by years of experience and intuition through self-organisation. A critical component of self-organisation is the absence of supervised and pre-determined patterns. However, a manager or head coach of a football team may attempt to instil these within their team. Nevertheless, in a match context, opponents will attempt to disrupt the flow of the team, causing individuals to adjust their position. A head coach is unable to simultaneously direct the movements of 11 players on the field. Consequently, players decision making processes are predominantly instinctual and derived from their perceived local environment.

The model of self-organisation supports a player centred approach to quantifying the performance of football teams. As such, considerations must be made as to where coaches fit into this framework. In situations where players are acting based upon their environment, the coach is composed as part of that environment (Woods et al, 2020). Thus, coaches can manipulate the environment of the player to achieve desirable behaviours and decisions from players. For a coach to best apply appropriate manipulations, an awareness and appreciation of the underpinnings of ecological dynamics would be beneficial (Aquino et al, 2017a). In this context, player behaviours emerge from the constraints placed upon the player, the task and the environment. The player gathers information through the perception of constraints and performs based on this information. As the action is performed, new information becomes available for a player to make another decision. This cycle continues and has been suggested to develop functional movement patterns in the branch of non-linear pedagogy (Machado et al, 2019).

Coaches are potentially able to refine the development and learning of player behaviours through manipulation of the constraints within practice sessions. Specific emphasis has been placed on a constraints-lead approach in non-linear pedagogy where focus is made on manipulating constraints within game-like activities. For example, some studies have investigated how adjusting the pitch size impacts player decision making and behaviours in games (Coutinho et al, 2018; Olthof et al, 2018; Clemente et al; 2018a). The number of players in a game has also been investigated (Aguiar et al, 2015), with several studies observing the manipulation of constraints focussing on small-sided games. In these small-sided games, coaches have the ability to include floating players creating overloads which can offer players alternative learning experiences (Praca et al, 2016). The number of goals has also been shown to adapt the behaviour of teams in small-sided games (Travassos et al, 2014a). When analysing 1v1 practices, Headrick et al (2012) found that the proximity of the duel relative to the attacker’s goal had impacts on the attacker’s approach. Similarly, it has been demonstrated that an attacker will adjust their behaviours dependent on whether the goal is to the right, left, or in front of an attacker (Laakso et al, 2017). In fact, even coach instruction has been identified as causing substantial differences in player behaviour during 1v1 bouts (Clemente et al, 2013c). This highlights the range of constraints that can be manipulated by a coach, yet, understanding how manipulation of these constraints influence individual and collective behaviours of players is required before using them effectively. This necessitates a method of measuring the collective behaviour of a team through variables or metrics with the ability to quantify team decision making. Such measurements may be used by coaches to observe and understand how their team learns and develops functional and effective behaviours. However, this posits the question “how do you quantify collective behaviour?”.

2.4 Measuring collective behaviour in football

Over the last decade, numerous metrics have been applied to football to measure collective behaviours of football teams. A wide range of variables have been explored in the hope of providing insight into football performance (Low et al, 2019). The following section will provide an overview of the approaches used within the growing literature base. In chapters 2 and 3, full systematic reviews will be conducted and examine the validity and applicability of the metrics.

2.4.1 Collective behaviour through spatial temporal metrics

Over the last decade, numerous metrics have been applied to measure the collective behaviour of football players and teams (Low et al, 2019). One of the most frequently used metrics conceptualises the team centre of mass and is often labelled the team centroid. Traditionally calculated through the mean position of all outfield players on the pitch, the team centroid naturally progresses towards the opponent’s goal when a team is attacking, and back towards their own goal when defending (Duarte et al, 2013a). Due to the naturally longitudinal nature of football, more emphasis is often placed on the team centroid in the x axis. When observing matches along this plane, a crossing of centroids can emerge which might be a useful performance indicator to forecast critical events. Frencken et al (2011) demonstrated that the overlapping in opposing team centroids was linked with goal scoring chances created by football teams playing small-sided games. However, other research observing this phenomenon did not find the same results when observing 11-a-side matches, yet, there is a clear coupling demonstrated between the centroids of teams when a match is played, particularly along the x axis.

Alternative approaches for calculating centroid have been attempted. Clemente et al (2013d) proposed a system for calculating centroid through the average distance of each player including the goalkeeper through adding a weighting system across each player based on their distance to the ball. Whereas Goncalves et al (2014) applied a group centroid based on the average position of each defender, midfielder and attacker separately. The team centroid can be applied in other ways to observe player behaviour. For example, the distance of each player to their team centroid or the opponents team centroid may provide information to the decision-making processes of players and how it relates to their own team and their opponents (Goncalves et al, 2016). When testing non-elite players before and after a training programme, the distance to centroid was the strongest variable in identifying players behaviours pre and post intervention (Sampaio et al 2012). The stretch index measurement takes this concept of distance to centroid even further. By averaging each outfield player’s distance to their own team centroid, the dispersion of a team can be observed. This variable can be used to quantify how contracted or expanded a team are (Silva et al, 2016a).

Other metrics have been created to measure the dispersion of teams. For instance, the surface area of a team is calculated through convex hull algorithms. This creates a polygon around the peripheral outfield players and calculates the area within the shape created (Olthof et al, 2019). Alternatively, team spread can be used, calculating the Frobenius norm of the distances between every player in a team (Bartlett et al, 2012). As with the stretch index, these mathematical models can be used to measure team dispersions. Multiple studies have identified that teams will naturally form a compact tight space when defending and become more expansive when attacking, attempting to create space (Moura et al, 2016; Castellano et al, 2013; Barnabe et al, 2016). Consequently, the dynamics of these variables might provide more interest in phases of transition between attack and defence and vice-versa. When analysing the coordination of opposing team spreads through vector coding from the start of attacks, Moura et al (2016) found that higher periods of anti-phase synchronisation were found in attacks which resulted in shots compared with attacks that were successfully broken down by their opponents.

Clear patterns have been found in the aggregation and distribution of players during football matches to create and deny space. An interesting model to categorise the dominant region of individuals during matches is the application of a Voronoi cell computation. Dominant region identifies the area closest to each individual player on the pitch based on their current position (Fonseca et al, 2012). Rein et al, (2017) applied Voronoi cell computations across 107 Bundesliga matches attempting to evaluate the decision-making process in players passing. Voronoi cells offer information on how much space a player is in at any given moment, as such, relating it to passing behaviours in football appears to be natural. However, the Voronoi cell computations assume that each player is stationary and able to move in any direction at the same speed. This is unlikely to be the case on a football pitch as players are constantly moving. Indeed, Filetti et al (2017) demonstrated a similar approach in measuring the technical ability of passing by observing a technical efficiency rating. However, their adaptation of the Voronoi cell computations integrated player trajectory as well. Such an analysis can provide practitioners with the opportunity to quantify a player's technical ability in their passing success, as well as their decision making.

Not all collective behaviour metrics analyse the team or the game as a whole. Sometimes analysis is made across player dyads considering the distances, angles, or speeds between two players. These appear to be used more commonly in understanding the dynamics of 1v1 situations. Clemente et al (2013c) states that the distance between an attacker and defender is critical to understand how one manages to overcome the other. Similarly, the angular relationship between two players can also be an important factor often used when understanding an attacker and defender's position in relation to the goal (Leser et al, 2019; Laakso et al, 2019). In this construct, as an attacker passes the defender, the angle will significantly increase (Clemente et al, 2013c). Shafizadeh et al (2016) conducted analysis on 1v1 shots between attackers and goalkeepers in the English Premier league. By measuring the distance between the players as well as the velocity of the goalkeeper, it was identified that in the saved shots, specific correlations (0.59-0.95) between these measures were identified at the moments 80 milliseconds and 760-480 milliseconds before the shot was struck. Such analysis might provide a deeper understanding of successful behaviours for strikers and goalkeepers in such critical moments. Consequently, practices could be designed in helping goalkeepers develop these behaviours to improve their performance.

2.4.2 Non-linear analysis of spatial temporal metrics

When considering dyadic relationships that are present in a football match, the coordination and movement patterns of these are an important aspect in understanding player behaviour. Consequently, innovative methods for measuring synchronisation have been used in football performance analysis research (Low et al, 2019). Relative phase appears to be the most common approach used, identifying synergistic behaviours from two different signals. If two players are moving in the same direction, their movements can be described as in-phase. Whereas players moving in opposite directions are in anti-phase. This offers analysts a means to identify when player dyads are moving synchronously. It has been suggested that higher synchronised behaviour of players is linked with higher performance. Folgado et al (2015) identified that teams demonstrated more synchronised behaviour when playing against stronger opposition (η^2 = 0.254 – 0.648). Additionally, it has been demonstrated that synchronisation between players will increase as players begin to play regularly with their teammates (Folgado et al, 2018a).

This analysis method is not just limited to comparing individual player movements. Other metrics can be used, for example, the synchronisation between two team centroids as applied by Travassos et al (2014b). Moreover, other methods have been used to measure synchronised behaviour including vector coding (Moura et al, 2016) and cluster phase amplitude (Lopez-Felip et al, 2017).

A separate branch of nonlinear analysis used within collective behaviour research in performance analysis observes the regularity of different metrics. Like synchronisation, these measures observe signals over a period of time offering perspectives of emergent behaviours helping solidify knowledge in the underlying processes involved in successful football performance (low et al, 2019). Within the literature base, approximate entropy (ApEn) appears most frequently, measuring the likelihood a pattern will repeat itself in a time series. This approach has been used on many of the aforementioned metrics such as distance to centroid, (Aguiar et al, 2017) distance between centroids (Low et al, 2018) and stretch index (Coutinho et al, 2019). Higher regularity is assumed to represent a high level of performance by indicating more consistent and intelligent movement patters are performed by teams (Sampaio et al, 2012). Other approaches including sample entropy (Barnabe et al, 2016) and Shannon entropy (Vilar et al, 2013) have also been applied by researchers. The range of spatial-temporal metrics, alongside inventive methods of analysis such as entropy measures provides promise for the advancement of knowledge in understanding complex behaviours performed by football teams. So far, success has been largely superficial, although, collective behaviour is not limited to the observations in space and time.

2.4.3 Collective behaviour through social network analysis

Network analysis offers an alternative to spatial-temporal analysis in measuring collective behaviour of football teams. Network analysis applies event data categorising players who pass to each other through the creation of graphs. These graphs can be weighted, often by identifying how often each player passes to one another. Additionally, graphs can be directed or undirected. In the case of passing, a directed graph or di-graph identifies the passer and receiver of a football during a passing event (Ribeiro et al, 2017). Certain network properties can be constructed from analysing the graphs and visualised commonly through schematic representations. Properties of individual nodes or players can be described through network analysis. In-degree centrality identifies the prominence of a player receiving a pass from a teammate and may indicate the prestige of the player among their teammates. Conversely, out-degree centrality identifies a player’s ability to pass the ball to a teammate and characterises a player’s ability to contribute to a team’s offense (Clemente et al, 2015a). These metrics require a di-graph to be used so that the direction of passes are present in the analysis. Moreover, a weighted or unweighted graph may result in different outputs and should be considered when applying network analysis.

Network analysis measurements also provide information on the whole team. A frequently used metric when applying network analysis to football is network density. This metric identifies how well connected a team is, a higher network density shows a more interconnected team, where players pass to all of their teammates. Indeed, there is some evidence to suggest that having a higher network density correlates with successful team performance (Clemente et al, 2015b; Peixoto et al, 2017). However, there appears to be a limited number of investigations successfully relating network properties with successful performance, nevertheless, Grund (2012) found higher network intensity was linked to an increased team performance, suggesting that coaches should focus on increasing the passing rate of a team. Network analysis does not have to be constrained to passing between players. Garrido et al (2020) examined passing networks between pitch zones. However, in this case, how an analyst separates each sector of the pitch can vary massively (Buldu et al, 2018). Nevertheless, social network analysis has potential to offer the performance analysis community within football and aid in understanding the complex dynamics of the team sport.

2.5 Summary

Performance analysis through largely simple notation of actions has been used to inform preparation for competitive sport since the 19th century. This form of analysis continues to be the dominant method of creating key performance indicators and evaluating performance. However, due to the fluid and dynamic nature of team sports such as football, increasingly complex analyses are being used to support the coaching process. Data collection integrating situational variables allows for more comprehensive understanding of events in matches. However, in recent years, football teams have been described as superorganisms. Consequently, understanding how each part of the collective interacts with one another forming a unit that organises itself intelligently and more importantly, how to train the superorganism to make more effective decisions becomes challenging for traditional methods of performance analysis to support. This has resulted in novel approaches applying spatial-temporal analysis and social network analysis to measure constructs within a team and sub-groups of a team. There is a breadth of literature investigating various mathematical models applying these concepts in football. So far, these approaches are often exploratory and the methods, and techniques used require systematic study to identify what is actually being measured, whether measurements are valid and reliable with regards to football, and what if any are the practical implications from the research to date. To answer these questions, the following two chapters will take the form of systematic reviews to investigate spatial temporal and social network analysis within football.

2.6 Aims and Objectives

Investigate the applicability of collective behaviour metrics that involve spatial temporal and network analysis

This aim will be achieved by completing the following objectives

1. Identify issues in translating current research into practice.
2. Evaluate validity and reliability of collective behaviour metrics.
3. Propose a system to support coaching practices.
4. Support practitioners in implementing spatial temporal and network analysis metrics.

CHAPTER 3

Spatial-temporal Metrics to assess Collective Behaviour in Football: A Systematic Review and Assessment of Research Quality and Applicability

Prelude

This chapter largely follows the manuscript accepted for publication in the journal of athletic enhancement. The systematic review identified and evaluated the majority of the literature base applying spatial temporal metrics to investigate collective behaviour in football. Included in this chapter is a summary of investigations and outcomes of research, an appraisal of study quality and an assessment of author quotes, relating to the spatial temporal metrics investigated with focus on their relevance to coaching concepts and their application. This research provides a foundation for future research investigating spatial temporal metrics.

3.1 Introduction

Performance analysis is an evolving discipline of sport science that aims to use innovative approaches to instrument coach decision making and athlete performance (Mackenzie and Cushion 2013; Sarmento et al. 2018a). Technologies such as global positioning systems (GPS) and semi-automatic video tracking have been used extensively in elite sport for a prolonged period and enable insights into performance to be captured (Cummins, Orr, O'Connor & West 2013; Castellano, Alvarez-Pastor & Bradley 2014). It has also been reported that GPS data have been used primarily to quantify physical outcomes and, to a lesser extent, identify external factors that influence the physical outcomes measured (Castellano et al. 2014). However, team sports are highly complex, and it is recognised that descriptions of simple behaviours such as number of sprints and total distance achieved provide limited insight into the functioning of a team across collective units (McGarry 2009). Contemporary perspectives that view team sports as complex systems identify a need to focus on interactions between players in different match and training settings. Interactions can routinely be described by the positioning and motion of players relative to each other (Travassos, Araujo, Correia & Esteves. 2010). More abstractly, Duarte, et al (2012b) described sports teams as superorganisms composed of teammates continually communicating to help the team function as a unit. Importantly, communication is not limited to routine verbal instruction, but also includes the interrelated dynamics of player motion. Based on these perspectives, player tracking technologies can be used within a systems framework and the generated spatial-temporal data used to provide insight into collective behaviours that may lead to more intelligent self-organisation (Low et al. 2019).

One sport where spatial-temporal assessment of player collective behaviour is developing rapidly is football (Sarmento et al. 2018a). Conventional performance analysis approaches such as frequency analysis can be considered simple methods that describe outcomes of collective behaviour. However, due to the lack of contextual information describing the underlying processes that led to these outcomes, conventional approaches are limited in their ability to inform decision making (Tenga & Sigmundstat 2011). As a result, integration of spatial-temporal data into collective behaviour approaches is increasingly being developed to explore models that best describe underlying processes and subsequent outcomes generated. However, with up to twenty-two players plus substitutes participating, a wide range of approaches to quantify and assess coordinated behaviour in football exists. This range reflects the diverse research produced by authors over the last decade (Bartlett, Button, Robins, Dutt-Mazunder & Kennedy 2012, Clemente, Couceiro, Lourenco Martins, Mendes & Figuiredo 2015c, Castellano, Fernandez, Echeazarra, Barreira & Garganta 2017, Coutinho et al. 2019b), with Sarmento et al. (2018a) identifying collective behaviour analysis and associated metrics as one of the most innovative and important trends for football analysts. The large range of approaches to quantify collective behaviour also appears to be influenced by the overarching theoretical framework adopted (e.g. dynamic systems theory, sociobioglogy) and the specific backgrounds of researchers involved. Applications of these should ultimately be valuable to practitioners and coaches providing relevant information that can be used to improve performance through adaptations to training design or match play (Memmert, Lemmink & Sampaio 2017, Sarmento, et al. 2018a).

To date there has been limited attempt to synthesise research investigating spatial-temporal metrics and assessment of collective behaviour in football. The first systematic review was conducted recently by Low et al. (2019) and provided a comprehensive overview of the empirical research. A total of 77 studies featuring a mix of observational studies (n = 34) and field-based experiments (n = 43) with mostly male professionals were included. Low et al. (2019) identified 27 unique spatial-temporal metrics that were separated into four categories (position; distance; spaces; and numerical relations). Additionally, the authors' delineated between linear analysis methods performed on these metrics (e.g. mean, standard deviation and coefficient of variation) and non-linear analysis methods quantifying either predictability (e.g. approximate entropy, sample entropy and dynamic overlap) or synchronisation (e.g. relative phase, cross correlation, cluster phase and vector coding). Finally, the authors reported that investigations of collective behaviours were analysed at all system levels including the dyadic, sub-group, team, and match levels. Collectively, the review produced by Low et al. (2019) provided a clear and effective framework to synthesise findings from what initially could appear to be disparate approaches of individual studies. However, as the primary focus of Low et al. (2019) was to develop a framework to describe previous analytical approaches, there was limited synthesis and evaluation of the applicability of the evidence base. Given the complexity of spatial-temporal metrics and non-linear analysis approaches in comparison to conventional performance analysis methods, there is a need for authors' to effectively describe metrics and analysis approaches, providing context and recommendations so that practitioners and coaches can identify the value of the information and make appropriate decisions. In addition, effective discussion of reliability and validity of different metrics and analysis approaches would further enhance the practical value of the research. Therefore, the purpose of this current systematic review was to synthesise and evaluate the applicability of research investigating spatial-temporal metrics to analyse collective behaviours in football. The review identified and evaluated authors' descriptions of metrics and clarity provided to facilitate uptake by practitioners. A similar process was included with authors' discussions of practical applications of study findings. Finally, evaluations of research quality and attempts to address validity and reliability of approaches in primary studies were also included.

3.2 Method

A systematic review of published studies investigating spatial-temporal metrics for collective behaviour in football was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA). An initial scoping review of the research base was conducted to familiarise authors with the key metrics used to generate appropriate keyword searches. Subsequently, five electronic databases including SportDiscuss, Embase, Medline, Web of Science, and Scopus were searched and considered publications from the 1st of January 2008 to the 20th of February 2019. The search strategy combined two levels, the first included the following terms combined with the Boolean operator OR: 'centroid', 'centre of gravity', 'stretch index', 'team spread', 'surface area', 'dominant region', 'approximate entropy', 'relative phase', 'dyad', 'voronoi', 'coordination', 'patterns of play', ‘performance analysis’, ‘tactical analysis’, ‘notational analysis’, ‘group behaviour’, ‘group behavior’, ‘collective behaviour’, ‘collective behavior’. Results from the first level were then combined using the AND operator with the second level comprising: ‘football’ OR ‘soccer’.

Inclusion criteria for retrieved studies included: 1) participants of any age or sex engaging in football competition or training; 2) reporting of spatial-temporal metrics comprising at least two position references that described collective behaviour; and 3) the full publication was available in English. Investigations published as conference abstracts were excluded. Two separate reviewers (MC and NB) screened article titles then abstracts. Full-texts were then read and inclusion criteria applied to complete the 3-stage screening process. Disagreements regarding inclusion were resolved with discussion at the end of the abstract and full-text stages. The primary purpose of the review was to synthesise and evaluate the included research by considering authors’ descriptions of metrics and their discussion of reliability, validity, and practical application of findings. Therefore, data extraction was completed using two different extraction forms. The first extraction included basic information such as population investigated, sample size, specific metrics applied, analysis approach and overarching findings. The individual metrics and analysis approaches that were identified were categorised according to the criteria identified by Low et al. (2019). The second extraction included information regarding authors’ descriptions of included metrics and comments regarding their validity, reliability and practical applications. Direct quotes were extracted from each study and where multiple appropriate quotes were present, all were documented. Quotes regarding practical applications were categorised as either: 1) broad - generic conclusions providing limited direct applicability; 2) moderate – conclusions linked to specific game or training aspects providing some direct applicability; 3) specific – clear recommendations with specific reference to the use of a metric or analysis method providing direct applicability. Extractions were conducted independently by two reviewers (MC and TC) with a final discussion including the categorisation of the application of metrics amongst the full research team to ensure consistency.

To evaluate the methodological quality of studies a risk-of-bias quality form was adopted based on a 12-item checklist adjusted from Sarmento et al. (2018b). Studies were assessed based on the following criteria: 1) study purpose; 2) background literature; 3) study design; 4) detail of sample used; 5) justification of sample size; 6) identification of ethical approval; 7) detail of methods; 8) application of appropriate inferential statistics; 9) application of relevant analysis methods; 10) generation of appropriate conclusions; 11) generation of appropriate practical applications; and 12) acknowledgement of study limitations. A binary scale was used to score each item with the percentage of positive items awarded a unit score to provide an overall quality rating. Three bins were created to group articles into low (≤50%), moderate (>50%, ≤75%), and high (>75%) quality (Sarmento et al. 2018b). Data extraction and risk of bias assessment were made in duplicate across three reviewers (MC, NB and TC) with a final discussion amongst the full research team.

3.3 Results

The initial literature search identified 2282 studies which was reduced to 1110 after deduplication using RefWorks software. Title and abstract screening reduced the number of studies obtained in full-text to 97. A further 12 studies were removed due to metrics not meeting the inclusion criteria specified (7), non-reporting of data (3), and inclusion of sports other than football (2) as shown in figure 3.1. The 85 included studies comprised a wide range of population groups (Table 3.1) with respect to game type (competition, friendly, training game, small sided games (SSGs) and 1v1 bouts), playing level (professional, youth, semi-professional, amateur, composite), and country of investigation (Australia, Austria, Brazil, England, Finland, Germany, Italy, Multi-national, Netherlands, Portugal, Spain and Switzerland). The mean number of matches analysed was 10.6 with a standard deviation of 18.1 (range: 1 to 103). Thirty-one studies investigated metrics across full 90-minute matches, whereas most studies investigated much shorter SSGs. Findings from the studies were varied (Table 3.2) and generally focused on relation of metrics to success in terms of offense or defence, or the effect of factors such as gender, age, formations, tactics, number of players or pitch dimension on metric values.

The research quality evaluation (Table 3.3) identified a single (1.2%) "low quality" study, 30 "moderate quality" studies (35.3%) and 54 "high quality" studies (63.5%). The studies were most susceptible to bias through a lack of sample justification with only 5 studies (5.9%) stating a reason for the population selected. The research quality evaluation also highlighted that most studies failed to acknowledge study limitations (62.7%) and many (21.9%)

failed to identify practical applications. Where studies did identify practical applications, these were most often categorised as being broad and providing limited clear applicability (54.0%; Table 3.4). Examples of practical applications categorised as being of moderate applicability (36.8%) generally focused on constraints that could be applied in training including manipulation of pitch size (Goncalves et al. 2018a) formations (Baptista et al. 2018) and SSGs (Praça, Folgado, De Andrade & Greco 2016; Goncalves et al. 2017a). A limited number of practical applications (9.2%) were categorised as specific and provided clear recommendations with target values for team centroid (Aguiar, Goncalves, Botelho, Duarte & Sampaio. 2015), field space (De Souza 2018, Goncalves et al. 2018b) and distance between players (Headrick et al. 2012) that could be directly applied by practitioners and coaches.

Across the 85 studies, 115 unique metrics and analysis approaches were identified across a total sample of 366. A total of 84 (23%) instances were identified where an equation was presented. In contrast, there were 99 (36%) instances where a metric or analysis procedure was reported with no equation or source provided to describe calculations. Similarly, there were 79 (21%) instances of metrics reported with no formal description or comment to provide context or understanding of the purpose of the metric. According to the framework presented by Low et al (2019), the most commonly reported metric category was space metrics (129 instances), followed by distance metrics (110 instances), position metrics (21 instances) and numerical relation metrics (17 instances). Non-linear analysis methods quantifying synchronisation (52 instances) were most frequently applied using the Hilbert transform and at the team level through assessment of team centroids to identify coordinated movement across teams (Gonçalves et al. 2014, Siegle & Lames 2013). Predictability analysis methods (37 instances) were most frequently applied using approximate entropy (ApEn), followed by sample entropy (SampEn) and Shannon entropy.

No explicit reference was made to reliability or validity of metrics in any of the included studies. Implicitly, authors assessed the validity of metrics through multiple approaches. The most common was to employ rank-order methods and compare metrics across age groups or playing levels with the implicit assumption that older players and those playing at a higher level or in stronger teams would demonstrate more effective collective behaviours. Barnabe, Volossovitch, Duarte, Ferreira & Davids (2016), Castallano et al. (2017), Olthof, Frencken & Lemmink (2015), Palucci et al. (2018) each identified positive relationships between metrics and age with greater width, surface area, team spread, team centroid distance and attack-defence synchronisation with older players. Similarly, Silva et al. (2014a) identified that players of the same age group but from a higher standard of competition worked together more effectively to explore greater amounts of available space. Additionally, Folgado Duarte, Fernandes & Sampaio (2014a) identified that competition against stronger teams resulted in increased time in synchronized behaviour for overall displacements and displacements at higher intensities.

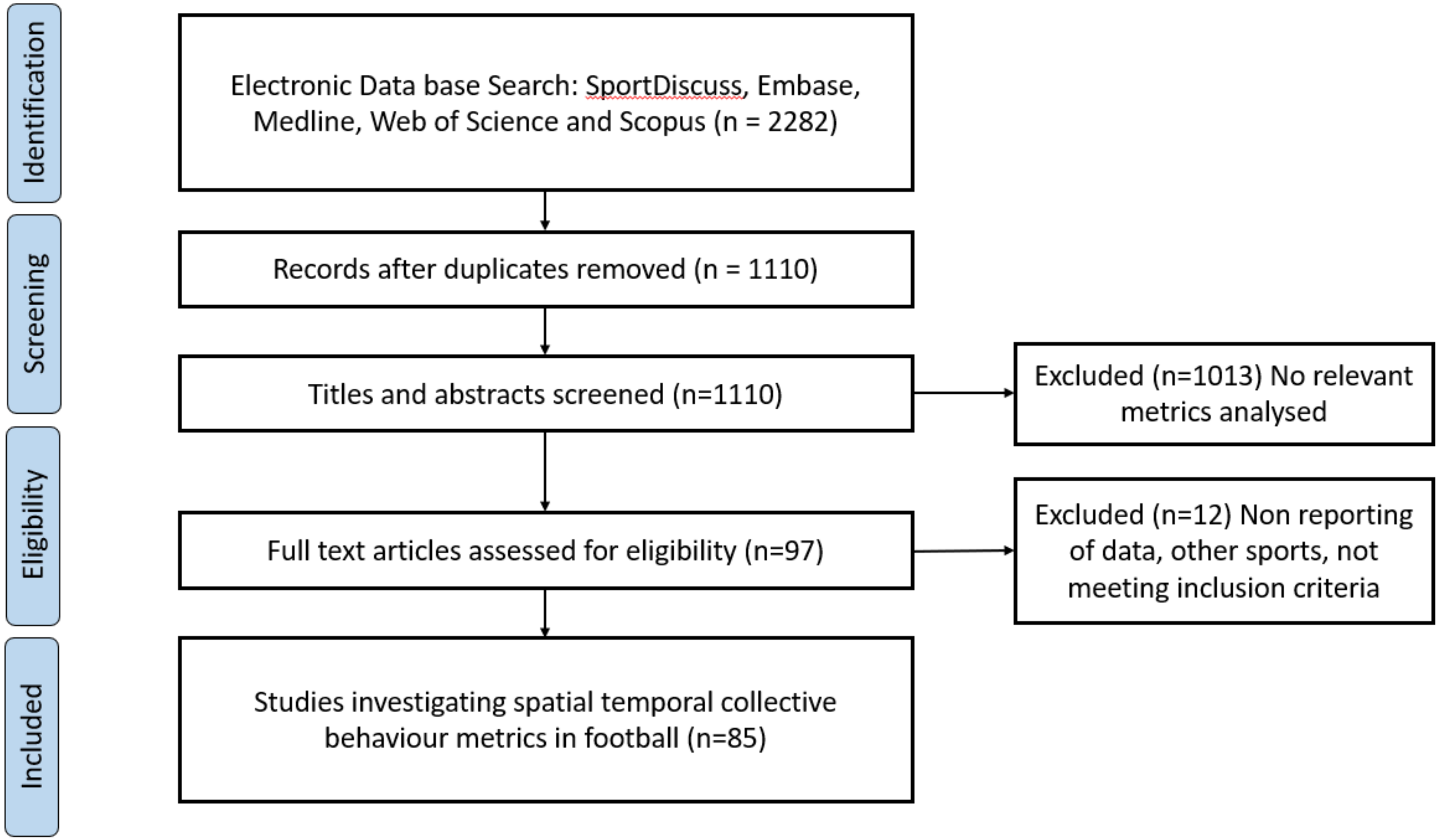


Figure 3.1 PRISMA diagram for collection of studies including spatial temporal metrics in footba

3.4 Discussion

The present study comprised a systematic review of research investigating collective behaviour in football through analysis of spatial-temporal data. The review identified that the area is rapidly growing and features a wide range of metrics (spaces, distances, positions and numerical relations), analysis methods (synchronisation and predictability), populations (primarily elite level males from U11 to adult) and game scenarios (e.g. competitive matches, SSGs, and 1v1 drills). Focus of the review was placed on authors' descriptions of the metrics generated and their discussion of reliability, validity, and practical applications of findings. The present review identified several areas for further development of the evidence base to increase the quality of the information and uptake in practice. An initial barrier was identified with regards to authors' descriptions of metrics with overemphasis placed on operationalised definitions and limited translation to game scenarios or coaching strategies. Additionally, when discussing practical applications, it was identified that authors frequently provided broad statements that restated results and did not provide clear recommendations with guidelines on relevant values or processes that could be used to generate team specific values. The following sections discuss in greater detail the data extracted from the review.

For practitioners to assess and apply a metric and analysis approach to their own data, an understanding of what the approach measures and how it relates to performance is required. However, review of the included studies identified that most authors' descriptions lacked conceptual overview, and instead focused exclusively on operationalised definitions. Common examples across the metric categories included: Spaces (surface area): “the convex hull formed by positions of the players in each team” (Castellano et al. 2017); Distance (distance between centroids): “the difference, longitudinally and laterally, between teams centroid positions” (Frencken, De Poel, Visscher & Lemmink 2012); Position (relative angle): “the relative angle (ɑ) between the centre of goal, defender and attacker” (Laakso, Davids, Liukkonen & Travassos 2017); and Numerical relations (Space Control Gain): “measured by the difference of space control percentage between pass initiation and pass completion modelled by utilising Voronoi diagrams of the pitch at each time frame” (Memmert, Raabe, Schwab & Rein 2019). Similarly, descriptions regarding the two main analysis methods provided limited conceptual understanding with representative examples including: Synchronisation (relative phase): “the relative phase of the time series corresponding to speed displacements of all dyads” (Goncalves et al. 2018b); and Predictability (ApEn): “measure was used to assess the complexity of the particular collective behaviours” (Duarte et al. 2013a). In contrast, there were a limited number of examples where metric descriptions also provided conceptual detail to relate to aspects of football: “The stretch index measures the compactness of a team on a given moment” (Clemente, Sequieros, Correia, Silva & Lourenco Martins 2018a); “[effective playing space] was calculated as the surface area of the convex hull of all players (excluding goalkeepers) as a measure of the playing area used by the players in a given situation” (Memmert et al. 2019) ; “[surface area] This variable expresses the relationship between the tactical forms (shapes) adopted and spaces exploited by both teams, to support analysis of how they varied over time” (Barnabe et al. 2016). In addition, primarily for numerical relations metrics there were examples where authors attempted to add context with regards to tactics and philosophy: “[offensive space ratio] The aim of this principle is to reduce the concentration of opponents in their central zone, thus attempting to open up some spaces to penetrate” (Clemente, Martins, Couceiro, Mendes & Figueiredo 2014a); “[pressure passing efficacy] aims to measure high quality through-balls by weighing passes with more than one outplayed opponent by the pressure on both pass initiator and receiver” (Memmert et al. 2019). Whilst operational definitions and associated equations are important to ensure consistency across analyses, authors should seek to combine this information with greater context as demonstrated by these latter examples.

The review of included studies identified no explicit reference to reliability or validity of metrics and analysis approaches generated. Several authors referred to the reliability of instrumentation used to collect position measures (Castellano et al. 2017, Laakso et al. 2017, Praça et al. 2016); however, no study investigated the extent to which noise in data influenced reliability or consistency of values generated within for example single sessions (where tactics may be expected to be somewhat consistent). Metrics and analysis approaches that are unduly influenced by noise or vary substantially across time periods where collective behaviours are expected to be consistent should not be recommended. Across the included studies a wide range of systems measuring position coordinates were identified with variation in accuracy expected. As a result, future research should seek to identify the positional accuracy required to generate reliable data across different metrics and analysis approaches to inform practice and measurement systems used.

Whilst no explicit references were made in the included studies to validity, implicit attempts to assess validity were made using rank order methods to compare for example metrics across age groups (Barnabe et al. 2016; Folgado Duarte, Fernandes & Sampaio 2014a; Olthof et al. 2015; Palucci et al. 2018) with the reasonable assumption that older players would demonstrate more effective collective behaviours. For example, Barnabe et al. (2016) identified significant differences in the surface area between U16, U17, and U19 teams. Similarly, Olthof et al. (2015) found significant differences in U17 and U19s lateral stretch index. However, for collective behaviour approaches to be more widely used by practitioners there is need for future research to establish the sensitivity of approaches to distinguish between strong and weak teams within leagues or distinguish between good and bad performances within a single team.

In addition to rank order methods, included studies also implicitly investigated validity by assessing whether approaches to assessing collective behaviour demonstrated longitudinal patterns that would be expected with regards to fatigue or increased experience. Goncalves et al. (2018a) observed the variation in teammate dyad synchronisation over 51 matches. Synchronisation of players increased when walking and decreased when jogging and running as the match progressed. Moreover, coefficient of variation values identified greater variation in jogging and running synchronisation as matches progressed which the authors attributed to mental fatigue. In the context of increased experience, Folgado, Duarte, Fernandes & Sampaio (2018a) investigated synchronous behaviours of a professional football team from the beginning of pre-season to the end of pre-season during 9v9 matches.

Analyses identified an increase in synchronisation between the first and last training sessions consistent with the hypothesis that familiarity and greater synchronised behaviours can emerge as team-mates obtain greater experience playing with each other. These and similar results generated highlight that tempo-spatial metrics may be sensitive to a range of external factors relevant to football. However, the results also highlight potential variability in metrics within and between games, which may have to be accounted for by practitioners when profiling data obtained.

A limited number of the studies included in this review also attempted to assess whether spatial-temporal metrics could predict critical events. Moura et al. (2016) applied vector coding to team spread and identified differences between inter-team coordination preceding shots on goal, and defensive tackles. The authors reported that attacking plays that ended in shots on goal presented greater anti-phase patterns in the early stage of the possession and that teams should attempt to present contrary behaviours to their opponent as soon as ball possession is regained. In a more focused aspect of game play, Shafizadeh, Davids, Wheat & Hizan (2016) analysed 1v1 situations between strikers and goalkeepers in the English Premier League. Results demonstrated that interpersonal distance and relative velocity between attacker and goalkeeper in the longitudinal direction influenced the probability of a goal being scored or not. Whist these findings and others may provide initial information regarding collective behaviours preceding important events, a challenge for researchers and practitioners is to determine what football related strategies should be employed when limitations are identified. Given the complexity of sports such as football where multiple confounding factors limit clear explanations (Carpita, Sandri, Simonetto & Zuccolotto 2016), development in this area may require greater integration between metrics and football strategies or identification of key constraints that can be manipulated to alter behaviours.

Based on recognition that collective behaviour analyses must ultimately provide practitioners with information to improve performance through adaptations to training design or match play (Memmert 2017, Sarmento et al. 2018a), clear discussion of practical applications of findings from research is required. Review of the included studies identified that a substantive portion (21.9%) did not make any direct reference to practical applications. Additionally, it was identified that when practical applications were made by authors these were most often broad, reiterating results of the study in slightly different contexts that demonstrated limited actionable qualities. Representative examples of broad practical applications (Table 3.4) included: "Varying individual playing area by manipulating number of players or by manipulating pitch dimension possesses different implications on emergent teams' behavioural patterns. Therefore, this evidence is an important aspect for coaches to consider" (Chung, Carvalho, Casanova & Silva 2019); "The results of this study can provide valuable tools for controlling player organisation on the pitch" (Moura et al. 2013); "The manipulation of informational constraints to shape tactical behaviour may be an asset" (Silva, Vilar, Davids, Araujo & Garganta 2016a); "The phase couplings and other spatial-temporal relations among players and teams reflects their tactical performances and should consequently be considered by coaches in the design and implementation of small-sided games." (Travassos, Vilar, Araujo &

McGarry 2014b). Collectively, examples of broad practical applications identified to practitioners and coaches procedures that may have potential to improve collective behaviours but provided limited information to enact changes in training or competition. Additionally, when combined with metrics and analysis procedures that are not well contextualised with regards to football specific actions or well understood tactics, broad practical applications made by researchers are unlikely to be implemented.

The next most common categorisation of practical applications was moderate where authors linked conclusions to specific aspects of competitive games or training drills. Consistent with the dynamic systems approach, many practical applications categorised as moderate focused on different types of constraints as a means of altering collective behaviours. Examples of manipulating organismic constraints were proposed by Folgado, Bravo, Pereira & Sampaio (2018a) and Figueira, Goncalves, Masiulis & Sampaio (2018) who recommended altering team make up of stronger and weaker opponents in training, with drills featuring stronger opponents tending to increase synchronisation. In contrast, Coutinho et al. (2019b) identified that manipulating environmental constraints in the form of spatial references with additional pitch lines altered collective behaviours. Inclusion of reference lines was found to increase team defensive behaviour with lower approximate entropy values and was recommended for situations where more structured patterns of play were desired (Coutinho et al. 2019b). In contrast, removal of reference lines was found to increase players movement synchronisation which was recommended for developing offensive movement patterns through lower structured playing styles to create instabilities (Coutinho et al. 2019b). Finally, multiple authors identified task constraints that were categorised as moderate practical applications focusing on SSGs. Manipulations of formations (Baptista et al. 2018) and overload situations (Praça et al. 2016, Goncalves et al. 2017a) were identified as strategies to enhance both effective attacking and defensive behaviours. Whilst practical applications identified as moderate provide coaches and practitioners with clearer guidance on potential constraints to manipulate and analysis procedures to adopt, the examples did not provide guidance on values expected and what may represent substantive changes. One of the most consistent conclusions across the research base was that greater synchronisation in spatial-temporal metrics tended to reflect more effective behaviours (Clemente, Santos-Couceiro, Lourenco-Martins, Sousa & Figueiredo 2014c; Clemente, Couceiro, Martins, Mendes & Figueiredo 2014d; Coutinho et al. 2019; Folgado Duarte, Fernandes & Sampaio 2014a; Folgado, Duarte, Fernandes & Sampaio 2018b; Goncalves et al. 2017a). However, recommendations and guidance identifying expected changes in values for different metrics may be required to facilitate greater uptake in practice.

In a small number of instances, it was identified that authors provided specific practical applications that included recommendations of metric values that were linked to relevant training situations and goals. Aguiar et al (2015) recommend that players should be approximately 5-6m from their team centroid during 3-a side SSGs to enhance availability of environmental cues and ability to pass effectively. Similarly, Gonclaves et al. (2018a) recommended that players adopt approximately 12 m^2 effective playing space for actions involving three team mates to enhance availability of environmental cues and control players' positioning while defending.

Additionally, multiple authors provided specific recommendations regarding pitch dimensions to enhance collective behaviours described by distances and positional measures in youths (Castellano et al. 2017) and female (Zubillaga et al, 2013) players. In contrast to researchers providing specific recommendations regarding metric values, it has been suggested that practitioners and coaches set their own values based on their philosophy and tactical approach to matches (Clemente, Couceiro, Martins, Mendes & Figuieredo 2013a). Additionally, it is recognised that given the extent to which the spatial-temporal research base is developing with novel metrics and analysis approaches that many studies are explorative and clear practical applications should not be expected. However, transfer of the approaches from the research domain to practical use will require more specific recommendations linked to football specific concepts, or clearer guidance on procedures that can be used by coaches and practitioners to develop their own values and monitoring processes.

Based on the risk of bias assessment, one of the key weaknesses in the included studies was authors' lack of justifying sample size, with only 5.9% of studies providing justification. The number of matches investigated in the included studies ranged from 1 to 103. Most studies investigated fewer than 10 matches, with similar metrics used for both analysis of 11v11 matches and SSGs. These findings suggest that most studies feature convenience samples rather than identifying likely effect magnitudes and performing analyses to identify sample sizes required to obtain appropriate statistical power. Alternatively, considerations could be made on the data an individual team might have available and justify the sample by grounding the research in a practical context. Broader consideration of populations investigated in the research identified a substantial skew towards male players, with only two studies incorporating females (Tenga, Zubilaga, Caro & Fradua 2015, Zubillaga et al. 2013). It is unclear whether collective behaviour as assessed by spatial-temporal metrics would be different between males and females. Tenga et al. (2015) identified similarities between the playing length and width of both male and female players. However, male players demonstrated higher levels of variation which was suggested to aid in creation of more space and passing opportunities. Further analysis should be conducted to identify clearer differences across a range of metrics between males and females. If clear differences are identified then further investigation in collective behaviour in women's football must be executed to provide gender disaggregated data.

A key area for future investigation that may assist with practitioners adopting assessment of collective behaviours includes addressing the link between competitive matches and training. Matches provide information on team dynamics within the performance context, however, official competitions are relatively fixed, whereas during training sessions, coaches are free to make large changes to constraints in attempts to alter behaviours and generate effective team dynamics. The information obtained during training sessions could then be used to inform strategies during matches and determine whether similar changes to team dynamics emerge. To date, studies have identified that subtle differences in collective behaviour metrics can be obtained by manipulating constraints such as pitch dimensions, number of players and player formations (Aguiar et al. 2015, Castellano, Silva, Usabiaga & Barreira 2016, Castellano et al. 2017, Coutinho et al. 2019b, Frencken, Van der Plaats, Visscher & Lemmink. 2013, Olthof, Frencken & Lemmink. 2019a, Praça et al. 2016, Silva et al. 2014b, Travassos, Goncalves, Marcelino, Monteiro & Sampaio 2014). However, due to the lack of understanding of ideal values for metrics, it is unclear whether these adaptations are desirable. Moreover, understanding how these manipulations translate from training into matches is a further abstraction that at present there is no evidence for.

3.5 Conclusions

There has been a substantial increase in the number of studies investigating collective behaviours using positional data in football over the last decade. Only 23 studies matching these criteria were identified between 2008 to 2013, whereas 62 studies were identified between 2014 to 2019. Additionally, more recent studies have more frequently included numerical relations (16/17), synchronisation (45/52) and predictability (32/37) metrics than studies prior to 2014. Across the included studies many metrics were analysed using a range of approaches producing extensive areas for future research and practitioners to implement. Whilst the research base highlights that collective behaviour analysis through spatial-temporal data may provide unique insights into performance in football, there are limitations and gaps in understanding that currently prevent the widespread use of the approaches in practice. Common limitations acting as barriers to implementation include reliance on purely mathematical descriptions of metrics at the expense of clear conceptual descriptions. Similarly, a lack of detailed practical applications including normative data and clear guidance on how player position and relative movement are best manipulated to simultaneously improve metric values and performance currently limits uptake. Greater conceptual clarity of metrics may be obtained by researchers incorporating the views and playing principles of coaches to align or adjust metrics. This process may enhance coach buy-in and as a result the likelihood of performance analysts conducting collective behaviour analyses as part of their reporting to coaches. In contrast, progressing to the stage where clear practical recommendations can be made is likely to require substantially more research. Important aspects for future research to consider include the assessment of reliability; establishing whether collective behaviour metrics are sensitive enough to explain performance differences between teams and stronger performances within a team; and whether manipulations in training can create changes in collective behaviours and their associated metrics that transfer to competition.

Table 3.1: Summary of study characteristics

Classification	Population type	Frequency
Game type	Official competition	30
	Friendly	3
	Training Game	9
	Small-sided conditioned game	37
	1v1 bouts	6
Playing level	Professional	33
	Youth	38
	Semi-professional	2
	Amateur	9
	Composite	3
Country observed	Australia	1
	Austria	1
	Brazil	8
	England	7
	Finland	1
	Germany	2
	Italy	1
	Multi-national	4
	Netherlands	4
	Portugal	18
	Spain	8
	Switzerland	1
	Unspecified	29

Table 3.2: Summary of study details, findings and applications.

Author	Sample	Metrics/Analysis	Findings	Applications
Aguiar, Goncalves, Botelho, Lemmink & Sampaio 2017	N= 6 small sided games (5v5) with U19 Males	Distances (2) Predictability (2)	Higher **ApEn** scores were related to short sequences of play suggesting higher irregularity.	- Longer sequences of play demonstrated higher regularity than short sequences of play. - Intra-team interactions had comparatively higher regularity compared with inter-team interactions in short sequences, however the reverse was found in long sequences. - Coaches should design game strategies to increase consistent intra-team or inter-team behaviours.
Aguiar, Goncalves, Botelho, Lemmink & Sampaio 2015	N=24 small sided games (2v2/3v3/4v4/5v5) with U19 Males	Distances (3) Predictability (1)	4v4 and 5v5 small sided games resulted in collective behaviours that appeared more regular than 3v3 and 2v2 small sided games.	- For SSGs an increase in players [from 2 to 5] increases rational space occupation and players' positional regularity suggesting training tactical performance requires more players. - For 3-a side SSG distance between players and team centroid should be 5 to 6 m.
Aquino et al. 2016a	N=4 matches of unspecified size with U16 males	Spaces (2)	A trend for increased **surface area** and **team spread** were observed from preseason phase to competition phase	- During pre-season, tactical performance should improve and can be measured by increases in size of surface area and team spread.
Aquino et al. 2016b	Single match unspecified size with U16 males	Spaces (2)	Significant increases in **team spread** and **surface area** were observed during 2nd half of matches.	-
Baptista et al 2020	N=6 8v8 matches with semi-professional males	Spaces (4) Distances (2) Predictability (1)	Application of different formations (particular number of midfielders) significantly alter game dynamics.	- During 8-a-side matches the 4:3:0 formation can be used to increase the spread of players, conversely with the 4:1:2 formation which encourages compactness within the team. A 0:4:3 formation can be used to find balance between this concept of space exploration.
Barnabe, Volossovitch, Duarte, Ferreira & Davids 2016	N=240 bouts of 6v6 with U16/U17/U19 males	Spaces (4) Predictability (4)	Greater **width**, **surface area**, and **attack-defence synchronisation** were observed in older players when attacking.	- Older and more experienced players demonstrated more intricate attacking plays by demonstrating greater use of the width of the pitch and having a larger surface area. - These values can guide coaches to adapt practice constraints to assist developing tactical behaviours in young players. - Coaches and sport scientists can manipulate constraints in matches to improve player organisation specific to age group.
Bartlett, Button, Robins, Dutt-Mazunder & Kennedy 2012	N=10 11v11 matches with professional males	Spaces (4) Distances (2)	Opposition **centroids** appear **synchronised** throughout matches. Clearer coordination patterns are obtained in the longitudinal direction.	- Team centroid metrics and dispersion metrics including surface area, team spread and stretch index do not appear to be sensitive enough for appropriate analysis of soccer team's performance.

Castellano, Alvarez, Figueira, Coutinho & Sampaio 2013	N=6 11v11 matches with professional males	Spaces (1) Predictability (1)	Regularity was observed in teams tactical strategies in attack and defence.	- When playing stronger teams, team length was measured between 40–55m describing direct attacks whereas against weaker teams more elaborate attacks were formed through smaller lengths of 25-40m. Coaches can use this information to develop effective tasks to develop players.
Castellano, Fernandez, Echeazarra, Barreira & Garganta 2017	N=8 7v7 matches with U13/U14 males	Spaces (8) Distances (1) Predictability (1)	**Surface area** and **stretch index** can be increased by creating longer pitches. Distance between **team centroids** were greater in U13 compared with U14.	- Setting up a 60m length pitch for U14 male players can demonstrate similar team dynamics than elite professionals.
Castellano, Silva, Usabiaga & Barreira 2016	N=6 6v6 matches with amateur males	Spaces (3) Distances (1)	Coaches can manipulate constraints to achieve desired player behaviour in training sessions.	-
Chung, Carvalho, Casanova & Silva 2019	N=6 3v3, 4v4, 5v5 matches with Under 15 males	Spaces (2)	Manipulating the number of players in each team can impact the **length** and **width** during small sided games.	- Coaches should consider manipulating the individual playing area by altering field dimension or number of players in a match to achieve the desired co-adaptations.
Clemente, Couceiro, Lourenco Martins, Dias, Mendes, 2013c	N=30 bouts of 1v1 with U18 males	Distances (1) Positions (2)	In 1v1 situations, **speed** and **angular position** of attacker relative to defender are critical to success.	- Increases in speed and angular variation are required to unbalance an attacker defender dyad to help create scoring opportunities in 1v1 situations. - Coaches can provide instructions to constrain players decision making processes and alter co-adaptive behaviours.
Clemente, Couceiro, Lourenco Martins, Mendes & Figueiredo 2015	Single 7v7 match with U13 males	Numerical relations(1) Predictability (1)	There are high levels of variability in the centre of the field when observing **numerical advantage**.	- Defending teams should have more players in the central zone in the defensive area to secure that area and make it tougher for opponents to score. - High levels of variability in the central areas of the park are used to destabilise the opponents to create goal scoring opportunities.
Clemente, Couceiro, Martins, Mendes & Figueiredo 2013a	N=3 11v11 matches with professional males	Spaces (3) Distances (1)	Fatigue influences collective organisation and causes **surface area** and **stretch index** to decrease in 2^{nd} half.	- Higher values in surface area, stretch index and effective area of play can potentially be linked with better performance as it suggested lower values are found in phases where teams have become fatigued. - Coaches should monitor these values in regular training practices and adapt tasks to manifest effective behavioural habits.
Clemente, Couceiro,	Single 7v7 match with U13 males	Spaces (2) Distances (2)	Attacking teams attempt to expand **effective play**	- Metrics such as surface area and stretch index can indicate how expansive or contracted teams want to be in certain situations. Coaches should set their own limits

Martins, Mendes & Figueiredo 2013b			**area**, whereas, defending teams look to decrease the metric.	for these aspects and other measurable factors such as optimal distance in defence triangulations based on their own philosophy.
Clemente, Couceiro, Martins, Mendes & Figueiredo 2014c	N=3 11v11 matches with professional males	Spaces (3) Positions (1)	**Weighted stretch index** and **surface area** are greater when teams are losing.	- Coaches can improve synchronisation of the team by observing variables including team centroid, dispersion and triangulations formed. - Opponent coaches can observe this data to identify strengths and weaknesses and adjust team tactics to exploit certain behaviours.
Clemente, Couceiro, Martins & Mendes 2013d	Single 7v7 match with U13 males	Spaces (2) Distances (3)	An inverse relationship exists between opposing team's **effective area of play**.	- Effective area of play can be used to measure team contraction and expansion and a system can be created to provide online feedback to coaches to quantify these principles.
Clemente, Couceiro & Martins 2012	Single 7v7 match with U13 males	Spaces (1)	Metrics quantifying collective behaviours can be generated live and used to improve decision making during training sessions.	- A visualisation strategy can be used to help coaches understand tactical changes in the team's behaviour. -The coach can use this information to manipulate the players behaviour in real time.
Clemente, Martins, Couceiro, Mendes & Figueirido 2014a	N=3 11v11 matches with professional males	Numerical relations(3)	Novel variables including **penetration ratio, offensive space ratio** and **unity ratio** were associated with offensive success.	- These novel variables have potential to provide real time information to coaches and augment their perception of the game. - Further improvements in augmented reality could provide improved visualisation techniques for coaches and players to learn from.
Clemente, Martins, Couceiro Mendes & Figueirido 2014b	N=3 11v11 matches with professional males	Numerical relations(4)	novel metrics of **Cover in vigilance** and **Depth mobility** demonstrated high regularity and success	- Cover in support or in vigilance occurred at a mean ratio of 0.78, which are suggested as important principles of play in football. - Coaches can measure the regularity of teams performing cover during training sessions and matches and additionally alter sessions to improve playing principles.
Clemente, Martins, Couceiro, Mendes & Figueiredo 2016a	N=3 11v11 matches with professional males	Spaces (4)	Novel metrics such as **defensive play area** which can be used to describe the organisation of the defence in different areas on the pitch	- Higher number of defensive triangulations in the midfield area are associated with positive match results.
Clemente et al. 2018b	N=2 11v11 30 min matches with amateur males	Spaces (1)	Use of full size pitch emphasises mobility, whereas, half-pitch	- Using a half-sized pitch can result in an increase of space exploration by players on the pitch. This information can be sent on to a coach mid game and help inform the coach in their decision making processes.

			emphasises teams working as a unit.	
Clemente, Santos-Couceiro, Lourenco-Martins, Sousa & Figueiredo 2014d	Single 11v11 match with professional males	Distances (3)	Opposing **team centroids** were synchronised in lateral and longitudinal directions. Additionally, successful teams maintained greater **team centroid** distance in defence.	- By adjusting relationships there can be in an increase in metric synchronisation such as weighted team centroid which is stated as being desirable for improved performance. - Variables can be used to measure tactical performance.
Clemente, Sequiros, Correia, Silva & Lourenco Martins 2018a	N=2 11v11 30 min matches with amateur males	Spaces (3) Distances (2)	Alteration of pitch size has significant effects on decision making and collective organisation as a team.	- Full size pitches are recommended to be used as a means of developing the team tactical principle of mobility.
Coutinho et al. 2019a	N=16 6v6 matches with U13/U15 males	Spaces (1) Synchronisation (2)	Player decision is influenced by pitch length.	- Functional behaviours such as better positional decisions might be best developed using varied pitch configurations.
Coutinho et al. 2019b	N=6 7v7 matches with U15 males	Spaces (1) Distances (1) Predictability (1) Synchronisation (2)	Spatial references cause teams to organise themselves in more regular but less **synchronous** structures.	- Spatial references can be used to improve the regularity of teams positioning which is suggested to help develop defensive organisation. - Removal of spatial references can help increase the synchronised behaviour between players which is suggested to improve the attacking threat of teams.
De Souza et al. 2018	N=4 11v11 matches with professional males	Spaces (2) Distances (1) Numerical relations(1)	Attacking sequences commonly involve the interaction of a large number of players.	- Guidelines for coaches creating 6v6 SSGs should lay out pitches with length 23m in length and 44m in width. This should motivate players to invade space. After an invasion the distance to the goal should be 33m.
Duarte et al. 2013b	Single 11v11 match with professional males	Synchronisation (1)	Both teams demonstrated high **synchrony** that had mutual influence over the others collective behaviour.	-
Duarte et al. 2012a	N=82 bouts of 1v1 with U13 males	Predictability (1) Synchronisation (1)	Successful attackers demonstrated high **synchrony** during 1v1 situations with the defender.	-
Duarte et al 2012c	N=20 bouts of 3v3 with U13 males	Spaces (1) Positions (1)	During goal scoring opportunities, the distance	-

			between **team centroids** decrease.	
Duarte et al. 2013a	Single 11v11 match with professional males	Spaces (4) Positions (1) Predictability (1)	**Approximate entropy** identifies that teams become more predictable and stable as a match progresses	-
Figueira, Goncalves, Masiulis & Sampaio 2018	N=3 training matches 11v11 with U15 and U17 males	Distances (1) Predictability (1) Synchronisation (1)	Movement **synchronisation** was different between age groups	- Mixing age groups, for example U15 and U17 can lead to potential improvements in performance through increased coordinated behaviours.
Filetti, Ruscello, D'ttavio &Fanelli 2017	N=70 11v11 matches using professional males	Numerical relations(2)	Training should focus more on developing players decision making as opposed to physical fitness	- The technical efficiency index is a strong predictor of team performance and result and each individual player should be monitored to measure improvements in this index. - If a team scores a technical efficiency index of >5 compared to their opponents the likelihood of winning is near 100%.
Folgado, Bravo, Pereira & Sampaio 2018c	N=18 5v5 matches with U15 males	Spaces (2) Distances (2) Synchronisation (2)	Adapting the pitch width and length can be used by coaches to achieve desirable behaviours from players during training.	-
Folgado, Duarte, Fernandes & Sampaio 2014a	N=6 11v11 first half friendly match with professional males	Synchronisation (2)	Players demonstrate higher levels of **synchrony** with team mates when playing stronger opponents	- Playing against stronger opponents in training and in matches increases the synchronisation between teammates which is indicative of improved performance.
Folgado, Duarte, Marques, Goncalves & Sampaio 2018b	N=4 11v11 matches using professional males	Synchronisation (2)	**Synchronisation** is a potential performance indicator to analyse tactical performance	- Successful teams demonstrated higher levels of synchronised behaviour compared with losing teams during matches. - An individual teams synchronisation values can me be measured across matches to provide understanding of team performance.
Folgado Duarte, Marques & Sampaio 2015	N=6 11v11 matches with professional males	Synchronisation (2)	Congested fixture lists can cause a decrease in **synchrony** between team mates.	- The lower synchrony between team-mates during congested fixture lists can be used to identify players who should be rotated. Alternatively, specific training practices could be used to improve synchronised behaviours during congested fixtures.
Folgado, Goncalves & Sampaio 2018a	Unspecified number of 9v9 matches with professional males	Synchronisation (2)	Large sided games can be used to develop **synchronous** behaviours	- More experienced players develop synchronous behaviours quicker compared with less experienced players in meaningful training sessions. - Coaches should monitor tactical development of their team during pre-season.

			with new team mates during pre-season training.	
Folgado, Lemmink, Frencken & Sampaio 2014b	N=6 small sided games (4v4/5v5) with U9/U11/U13 males	Spaces (3) Distances (1)	Applying different constraints to small sided games can influence metrics such as **centroid** and **length per width ratio**.	-
Frencken, De Poel, Visscher & lemmink 2012	Single 11v11 match with professional males	Distances (6)	Variation in distance between **team centroids** is linked to critical events in a match.	-
Frencken, Lemmink, Delleman & Visscher 2011	N=3 5v5 matches with U19 males	Spaces (3) Distances (3)	Movement of **team centroid** is associated with creation and scoring of goals.	- Overlapping centroids may have some bearing on identifying goal scoring opportunities and should be monitored by sports scientists to provide the coach with appropriate tools to set more effective training tasks for the team.
Frencken, Van der Plaats, Visscher & Lemmink 2013	N=4 5v5 matches with amateur males	Spaces (2) Distances (2)	Altering pitch length and width in small sided games influences distance between **team centroids.**	- Coaches should carefully think about setting out SSGs as pitch dimensions can vary team interactions significantly.
Frias & Duarte 2014	N=2 6v6 matches with U17 males	Spaces (3) Distances (1)	Man-to-man marking system are more variable in dispersion metrics such as **surface area** and **stretch index** compared with teams that use zonal marking.	- Coaches instructing teams to use zonal marking as a defensive strategy should expect more compact surface area and stretch index, and less variability. Alternatively, a greater and more variable inter-team distance will also be present in the team dynamics.
Goncalves et al. 2019	N=12 11v11 matches using professional males	Spaces (4)	Measuring spatial-temporal patterns can provide guidance for designing more effective training sessions	- An example base structure is provided involving a play area of length 36m and width 48m placed 18m from the goal from which training tasks can be designed around.
Goncalves et al 2018a	N=51 11v11 matches with professional males	Synchronisation (3)	Gradual decrease in **synchronisation** among players occurs throughout a match based potentially on fatigue.	- Coaches should prepare training practice which increase players’ ability to perform under fatigue and mental fatigue
Goncalves et al. 2017a	N=2 11v10 matches with professional males	Distances (2) Predictability (1)	Greater focus on tactics employed can be achieved	- Unrestricting player movement will increase the synchronisation between players in a team.

		Synchronisation (2)	when restricting playing areas for each player.	- Coaches are encouraged to increase the levels in synchronisation in a team.
Goncalves, Figueira, Macas & Sampaio 2014	Single 11v11 50 min match with U19 males	Distances (9) Predictability (9) Synchronisation (6)	**Group centroids** based on position demonstrate high **synchrony**. Midfielders demonstrate the highest **synchrony**.	-
Goncalves et al. 2018b	N=6 11v11 first half matches with professional males	Spaces (1) Predictability (1)	Variation identified in **effective playing space** suggests that coaches should design practices carefully to focus on specific match scenarios.	- Field dimension will impact the environment cues from which players adapt and learn from. - Approximately 12m^2 effective field space is stated as enhancing players perceptions for actions involving 3 teammates
Goncalves, Marcelino, Torres-Ronda, Torrents & Sampaio 2016	N=24 overloaded small sided games (4v3/4v4/4v7) with amateur and professional males	Spaces (1) Distances (3) Predictability (1)	Training sessions that involve overloaded situations emphasise use of local information in the decision making process to develop organised behaviours.	- SSGs with unbalanced situations such as a 7v4 can help players develop organisation behaviours and help players learn defensive principles of play including concentration and balance.
Headrick et al. 2012	N=144 bouts of 1v1 with U17 males	Distances (2)	During 1v1 situations, the proximity to goal influences decision making of both attacker and defender.	- During 1v1 bouts, practices should be designed with an intention of keeping the defender to ball distance at 2m which is stated as a critical distance which a stable state of the defender to ball distance appeared.
Laakso, Davids, Liukkonen & Travassos 2019	N=142 trials of 2v1 with U15 males	Distances (1) Positions (2)	Pitch location can constrain **interpersonal coordination** tendencies in 2v1 situations	-
Laasko, Travassos, Liukkonen & Davids 2017	N=129 bouts of 1v1 with U15 males	Distances (1) Positions (1)	Location on the pitch and favoured foot are constraints that can influence **interpersonal coordination.**	- Attackers and defenders can be provided practices on varying parts of the pitch to achieve specific behaviour developments to exploit space or nullify use of a stronger foot based on the position related to the goal.
Leser et al. 2019	N=279 trials of 3v2 with U16 males	Distances (2) Positions (2)	Sequences that were played wide and close to the goal line appeared to	-

			have more chance of resulting in a goal	
Leser et al. 2015	N=75 trials of 1v1 with U15 males	Distances (1) Positions (2)	Spatial-temporal processes have a high potential in providing useful information to coaches	- The acceleration of an attacker when attempting to overcome a defender is the biggest indicator of success in a 1v1 bout, the higher the acceleration, the more likely a successful dribble will be completed. Therefore, coaches should attempt to improve this attribute in players attempting 1v1s.
Lopez-Felip, Davis, Frank & Dixon 2019	Single 11v11 friendly match using professional males	Distances (1) Synchronisation (1)	**Distance to centroid** and **Synchrony** are both appropriate to be used when modelling the dynamics of a soccer match	-
Low et al 2018	Single 10v11 training match using professional males	Spaces (2) Distances (3) Predictability (3) Synchronisation (2)	Applying a high press strategy resulted in smaller **distances between opponent centroid** and higher regularity in **player distances to centroid**	- Coaches are recommended to use a high press strategy during matches where they have a numerical superiority as this will result in higher regularity in tactical patterns.
Memmert, raabe, Scwab & Rein 2017	N=144 bouts of 11v11 with amateur males	Spaces (2) Distances (1) Numerical relations(2)	Changing formation in match like practices can influence metrics such as **length per width ratio**.	- The 3-5-2 formation demonstrated higher pressure passing efficiency which suggests a stronger through ball passing ability compared to a 4-2-3-1 which can help support decision making.
Menuchi, Moro, Ambrosio, Pariente & Araujo 2018	N=4 4v1 possession games with U13/U15/U17/U19 males	Distances (5) Positions (2)	The coupling of players during a rondo game are constrained by the age of participants and the spatial-temporal positioning of the participants	- The marking coupling can be influenced by the playing space of the rondo game. This can help coaches guide players to harmonious collective behaviours from their players.
Moura, Barreto Martins, Anido, Leite De Barros & Cunha 2012	N=8 11v11 matches with professional males	Spaces (2)	Higher values of **team spread** and **surface area** were found in attacking teams compared to defending teams.	-
Moura et al. 2013	N=10 11v11 matches with professional males	Spaces (2)	**Surface area** and **team spread** demonstrate a low frequency time series that decreased during the second half of matches.	- Surface area and team spread are valuable tools that can be observed and controlled by coaches in training and competition.
Moura et al. 2016	N=10 11v11 matches with professional males	Spaces (1) Synchronisation (1)	Teams that were slow to **synchronise** their **team**	- Coaches should focus on aspects of play that occur directly after a transition.

			spread with their opponents were more likely to concede a goal scoring chance.	- Higher levels of anti-phase and attacking team coordination in team spread were found directly after transitions, so coaches should seek to increase these values after winning the ball to improve likelihood of scoring.
Olthof, frencken & lemmink 2015	N=24 5v5 matches with U17/U19 males	Spaces (3) Distances (2)	Older age groups demonstrate greater **width** compared to younger age groups.	-
Olthof et al. 2019a	N=90 5v5 matches with U13/U15/U17/U19 males	Spaces (4) Distances (2)	Playing space relative to match sizes influences the tactical performances with greater space to explore resulting in alternative collective behaviours.	- Match derived pitch areas might be an appropriate constraint to develop players in a representative learning environment.
Olthof et al. 2019b	N=20 11v11 matches 5 competitive, 15 training using U17 and U19 males	Spaces (6)	Training games demonstrate small tactical differences to competitive matches	- Coaches are recommended to use 11v11 matches in training to replicate as closely as possible the tactical behaviour of competitive matches.
Palucci Vieira et al. 2018.	N=12 11v11 matches with U11/U13/U15/U17/U20 and professional males	Spaces (2)	Older age groups demonstrate greater **surface area** and **team spread** compared to younger age groups.	-
Praca, Folgado, de Andrade & Greco 2016	N=36 split evenly between 3v3, 4v3 and 3v3+2 with U17 males	Spaces (3) Distances (1)	3v3+2 small sided game resulted in a significant increase in width	- 3v3+2 SSGs appear to be useful for direct and counterattacking teams compared with the 4v3 condition which is recommended for teams who create chances through lateral movement.
Rein, Raabe & Memmert 2017	N=103 11v11 matches with professional males	Spaces (1) Numerical relations (2)	There appears to be a positive relation between team performance with **outplayed defenders** and **space dominance**	- An increase in space dominance or outplayed defenders could improve team performance in matches. - Coaches can use these variables in preparation for matches as well as post-match analysis and feedback.
Ric, Torrents, Goncalves, Sampaio & Hristovski 2017	Single 11v11 friendly match using professional males	Spaces (4) Positions (2)	Measuring collective behaviour can help coaches observe if teams are developing over time	-
Sampaio, Lago, Goncalves, Macas, & leite 2014	N=28 5v5 matches with amateur males	Distances (1) Predictability (1)	Variables such as match status and game speed can influence emergent team behaviour and situations involving overloads.	- Coaches should manipulate game pace, match score and numerical superiority to achieve desired co-adaptive behaviours from their players and use measures such as distance of player to team centroid to evaluate team behaviour.

Sampaio & Macas 2012	N=2 6v6 matches with amateur males	Distances (5) Predictability (1) Synchronisation (1)	Spatial-temporal data can be used as a collective to measure tactical patterns.	- More regular patterns of player distance to centroid when travelling slower than 13 km/h was identified as the biggest predictor after intervention suggesting this was indicative of better team tactical performance. Providing this information can improve the decision making of coaches.
Santos et al. 2018	N=8 6v6 matches using U13 and U15 males	Distances (3) Predictability (3)	A differential learning approach resulted in more regularity in players positioning	- Coaches apply differential learning variations in SSGs with young players to develop creative and tactically efficient behaviour.
Santos, Lagos-Penas & Garcia-Garcia 2017	N=13 11v11 matches using professional males	Distances (2)	Changing match conditions results in players adapting their behaviour and decision making	- Coaches can set objectives for players in matches and practice sessions based on potential changes in the organisation of defending teams.
Serra-Olivares, Garcia Lopez & Goncalves 2019	N=22 small sided games (7v7/8v8) using U11 and U12 males	Spaces (6) Distances (6)	Older players were more able to organise themselves tactically to make use of the available space effectively	- Eight-a-side games are recommended for older and more experienced players while seven-a-side games are recommended for younger and less experienced players.
Shafizadeh, Davids, Correia, Wheat & Hizan 2016	N=42 bouts of 1v1 with professional males	Positions (2)	Goalkeepers use informational constraints of **interpersonal distance** in 1v1 situations to intercept attackers.	- Learning goalkeepers and attackers should attempt to identify functional perception-action couplings in 1v1, task constraints can be manipulated to help with the learning process.
Siegle & Lames	Single 11v11 match with professional males	Synchronisation (4)	**Relative phase** analysis shows potential to identify valuable information during matches.	-
Silva et al. 2016b	N=13 11v11 30 min matches with amateur males	Synchronisation (3)	Synergistic behaviour is developed and enhanced through repeated practice.	- Higher levels of coordinated behaviour, both in phase and anti-phase are suggested to be increased to improve performance.
Silva et al 2014a	N=6 small sided games (5v5/5v4/5v3) using regional and national standard under 19 males	Spaces (2) Distances (1) Predictability (1) Synchronisation (2)	National standard players adapt to constraints in small sided conditioned games differently to regional players. Identified that national standard players explore more of the available space.	- Field dimensions and other constraints can be intentionally manipulated by coaches to achieve desired team tactical organisation in training sessions to develop performance.

Silva et al 2014b	N=6 small sided games (5v5/5v4/5v3) with U19 males	Spaces (1) Distances (5)	Small sided games can be constrained by overloads and conditions to manipulate player and team coordination.	- Numerical superiority is a constraint that coaches can manipulate to alter the behaviour and development of their team.
Silva, Vilar, Davids, Araujo, Garganta 2016a	N=3 small sided games (3v3/4v4/5v5) with U15 males	Spaces (3) Distances (1) Positions (2) Synchronisation (2)	By altering the ratios of players in attack and defence, desired characteristics of play can be achieved through the space and time players have to make decisions.	- A time delay of 0.5s or more in centroid synchronisation could be enough to destabilise opponents and create dangerous opportunities. This information can be used by practitioners to adapt task constraints to achieve specific behavioural adaptations.
Steiner, Rauh, Rumo, Sonderegger & Seiler 2019	N=5 11v11 first halves of matches using U18 males	Positions (1) Numerical relations(1)	Positional data can augment the understanding of passing decision making	- understanding and evaluating situations in which riskier passes should and should not be made can be used to understand good and bad decision making of players at a deeper level.
Tenga, Zubilaga, Caro & Fradua 2015	N=8 11v11 matches with males and females	Spaces (2)	There were key similarities in patterns of play in male and female matches. However, male patterns of play demonstrated greater variability.	- Tactical patterns from game specific structures can be used to develop collective tactical organisation by coaches.
Travassos, Goncalves, Marcelino, Monteiro & Sampaio 2014a	N=12 small sided games (5v5/5v4) with amateur males	Spaces (2) Distance (1)	Small sided games with either equal or unequal numbers influence collective behaviours.	- Coaches should consider using relative phase couplings and other spatial-temporal measures such as surface area and team centroid when designing practices.
Travassos, Vilar, Araujo & McGarry 2014b	N=8 5v5 matches with professional males	Spaces (3) Distances (3) Synchronisation (8)	Increasing the number of goals in a conditioned game increases the distance between **team centroids**	- Using conditioned games with multiple small goals can increase the perception and breadth of attention from players altering their tactical performance.
Vilar, Araujo, Davids & Bar-Yam 2013	Single 11v11 match with professional males	Numerical relations(1) Predictability (1)	**Numerical advantage** influences team stability/instability and collective behaviours that emerge.	- Practitioners can use numerical advantage to understand team performance quantitatively.
Zubilaga et al. 2013	N=4 11v11 matches with semi-professional females	Spaces (3) Distances (4)	Team **width** decreases and team **length** increases as the ball moves closer to either goal.	- The individual player area for a female footballer in practice sessions should not exceed 110m^2

Table 3.3: Research quality evaluation

Quality Item	Success rate
A clearly stated study purpose	100%
Relevant background literature used	100%
Appropriate design for the research question	98.8%
Detailed reporting of the sample size	96.5%
Justification of the sample size	5.9%
Informed consent or ethical permission	87.1%
Detailed reporting of methodology	97.6%
Used inferential statistics related to the aim	89.4%
Appropriate analysis methods considering the study aim	94.1%
Appropriate conclusions stated relating to study methods	96.5%
Stated practical applications derived from study results	63.5%
Acknowledged and described study limitations	37.7%

Table 3.4: Summary of practical applications identified in included studies

Application type	Author	Year	Representative quotes
Broad (generic conclusions providing limited direct applicability)	**Chung**	2019	"this evidence is an important aspect for coaches to consider when planning SSCG tasks, since manipulation of [individual playing area] through different constraints manipulation (i.e., pitch dimension or number of players) promote different contextual information as well as new affordances.
	Clemente	2014c	"players should be repositioned to ensure the in-phase relationship and adjust the distance between the centroids."
	Coutinho	2019a	"using different pitch configurations might help players to improve their ability to identify the most relevant cues that support the emergence of functional behaviours."
	Frencken	2013	"Coaches must carefully choose the type of small-sided game in training, as interaction patterns vary depending on pitch dimensions."

	Goncalves	2018b	"Coaches should prepare physical and mental fatiguing practice tasks to increase players ability to adapt and perform under these scenarios."
	Moura	2012	"Automatic tracking methods during training sessions allow coaches to calculate the same variables proposed in the present research, and based on this information, they can precisely control their players' organisation on the pitch and systematise tactical strategies"
Moderate (conclusions linked to specific game or training aspects providing some direct applicability)	**Castellano**	2013	"The surface area may help to explain the defending flow"
	Clemente	2015d	"can be a useful information to coaches in order to control the superiority or inferiority zones, reorganizing a team's strategies according to its weaknesses or strengths
	Clemente	2013b	"the speed and angular positioning of the attacker are key factors when trying to unbalance the attacker-defender dyad."
	Folgado	2014a	"Selecting stronger opponents for matches during the pre-season seems to promote more synchronized behaviors between players."
	Siegle	2013	"perturbations can be used to identify playing situations in which one team attacked in a way, which the defending team was not able to answer."
	Vilar	2013	"This method captures how teams explored different regions to maintain backward stability and create forward instability, in accordance with the shape and location of the area of play."

Specific (clear recommendations with specific reference to the use of a metric or analysis method providing direct applicability)	**Aguiar**	2015	"For example, in a 3-a side SSG, these distances [player to team centroid] should be around 5 to 6 m and, therefore, require the optimisation from the focus on environmental cues, passing performances and explosive strength and power within these limits."
	De Souza	2018	"As for practical recommendations of our paper, coaches may create, for instance, 6 × 6 SSGs with the spaces presented in the study: about 23 m in length and 44 m in width, with the objective of motivating players to invade the space."
	Headrick	2012	"player-to-ball relationships can be used to design practice tasks by positioning the players and ball within critical distances of each other. For example, a practice game could be designed with a D-Ball distance of 2m, representing the range at which the stable state of D-Ball distance appeared in this study."

CHAPTER 4

Network Metrics to Assess Collective Behaviour in Football: A Systematic Review

Prelude

This chapter largely follows the manuscript accepted for presentation at the 8th International Workshop and Conference of the International Society of Performance Analysis of Sport. The work was conducted in parallel with the systematic review presented in chapter 3 and included similar methods. In this chapter a summary of research applying social network analysis to passing behaviour in football is conducted. An appraisal of the study quality and author quotes relating to the application and conceptual clarity of network analysis metrics in football.

4.1 Introduction

The collection and intelligent use of data has become a central focus in many disciplines to improve performance. This phenomenon has now spread to sport where coaches can be presented data from a range of different sources to inform their decision making (Nicholls et al 2019). In elite team sports, standard data generating sources include time motion analysis, individual player fitness, health and wellbeing monitoring. Notation analysis is also commonly applied in sport to identify technical-tactical aspects of individuals and teams. Analysts can apply this information by providing coaches with objective information that identifies strengths and weaknesses. Although, within sports, there appears no consensus on how to use these key performance indicators. In team sports such as football, performance is highly complex and involves the continuous interaction of many groups and subgroups of players (Ramos et al 2017). These complex interactions are also compounded by the range of contextual variables such as opponent strength, match location, and match score (Mackenzie and Cushion 2013).

Historically, coaches in team sports have relied on video analysis as the primary data generating source to better understand opponents and their own team which subsequently inform decisions. The progression of software packages within video analysis has allowed analysts to provide immediate video feedback to coaches of match events. This provides the coach with a more reliable observation of team behaviour (O'Donoghue and Mayes 2013). Furthermore, video clips can also be useful in supplying feedback to players. However, this method of analysis is highly intuitive and is at risk of observer bias through the coach cherry picking instances which support their outlook, and selection bias of observing players during key instances (O'Donoghue and Mayes 2013). Additionally, football is a particularly complex team sport with up to 22 individuals interacting with each other simultaneously through a range of technical actions including dribbling, shooting, tackling, off-ball movements and set pieces. In addition to the potential biases that are present with video analysis, there are challenges in summarising and presenting the complexity of collective team performance. There is a desire for coaches to understand how players link with each other to create effective and cohesive team principles which leads to success.

Recently, there has been a surge in research applying social network analysis to data collected during football matches and training (Ramos et al 2018; Ribeiro et al 2017; Mclean et al 2018; Castellano et al 2019). Social network analysis applies the mathematical principles of graph theory. In graph theory, pairwise connections between nodes are observed (Mclean et al 2018). Translated to football, the nodes are representative of players, and the connections between them, which can also be called arcs, or vertices, are representative of the successful passes between players (Clemente et al 2014e). This alternative quantitative analysis method has potential to create objective information helping coaches to inform decision making. The conceptual underpinnings of social network analysis are derived from dynamic systems theory. Previous literature has stated that dynamic systems theory offers a framework from which performance analysts can use to better investigate the complex game of football (Ramos et al 2017). The passing patterns which are self-organised, emergent from the team and demonstrates interpersonal coordination fit well into the dynamic systems theory framework (Glazier 2010). The networks are usually constructed over a predetermined period, usually a match and identify which players are connected through passing, how often are they connected, and in which direction. Such networks identified in match play demonstrate the strategic approaches used by teams in an attempt to overcome their opponents through offensive manoeuvres. Patterns can be established from this self-organised behaviour of passing to team mates. Whilst communication between two team mates is not limited to passing (Ribeiro et al 2017), this form of communication is the central mechanism studied at present. Nevertheless, passing networks can still display useful patterns which represent either positive or negative characteristics of a team that a coach can use to tailor practice to improve team performance. The methodology used for social network analysis does appear to highlight a principal weakness of the approach, as the network describes the match as a cross-section in time. However, the flow of a match changes regularly through score changes, tactical alterations, and substitutions. As such, features of the networks within individual phases of play are not fully evident.

The literature base applying social network analysis to football is rapidly growing and there is a need to appraise the emerging research, both in terms of study quality, and the metrics used. Therefore, the purpose of this research was to conduct a systematic review of published articles investigating network metrics in football and in the process, organise common approaches and synthesise their findings. To assess the practical relevance and accessibility of the metrics investigated, author's descriptions and comments regarding potential application of each metric were recorded and synthesised. Finally, the overall methodological quality of the research base was appraised.

4.2 Method

A systematic review of peer reviewed studies applying social network analysis to offensive phases of football was implemented conforming to the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA). Five electronic databases including Embase, Medline, Scopus, SportDiscuss, and Web of Science were used to perform the search strategy. A two level strategy was employed, where the initial level contained the following social network search terms combined with the Boolean operator OR: 'indegree', 'outdegree', 'closeness centrality', 'betweenness centrality', 'betweeness centrality', 'eigenvector', 'network density', 'clustering coefficient', 'total links', 'degree centrality', 'degree prestige', 'page rank', 'network intensity', 'network centralisation', 'cohesion', 'sociometric status'. The second level search terms related to the sport and included 'football' OR 'soccer; The output from two levels were amalgamated with the AND Boolean function. The search was limited to journals published between the 1st of January 2008 and the 20th of February 2019, the date of the search.

Criteria for inclusion in the review comprised (1) generation of passing data from 11-a-side football matches or small sided games; (2) application of social network analyses to the data; (3) full publication accessible in English. Results from conference abstracts or books were excluded

Articles were screened separately by the reviewers (MC, TC) following a 3-stage process. Firstly, article titles were reviewed, then abstracts. The remaining articles were read in full and the inclusion and exclusion criteria were applied. Any disagreements were resolved through discussion between the reviewers at the abstract and full text stage. A 3-step data extraction process was conducted on the articles included in the review. The first step centred on general article information including, population investigated, sample size, and the primary conclusion of the article. The second step involved the extraction of quotes directly from the publication which centred on metric description, application and validity. In cases where multiple quotes were relevant, every quote was recorded. This extraction phase also documented the equation stated in the article relating to each metric to identify if metrics were calculated using alternative formulas. The final extraction phase applied a 12-item checklist adapted from Sarmento et al (2018b) to assess study quality. The items were based on study purpose (1), background literature (2), study design (3), detail of sample used (4), justification of sample size (5), ethical approval (6), method detail (7), applied inferential statistics (8), relevant analysis methods, (9), appropriate conclusions (10), practical applications (11), and acknowledged limitation (12). Each item was scored using a binary scale with positive items awarded a unit score. Article quality was classified using 3 groups. Low (<50%), moderate (>50%, ≤75%) and, high (>75%) quality studies were based on the percentage of successful items identified in the publication (Sarmento et al 2018b). The extraction process was executed twice collectively by three reviewers (MC, NB, TC), and a final discussion amongst the full team was performed to ensure accuracy.

4.3 Results

A total of 1978 articles were identified across the five electronic databases. Duplicates were removed reducing the number to 1170. The remaining articles were screened through the title and abstract leaving 48 articles that were obtained in full text. The articles were read in full and data extraction was completed on them. A further nine articles were removed due to not meeting the requirements of including data from football competition or training (5), being published in a book or a conference abstract (2), not applying network analysis (2).

Of the 38 studies included in the review, 30 comprised analyses of elite adult football matches. The remaining studies included analyses on data collected from youth male footballers (5), both adult and youth male footballers (2), or adult amateur footballers (1). The majority of studies analysed data from competitive matches (34) with a limited sample (4) focussing on data obtained during training. The domestic leagues analysed included Australia (1) Brazil (1), England (4), Portugal (12), and Spain (1). Also, publications comprised of the other competitions including the UEFA Champions league (6), the FIFA World Cup (7) and international tournaments including the UEFA European Championship and Copa America (2). The number of matches used in articles ranged from 1 to 760 with mean ± sd of 108±208 games analysed. 35 studies performed network analyses across all events in a game; whereas, 3 studies only considered network metrics preceding a goal or shot.

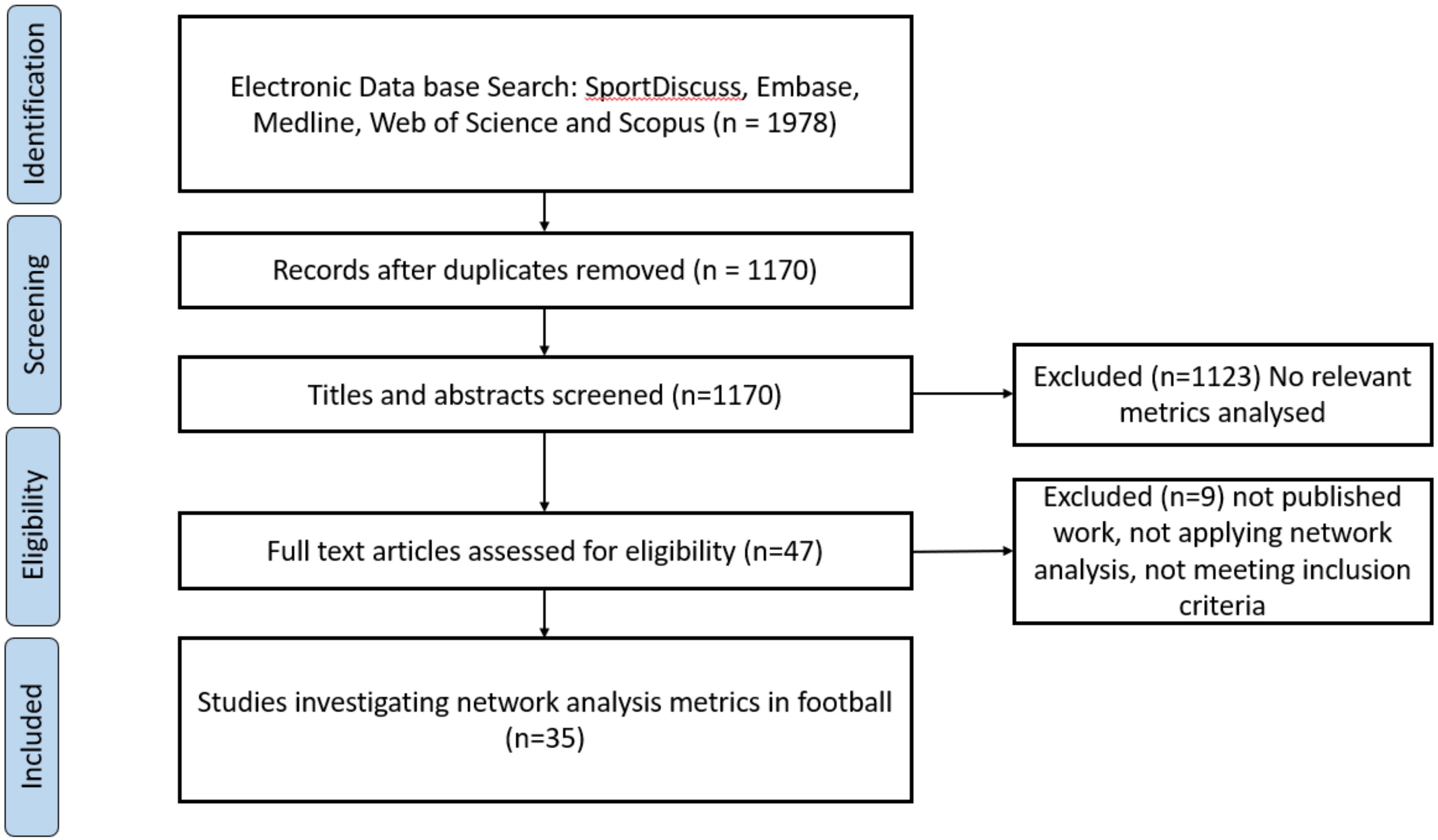


Figure 4.1 PRISMA diagram for collection of studies including network analysis in football

The included studies featured 37 distinct network metrics across a total sample of 154 reporting instances. An equation or clear algorithm describing the metric was presented on 93% of studies of occasions. Similarly, 89% of metrics were accompanied with a descriptive or conceptual statement conveying the relevance of the metric. In contrast, statements regarding practical applications of a specific metric were only stated in 14% of cases. Moreover, validity of metrics was only established in 8% of cases.

Analysis of quotes regarding the relevance or conceptual underpinning of metrics used identified 2 broad themes including 1) individual network metrics describing characteristics of player or node within the network; 2) team network metrics describing the overall behaviour of the team (table 4.1). Individual network metrics were observed most frequently (93 instances used across 21 separate metrics) and included metrics such as betweenness centrality, degree prestige, closeness centrality, and degree centrality to describe a player's performance and their importance to the team. For example, representative quotes from authors included [betweenness centrality] "represents how the ball-flow between other players depends on that particular player" (Castellano et al 2019); [indegree centrality] "measures the prestige of a player" (Clemente et al 2017a) and [clustering coefficient] "calculates the capacity of a player to promote clusters or union in a team" (Clemente et al 2017b). Team network metrics were observed with a similar frequency as individual network metrics (52 instances across 16 metrics), and generally described how a team communicated with each other and organised their attacking phases of play. Representative examples included network density and heterogeneity which were explained as [network density] “measures the overall affection between teammates”, and [network heterogeneity] “is closely related to the variation of connectivity across players” (Clemente et al 2015e).

Individual network metrics	**Team network metrics**
Betweenness centrality	Arc reciprocity
Breadth	Average path length
Centrality	Clique
Centroid	Cohesion
Closeness centrality	Compactness
clustering coefficient	Consecutive passes
Degree centrality	Network density
Degree prestige	Heterogeneity
Egonet density	Network centralisation
Eigenvector	Network diameter
Geodetic distance	Network experience
Global rank	Network intensity
In degree	Number of connections
Individual sociometric status	Overlapping cliques
Out degree	Total links
Page rank	
Scaled connectivity	
Sociometric status	
Dependence	
Within centrality	
Clustering coefficient	

Validity of metrics were assessed in a limited number of studies (8%) and generally took the form of predictive validity. Multiple authors quantified relationships between a particular network metric and goals scored, shots attempted successful attacks, and match result. However, across studies there were examples of contrasting findings with for example both an increase and a decrease in total links and network density related to a positive outcome measure. In 2 study's authors assessed rank order validity by identifying whether the metrics could differentiate between more skilled players (Praca et al 2018a) and the tournament elimination stage of a team (Clemente et al 2015b).

The assessment of research quality was varied with 3 studies categorised as low quality, 23 categorised as moderate quality and 11 studies categorised as high quality. The most common limitation in research quality included a lack of justification regarding sample size, with only 8% of studies providing justification. Less frequent limitations of research quality included a lack of acknowledging limitations (72%), lack of relevant inferential statistics (72%) and a lack of detail in the method (78%). Forty eight percent of studies included a section on practical applications. Many of these suggest that increasing a specific metric should lead to an improvement in performance. For example, "higher values identify a better homogeneity of interactions between players of the same team; this may be related to team success" (Aquino et al 2018), [total links] "may also be correlated to offensive success" (Peixoto et al 2017). When speaking about individual network metrics, characteristics of players could be identified and interpreted as a practical application [network centralisation] "a value closer to 1 suggests that one player can be the 'master' of the team, being the playmaker" (Clemente et al 2015), [outdegree] "players with larger ODC scores are those who contributed more to their team's offensive attempts" (Clemente et al 2015c).

4.4 Discussion

The populations analysed were largely performing at a very high level. Teams composed of professional players were used in 79% of the studies. This highlights the limited publications which use youth or amateur players. However, professional teams are more likely to have access to athlete data such as this. Therefore, identifying practical applications to these groups are perhaps more realistic. However, no articles were found to have performed a network analysis on female players. Women's football is a growing sport and applying social network analysis to their matches may provide more specific insight into the collective behaviour of females playing football. As expected, official matches were almost always used. These were regularly observed in national league competitions, or elite tournaments such as the World Cup and the Champions League.

An alternative approach to this was to analyse passing networks in training matches These types of observations were conducted 4 times across the literature base. This posits an interesting question about how training matches may alter the network patterns which emerge from the play. Indeed, coaches may apply constraints or alterations to training matches to manipulate the networks which form through the matches. Coaches altering the pitch size, the number of players on each team, and the number of goals may subtly change the team dynamics. However, the few studies exploring these adaptations have only scratched the surface and an extensive research base would be required to understand this phenomenon.

There was a wide variety of sample sizes used when considering the publications which observed matches. Indeed, the sample selected was seldom justified which suggests that researchers used a convenient sample size. Nevertheless, some still appeared reasonable despite not applying a power calculation to ascertain the desired sample size. For example, several studies considered every match in a league season or a tournament. Conversely, eight studies used 5 or fewer matches, often only considering one team. The conclusions of such studies are unlikely to be generalisable, and as such, should be used with caution. A separate issue pertaining to article sample size is that not all articles considered every passing network in a match. Some articles only considered the networks preceding goals, or shots at goal. This inclusion criteria drastically reduced the number of networks analysed. Therefore, articles that identify the number of attacks used would offer a more transparent methodology.

Regarding the authors description of the network metrics, studies largely appeared in agreement on the underpinning concepts demonstrated. However, Clustering coefficient was usually used as a measure of an individual player. Whereas, on one occasion, the equation was adapted to “measures the level of clustering in the entire network” (Clemente et al 2017b). This may cause confusion for analysts applying these metrics when adapted metrics which represent different constructs share the same name. However, there appears to be conceptual clarity on the network analysis metrics applied in football.

The collected quotes highlight that validity is not fully established in many of the metrics used within the literature. For example, Clemente et al (2017b) identified that “goals scored showed a small positive correlation with network density” (p = 0.003; ES = 0.088). On the other hand, Peixoto et al (2017) identified “Total links and network density were significantly lower in successful than in unsuccessful offensive units” (p = 0.031; d = 0.691). This may be resultant from the alternative populations, with one study analysing shots, the other analysing the network of a whole game. Indeed, the constructs which these metrics have been validated against may be subtly different. One identifies the network characteristics that emerge from teams who score more goals, the other identifies the network characteristics that lead to more successful shots. Collectively, the lack of valid network metrics is perhaps a result of the descriptive nature of publications with few attempting to establish validity through relating it to a reasonable performance measure, or rank order. Conversely, many of these metrics may not be a valid representation of the concepts they are supposed to represent.

The lack of validity may also be linked with the limited practical applications provided in studies. This finding was highlighted in both the author quotes, and the quality appraisal. A case where increasing or decreasing a metric clearly relates to an increase in performance is ideal. However, this is rarely seen, with few studies identifying a clear relationship and in studies that do, the result appears inconsistent (Clemente et al, 2017b; Peixoto et al, 2017). The use of individual network metrics appears to identify differences within teams, instead of between teams. These metrics can be applied to identify key players in the offensive phase. This information may be more useful to opposition coaches in preparing teams to limit the effectiveness of the player.

The quality appraisal study also identified several other weaknesses in the literature base. A large number of studies did not gain consent nor state whether ethical approval was gained when performing the research. This is likely an effect of gathering the data through means such as television observation and not having direct access to the performing athletes. Another weakness identified from the appraisal was that some publications lacked enough methodological detail to repeat the procedure. In fact, it was an occasional occurrence to use metrics in which equations were not present, nor was a direct reference to the equation available. The final recurring weakness in the literature was a lack of study limitations presented by the authors. This could suggest a lack of transparency in the methodology

An interesting issue when applying network analysis, is how substitutions are accounted for. Considering that when a substitution occurs, a new node becomes available, whilst an old node becomes inaccessible. Articles have approached this problem in different ways. For example, some studies stop recording a network after the first substitution has been made (Trequattrini et al, 2015). Others simply add the nodes and continue as normal. An interesting approach by Grund et al (2016) only included the 8 most active players during the game in the network analysis. This is not an issue when networks of pitch zones are used instead of players, however this approach appears to remain in the minority. Another consideration relates to substitutions as coaches may employ a tactical switch which the oncoming player might play in a different position to the previous player. This could impact the true effect of a network. So far, it is unclear how best to accommodate for replacements in network analysis.

4.5 Conclusions

This systematic review has identified that researchers in performance analysis of football currently use a range of network metrics to describe collective behaviour amongst players. The metrics reported can be categorised into individual network metrics or team network metrics. However, there is often inconsistency in calculations used by authors with metrics of the same name. Furthermore, articles fail to present rigorous conceptual definitions or purposefully align the metric with common principles of football that coaches frequently adhere to. Whilst the methodological quality of individual research studies was classified as "moderate quality" and "high quality" for the majority of studies, at present the limitations of the collective evidence base are likely to prevent widespread usage of the metrics by practitioners.

An additional limitation identified in the reviewed literature was a lack of discussion regarding practical applications and provision of clear recommendations. When authors discussed use of metrics, comments generally inferred that performance may be related to either higher or lower values of a specific metric. However, this is rarely seen with only small differences observed between successful and unsuccessful teams. This is potentially a result of the over simplification of applying network analysis to a complex and situationally volatile sport. Future studies should seek to integrate network metrics with either spatially relevant information such as number of opponents outplayed or location of pass. This could assist in grounding this analysis in the way coaches can understand increasing the likelihood of buy in, and also monitoring, understanding, and predicting subtle team dynamics that have significant relationships to performance and outcome.

Table 4.1: Summary of study characteristics

Classification	Population type	Frequency
Game type	International competition	9
	Professional league	21
	Youth league	1
	Amateur league	1
	Small-sided conditioned game	4
	Composite game types	2
Playing level	Professional	30
	Youth	5
	Amateur	1
	Composite	2
Country observed	Australia	1
	Brazil	3
	England	4
	Germany	1
	Multi-national	12
	Not stated	2
	Portugal	12
	Spain	2
	Switzerland	1

Table 4.2: Summary of study details, findings and applications.

Author	Sample	Metrics/Analysis	Findings	Applications
Aquino et al, 2018	N=18 Brazilian 3rd division matches	Indegree, Outdegree, Closeness centrality, betweenness centrality, eigenvector, Density and clustering coefficients	Matches played at home or against "weaker" influenced network analysis and Central/external midfielders reported greater **closeness/ betweenness centrality**, **outdegree**, and **eigenvector** compared with central/external defenders and forward	- Practitioners should account for match opponent and location when designing practices as greater demands are put upon the individuals in a tactical, technical and physical sense when playing at home or against a weaker team
Castellano and Echeazarra 2019	N=36 matches from La Liga	Degree Prestige, Degree Centrality, Betweenness Centrality, Page Rank and Closeness Centrality	Teams demonstrate different tactical approaches based on their passing networks.	- By using Network metrics along with other measures, coaches can design practices that replicate values in these network measures to make practices more specific
Cho, Yoon and Lee 2018	N=434 Champions League matches	degree centrality, Weighted Degree Centrality, Eigenvector Centrality, Betweenness, Centrality, Closeness Centrality, Average Path Length, Network Diameter, Gradient Boosting	Social network analysis can be used to create an accurate prediction model	- The system allows coaches to be informed of a need to change strategy based on scores in performance indicators as well as identify issues with individual players
Clemente Martins, Couceiro, Mendes and Figueiredo 2014f	N=1 Portuguese premier league match	density, heterogeneity, Centralization, Scaled Connectivity, Clustering coefficient, centroid player	Defenders contribute greatly to the building of attacks and **centroid players** are the biggest contributors to attacks	- Network analysis offers an alternative tool that traditional analytical methods to characterise team performance and assist the coach in match analysis
Clemente, Couceiro and Martins, 2014e	N=5 Portuguese Premier league matches	Scaled connectivity, Clustering coefficient, Centroid value and topological overlap (or dependence)	full backs, centre backs and midfielders are the players with the highest connectivity across the team and are important for creating attacks	- Using centroid players can help teams identify critical players in the opposing teams attacking strategy
Clemente, Martins and Mendes 2015f	N=15 youth and amateur matches	indegree, Outdegree	Match score related with **outdegree** values suggesting that higher technical ability leads to stronger tactical performance	-
Clemente, Martins, Kalamaras, Wong and Mendes 2015b	N=64 world cup matches	total links, network density, clustering coefficient, network diameter	successful teams demonstrated higher levels of **total links, network density** and **clustering coefficient**	- Coaches should seek to increase the connectivity within their teams players
Clemente, Martins,	N=64 world cup matches	Indegree Centrality, Outdegree Centrality,	**centrality** measures indicate where role of each position when attacking. The central	- Helps coaches identify the players most central in building attacks in post-match analysis

Kalamaras, Wong and Mendes 2015c		closeness centrality, Betweenness Centrality	midfielders are identified as the most important in the attacking strategy	
Clemente et al, 2015a	N=4 world cup matches	total links, network density, degree centrality, degree prestige	the primary strategy used by the team was to build attacks through the defence and midfield layers	- The system used offers information on a team's style of play
Clemente, Couceiro, Martins and Mendes, 2015e	N=5 Portuguese premier league matches	Network density, Network heterogeneity, network centralisation	Network analysis metrics including **density**, **heterogeneity** and **centralisation** can be used to identify traits of a team's collective behaviour	- Network measures can be used to support coach decision making through the design of training practices
Clemente , Martins and Mendes 2016d	N=36 Portuguese premier league matches (goals only)	Clustering coefficient, indegree centrality, outdegree centrality, closeness centrality, betweenness centrality	wingers and attacking midfielders were most prominent in attacks and goals were scored or conceded from the attacking midfield zone	- Network analysis offers assistance to coaches in providing analysis of player interactions during matches
Clemente, Silva, Martins, Kalamaras and Mendes 2016b	N=7 world cup matches	total links, network density, degree centrality, degree prestige, betweenness centrality	The German national team relied most upon their midfielders for building attacks but used all players and were patient in build up as opposed to using counter attacks	- Network analysis and metrics can be used as part of a user-friendly feedback system
Clemente, Figueiredo, Martins, Mendes, Wong 2016c	N=10 Amateur matches	Degree centrality, betweenness centrality, degree prestige	There may be some correlation between physical and technical ability and social network analysis metrics measuring tactical performance.	-
Clemente and Martins 2017a	N=109 Champions league matches	total links, network density, Clustering coefficient	network metrics showed little correlation to classical measures of performance in football including shots and goals	- Coaches might want to try to decentralise attacks and increase network density and clustering coefficient scores to improve results but should exercise caution as the evidence is limited
Clemente and Martins 2017b	N=109 Champions League Matches	indegree centrality, Outdegree Centrality, Betweenness Centrality	Central midfielders are the most important players for teams in the Champions league in terms of building and creating opportunities	- Network analysis metrics such as the ones suggested in this study can be used by coaches to understand the dynamics in training or matches and make decisions to adapt strategy accordingly
Clemente 2018c	N=64 World cup Matches	total arcs, density, group clustering, arc reciprocity, dyad reciprocity	Network measures including **total arcs, density** and **reciprocity** had small to moderate correlations with match outcomes	- Network metrics can characterise variation in team performance
Cotta, Mora, Merelo, Merelo-Molina	N=3 World cup Matchers	Network intensity, consecutive passes, clustering coefficient, centrality	Network metrics can be used to describe the teams nature and characteristics in their attacking play	- Networks can have use for identifying nodes on the network that a team is most vulnerable at and an opponent may target.
Gama et al 2014	N=6 Portuguese Premier league matches	In degree centrality, Outdegree centrality	player decisions emerge through the constraints and complexities of the game	-

Gama, Couceiro, Dias and Vaz 2015	N=30 Portugues premier league Matches	scaled connectivity, clustering coefficient, Global rank, centroid player,	**centroid players** are important in organising the attacking strategy through both receiving and playing passes	- Network metrics can be used to help coaches understand how teams self-organise themselves
Gama et al, 2016a	N=2 Portuguese premier league matches	Clustering coefficient, centroid player, global rank	**Centroid players** are critical for how teams self-organise themselves	- Observing team performance through networks can be helpful for coaches understanding how their teams organise
Gama, Dias, Couceiro, Sousa and Vaz 2016b	N=30 Portugues premier league Matches	scaled connectivity, clustering coefficient, Global rank, centroid player,	more successful passes go through midfield area and are focussed on the **centroid player** in the team, whereas less successful passes occur closer to the opponents goal	- Network analysis can complement existing analytical procedures when measuring team and individual performance
Goncalves et al, 2017b	N=2 Portuguese youth league matches	Closeness centrality, betweenness centrality	teams with lower dependency on an individual player and havei well balanced and connected networks may optimise team performance	- Training design can use social network analysis to maximise specificity in identifying critical features of a team
Grund 2012	N=760 Premier league matches	network intensity, network centralisation	higher network intensity values indicate higher team performance, whereas high network centralisation indicate lower team performance	- Using these metrics can be useful for observing player performance trends
Grund 2016	N=760 Premier league matches	network intensity, network centralisation, indegree centrality, outdegree centrality, network experience	team mates who have played together for a longer period of time are more likely to perform at a higher level and perform intelligent interactions with one another.	-
McHale and Relton 2018	N=380 Premier league matches	Exponential centrality, exponential betweenness centrality	A methodology is presented to measure player contributions and skill when their team is attacking	- Assist in owners and coaches recruiting players
Mclean, Salmon, Gorman, Naughton, Solomon 2017b.	N=83 matches from EURO's and Copa America (goals only)	density, cohesion, sociometric status, individual sociometric status, number of connections	There were no clear differences between international teams in Europe and north/south America when analysis network metrics	- Social network analysis could be used as a tool for analysing interactions in team sports
Mclean, Salmon, Gorman, Stevens, Solomon 2018.	N=51 European championships matches (goal only)	Density, cohesion, sociometric status, indegree centrality, outdegree centrality, withindegree centrality	goal scoring passing networks were varied and did not determine successful and unsuccessful teams. Match status did impact network structure	-
Mclean, Salmon, Gorman, Dodd, Solomon 2018.	N=22 matches from the A-League	Network Density, Sociometric status, Cohesion	teams are connected more through communication than by passing and that teams do not communicate as much when attacking compared to defending	- Coaches can use unstructured defensive situations to improve communication

Mendes, Clemente and Mauricio, 2018	N=132 Portugues premier league matches/ Uefa youth league	total links, network density, indegree centrality, outdegree centrality, betweenness centrality	a correlation was found between general network properties and the final score and goals conceded. Networks were also impacted on match location younger layers were more likely to be heterogeneous in their passing structures	-
Oliveira and Clemente 2018	N=6 Champions league matches	Network Density, total links	**total links** had a small correlation with total distance covered	-
Peixoto, Praca, Bredt and Clemente, 2017	N=64 World Cup matches	total links, network density, indegree centrality, outdegree centrality, betweenness centrality	**total links** and **network density** were lower in attacks that were deemed successful compared with unsuccessful	-
Pina, Paulo and Araujo 2017	N=12 Champions League matches	Network density, clustering coefficient, degree centralisation	**network density** scores were negatively correlated with successful offensive plays but, networks with higher **density** would suffer fewer possession losses before reaching dangerous areas on the pitch	- Task constraints can be manipulated in training session to allow for teams to develop the ability to enter dangerous zones through different space channels
Praca, Clemente, Pereira De Andrade, Morales and Greco 2017	3v3 small sided games with conditions (Number not stated)	total links, network density, clustering coefficient, degree centrality, degree prestige, page rank	both conditions on the game and players position impact the network properties of that emerge	- Depending on coach desires, the coach should change condition to suit learning outcomes, e.g. to train goal scoring or ball possession 4v3 seems more suitable
Praca et al, 2018a	N=12 3v3 small sided games	total links, network density	players with high skill demonstrated larger scores for **total links** and **network density**	-
Praca et al 2018b	3v3 small sided games (number not stated)	Network density, clustering coefficient, degree centrality, degree prestige, page rank	U13 and U14 players demonstrated similar network properties. U14 players appeared to demonstrate higher defensive prominence	-
Praca, Sousa and Greco 2019	3v3 small sided games with conditions (number not stated)	Network Density	aerobic power appears to have a low impact of the network properties emergent in 3v3 games	- Small sided games of 3v3 lasting 4 minutes with 4 minutes of recovery are recommended to develop tactical behaviours of players
Ramos, Lopes, Marques, Araujo 2017	N=5 Premier league matches	Simplice level N+1, Simplice level N+2, Simplice level N+3	the most common simplices were 1v1, 2v1, 1v2, 2v2 and 3v1. unbalanced **simplices** are distributed across the regions of the pitch unevenly,	- Training and playing strategies can be refined by understanding the importance of specific simplices and their likelihood and frequency of occurrence across the pitch locations
Trequattrini, Lombardi and Battista 2015	N=1 Champions league match	clustering coefficient, Density, Egonet Density, Geodesic distance, Average distance, Compactness, breadth, In degree centrality, Outdegree centrality, Betweenness Centrality, Clique, Overlapping cliques	social network analysis can be used to model the interaction between team mates on the pitch	- Models using network analysis support coach decisions in matches and at training to maximise performance

Table 4.3: Research quality evaluation

Quality Item	Success rate
A clearly stated study purpose	92.1%
Relevant background literature used	94.7%
Appropriate design for the research question	92.1%
Detailed reporting of the sample size	94.7%
Justification of the sample size	7.9%
Informed consent or ethical permission	28.9%
Detailed reporting of methodology	76.3%
Used inferential statistics related to the aim	73.7%
Appropriate analysis methods considering the study aim	92.1%
Appropriate conclusions stated relating to study methods	92.1%
Stated practical applications derived from study results	47.4%
Acknowledged and described study limitations	73.7%

CHAPTER 5

Reliability of spatial-temporal metrics used to assess collective behaviours in football: An in-silico experiment.

Prelude

This chapter largely follows a manuscript under review at Medicine and Science in football investigating the reliability of novel approaches discussed in chapter 3. It was identified in the systematic review that the reliability of metrics was important but never measured appropriately. Many authors have assumed that tracking systems are reliable due to previous work investigating how they perform when evaluating physical measures and that the results will transfer to more complex spatial-temporal metrics. This assumption, however, may not be appropriate and research is required to identify the practical reliability of applying spatial temporal collective behaviour metrics. Some variables may currently be too sensitive to errors within the positional tracking system and therefore should not yet be considered applicable for analytical procedures in football. This chapter will seek to explore the boundaries of spatial-temporal metrics when dealing with positional data that includes errors of different magnitudes.

5.1 Introduction

Player tracking has become a staple of performance analysis in elite team-based sports (Buchheit and Simpson 2017) and has traditionally been used to help coaches understand the physical demands and activity profiles of players (Cummins et al, 2013; Sarmento et al, 2018a). However, a focus on individuals and physical outputs provides limited information regarding the overall functioning of the team. Consequently, collective behaviour metrics quantifying team behaviour and decision making have emerged as an alternative application of player tracking data. In team sports, player tracking is commonly achieved by Global Positioning Systems (GPS), semi-automatic video tracking and radio based local position measurement systems. Data obtained provides information on players' position in space at relatively high frequencies (e.g. 5 to 45 Hz), hence is referred to as spatial-temporal data and has the potential to describe collective behaviour once transformed by mathematical models integrating data from multiple players (Low et al, 2019). A range of sports and mathematical models have been explored and are currently being developed (Sampaio et al, 2012; Goncalves et al, 2016; Laakso et al, 2017; Moura et al, 2016; Goncalves et al, 2019). One of the most popular sports to apply collective behaviour metrics is football, where the complexity of the game and relatively low scoring opportunities has limited the success of traditional performance analysis methods, such as frequency analysis, in understanding team behaviour (Sarmento et al, 2018a). A range of potential applications have been proposed for spatial-temporal data including talent identification (Low et al, 2019), evaluation of complex decisions made by players (Steiner et al 2019), development of youth players (Barnabe et al, 2016), enhanced training sessions through constraint manipulation (Silva et al, 2014a), and even providing live data to inform coach decision making during matches (Clemente et al, 2013d). However, extensive theoretical and practical development is required before approaches are widely used within football teams.

Only recently has research investigating spatial-temporal data and associated team collective behaviours in football been systematically reviewed. In the review, Low et al (2019) created a taxonomy with 6 categories to delineate the most common spatial-temporal metrics reported. Four of these categories considered tactical variables describing characteristics of player and team organisation. These categories included position, distance, space, and numerical relations that referred to coordinate location, distances between positions, team dispersion and areas of superiority or inferiority, respectively. The remaining two categories (synchronisation and predictability) identified by Low et al (2019) represent non-linear methods used to analyse metrics from the first 4 categories. Low et al (2019) identified 27 distinct tactical metrics presented in research investigating collective behaviours in football with spatial-temporal data. However, these distinctions may fail to account for subtle differences between metric calculations that may ultimately influence findings and practical applications.

In addition to the large number of spatial-temporal metrics and analysis methods that can be applied, previous research also varies substantially in the approaches used to process and report analyses. For example, analyses can be processed across different levels of team structure (e.g. dyads, sub-groups, the team or across both teams); different periods of competition (e.g. every 10 minutes, per half, or full game); or according to different game scenarios (e.g. all play, continuous play, attacks, or successful attacks). Generally, researchers have recorded metrics at high frequencies and summarised values through means and standard deviations over a range of time intervals including 30 seconds (Clemente et al, 2018b), 5 minutes (Moura et al, 2016; Moura et al, 2013), 10 minutes (Moura et al, 2012) and 15 minutes (Duarte et al, 2013b; Goncalves et al, 2018b). Statistical analyses are then frequently performed by comparing samples across independent variables that may indicate expertise such as age (Barnabe et al, 2016; Olthof et al, 2015; Menuchi et al, 2018; Aquino et al, 2016a). The most common metrics analysed using this approach include those describing behaviour at the team level such as team centroid (Barnabe et al, 2016; Moura et al, 2012; Frencken et al, 2011; Castellano et al, 2016; Olthof et al, 2019; Praca et al, 2016), Surface area (Frencken et al, 2011; Aquino et al, 2016a; Baptista et al, 2020; Castellano et al, 2017; Frencken et al 2013; Travassos et al, 2014b) and stretch index (Olthof et al, 2019; Bartlett et al, 2012; Frias et al, 2014; Silva et al, 2014). Additionally, common data processing approaches include removal of segments where the ball is out of play and set plays such as corners and free kicks. Alternative approaches have included calculation of samples across discrete sequences within games such as attacks. This approach has often featured in small-sided games (SSGs) or sub-phases of full games to analyse 1v1 or 2v1 situations with metrics calculated at the dyad level (Headrick et al, 2012; Laakso et al, 2019; Leser et al, 2015). Bartlett et al (Bartlett et al, 2012) also excluded data from attacks starting from set plays such as corners and free kicks. Data processing strategies restricting analyses to the longest attacks (Chung et al, 2017) or successful attacks (e.g. possession resulting in a shot or goal (Moura et al, 2016; Bartlett et al, 2012;

Shafizadeh et al, 2016) have also been used in attempts to assess behaviours over patterns of play that are deemed most meaningful. Clear processes and guidelines have yet to be established due to the developing nature of the research topic.

An important preliminary concern that is yet to be addressed with analysis of spatial-temporal metrics is reliability given errors in positional data. Where reliability has been discussed previously this has been restricted to errors in individual players' x-y position which has been shown to range from 2 (Frencken et al, 2010) to 470 cm (Siegle et al, 2013). However, the extent to which these individual position errors influence reliability of specific collective behaviour metrics is unknown. Moreover, it is important to determine the effects of data processing approaches such as analysis over different time periods versus analysis across specific game scenarios on reliability. Given the intrinsic nature of variability in dynamic systems such as collective behaviours in football, general conceptions of reliability as the ratio of variances representing signal relative to noise (Hämmerer et al, 2013; Ryan et al, 2020; Lacome et al, 2019) are required to establish the influence of positional errors. For common approaches to be appropriate, the variance induced by positional errors should be substantively lower than variance exhibited by metrics across the analysis framework selected (e.g. time or specific game scenarios). An efficient and powerful method to investigate reliability is through in silico experiment where positional errors can be systematically introduced and manipulated to reflect expected errors based on current technology. Variance caused by this manipulation can be compared to variance across different frameworks and the reliability quantified. Therefore, the purpose of this study was to provide a novel reliability assessment of common spatial-temporal metrics used in football and assess the potential moderating effects of procedures such as time periods and game scenarios selected to perform calculations.

5.2 Method

5.2.1 Data collection

Data were collected from seven competitive international matches during the UEFA European championships 2020 qualifying stage. The data comprised player x-y position information gathered at 25 Hz using the TRACAB optical tracking system (Chyronhego, New York). Player coordinates from both teams were recorded longitudinally from -5250 to 5250 cm, laterally from -3400 to 3400 cm and were subsequently transformed into metres. Additionally, time stamped event data were tagged separately and integrated to identify how the attack started, ended and which team was in possession.

5.2.2 Study protocols

Eight frequently investigated spatial-temporal metrics were selected, comprising two metrics per category identified by Low et al (2019). The metrics included $centroid_x$ and $centroid_y$ (position); distance between team centroids and team length (distance); surface area and team spread (spaces); length per width ratio and space control (numerical relations). Analyses were performed on samples generated across predetermined time intervals (1, 5, or 15 minutes) and across possessions separated by the attacking team. Three criteria were used to identify the start of possessions: (i) where a player takes three consecutive touches of the ball; (ii) a player makes a successful pass to a teammate; or (iii) a player takes a shot (Santos et al, 2019). A possession lasted until the ball went out of play or the opposition successfully recovered the ball. When analysing across possessions, data of players were removed when the ball was out of play.

Centroid coordinates were calculated using the following equation:

[eq1] $\text{Centroid}_{x,y}(t) = \left(\frac{\sum_i^N p_{xi}(t)}{N}, \frac{\sum_i^N p_{yi}(t)}{N}\right).$

Where N is number of outfield players in the team, $p_{xi}(t), p_{yi}(t)$ represent the x-y position of player i at time t (t). The radial distance between centroids was calculated as:

[eq2] $\text{Distance between centroids } (t) = \sqrt{(xC(t)_a - xC(t)_b)^2 + (yC(t)_a - yC(t)_b)^2}.$

where $C(t)_a$ and $C(t)_b$ refer to the centroid coordinate of team a and b, respectively. Team length was calculated as the maximum distance along the x axis between two outfield players in a team.

Surface area was calculated by creating a convex hull around the outfield players of each team and calculating the area inside (Palucci et al, 2019). Team spread was measured by calculating the Euclidean distances between all outfield players, arranging the values in a 10 by 10 matrix. The Frobenius norm was calculated using the lower triangular matrix as shown in equation 3:

[eq3] $\| \text{Team spread}(t) \|_F = \sqrt{\sum_{i=1}^N \sum_{j=1}^N |l_{ij}|^2}$

Where l_{ij} is the distance between each pair of teammates and N is the number of outfield players (Moura et al, 2016). Length per width ratio (LPW) was calculated as the ratio of team length (maximum distance between teammates in the x direction) relative to team width (maximum distance between teammates in the y direction).

Space control was measured using Voronoi cell computations that identify the area on the pitch closest to each individual. This describes a player's dominant region, where space control equals the ratio of the sum of team a's dominant regions relative to the sum of team b's dominant regions.

For each game, the eight metrics were calculated at a frequency of 1Hz using the positional information from each player. These data were used to calculate the "natural variance" in metrics (e.g. signal). To complete the in silico experiments positional errors were added to each players' x-y position at each time recording (e.g. noise). Positional errors comprised random draws from a Gaussian distribution with mean of 0 and standard deviation selected to represent the typical error magnitude of positional tracking equipment. Three error magnitudes were selected (small, medium, and large) equal to 0.5, 2 and 4 m (Frencken et al, 2010; Siegle et al, 2013; Ogris et al, 2012; Linke et al, 2018; Linke et al 2020). The process of adding positional errors to all players positional data was completed 1,000 times, such that metrics were calculated on the actual game data and across 1,000 in silico experiments. A schematic overview of the entire in silico approach is illustrated in figure 5.1. Reliability values were calculated at each instance of the analysis framework selected (e.g. time period or game scenario) by comparing the variance in the metric from *within* the raw data (e.g. signal), relative to the variance in the metric *between* the in silico experiments (e.g. noise). Reliability was calculated as the within variance relative to the total variance (within + between) providing a value between 0 and 1. The within variance would be measured through the simulated data across each instance. Values higher than 0.909 were deemed highly reliable (Beltrame et al, 2020).

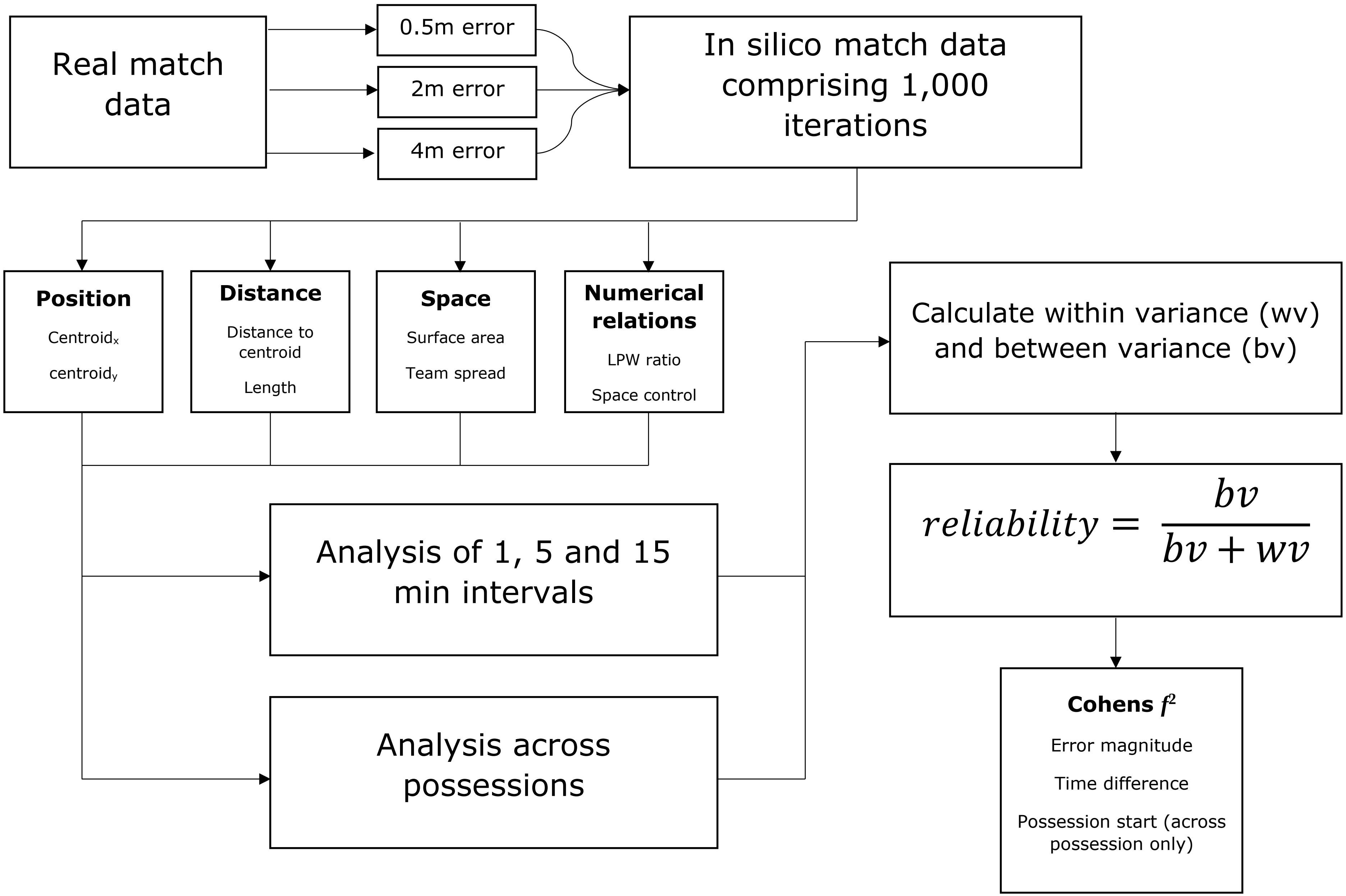


Figure 5.1: Schematic of in silico experiment process

5.2.3 Statistical analysis

For each metric, the effects of error magnitude (0.5, 2 and 4 m), and time of analysis (playing time: 1, 5 and 15 min; attack duration: short [<10 s], medium [10-20 s], long [≥20 s]) were evaluated by fitting linear mixed effects models with reliability values as the dependent variable and random effects fitted to each game to account for covariance of values made within the same match. Cohens f^2 effect sizes for mixed effects models were used to quantify and compare the influence of each independent variable (Nakagawa and Sheilzeth 2012) with threshold values of 0.02, 0.15 and 0.35 used to categorise effects as small, medium and large, respectively.

5.3 Results

An initial overview of the results was obtained by aggregating reliability values across all eight metrics and presenting the information relative to the error magnitude and time of analysis (table 5.1). A visual distribution of reliability values for individual metrics are presented for different error magnitudes in figure 5.2. High mean reliability values (>0.909) were obtained for all metrics with 0.5 m errors. Across all error magnitudes, three clusters were observed with the greatest mean reliability obtained for centroids (x direction: 0.984 to >0.999; y direction: 0.961 to >0.999). In the second cluster, high mean reliability values were obtained with the spaces metrics (team spread and surface area) for both 0.5 and 2 m errors. In contrast, poor mean reliability was obtained for all other metrics in the third cluster, with values as low as 0.6 for distance between centroids under 4 m errors.

Table 5.1: Mean (standard deviation) reliability values pooled across metrics and presented by error magnitude and time grouping.

error	1 minute	5 minutes	15 minutes
0.5	0.984 (±0.027)	0.992 (±0.012)	0.992 (±0.012)
2	0.865 (±0.136)	0.924 (±0.077)	0.936 (±0.062)
4	0.706 (±0.215)	0.811 (±0.155)	0.836 (±0.13)

When quantifying the effects of error magnitude and time of analysis on reliability values (table 5.2), the relative effects of error magnitude were substantively greater for all metrics except centroids where both variables exerted medium effects (f^2 = 0.18 to 0.32). Time was found to exhibit small effects on LPW ratio, length, and surface area (f^2 = 0.08 to 0.13); and medium effects on team spread, space control and distance between centroids (f^2 = 0.15 to 0.22). Error magnitude was found to exert large effects on all metrics except centroids (f^2 = 0.89 to 6.02)

Figure 5.2: Distribution of reliability values from in silico experiments created with error magnitudes of 0.5, 2 and 4 metres.

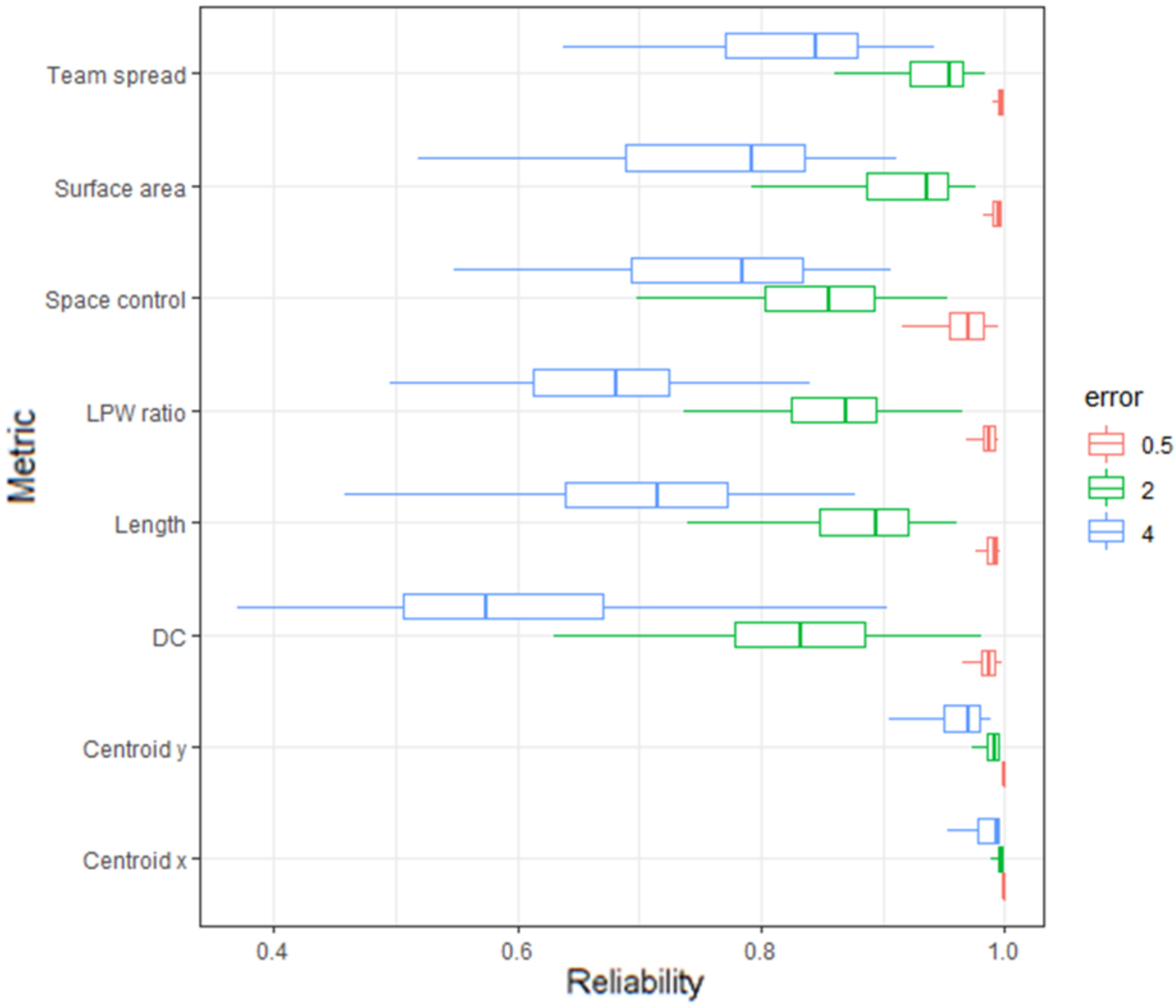


Table 5.2. Effect sizes of error magnitude and time grouping (1, 5 and 15 minutes) on reliability values

Metric	Time f^2	Error magnitude f^2
DC	0.22	0.89
Length	0.08	4.22
LPW ratio	0.06	6.02
Surface area	0.13	2.91
Team spread	0.15	2.45
Space control	0.16	1.84
Centroid$_x$	0.30	0.32
Centroid$_y$	0.22	0.18

DC = distance between centroid. LPW ratio = length per width ratio.

Over the seven games 1189 attacks were analysed, 152 (13%) of which ended in shots, with 21 goals scored. A total of 514 (43%) attacks were categorised as short (<10 s), 294 (25%) were categorised as medium (10 - 20 s), and 381 (32%) were categorised as long (>20 s). As an initial analysis, reliability values were pooled across all metrics and the effects of error magnitude and duration of attack visualised (Figure 5.3). The analysis illustrates that as attack durations decrease, the distribution of reliability values spreads substantively for even 0.5 m errors. In general reliability values are poor for all conditions except medium and long duration attacks with 0.5 m errors.

A similar pattern of findings was obtained when quantifying the effects of error magnitude and time when attack duration was assessed compared with simply continuous epochs (Table 5.3). The relative effects of error magnitude tended to be substantively greater for all metrics except centroids and in this instance space control. Medium effects were obtained for attack duration on all metrics (f^2 = 0.18 to 0.32) except centroid$_y$ where a large effect was measured (f^2 =0.37). Error magnitude was found to exert large effects on all metrics (f^2 = 0.39 to 1.17) except centroids where small to medium effects were measured (f^2 = 0.13 to 0.18).

Table 5.3. Effect sizes of error magnitude and attack time grouping (short: <10 s, medium: 10-20 s, long: >20 s) on reliability values.

Metric	Attack time f^2	Error magnitude f^2
DC	0.18	1.17
Length	0.17	0.84
LPW ratio	0.19	0.96
Surface area	0.18	0.83
Team spread	0.20	0.67
Space control	0.32	0.39
Centroid$_x$	0.27	0.13
Centroid$_y$	0.37	0.18

DC = distance between centroid. LPW ratio = length per width ratio.

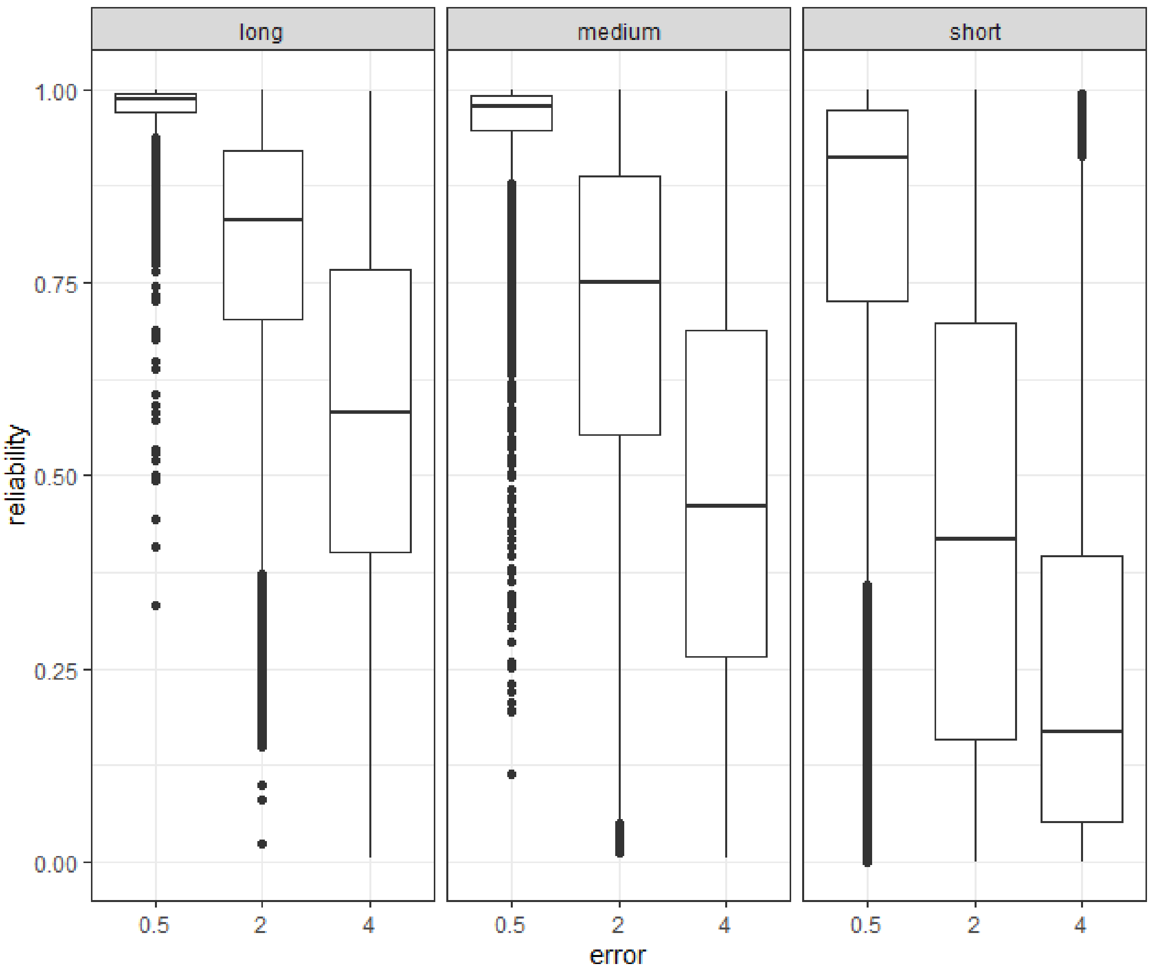


Figure 5.3: Distribution of reliability values pooled across all metrics from in silico experiments created with error magnitudes of 0.5, 2 and 4 metres and attack time grouping (short: <10 s, medium: 10-20 s, long: >20 s).

5.4 Discussion

The purpose of this study was to investigate the reliability of spatial-temporal metrics used to quantify collective team behaviour in football. Reliability was assessed using in silico experiments with errors representative of that induced by current positional tracking technologies. Collectively, the results demonstrate that technologies which generate errors in position tacking of 0.5 m or less should be expected to produce spatial-temporal metrics with high reliability. However, technologies that generate error values of around 2 m may have substantive effects on metrics such as distance between centroids, length, length per width ratio and space control, reducing reliability and limiting the accuracy of conclusions drawn with limited data. Similarly, technologies generating larger positional errors such as 4 m can lower reliability to where noise and signal are approximately equal, severely limiting insights. Whilst researchers and practitioners may seek to analyse spatial temporal metrics over crucial sequences of plays such as attacks, the findings from this study indicate that this strategy may results in very low reliability, due to the short duration of most attacks. Unacceptably low reliability is likely to occur where analyses are conducted over sequences of plays such as attacks using tracking technologies that generate moderate to large positional errors (e.g. 2 to 4 m).

Team centroid representing the weighted position of players was identified as the most reliable metric and demonstrated high values even when positional errors were large, and time of analysis was short. The team centroid concept has been applied regularly throughout the literature base (Low et al, 2018; Olthof et al, 2018; Goncalves et al, 2014; Frencken et al, 2012). Centroid values across the x and y axis have been demonstrated as relevant in critical situations such as teams reorganising after loss of possession or goal scoring opportunities. Distance between centroids was identified as the least reliable metric with noise variance frequently exceeding that of the signal when subjected to large positional errors. Previous research has demonstrated that distance between centroids have the potential to cross before teams score in small-sided games[22] and that team centroids are generally coupled with oscillations tightly synchronised (Frencken et al, 2012). Siegle and Lames (2013) identified that perturbations in the distance between team centroids related to critical moments in the match. However, the findings from the current study highlight that the metric may not be reliable in measurement contexts of large positional errors and when investigating critical moments such as short duration attacks, thereby warranting care in future analyses.

Space control has recently become a popular metric with some professional football clubs (Fernandez et al, 2018; Spearman et al, 2017) where it is used to evaluate passing ability and decision making (Spearman et al, 2017; Rein et al, 2017; Filetti et al, 2017). However, similar to distance between centroids, space control also demonstrated poor reliability in the in-silico experiments when calculated with 2 and 4 m errors. Various approaches have been used to measure dominant regions of each player including Voronoi cell computations(58) and calculations including player speed in attempts to predict areas on the pitch a player will reach first (Filetti et al 2017; Brefeld et al, 2019). It is likely that inclusion of player speed in computations would further decrease reliability where there are large positional errors.

The analysis identified strong reliability for the spaces metrics (surface area and team spread) when calculated with 0.5 and 2 m errors. Surface area and team spread are generally used as measurements of team dispersion and aggregation (Bartlett et al, 2012), with teams commonly expanding when attacking and demonstrating compactness when defending (Moura et al, 2016; Duarte et al, 2013a). Clear shifts when transitioning from attacks to defending and the large variability this may contribute towards a strong signal and as a result the high reliability values. Oscillations in the metrics have been used by researchers to better understand dynamics of transitions, with attempts to measure the coordination of opposing teams surface area and team spread using non-linear analysis methods such as relative phase and vector coding (Low et al, 2020; Moura et al, 2016). Further research is required to investigate the effects of non-linear analyses on reliability and as a result the accuracy of conclusions that may be drawn.

Across all metrics, results were consistent in demonstrating that reliability substantively increased when analyses were conducted over longer periods of time. Most metrics demonstrated low reliability values when analyses were made over a period as short as one minute. Whilst research has been conducted analysing spatial-temporal metrics over periods of time (Moura et al, 2016; Clemente et al, 2018b; Goncalves et al, 2018b; Duarte et al, 2013b), it is more common to analyse values across possessions and sequences such as attacks (Laakso et al, 2017; Barnabe et al, 2016; Bartlett et al, 2012; Headrick et al, 2012; Duarte et al, 2012a; Castellano et al, 2013). However, the results of the present study show that attacks generally last short periods (e.g. <10 s) and as a result, reliability of spatial temporal metrics may be limited even when using measurement technologies with positional errors as low as the 0.5 m error. Therefore, researchers and practitioners should remain cautious when analysing spatial-temporal metrics across short possessions.

The present study includes multiple limitations that should be considered when interpreting the findings. The primary limitation of the analysis is the relatively simple error model used for the in-silico experiments. The independent Gaussian errors do not match some of the properties of positional tracking methods including a relationship between player speed and position error, and type of action (e.g. accelerations and change of direction) and position error (Siegle et al, 2013) These types of movements may be more likely to occur during critical events within a football match such as attacks influencing one of the central contexts studied in this domain. In addition, more advanced Markov models may be more appropriate to describe positional error and autocorrelation which is likely to occur. As an initial attempt to quantify reliability in this rapidly growing area, the study only considered traditional linear analyses, however, non-linear approaches including relative phase and approximate entropy are commonly applied to spatial-temporal metrics in football. Future research is required to quantify the reliability of these approaches when applied to the metrics evaluated here and others that are commonly used.

5.5 Practical applications

The results of the present analysis highlight that researchers and practitioners using spatial-temporal metrics to analyse collective behaviours in football should carefully consider the tracking technologies used to obtain the data, the errors that they are likely to introduce, and the analysis procedures implemented. Practitioners should be cautious over their application. A range of metrics including those belonging to distances and numerical relation categories may suffer from unacceptable reliability, particularly when analyses are completed across relatively low samples, and analysed across short periods of times. Whilst researchers and practitioners may wish to analyse spatial-temporal metrics across discrete sequences of plays such as attacks, they should be aware that short attacks including those lasting ten seconds or less may be unsuitable to draw conclusions from based on low reliability. As tracking technology develops, it is likely that more systems will be able to reliably measure a greater range of spatial-temporal metric over shorter periods of time.

CHAPTER 6

Novel collective behaviour measures translated to principles of play and concepts as understood by football coaches

Prelude

Reliability is a pivotal aspect of research within collective behaviour metrics. However, the largest limitation identified within the systematic analysis of the literature base was the lack of conceptual clarity. What is being measured by collective behaviour metric is often described in a mathematical sense but seldom grounded in principles of play and explained why such a value is tactically relevant. Investigations into how coaches perceive and understand collective behaviour metrics and how they relate to their own philosophies and strategies when organising and training their teams are sparse. As such, this chapter will seek to establish the collective behaviour measurements that align with coaches own perception of tactical concepts through an iterative process that seeks to reflect the role of a performance and data analyst in a professional setting to create, adapt and refine collective behaviour metrics that resonate with current philosophy and pedagogy in football.

6.1 Introduction

With increasing data collection in elite football, more sophisticated approaches are being developed to derive knowledge and insight information (Goes et al, 2021). Traditional approaches to data analysis have focussed on players physical performance through information obtained by movement analyses or on team's overall performance measured by technical or tactical event frequencies occurring in matches such as passes or dribbles (Sarmento et al, 2014). More wide-ranging analyses have also sought to incorporate contextual information including location and situation both relating to the instance of the event occurrence (i.e. position on the pitch the action took place) and features at the match level (i.e. home advantage, match status, opponent strength). However, many of these factors are not collected in the processes used to capture data (Mackenzie & Cushion 2013). Indeed, researchers still actively pursue game actions as the primary means of performance analysis in both football and other sports with methods including collective behaviour, movement patterns and social network analysis gaining more popularity over the last decade (Lord et al, 2020). A variety of analysis techniques have also been applied to the data collected, most prominently inferential statistics that attempt to explain differences between successful and unsuccessful performance (Lord et al, 2020). Predictive models have also increased in popularity over the last decade with implementations frequently including clustering techniques comprising factor analysis and machine learning (Goes et al, 2021). There is evidence that the widespread adoption of both data analysis and performance analysis in elite football has shaped the coaching process and development of the sport (Bush et al, 2016; Zhou et al, 2020). However, the true value and influence of the systems implemented is still unknown.

The increased use of data in sport is not unique to football as identified by the incorporation of data scientists within a range of professional sports teams and organisations (Fernandez and Bonn, 2018; Spearman et al, 2017). However, there are several factors that make football performance challenging to quantify and analyse including primarily the low scoring nature causing fine margins to frequently separate winning and losing teams (Horvat and Job, 2020). Subsequently, individual moments in football can greatly influence the match outcome and can result in more frequent victories by teams whose overall performance is perceived to be below that of their losing opponents (Brechot and Flepp, 2020). Additionally, the continuous nature of football creates a highly dynamic environment where each player is constantly moving and adjusting based on the positions of their teammates and the ball (Grehainge and Godbout, 2021). The complexity can be challenging to summarise coherently such that performance analysts in football have largely supported coaching staff through video analysis supplemented with basic descriptive statistics (Wright et al, 2013). The provision of instructive video clips helps centre the analytical process around the coach’s own principles and philosophies, however, there may be limited use of such methods in objectively summarising the state of play.

To effectively quantify dynamics in football that are most informative to coaches, performance analysts may choose to construct analysis systems that align with a coach’s underpinning concepts and tactics. The traditional classification of principles of play in football identifies 10 components that relate to performance. The principles include attacking concepts: 1) penetration, 2) support, 3) width, 4) mobility and 5) creativity; and defending concepts: 6) delay, 7) depth, 8) compactness, 9) balance and 10) discipline (Costa et al, 2009; Prickett 2021). However, these principles of play are not necessarily exhaustive, for example, they ignore transitions between attacking and defending phases. Additionally, coaches may have different interpretations of each of the concepts and favour their own set of principles when providing tactical instruction to their team, disregarding others (Clemente et al, 2014g). Regardless, adoption of metrics within a framework that they assess and quantify these different principles may provide a deeper understanding of player and team performance within the dynamic fluid nature of a match grounded within concepts understood by the coach.

Recent advancements in data collection paired with advanced analytical procedures offer the potential for principals of play to be measured more accurately (Clemente et al, 2014g). Applications of spatial-temporal and network analysis on position and passing data shows promise in evaluating such tactical concepts. Many of the traditional principles of play have been directly measured. For instance, many authors have measured the width of a team as the distance between the two players in a team closest to each side of the pitch (Baptista et al, 2018; Folgado et al, 2018c; Chung et al, 2019; De Souza et al, 2018 Goncalves et al, 2019; Olivares et al, 2019; Olthof et al, 2019). In a similar manner, researchers have used various methods to measure the defensive concept of compactness through the dispersion and aggregation of the team using metrics including stretch index, surface area and team spread (Barnabe et al, 2016; Moura et al, 2012; Moura et al, 2016). These different methods all show a clear expansion and contraction pattern when teams are in and out of possession. However, they may not be fully representative of how a coach conceptualises compactness when coaching defensive structure to their team.

In some instances, researchers have constructed metrics that directly seek to calculate a principle of play, for example, Clemente et al (2014a) investigated penetration by observing two conditions every second: 1) if the centre of the game progressed towards the goal and 2) if the numerical balance of players within a 5m region of the ball did not deteriorate. In this instance, the metric was designed without direct consultation with coaches and how they may interpret the principles of play. In a rapidly growing field of research, there seems to be little collaboration with coaches regarding how the metrics used in this field can be applied to the coaching process. Gudmundsson and Wolle (2014) created tools while in close contact with coaches and analysts to help focus on issues that are relevant. This included analysis techniques that observed passing, passing sequences and movement patterns. However, their relevance remains largely theoretical as the practical relevance for presented metrics were not explored. This limitation appears common in a large section of the literature base with research independent of coaches' perceptions. Consequently, the purpose of this research was to explore how novel and increasingly popular measurements of collective behaviour could be adapted to fit into a coach's philosophy and principles of play. The study comprised an iterative approach working with elite football coaches to present contemporary collective behaviour metrics, explore the coaches' interpretations and their own philosophies and principles through qualitative interview and subsequently refine the metrics used.

6.2 Method

6.2.1 Study design

A framework for creating a tailored system to augment coach decision making through performance data analysis and visualisation was explored in this study. The framework comprised an iterative process that began with standard collective spatial temporal and network metrics, and through discussion with coaches, selects and modifies metrics that align with their philosophy. A visual representation of the process is presented in figure 6.1.

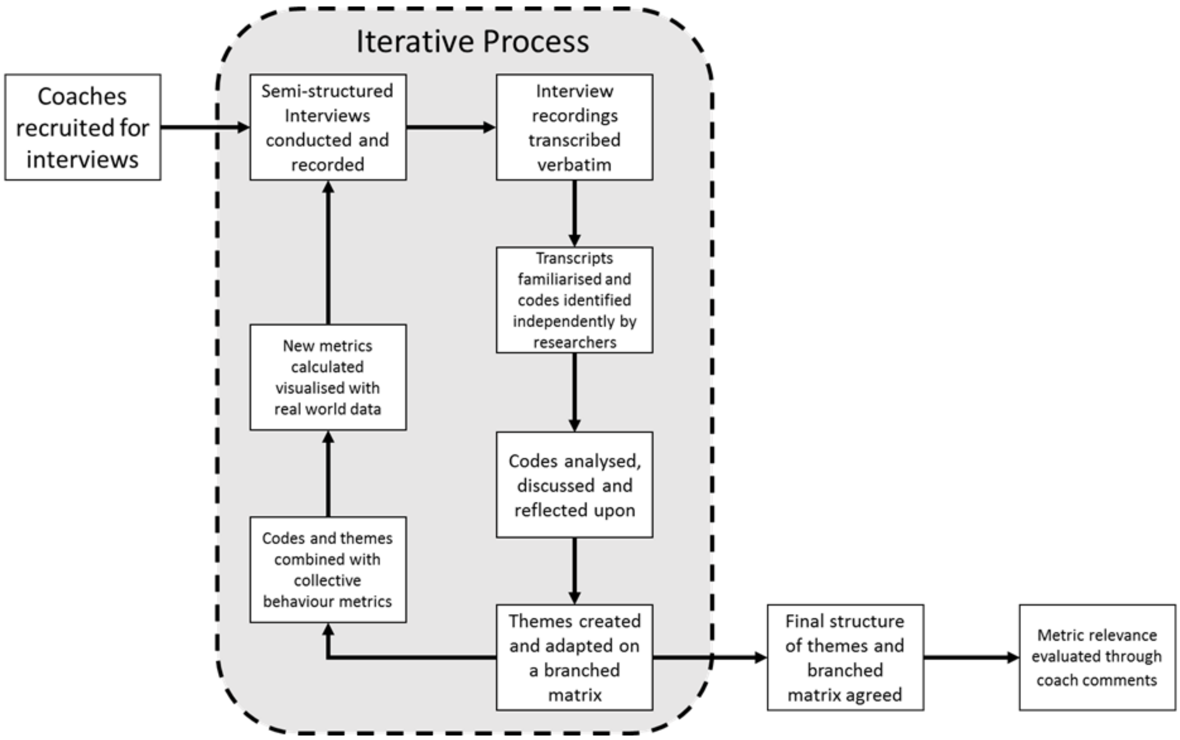


Figure 6.1: schematic overview of interview process

6.2.2 Participants

Purposive sampling was used to recruit three elite football coaches currently employed at an international level within Scottish football. The coaches had between 8- and 28-years coaching experience, and each held the UEFA Pro License. Prior to data collection, ethical approval was granted from Robert Gordon University, and coaches provided informed consent before participating in the research. Between the initial and follow up interview, one coach moved job, leaving a total of five interviews throughout the iterative research process.

6.2.3 Data Collection

Semi-structured interviews were selected for the research project as they provide open ended questions which allowed for the concepts to be fully discussed whilst the interviewer retains control of the direction of the interview, guiding interviewees to explore aspects relevant to the interviewer (Deterding and Waters, 2021). Interviews lasted approximately one hour and consisted of an initial fifteen-minute presentation where the researchers presented standard or adapted spatial-temporal and network metrics quantifying aspects of collective behaviour. Presentations were included in the first instance to provide context to collective behaviour and the approaches that are commonly used in the football literature to quantify these dynamics, so that coach's philosophy and principles of play could be explored in this context. In the initial interviews, questions primarily focused on the concepts of attacking, defending and transitions. Beyond this, additional questions centered around spatial temporal principles including position, distances, spaces, and numerical relations along with network metrics seeking to gain further understanding of passing sequences. An interview schedule (figure 6.2) was constructed to constrain the discussion to the most relevant aspects. Between interviews, metrics were adapted or created based on coach comments. Follow up interviews were initiated with a second presentation of the development of metrics and interviewees were provided quotes and interpretations of the initial interview and asked to comment on whether the calculated and visualised metrics were accurate and relevant and if important concepts were missing from the measurements.

6.2.4 Data analysis

Reflexive thematic analysis according to the methods of Braun and Clarke (2019) was used by two researchers to generate nuanced themes relevant to the individual coach. Initially, interviews were recorded and transcribed verbatim. A familiarization process was then conducted where each transcript was read and re-read by the two researchers to fully understand the data collected (Bertolio, Saltarelli and Robazza 2009). Both researchers then coded each of the transcripts independently and noted important information for further analysis (Braun and Clarke 2019). The researchers then collaborated to identify raw codes that provide a foundation to create sub themes, themes, and main themes (Taylor and Bogdan 1989). Within this stage, many codes were discussed and reflected upon by both researchers, resulting in several codes combining and others being removed (Deterding and Waters, 2021). Following this, a branched matrix was created where sub-themes fed into themes that fed into main themes (Vaismoradi 2013).

This process identified the critical aspects important to coaches from a tactical point of view. Based on the branched matrix and the interviewee quotes, systems were created to measure the tactical concepts and principles of play that coaches highlighted as important. These were then computed using real world data and visualized using the programming software R studio and presented back to the participating coaches. This step acted as both a member checking process to strengthen the credibility and trustworthiness (Podlog and Eklund, 2009) as well as forming the iterative process whereby domain expertise and evidence-based research were combined to create a robust process that could inform practice (Donaldson et al, 2016). The follow up interviews were also recorded, and the same reflexive thematic analysis procedure followed (Braun and Clarke 2019) to establish themes between the collected data. This allowed interviewee quotes to be scrutinized, identifying required adaptations to metrics and visualisations initially presented to better suit coach conceptualization. A grading system was used to categorise metrics into identifying which metrics had the greatest likelihood of aligning with the coaching philosophy (resonant), which had potential but required further investigation or adaptation (relevant) and which coaches believed had significant disconnect from their own ideas and tactics (hesitant).

Figure 6.2. Interview schedule, providing core questions for coaches

6.3 Results and Discussion

This section will provide an overview of the data derived from the iterative interview procedure along with discussing the thematic analysis that was created and adjusted based on the coach comments. Additionally, the coach perceptions of proposed metrics and visualisations describing the principles of play are discussed, identifying the most promising metrics for direct application. The tagged coach comments are shown in appendix I. Finally, a discussion on how these metrics can be further developed and integrated to support the coaching process for developing players and maximising team performance will conclude this section.

6.3.1 Iterative Thematic Analysis

Questions in the first interview were structured around 4 individual phases in a game including, attacking, transition to defence, defence, and transition to attack. Naturally, these concepts were represented in the main themes identified from the initial thematic analysis: *attacking, defending,* and *team performance.* Similarly, the 12 themes that feed into *attacking* and *defending* main themes share strong similarities with the traditional principles of play found in football literature. (Costa 2009; Clemente et al, 2014g; Prickett 2021). As stated by Prickett (2021) these include 5 attacking principles: i) penetration, ii) support, iii) width, iv) mobility and v) creativity, and 5 defensive principles: i) delay, ii) depth, iii) compatness, iv) balance and v) discipline. attacking transition and defending transition were also identified as themes, two concepts that are sometimes mentioned when discussing the traditional 10 principles (Prickett 2021; Costa 2009). These 10 principles of play were identified in the thematic analysis with interview questions being designed without considering these concepts. The interviewees all recognised these concepts with coach 1 stating.

"I'm one that very much strives to stick to the principles of the game, you know, those are the constant strains."

This suggests that the principles of play previously identified are robust and exhaustive. However, this may be caused by coaches going through education systems where these 10 principles are core features of the programme. The traditional principles also suffer from inconsistency in terminology used. This is demonstrated by coach 1 who lists the attacking principles as.

"depth, width, mobility, improvisation, penetration for your attacking ones."

The five principles highlighted by the coach align conceptually with the previously stated principles, however, use different names. Inconsistency in the terminology could lead to different interpretations. Moreover, coaches will have differing opinions on how to implement tactical strategies, underpinned by the traditional principles. The initial thematic analysis, integrated across all first stage coach interviews can be seen in table 7.1.

The third and final main theme identified from the first iteration of interviews was labelled as *team performance*. This branched into two themes: match preparation and player development. These factors related to how performance analysis provision can support the coaching process. Player development focussed more on how training can be shaped to maximise development with sub-themes including learning experiences, available coaching time and barriers to development. These relate more generally to the holistic improvement of players and teams. Whereas match preparation identified how changing contexts match to match can impact desirable aspects of team performance. Flexible tactics are often applied by coaches due to variables including the pitch size or the opponent ability and a performance analyst support system must be able to adapt to any changes.

sub themes	themes	main themes
disrupting opponent	penetration	Attacking
creating space		
diamonds and triangles	support	
balance		
control		
overloads	width	
attacking shape		
speed of play	mobility	
movement		
attacking risk	creativity	
patterns of play		
decision making	attacking transition	
counter attacking		
defensive shape	delay	defending
pressure		
team length	depth	
lines		
cover	balance	
adjusting		
Concentration	Compactness	
distances		
triggers	discipline	
working as a team		
reaction	defensive transition	
prediction of transition		
barriers to development	player development	team performance
learning styles		
learning experiences		
available coaching time		
flexible tactics	match preparation	
pitch size		
opponent ability		
game context		

Table 6.1: initial themes identified and organised

After completing the initial thematic analysis, metrics from the literature were selected and adjusted based on the coach comments. These were presented back in a second interview to confirm the interpretation of the coaches comments were accurate and evaluate how representative the metrics are. From the transcripts of the 2nd interview process, the thematic analysis was adapted further. The biggest difference was changing the main themes of attacking and defending to *penetration* and *delay* respectively. These were changed as *penetration* appeared to comprehensively describe attacking while *delay* broadly matched every principle of defending. The transition themes were also removed from the second iteration of the thematic analysis due to their incompatibility in applying novel spatial temporal or network measures that are directly related. Instead, aspects of transition were knitted into measurements describing other concepts. The changes were not limited to the removal of transitions from the themes and the promotion of penetration and delay. Of the 33 original sub-themes identified, only 16 (48%) remained unchanged in the second iteration of the table. Some of these changes were minor and caused by the removal of the attacking and defending transition themes whereby sub themes were moved into other relevant themes. For example, prediction of transition moved from defensive transition to compactness and was renamed to anticipation to better suit the terminology used by coaches. Only 4 (12%) sub themes were rephrased and another 4 (12%) were removed completely where words were either too similar to the themes they were allocated or were too broad and as a result not informative. For instance, 'decision making', could be perceived as relevant in each theme and was consequently removed to avoid sub-themes bleeding across the thematic analysis.

sub themes	themes	main themes
diamonds and triangles	support	Penetration
passing options		
angles		
team mate distances		
coordination		
overloads in wide areas	width	
creating space		
disrupting opponents		
attacking shape		
passing speed	mobility	
contact time		
movement		
risk	creativity	
breaking lines		
patterns of play		
deception		
1v1		
defensive shape	compactness	Delay
reaction		
recovery		
controlling opponents decisions		
anticipation		
length	depth	
lines		
cover	Balance	
overloads near the ball		
adjusting	Discipline	
triggers		
time		
distance to opponent		
working as a team		
pressure		
barriers to development	player development	team performance
learning styles		
learning experiences		
available coaching time		
flexible tactics	match preperation	
pitch size		
opponent ability		
game context		
player strengths		

Table 6.2: Second iteration of themes organised into a table

Such an effect is expected when evaluating tactical principles in a complex dynamical system such as a football match. Indeed, all of these concepts are interconnected, naturally causing some of the initial sub-themes to bleed into multiple themes. To minimise the impact of this effect, 3 (12%) of the original codes were split into 6 (14%) of the 41 total sub-themes identified in the second iteration of the thematic analysis. For example, distances were commonly referenced in the initial interviews. However, after devising the tools and presenting them to coaches, it appeared that the distances occupied two distinct themes: discipline and support. Consequently, distance to opponent and distance to teammates were placed in the themes, respectively. Finally, a total of 9 new sub-themes were added to the thematic analysis based on the coaches' comments in the second interviews that related distinctly to each principle of play.

6.3.2 Coach perceptions of collective behaviour measurements

This section discusses the coaches' perceptions of metric and visualisations presented to them in the second interview that were constructed and adapted from approaches in the literature based on the comments made in the first interview. A grading system was used to categorise how coaches responded to each metric. If coaches demonstrated enthusiasm towards a visualisation or identified that the measurement was fully descriptive of a principle in football, then it was labelled as *resonant.* If the metric was identified as accurately describing a concept, however, the coach identified limitations or aspects that needed improved, then it was labelled as *relevant*. Finally, if a coach was sceptical of how useful a metric would be in practical settings, or identified situations where the model was inaccurate at representing the principle then it was labelled as *hesitant*. Table 7.3 provides an overview of the 9 visualisations presented to the coaches, with summary quotes supporting the categorisation of each metric.

Metric	**principle**	**Coach perception**	**Coach quotes**
Network intensity	Mobility	Resonant	"I love it, I think it's absolutely brilliant and so critical in terms of player development, team development, winning games."
Distance between defenders	Discipline	Resonant	"the whole team needs to get back out. It's, for me really, really important to get those adjustments. And always, you can't… you can't take risks."
Triangles	Support	Resonant	"The distances are really important. But I also think it's the players that that need to sort understand that, you know, you don't just move to support the ball, if you're part of the Midfield three like that."
Team length and distance between deepest defender and goal line	Depth	Resonant	"I personally, coach my team's in a similar way. If we were under pressure, then I would want in that scenario, I would want my striker to be back as well"
Surface area	Compactness	Relevant	"I think the only thing I would add to that [author] is on the tactical instruction of the coach and the team, knowing whether on those transitions…"
Team width	Width	Relevant	"at a higher level of the game, they'll start to do things that are very different and much more complex"
Distance dyads, time to contact, and passing lane	Delay	Relevant	"I agree with your description of the pressures. What I would add I'm sure you're aware of it is, in my opinion, it's the decision from the [Team1] central defenders not to pressure once the transition happens."
Numerical Advantage	Balance	Hesitant	"I would say probably needs a little bit. A little bit of work."
Pitch control and number of outplayed opponents	Penetration	Hesitant	"You can show lots of pictures of good examples. But at the end of the day, it comes down to quick time decision making and execution."

Table 6.3: summary of coach responses to metrics.

6.3.3 Resonant Metrics

Resonant metrics were characterised by their often simple to understand measurements, relating to individual dimensions of space, time distance and speed

6.3.3.1 Mobility

The mobility principle was discussed several times in the initial interview phase. Naturally, mobility relates to the movement of the players and was suggested as being intrinsically linked to the concept of support, where teammates must move into the appropriate positions to provide passing options. However, mobility can also relate to actions on the ball and how a team is able to move the ball at pace. Coach 3 emphasised the importance of this:

> *"that's what the top players can do, they can, they can play at speed, they can do everything quickly, control the ball pass the ball, turn."*

This relates closely with the measurement of network intensity, explored by Grund (2012), measuring the speed at which teams pass the ball. This was presented to the coaches as a mean across the whole match, as well as during individual attacks with comparisons between and within matches. This measure received a positive reaction with coach 3 stating

> *"I love it, I think it's absolutely brilliant and so critical in terms of player development, team development, winning games."*

Evidence from Grund (2012;2016) has suggested that there is a strong link between successful teams and high network intensity. However, more investigation in this metric is still required to fully utilise a measure to guide training. Despite reacting positively, coach 3 provides more detail.

"…it's not just the speed of the pass, it's the contact time in between, you know, the amount of time it takes a player to control the ball and play the ball."

This suggests that in its simplest form, network intensity may not fully describe the team's ability to move the ball quickly. By splitting passing actions into control time and pass time, and incorporating starting and ending positions of passes, a deeper understanding might be obtained.

6.3.3.2 Discipline

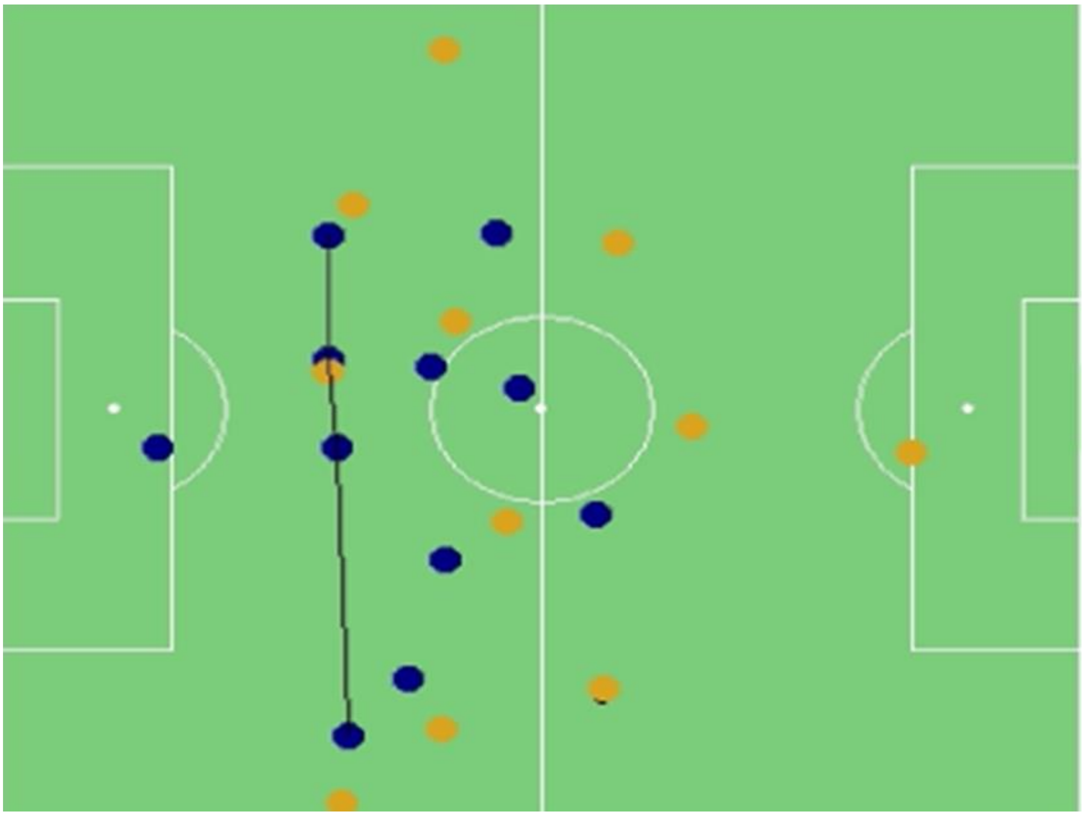


Figure 6.3. Discipline principle measured by distance between defenders.

Another measurement that the coaches responded positively to measured the distances between defenders. This was related to discipline, a principle that focuses on a strategy of when to delay the opponent through structure of the defensive unit or when to try and regain possession. Trigger points were identified as a sub-theme relating to discipline as coach 3 states.

"We speak about where we're going to engage with the opposition, whether it's at the top of the circle, whether it's the halfway line, the distances from side to side, are as important as from front to back and back to front… it comes from, from practice, and players being good enough to do what they've been asked and recognise it. And also disciplined enough to do it as well."

Through discussions with a coach, these can act as transition between defensive states of organisation and pressure. However, the measurement presented to the coaches focuses on the concept of adjusting, a separate sub-theme identified. In the visualisation presented (figure 6.3), coach 3 believes the players are not adjusting properly.

"I don't think that's correct. I personally don't think the [Team 1] players are adjusting enough. Like for me, they need to be adjusting more aggressively, especially in the right back."

Interestingly, there was a difference of opinion between coach 2 and coach 3 in the example shown. Coach 2 is happy with large gaps appearing based on contextual information.

"we're quite happy for the distance between the centre back and the fullback to be there, because we know that that central midfielder can drop in there as well."

This suggests the need for systems based at clubs to be tailored to the coach's own principles and philosophies as there is no universal agreement on nuances held within each concept.

6.3.3.3 Support

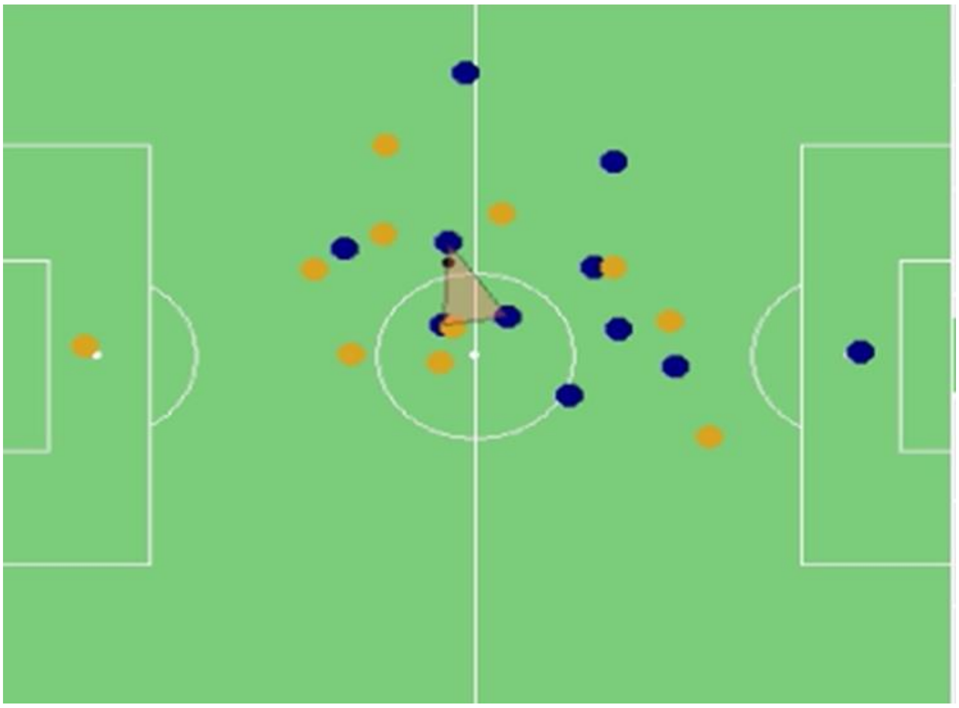


Figure 6.4: triangle of 3 midfield players.

The principle of support centres around how teammates position themselves as passing options to the player on the ball. Coach 1 highlighted the importance of angles.

> *"we play on those sorts of angles, you know, you've got that ability, you know, to see where the balls coming from, if it's coming from a deeper position and also the goal you want to attack so you can make a decision on how to use the ball next."*

This connects to another sub-theme identified as diamonds and triangles. The coaches emphasised these concepts are important structures created by the players on the pitch to help teams. Angles and distances have been used in multiple investigations, researching the coordination of player actions (Clemente et al, 2013c; Laakso et al, 2017; Laakso et al, 2019; Leser et al, 2019). Conceptualising players in groups of 3 and calculating the properties of the triangles their positions on the pitch form including distances, angles, areas and positions on the x axis were identified by the coaches as important for successful team cohesion. These properties are visualised in figure 6.4. Whilst measurements of distances and angles have predominantly been identified through dyadic relationships (Leser et al, 2017; Clemente et al, 2013c). Coaches agreed that the dynamics of teams in subgroups and triangle formation was an important aspect of their team's performance with specific emphasis on the distances and angles between the players.

The triangle described by three central midfielders was presented and was identified in the follow up interviews as the most critical triangle in the formation, however other triangles were also stated as useful. Coach 3 also highlighted that the triangular shape in the centre midfield can also be important when defending.

"…whether the triangles match, because not sometimes it's just say, my team are playing two holding midfielders and the number 10. So, in my, the way, I see the game that's triangle up and the other team might be playing triangle up as well, which means there's not, it's not man for man, the triangles don't match."

Therefore, triangles, and their relationship between attacking and defending teams may be important, however, specific measurement for how these are related to each other and what constitutes successful and unsuccessful organisation needs to be identified. Clemente et al (2013b) have previously investigated defensive triangles, specifically looking at the area. However, these measures have not been comprehensively explored.

6.3.3.4 Depth

The final theme and visualisation that resonated with the coaches was depth. this relates to the position of players along the pitch. In this sense, many coaches perceive "*lines*" in their team. Indeed, this aspect was presented to them in the initial interviews through group centroids along the x axis as shown in figure 6.5. This visualisation received positive feedback, however, coach 1 mentioned an alternative measurement that appears in the literature often named length or team length (Castellano et al, 2015; Praca et al, 2016; Tenga et al, 2016).

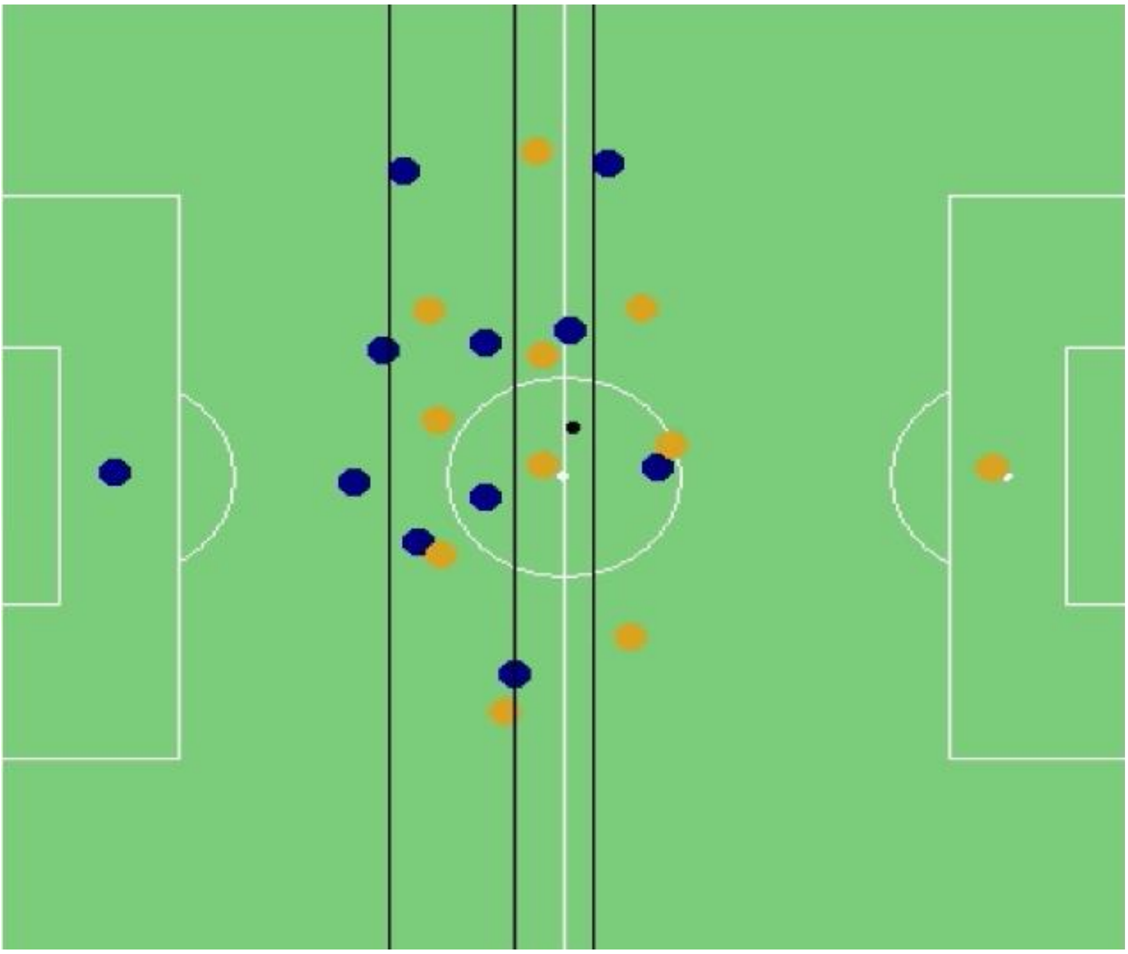


Figure 6.5. Group centroids representing lines across the pitch and team depth

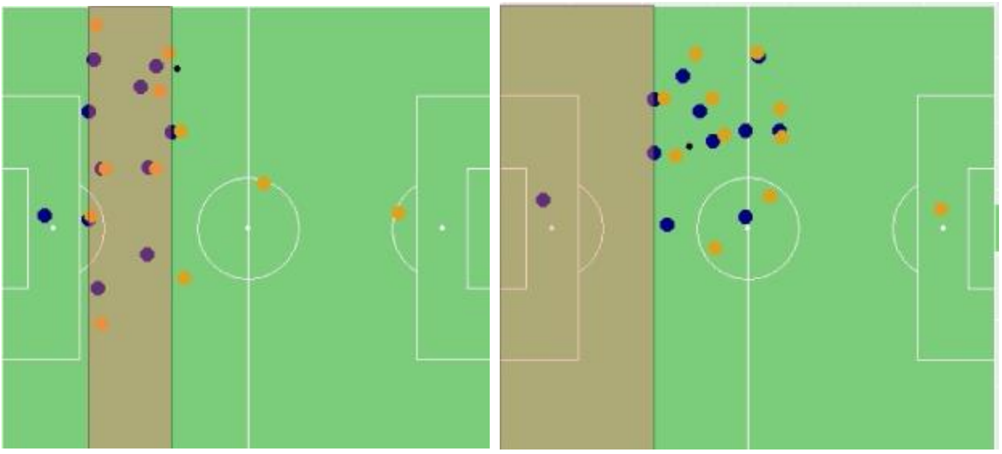


Figure 6.6 Team length and space in behind visualisations.

"I've seen similar ones where they kind of always have a constant distance from the deepest defender, you know, maybe one of your centre backs is behind the rest of the line. And the furthest forward player, you know, is that at 35 or 40 metres."

This measurement, taken as the distance between the furthest back and the furthest forward player was presented to coaches accompanied by the distance between the deepest defender and the goal line (Zubillaga et al, 2013) as shown in figure 6.6. In the second round of interviews, the coaches stated they actively coached this concept and that both visualisations aligned with their perception of the principle and were related to each other.

6.3.4 Relevant Metrics

Relevant metrics often involved more complex measurement, however, and were limited in their inability to fully encapsulate a concept or principle.

6.3.4.1 Compactness

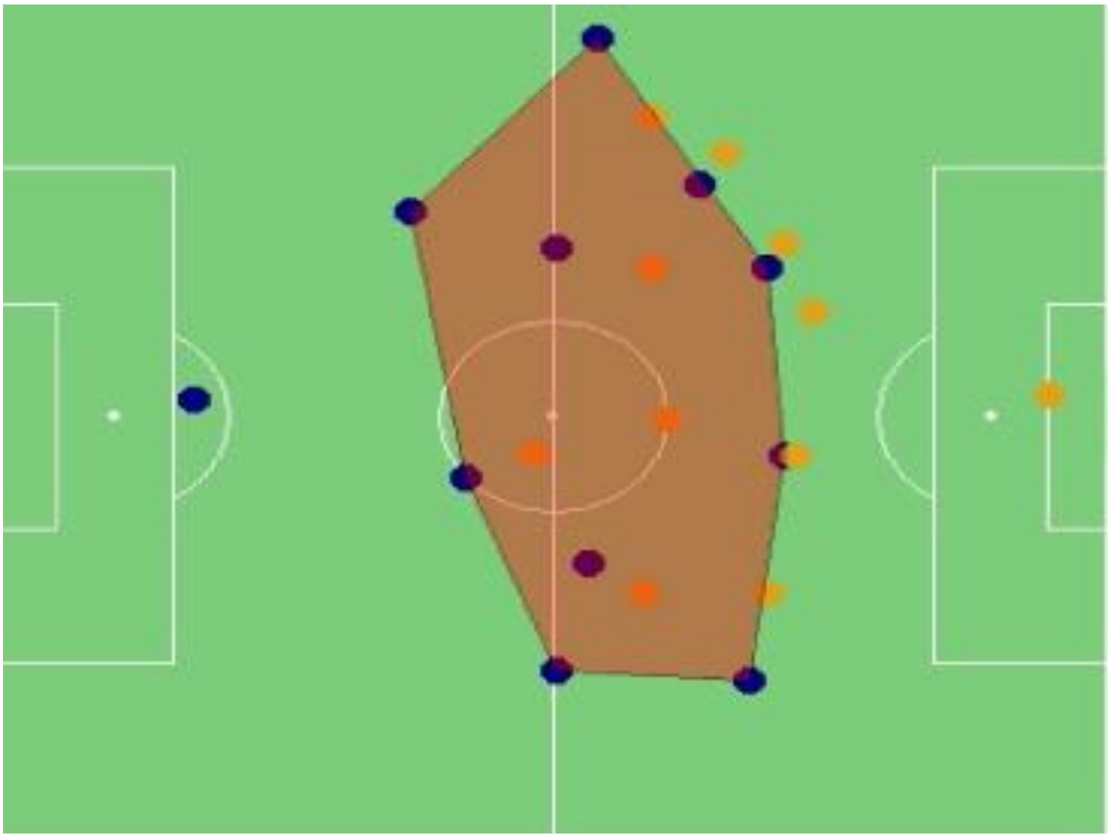


Figure 6.7: Surface area of outfield players

Common measures to evaluate the compactness of a team include surface area, stretch index and team spread (Aquino et al, 2016b; Baptista et al, 2018; Barnabe et al, 2018; Moura et al, 2012; Olthof et al, 2018). These metrics have demonstrated similarities in measurement patterns even when observing intricate attacks (Bartlett et al, 2012). The sub-theme of defensive shape was identified as a component of compactness, therefore, surface area was selected due to its alignment with this term (figure 6.7). However, simple analyses of surface area along with other measures have demonstrated that they are not sensitive enough to differentiate between successful and unsuccessful team compactness (Bartlett et al, 2012). To measure this principle in a meaningful way, the coaches' conceptualisation of it must be understood. Coach 3 highlighted the importance of returning into a defensive shape quickly after a transition.

"how quickly you can get back in shape after you lose the ball. And that is something that we coach."

When discussing these concepts, the coaches emphasised the importance of "anticipating" and "reacting to" the loss of possession. Therefore, these aspects are likely relevant when performing analysis on team compactness and defensive shape through surface area. The output signal of this measurement is the area encompassed by the outfield players as described in figure 6.7. Anticipation was measured by the difference between surface area at the loss of possession and 1 second before. Reaction was measured as the difference between the surface area at the loss of possession and 2 seconds afterwards. Finally, the time to get into a defensive shape was also recorded and measured as the time between losing possession to reaching a surface area of 600m^2. This value was selected based on previous data examining international teams surface area. The coaches agreed that this model makes sense, however, this value requires additional contextual information to be fully representative as coach 3 highlights that an immediate return into a defensive shape is not always desired by the coach.

"what is the objective? to get back into shape, and be compact as quickly as possible, like you're speaking about, or is it to try and win the ball back immediately and to actually counter press?"

6.3.4.2 Width

Width is a principle simplistically measured in the literature base (Baptista et al, 2018; Bartlett et al, 2012; Duarte et al, 2013a; Folgado et al, 2015; Tenga et al 2015). This metric is directly taken as the distance across the y axis from the player furthest right on the pitch and the player furthest left. This output is often combined with the team length measurement already discussed in the defensive principle and shown in figure 6.8.

Figure 6.8: Width of blue team described by distance between 2 widest players

Presenting this visualisation to coaches, they believed this was an important aspect of the game when attacking. However, the example provided was specifically chosen to be a situation where the team were not demonstrating high levels of team width but were still successful in scoring a goal. Coach 2 believed that they were performing complex actions due to the tactical set up of the opposition.

"…you can be really expansive in terms of your width and stuff like that. But if they sit in and are happy just to defend whatever comes in, then you have to start going in and trying to manipulate and get movements."

This suggests that applying width directly through players positioning themselves close to the edge of the pitch was not having the desired effect. One of the sub-themes of width is creating space, and having players in wide areas should facilitate the creation of space in central areas. Coach 3 highlights that in this situation, the attacking team still have the space to create viable passing options:

"…even though [team 2] are compact, there are still pass options through them available."

This proposes that width as measured in this example is not comprehensively evaluating the success of a team in destabilising the opponent. Incorporating the other sub-themes identified into this metric might provide a complete evaluation of how a team is using width to create space and penetrate through the opposition defence. Alternatively, incorporating another sub-theme of overloads in wide areas may evaluate a team's ability to penetrate in the wide areas.

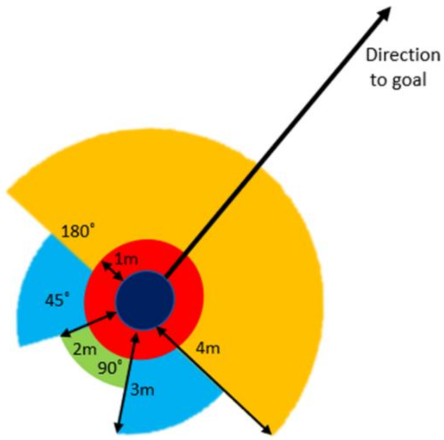
6.3.4.3 Delay

After the initial interviews, delay was identified as a theme. In the subsequent interviews and transcript analysis, delay was removed, however the following metric is still relevant to performance and fits closely into the sub-themes of discipline after the second iteration of the thematic analysis. The metric was initiated based on comments identifying the role that applying pressure plays in delaying the opponent. Coach 1 states.

"the first thing we have to do is delay the opposition from progressing towards our goal. So again, different applications doing that, you can apply pressure to, you know, the opponent…"

Across the three interviews, the coaches highlighted three ways in which a player on the ball can be placed under pressure. Most prominently, the distance between the players was emphasised as critical in delaying the opponent. However, other factors including the time a player had on the ball as well as how the out of possession team can dictate where the opponent plays through denying passing options. Three models were used and adapted to evaluate the total pressure that was being applied to the player. To evaluate the space pressure, the system devised by Link et al (2016) to measure pressure relating to danger was used. Time pressure was evaluated by the time it would take the closest defender to reach the player on the ball based on their current speed, similar to the time to contact calculation performed by Shafizadeh (2016). Finally, decision pressure was identified by how many simple passes to team mates were available. This was identified using Steiner's passing lane calculation whereby a simple pass required an angle > 10° for each player. Diagrams describing the calculations of space pressure and decision pressure can be seen in figure 6.9.

a)



b)

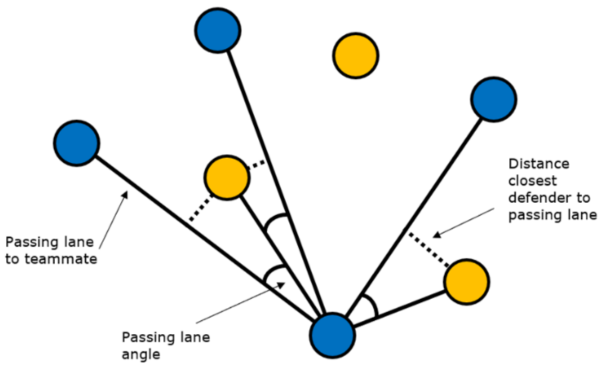


Figure 6.9: a) model to calculate space pressure. b) model to calculate decision pressure

The three measurements were scaled to represent very high pressure as the value approached 1 and very low pressure as the value approached 0. A weighting procedure was then intuitively applied where space, time and decision pressure values were multiplied by 0.7, 0.2 and 0.1 respectively before summing together to output the total pressure. An animated bar graph was presented to the coaches with the accompanying video footage and top-down x y coordinates of the players and ball on the football pitch (figure 6.10). The coaches stated that this made sense and agreed with the model as accurately describing the pressures on the pitch. However, the angular threshold of 10° for the decision pressure variable along with the weightings are not empirically supported and further data analysis would be required to refine this technique to be fully representative. These concepts can then be used to accurately understand the pressure that players are under when playing in matches and consequently tailor training to replicate what they will experience in the matches.

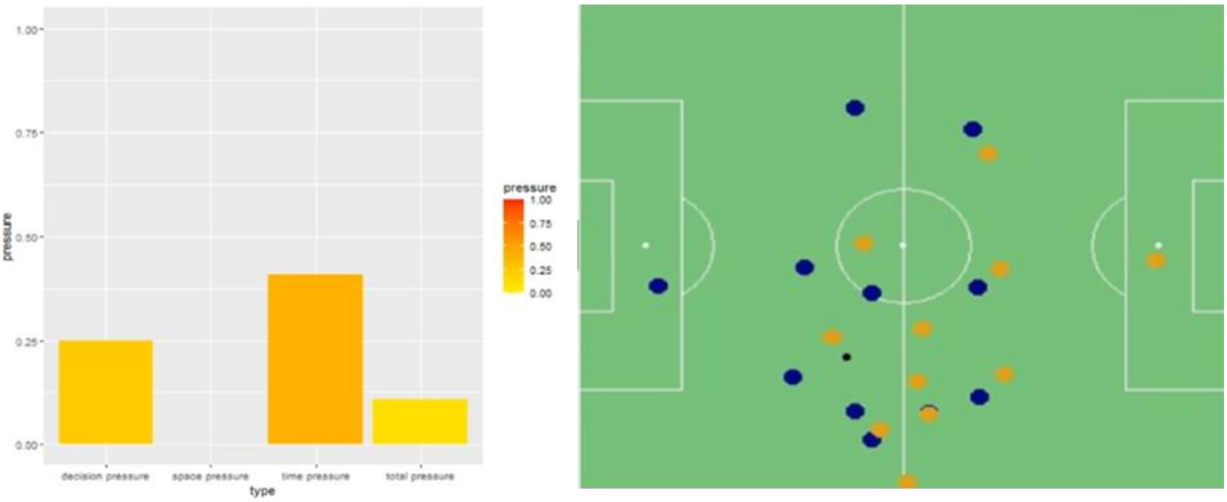


Figure 6.10: pressure chart along with top-down view of match at an instance.

6.3.5 Hesitant Metrics

Hesitant metrics focused on regions or zones on the pitch and were limited by their potential inaccuracy.

6.3.5.1 Penetration

Like delay, penetration was also changed from a theme to a main theme. Similarly, the proposed metrics may still be relevant, although likely needs adapted further as coaches were sceptical of its use. The number of outplayed opponents was used to describe actions that were penetrative adapted from Rein et al (2017). In their analysis, passes were examined to identify the difference in number of defenders closer to the goal line at the start and end of a pass. However, this value is outcome orientated and does not explain how a team has successfully progressed through the opposition and was used as a guide to identify instances deserving further analysis. Voronoi cell computations have been used to examine passing actions and behaviours of high-level teams when successfully penetrating opponents through creating space (Rein et al, 2017). This mathematical model identifies the areas on the pitch closest to each individual and its relevance aligns with a comment from coach 1.

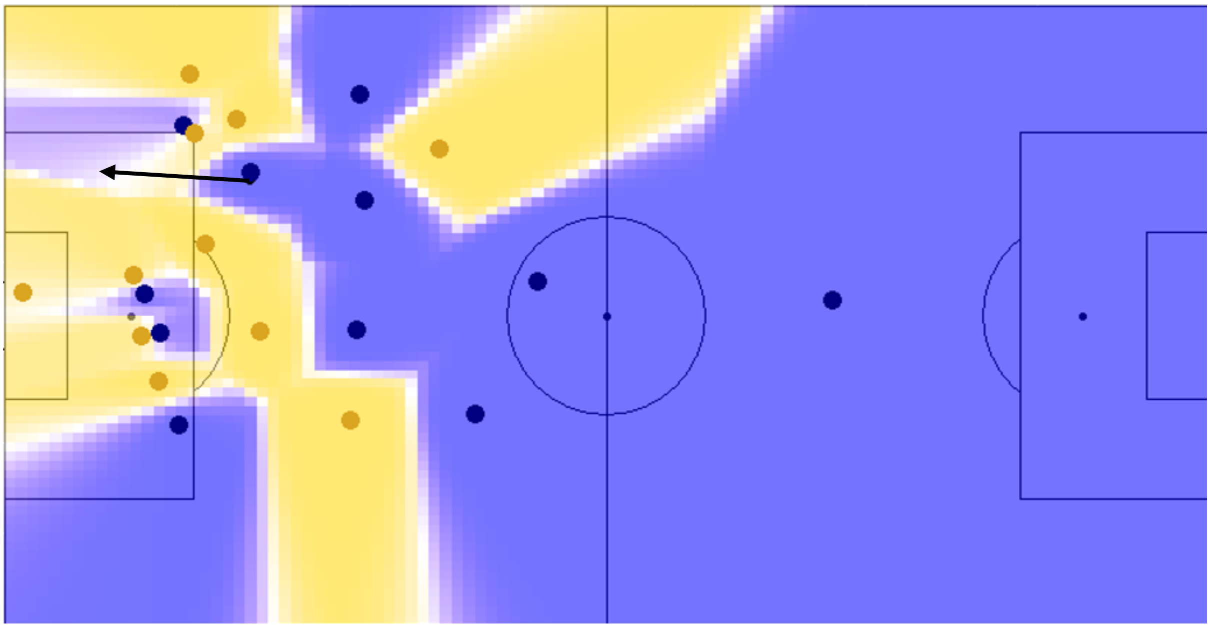


Figure 6.11: pitch control model showing a successful penetrative pass

"how can we get runs that will, in a sense destabilise, the opposition's organisation, and then use the ball to find those spaces or opportunities to penetrate."

Voronoi cell computations, or variations of the calculation termed as pitch control have been suggested as a means to identify likelihood of pass success based on the position a player is in and the space they occupy relative to everyone else (Spearman and Basye, 2017). Several unique calculations of Voronoi cell computations have been implemented across the literature, whereby player movement speed, player characteristics, the offside line and the ball trajectory have previously been implemented to evaluate actions such as passes (Brefeld et al, 2019; Fernandez et al, 2018; Filetti et al; 2017; Rein et al, 2017; Spearman and Basye, 2017). In the case presented to the coaches, a simple model whereby player speed was layered on top of positional data to identify areas of the pitch a player can pass the ball to successfully find a team mate. Figure 6.11 demonstrates the output of this model while estimating the probability of a successful penetrative pass that outplayed 6 opponents with a 55% likelihood.

When presenting this to the coaches, coach 3 was surprised by how low the success percentage of the pass was based on the calculation and how they perceived the pass in the video. That might indicate that a more sophisticated model is required to accurately predict the success rate of this pass. Moreover, the actual usefulness of this model for informing training practices is not obvious. Coach 2 emphasises that identifying and showing previous situations where this is done effectively or ineffectively does not necessarily translate to a player being able to identify opportunities and executing penetrative actions successfully.

"You can show lots of pictures of good examples. But at the end of the day, it comes down to quick time decision making and execution."

6.3.5.2 Balance

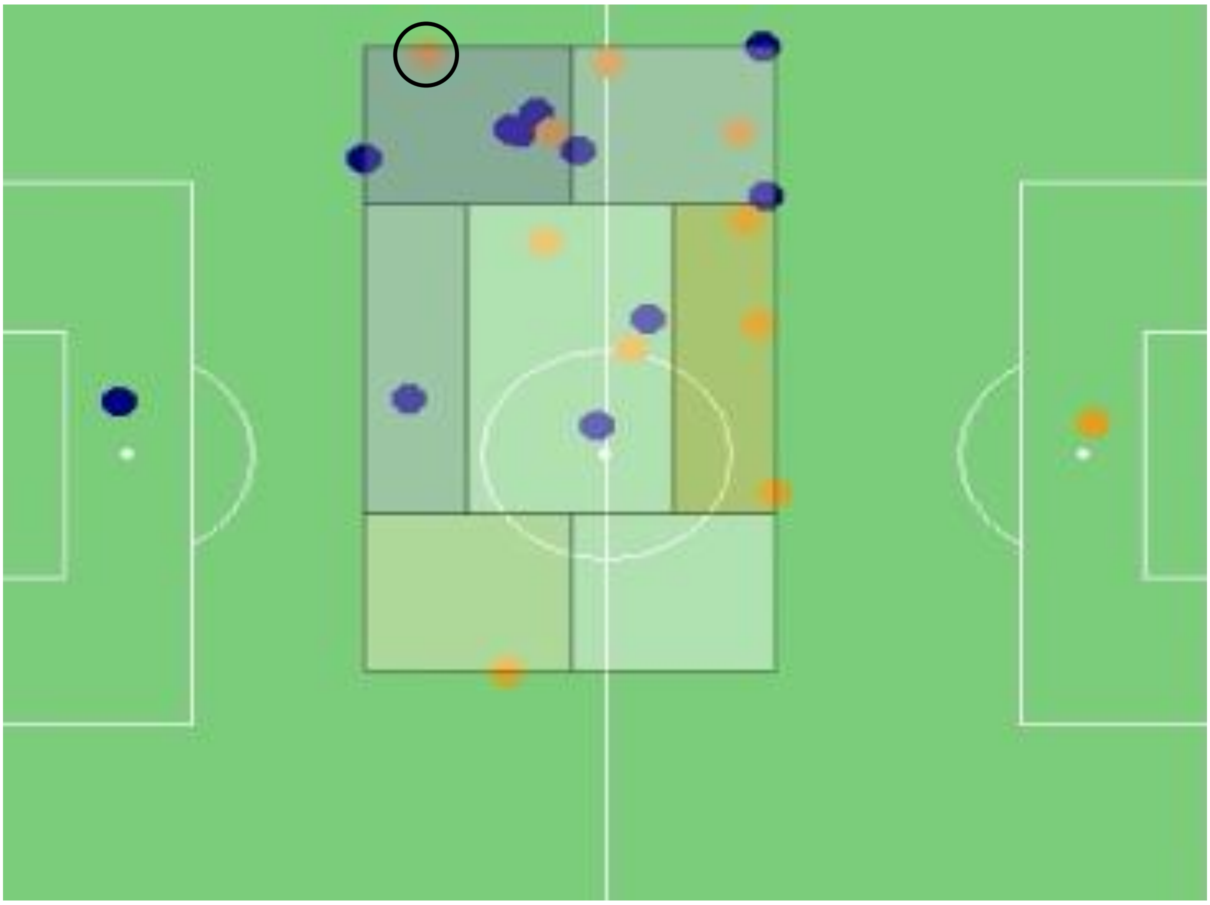


Figure 6.12: Balance measured through 7 dynamic zones based on numerical relations in each area

From the interviews coaches frequently discussed “*overloads”* as a tactically relevant concept. This occurs when a subgroup of players in a section of the pitch form numerical superiority in a game situation, for example creating a 2v1 or 3v2. This relates to the defensive principle of balance, where the defending team seeks to distribute their players so that the opposition is unable to create a numerical advantage. All coaches specifically mentioned trying to create overloads in the wide areas as an effective tactic to creating dangerous chances.

The coaches also identified that overloads in the middle of the pitch were also desirable but more challenging to create. Several different models of classifying zones for numerical advantage have been applied in the research. Clemente et al (2015d) used 12 static zones with 4 sections along the pitch and 3 sections across the pitch. However, the model selected to show to the coaches used 7 dynamic zones that shifted across the pitch relative to the positions of the outfield players (Vilar et al, 2013) as shown in figure 6.12.

The coaches believed this model was not fully representative of the situation presented to them. In the example shown in figure 6.12, where the highlighted player in the gold team has the ball, the numerical advantage is identified as a 4v2 in favour of the blue team. However, the coaches identified that they perceive this situation to be representative of a 1v1, which is recognised as precarious considering the location on the pitch. Although, coach 2 suggested that it may be useful with some refinement.

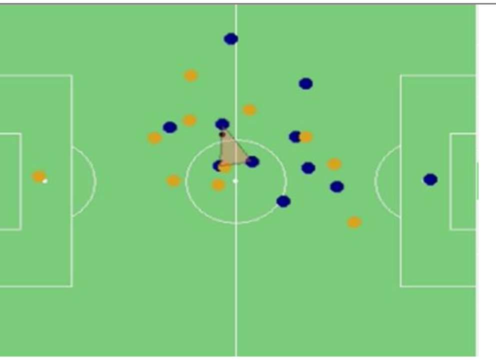
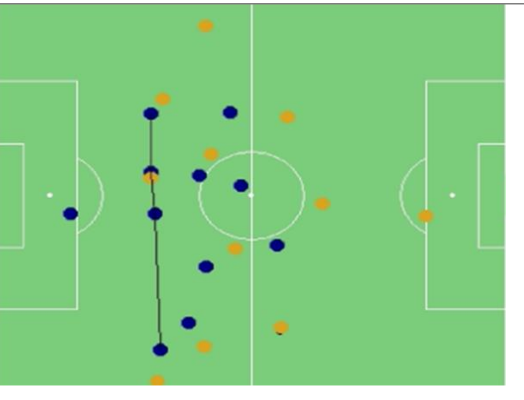
"I like the thought process of it. It's more of an active zone as opposed to static zones"

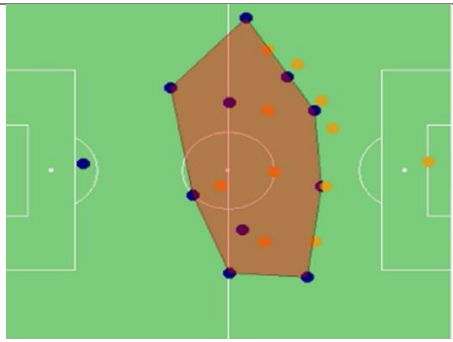
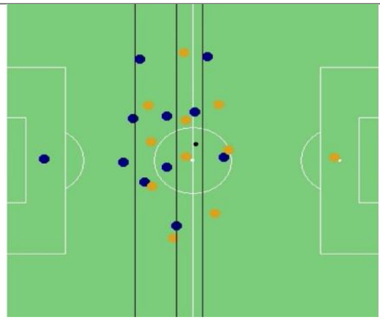
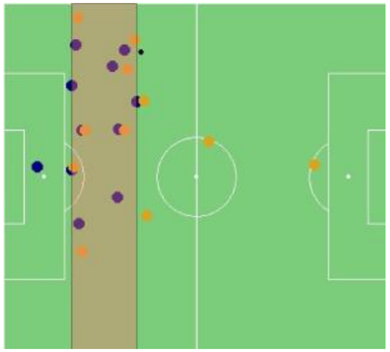
6.3.5.3 Creativity

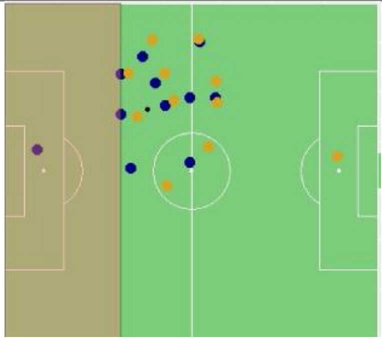
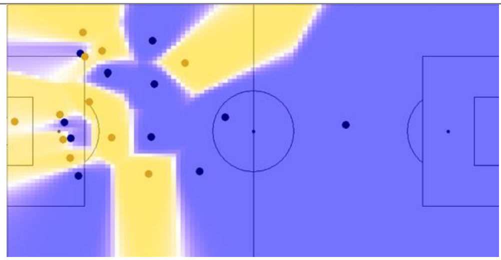
Creativity was a recurring theme throughout the interview process. However, from the first stage of interviews, there was no clear method of quantifying or representing the principle. In the second interviews coaches were asked to expand on the principle of creativity. In turn, the coaches identified that creative behaviours often lead to penetrative behaviour. Coach 2 states.

"I think when something is creative it penetrates a backline or the end result as potentially maybe getting in behind or creating an overload situation."

This indicates that perhaps metrics used for penetration might be helpful in quantifying some aspect of creativity. However, the coaches were hesitant on their value, so would require adaptation. Based on the sub-themes identified, other measurements could investigate the dynamics of 1v1 situations, as some research has already investigated (Duarte 2012a; Clemente et al, 2013c; Leser et al 2017). Additionally, the sub-theme of deception, might provide some insight into a team or groups ability to play through the rate of change in distance between team centroids (Memmert et al, 2016). Although, such a metric may not align with how coaches conceptualise such a principle.

Principle	Base metric	New Metric	New Metric description	Future work
Support	Angle and distance dyad, (Clemente et al, 2013c), and surface area (Duarte et al, 2012c)		Calculates distance, angles and area described by three selected players	Understand effective dynamics that relate to triangle characteristics
Delay	Pressure (Link et al, 2016), passing lane (Steiner et al, 2018), and time to contact (Shafizadeh et al, 2016)	No visualisation of metric	Calculate pressure based on distance of defenders to player on the ball, time the player has on the ball and the passing options a player has	Refine the weightings for each aspect to calculate total pressure and integrate with surface area measure to evaluate organisation after loss of possession
Discipline	Distance dyad (Clemente et al, 2013c)		Distance between defenders as they are organised across a line	Identify appropriate distances between defenders depending on pitch size and match situation.

Compactness	Convex hull (Bartlett et al, 2012)		Calculate area described by convex hull of outfielders. At loss of possession, identify how long it takes to recover to an area of 600m^2	Integrate with pressure metric to identify how well teams are able to counter press, or recover into shape after loss of possession, also investigate selected area measurement that describes teams being in shape
Depth	Group centroid (SIegle et al, , 2013) team length, and distance to goalkeeper (Zubilaga et al, 2013)	 	Average position on the x-axis of defenders, midfielders and attackers. Distance between furthers forward attacker and deepest defender, and distance between deepest defender and goal line	Investigation of how these variables interact with one another.

				
Mobility	Network intensity (Grund 2012)	No visualisation of metric	Time in seconds per pass of a possessions	Integrate factors including distance of pass and match situation (for example what third of the pitch the ball is in) to analysis
Width	Team width (Folgado et al, 2018c)		Distance between the furthest player right and the furthest player left in a team	Investigate the space available in the wide areas of the pitch and integrate these measures.
Penetration	Pitch control (Spearman et al, 2017)		Zones of a pitch, where if the ball is put into the region, a number estimates the likelihood of each team retaining possession	Coaches were hesitant of metric

Balance	Numerical advantage (Vilar et al, 2013)		Numerical difference of player numbers in Dynamic areas of pitch based on outfield player positions	Adjust edge of zone where ball is, players behind ball are not in the same zone
Creativity	N/A	N/A	N/A	Potentially investigate Entropy calculation and how they related to predictability

Table 6.4. adjustments of metrics, visualisation and potential future ways to improve further

6.4 Future applications

This methodology identified that novel metrics evaluating collective behaviour are representative of some of the concepts as understood by coaches. A critical question remains, how can these be used in a practical setting to inform training practices and improve performance? In a professional team setting this would involve establishing normative values in matches previously played. Working closely with the coaches, measurement of the selected metrics will be made during relevant training sessions and adapted accordingly. An understanding of how each of the metrics are manipulated by adjusting constraints is required to fully utilise this performance analysis tool. Long term observations can become relevant for nurturing youth players, creating pathways and learning experiences that prepare players appropriately for competing at the highest level possible. For future research involving the construction of novel metrics, researchers should make sure the concepts that are measured are related to coach perceptions. Collaborations of researchers and practitioners working with professional teams may help direct novel research output towards applicable results.

Across the interviews, the four measures that evoked the strongest positive reaction from the coaches appeared to be from mobility, discipline, support and depth. These relate to the metrics of network intensity, distance between defenders, triads, team length and space behind the defence. However, from the second iteration of the interviews, despite understanding and agreeing with the principles, further information was given to help shape and refine them to represent their principles more accurately. The process suggested in this investigation is highly iterative in a practical setting. The next steps in the development and implementation of the measurements require applying them to improve them further. From each mathematical model computing a concept, the output is a number, it is important that the number is meaningful to allow for decision making processes to be guided.

In a game such as football, the signal of the mathematical models is highly contextual and changing all the time. Understanding the subtleties of such a complex game from these measures are important for them to be applied correctly. Recording and filtering through contextual variables such as match score, opponent ability, pitch size and the team strategy may be able to explain some of the variance. Additional contextual factors including the current position and organisation of the team are also likely relevant. From the interviews and concepts discussed, many of the principles are interlinked. Therefore, it is likely each of the variables fluctuate based on each other. It could be suggested that measurements attempting to quantify the team's defensive organisation may indicate the success of the opposing attacking team. If their values are perceived as poor by the normative data and thresholds set, then it may suggest their opponents are attacking effectively. On the other hand, if the defensive values are stronger then it may suggest the attacking team are ineffective in their attempts to destabilise the defenders. Other solutions include the application of analysis techniques that observe the signal across time. Such approaches already appear in the literature with methods involving approximate entropy, relative phase and vector coding. Practitioners must be cautious when using these to make sure the underlying mathematical procedures output values that are meaningful and relevant to principles of play and team performance.

6.5 Conclusion

This investigation demonstrates a methodology for collaborating with coaches to create a unique and tailored performance analysis system that integrates novel measurements applying social network analysis and spatial temporal metrics to quantify principles of play. Coaches suggested that network intensity, distance between defenders, team length, space behind the defence and triads were the most promising measures. From the interviews with the coaches, it appeared that these models can be useful for improving team performance with specific emphasis on enhancing training sessions. Further iteration and practical application of the systems being used are required to maximise the utility of applying novel collective behaviour systems. To be fully successful, the models require integration with contextual variables to comprehensively describe and explain the decision-making processes in football.

CHAPTER 7

Pedagogical Support for Analysts and Researchers

This chapter provides a detailed overview of the processes involved in measuring the metrics discussed in this thesis. Current educational programmes such as sport and exercise science degrees may not fully prepare graduates to perform contemporary roles in fast paced and high-pressure environments of professional sports teams. Experience using programming languages have become vital for effective performance analysis and the ability to perform valuable procedures for coaching staff. The following chapter provides a detailed overview of the coding process to measure the metrics discussed in this thesis, providing information and guidance to practitioners, helping them implement and learn to code using R studio. This includes a summary of the packages and structure of the data sets used to perform the analyses, whereby many of the analyses should be understandable and actionable by those with limited experience in coding languages. A step-by-step guide to create top-down visualisations (a 2-dimensional overview) will be provided as well as instruction on how to animate metrics into animations. The data analysis procedures, written in R Studio, calculate variables of collective behaviour and describe principles of play through position, distances, spaces, numerical relations and network analysis along with combinations of these measurements. The code provided can be replicated to analyse a match, or part of a match to calculate tactical variables and create visualisations. R code blocks are included in this section so that practitioners can implement and adapt these methods for themselves. Before discussing the data and methodologies to create visualisations, a list of packages required for replicating the below code are provided along with some additional packages that maybe useful for similar further analysis.

```
#packages useful for manipulating, tidying and organising the data s
ets
library(dplyr)
library(tidyverse)
library(zoo)
library(purrr)

#used to calculate triangulations of polygon and surface area equati
on
library(deldir)

#packages used for plotting and animation
library(ggplot2) # standard plotting package
library(gganimate) # animation of multiple plots across additional v
ariable
library(ggsoccer) # creation of the football pitch diagram
library(ggalt) # required to use geom_encircle
library(av) # used to save visualisations as a video file

#other potential useful packages for different analysis
library(igraph) # used for network analysis and graph theory
library(TSEntropies) # used for entropy analysis across time
```

7.1 Data

This section provides an overview of the structure of the data sets used in the analysis. Predominantly, these are arranged in long form, an example of this format is shown below with the first 25 rows of a dataframe. Each row is represented by a player that is identified by their player.id and team. The home team are identified as `team` 1 and the away team are identified as `team` 0. The ball is identified with a `player.id` of 99 and is assigned team 4. The match state is classified as either `alive` where the ball is in play, or `dead` where the ball is out of play. The x and y coordinates of each player are described in the x and y columns respectively with the origin located in the bottom left corner of the pitch. The frequency of the data collection is 25Hz such that every 23 rows the time value increases 0.04s. This data format is useful for applying animations in ggplot.

```
head(longdata, n=25)

##                x     y time state team player.id
## 1559999  52.61 34.11 0.00 alive    4        99
## 11        3.69 34.19 0.00 alive    1         1
## 9        38.40  8.49 0.00 alive    1         2
## 18       40.58 55.74 0.00 alive    1         3
## 21       43.21 30.43 0.00 alive    1         4
## 13       33.74 29.20 0.00 alive    1         5
## 15       34.14 42.54 0.00 alive    1         6
## 20       51.76 61.32 0.00 alive    1         7
## 22       52.82 34.03 0.00 alive    1         8
## 10       52.53 44.94 0.00 alive    1         9
## 1        42.84 39.44 0.00 alive    1        10
## 5        52.24  7.14 0.00 alive    1        11
## 16       73.99 33.83 0.00 alive    0         2
## 2        62.19 28.34 0.00 alive    0         6
## 3        53.20 43.08 0.00 alive    0         7
## 19       56.64 24.55 0.00 alive    0         9
## 8       102.52 34.48 0.00 alive    0        12
## 12       61.65 35.41 0.00 alive    0        14
## 17       74.22 44.11 0.00 alive    0        15
## 4        65.42 58.76 0.00 alive    0        16
## 6        60.87 13.36 0.00 alive    0        18
## 14       53.49 49.53 0.00 alive    0        19
## 7        72.95 21.79 0.00 alive    0        23
## 2310000  52.10 33.72 0.04 alive    4        99
## 33        3.67 34.20 0.04 alive    1         1
```

Data arranged in the wide format can also be useful and is applied in a handful of metrics. Each row in wide format represents a different time measurement with columns representing the x or y coordinate of a player (with the preceding *a* or *h* identifying if the player is in the home or away team) or the ball.

```
head(widedata[,c(1:5,26,27,46,47)])

##      time  a12.x a12.y a14.x a14.y h10.x h10.y     x     y
## 1    0.00 102.52 34.48 61.65 35.41 42.84 39.44 52.61 34.11
## 23   0.04 102.52 34.50 61.65 35.40 42.84 39.44 52.10 33.72
## 45   0.08 102.52 34.53 61.64 35.39 42.84 39.45 51.59 33.32
```

```
## 67  0.12 102.52 34.54 61.62 35.39 42.85 39.44 51.13 33.13
## 89  0.16 102.52 34.55 61.62 35.38 42.85 39.43 50.69 32.94
## 111 0.20 102.51 34.56 61.61 35.38 42.85 39.42 50.24 32.74
```

Multiple variations of the long format data also exist and are shown below. *homedata* and *awaydata* contain the data for the players of the home or away team, respectively. *playerdata1* combines these and includes players from both teams but not the ball. Finally, *matchdata* has a row for all the players on the pitch, however, the ball data is also encapsulated in this dataframe through the columns *ballx* and *bally*.

```r
head(datahome, n = 12)
```

```
##        x     y time state team player.id
## 11  3.69 34.19 0.00 alive    1         1
## 9  38.40  8.49 0.00 alive    1         2
## 18 40.58 55.74 0.00 alive    1         3
## 21 43.21 30.43 0.00 alive    1         4
## 13 33.74 29.20 0.00 alive    1         5
## 15 34.14 42.54 0.00 alive    1         6
## 20 51.76 61.32 0.00 alive    1         7
## 22 52.82 34.03 0.00 alive    1         8
## 10 52.53 44.94 0.00 alive    1         9
## 1  42.84 39.44 0.00 alive    1        10
## 5  52.24  7.14 0.00 alive    1        11
## 33  3.67 34.20 0.04 alive    1         1
```

```r
head(dataaway, n = 12)
```

```
##         x     y time state team player.id
## 16  73.99 33.83 0.00 alive    0         2
## 2   62.19 28.34 0.00 alive    0         6
## 3   53.20 43.08 0.00 alive    0         7
## 19  56.64 24.55 0.00 alive    0         9
## 8  102.52 34.48 0.00 alive    0        12
## 12  61.65 35.41 0.00 alive    0        14
## 17  74.22 44.11 0.00 alive    0        15
## 4   65.42 58.76 0.00 alive    0        16
## 6   60.87 13.36 0.00 alive    0        18
## 14  53.49 49.53 0.00 alive    0        19
## 7   72.95 21.79 0.00 alive    0        23
## 38  73.98 33.81 0.04 alive    0         2
```

```r
head(playerdata1, n = 25)
```

```
##          x     y time   state team player.id
## 1   42.84 39.44 0.00   Alive    1        10
## 2   62.19 28.34 0.00   Alive    0         6
## 3   53.20 43.08 0.00   Alive    0         7
## 4   65.42 58.76 0.00   Alive    0        16
## 5   52.24  7.14 0.00   Alive    1        11
## 6   60.87 13.36 0.00   Alive    0        18
## 7   72.95 21.79 0.00   Alive    0        23
## 8  102.52 34.48 0.00   Alive    0        12
## 9   38.40  8.49 0.00   Alive    1         2
## 10  52.53 44.94 0.00   Alive    1         9
## 11   3.69 34.19 0.00   Alive    1         1
## 12  61.65 35.41 0.00   Alive    0        14
## 13  33.74 29.20 0.00   Alive    1         5
## 14  53.49 49.53 0.00   Alive    0        19
## 15  34.14 42.54 0.00   Alive    1         6
## 16  73.99 33.83 0.00   Alive    0         2
## 17  74.22 44.11 0.00   Alive    0        15
## 18  40.58 55.74 0.00   Alive    1         3
## 19  56.64 24.55 0.00   Alive    0         9
## 20  51.76 61.32 0.00   Alive    1         7
## 21  43.21 30.43 0.00   Alive    1         4
## 22  52.82 34.03 0.00   Alive    1         8
## 23  42.84 39.44 0.04 Alive;:    1        10
## 24  62.18 28.29 0.04 Alive;:    0         6
## 25  53.19 43.10 0.04 Alive;:    0         7
```

```r
head(matchdata, n = 25)
```

```
##            x     y time   state team player.id ballx bally
## 1     42.84 39.44 0.00   Alive    1        10 52.61 34.11
## 1.1   62.19 28.34 0.00   Alive    0         6 52.61 34.11
## 1.2   53.20 43.08 0.00   Alive    0         7 52.61 34.11
## 1.3   65.42 58.76 0.00   Alive    0        16 52.61 34.11
## 1.4   52.24  7.14 0.00   Alive    1        11 52.61 34.11
## 1.5   60.87 13.36 0.00   Alive    0        18 52.61 34.11
## 1.6   72.95 21.79 0.00   Alive    0        23 52.61 34.11
## 1.7  102.52 34.48 0.00   Alive    0        12 52.61 34.11
## 1.8   38.40  8.49 0.00   Alive    1         2 52.61 34.11
## 1.9   52.53 44.94 0.00   Alive    1         9 52.61 34.11
## 1.10   3.69 34.19 0.00   Alive    1         1 52.61 34.11
## 1.11  61.65 35.41 0.00   Alive    0        14 52.61 34.11
## 1.12  33.74 29.20 0.00   Alive    1         5 52.61 34.11
## 1.13  53.49 49.53 0.00   Alive    0        19 52.61 34.11
## 1.14  34.14 42.54 0.00   Alive    1         6 52.61 34.11
## 1.15  73.99 33.83 0.00   Alive    0         2 52.61 34.11
## 1.16  74.22 44.11 0.00   Alive    0        15 52.61 34.11
## 1.17  40.58 55.74 0.00   Alive    1         3 52.61 34.11
```

```
## 1.18  56.64 24.55 0.00   Alive    0         9 52.61 34.11
## 1.19  51.76 61.32 0.00   Alive    1         7 52.61 34.11
## 1.20  43.21 30.43 0.00   Alive    1         4 52.61 34.11
## 1.21  52.82 34.03 0.00   Alive    1         8 52.61 34.11
## 23    42.84 39.44 0.04 Alive;:    1        10 52.10 33.72
## 23.1  62.18 28.29 0.04 Alive;:    0         6 52.10 33.72
## 23.2  53.19 43.10 0.04 Alive;:    0         7 52.10 33.72
```

Some of the analysis techniques used require this additional data that includes every instance a player touches the ball, the time codes for identifying which team is in possession or when the ball is out of play, and instances where a pass is successfully received. These provide necessary context for some of the metrics discussed later in the chapter. The `possession` dataframe is a csv file from performance analysis software nacsport. The critical aspects of this dataframe include the Category column, that identifies the instance a player touches the ball, or whether a team is in possession of the ball, or the ball is out of play. The Start column indicates the instance or start of an action. A touch is perceived as instantaneous, whereas possession has a relevant start and end time. The format of this time is in (MM:SS:MS) representing the time on the video clip the moment was tagged.

```
head(possession, n = 20)
```

```
##    N.  Category Descriptors    Start    Click      End Notes Dura
tion
## 1   1     touch          NA 47:41:96 47:41:96 47:42:96    NA 00:0
1:00
## 2   1 team2 pos          NA 47:41:96 47:41:96 48:02:48    NA 00:2
0:52
## 3   2     touch          NA 47:42:88 47:42:88 47:43:88    NA 00:0
1:00
## 4   3     touch          NA 47:45:44 47:45:44 47:46:44    NA 00:0
1:00
## 5   4     touch          NA 47:46:20 47:46:20 47:47:20    NA 00:0
1:00
```

```
## 6   5      touch           NA 47:48:36 47:48:36 47:49:36    NA 00:0
1:00
## 7   6      touch           NA 47:52:16 47:52:16 47:53:16    NA 00:0
1:00
## 8   7      touch           NA 47:53:00 47:53:00 47:54:00    NA 00:0
1:00
## 9   8      touch           NA 47:55:36 47:55:36 47:56:36    NA 00:0
1:00
## 10  9      touch           NA 47:57:48 47:57:48 47:58:48    NA 00:0
1:00
## 11 10      touch           NA 47:58:24 47:58:24 47:59:24    NA 00:0
1:00
## 12 11      touch           NA 48:00:80 48:00:80 48:01:80    NA 00:0
1:00
## 13 12      touch           NA 48:01:76 48:01:76 48:02:76    NA 00:0
1:00
## 14 13      touch           NA 48:02:48 48:02:48 48:03:48    NA 00:0
1:00
## 15  1 team1 pos            NA 48:02:48 48:02:48 49:26:88    NA 01:2
4:40
## 16 14      touch           NA 48:03:12 48:03:12 48:04:12    NA 00:0
1:00
## 17 15      touch           NA 48:03:64 48:03:64 48:04:64    NA 00:0
1:00
## 18 16      touch           NA 48:04:92 48:04:92 48:05:92    NA 00:0
1:00
## 19 17      touch           NA 48:06:92 48:06:92 48:07:92    NA 00:0
1:00
## 20 18      touch           NA 48:07:64 48:07:64 48:08:64    NA 00:0
1:00
```

The *contact* dataframe has 3 columns, the time column identifying the time at each instance, the contact column, identifying if a contact was made at that time (1) or not (0) and the *contact.num* column indicating how many contacts of the ball had happened up to that point.

```
head(contact, n = 10)

##    time contact contact.num
## 1  0.00       1           1
## 2  0.04       0           1
## 3  0.08       0           1
```

```
## 4  0.12       0           1
## 5  0.16       0           1
## 6  0.20       0           1
## 7  0.24       0           1
## 8  0.28       0           1
## 9  0.32       0           1
## 10 0.36       0           1
```

The *passing* dataframe identifies the time when a teammate successfully receives an attempted pass, this is identified in each row, with both teams accounted for along with the Start column identifying the instance when this happened, again in the (MM:SS:MS) format.

```
head(passing, n = 10)

##    N.            Category Descriptors    Start    Click      End
Notes Duration
## 1   1 team2 pass received          NA 47:37:88 47:42:88 47:47:88
NA 00:10:00
## 2   2 team2 pass received          NA 47:40:32 47:45:32 47:50:32
NA 00:10:00
## 3   3 team2 pass received          NA 47:43:36 47:48:36 47:53:36
NA 00:10:00
## 4   4 team2 pass received          NA 47:47:12 47:52:12 47:57:12
NA 00:10:00
## 5   1 team1 pass received          NA 47:57:48 48:02:48 48:07:48
NA 00:10:00
## 6   2 team1 pass received          NA 48:01:84 48:06:84 48:11:84
NA 00:10:00
## 7   3 team1 pass received          NA 48:04:36 48:09:36 48:14:36
NA 00:10:00
## 8   4 team1 pass received          NA 48:07:96 48:12:96 48:17:96
NA 00:10:00
## 9   5 team1 pass received          NA 48:10:68 48:15:68 48:20:68
NA 00:10:00
## 10  6 team1 pass received          NA 48:19:68 48:24:68 48:29:68
NA 00:10:00
```

The final dataframe used in this analysis is *timetouch*, this includes 3 columns, X as a reference for each individual touch, the time, referring to the time of each touch and the touch column, indicating a touch occurred at this time.

```
tail(timetouch, n = 5)
```

```
##          X    time touch
## 1146 1146 2761.28     1
## 1147 1147 2772.24     1
## 1148 1148 2774.24     1
## 1149 1149 2776.72     1
## 1150 1150 2778.72     1
```

7.2 Visualisation

To create visualisations of the metrics, the pitch, the players, the ball, and the metric illustration must be animated and layered correctly. The football pitch is the first step, and a custom field is created using the following code. The measurements below provide spatial references for the pitch in meters as this is how the x and y positional data is measured.

```
# Create a pitch plot
pitch_custom <- list(
  length = 105,
  width = 68,
  penalty_box_length = 16.5,
  penalty_box_width = 40.3,
  six_yard_box_length = 5.5,
  six_yard_box_width = 18.3,
  penalty_spot_distance = 11,
  goal_width = 7.32,
  origin_x = 0,
  origin_y = 0
)
```

To perform a visualisation, a section of the data is extracted into a new dataframe

```
#use 30 seconds of long form data
data <- longdata %>% filter(time >= 886.56 & time <= 916.56)
```

The next section of code instructs R on how to plot the data. The first line identifies that the dataframe *data* as created above is to be used. The columns of interest are labelled *x* and *y*. The *annotate_pitch* line uses the previously described custom pitch and instructs R to fill the field palegreen3 and colour the lines in white. The *geom_point* command transforms the x and y coordinates to dots and sets the colour and size of the points dependent on the team value for each row of the dataframe. The next two lines describe how to attribute colour and size to the points. *scale_color_manual* fills the players with team value of 0 as *goldenrod*, 1 as *navyblue* and the ball, which is team 4 as *grey1*. The *scale_size_manual*, then does similar, giving values of 5 for players with a team value of 1 or 0, and giving the ball, which has a team value of 4 a smaller size of 2. Finally, the legend is removed from the visualisation in the last line of code. Notice that the plot command is assigned to the element p.

```
#create graph of points that
p <- ggplot(data, aes(x, y)) +
  annotate_pitch(dimensions = pitch_custom, fill = "palegreen3", col
our = "white") +
  geom_point(aes(color = team, size = team)) +
  scale_color_manual(values=c("goldenrod", "navyblue", "grey1")) +
  scale_size_manual(values=c(5, 5, 2)) +
  theme(legend.position = 'none')
```

The *p* element is then animated by instructing r to run the plot *p* where the *transition_time* is identified as the time column in the dataframe. A title label describing the time is placed using the 2nd line of code and the final line fixes the plot so there is no movement across it. The animation is created using the animate function, providing the frame per second as 25 mirroring the data collection frequency and the total frames as 750 (30 seconds * 25 fps). The renderer *av_renderer()* is used to save the visualisation as a video clip.

```
#animate graph
anim <- p + transition_time(time) +
  labs(title = "time: {frame_time}") +
  view_follow(fixed_y = TRUE)
```

```
#creat animation
animate(anim, nframes = 750, fps = 25, renderer = av_renderer())
```

The above code renders a 30 second video clip of the football pitch, the players and the ball. Several frames of the video is shown below and provides the basis for the majority of future visualisations.

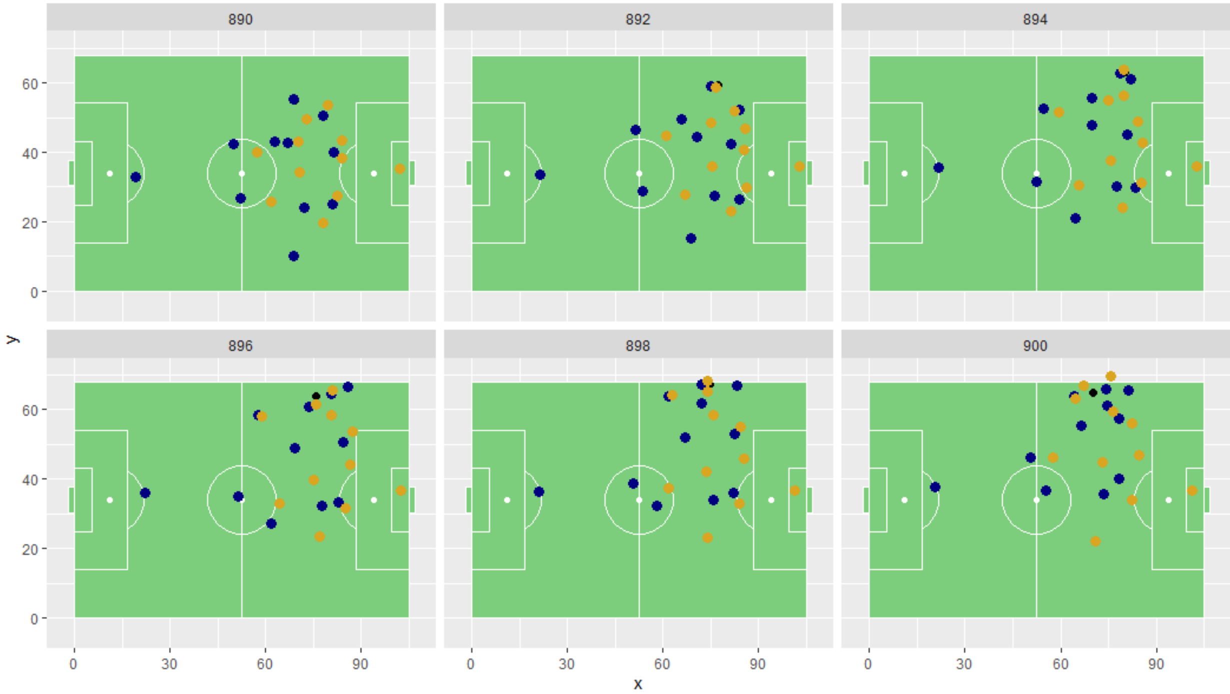


Figure 7.1. Top-down view of pitch showing players of two teams (blue and gold) and the ball in black across 2 second intervals.

7.3 Collective Behaviour metrics

The following section will provide guidance on how to calculate metrics. Coaching principles are briefly described and step by step instructions are provided for practitioners to replicate metrics that are representative of the concepts. These will be ordered through variables applying position, distances, spaces, numerical relations, network analysis and finally metrics that incorporate a combination of these properties in the calculation

7.3.1 Position

7.3.1.1 Team Length & Team Depth

In football, the team length and depth relates to the positioning of players along the pitch. One method that was identified as relevant for coaches was the difference in position on the x axis between the furthest back player and furthest forward player in a team. Additionally, the position of the deepest defender relative to the goal line is also calculated. The output is shown in figure 7.2 and a simple process can be applied to measure this metric. The following section of code extracts home player data and removes the goalkeeper data. The next step identifies the position of the furthest forward attacker (`max`) and deepest defender (`min`) and calculates the difference between them

```
# Depth

#length as furthest distance between players on x axis (outfield) and dept
h  as distance behind defender
length <- datahome %>% filter(player.id != 1) %>% group_by(time) %>%
  mutate(length = max(x)) %>% mutate(depth = min(x)) %>%
  mutate(teamlength = max(x)- min(x))
```

To animate this on the top-down pitch view, 4 points are required to describe the areas as a rectangle on the pitch.

```
length <- length[rep(seq(1,nrow(length),10), each = 4),] # get only 4 data
points per time
length$ref = rep(seq(1,4,1), nrow(length)/4) # create reference values for
data
```

3 new columns are created. TLx identifies the x value for the team length in the rows referenced as 1 and 2 and the depth value in rows referenced 3 and 4 for each instance. DTGK identifies the x value as 0 (the goal line) for row references 1 and 2 and again the depth value for row references 3 and 4. Finally, the edge mutation creates a column of repeating 0's in row references 1 and 3 and 68 in row references 2 and 4. This allows for the visualisations to create spaces for the team length and the space in behind the goalkeeper.

```
#adjust data so for visualisation - make calculation of points for x and y
of spaces
length <- length %>% mutate(TLx = case_when(ref == 1 | ref == 2 ~ length,
                                            T ~ depth)) %>%
  mutate(DTGK = case_when(ref == 1 | ref == 2 ~ 0,
         T ~ depth)) %>%
  mutate(edge = case_when(ref == 1 | ref == 3 ~ 0,
                          T ~ 68))
```

To analysing a specific period of play, data is extracted for the time of the attack.

```
length <- length %>% filter(time >= 1819.56 & time < 1869.56) # data for
period of play analysed
data <- longdata %>% filter(time >= 1819.56 & time < 1869.56)
```

The following two animations create similar plots where the annotate pitch function creates a pitch, *geom_point* adds the players and ball as a dot on the pitch. *geom_encircle* creates a space around the points in the data set to create an area describing the team length or the space in behind the defence.

```
#visualisation of team length
p <- ggplot(data, aes(x, y)) +
  annotate_pitch(dimensions = pitch_custom, fill = "palegreen3", colour =
"white") +
  geom_point(aes(color = team, size = team)) +
  scale_color_manual(values=c("goldenrod", "navyblue", "grey1")) +
  scale_size_manual(values=c(5, 5, 2)) +
  geom_encircle(data = length, aes(x = TLx, y = edge, fill = team),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
```

```
FALSE) +
  theme(legend.position = 'none')



anim = p + transition_manual(frames = time) +
  labs(title = "team length")



animate(anim, nframes = 1250, fps = 25, renderer = av_renderer())

### visualisation of distance to goalkeeper (area behind player)
p <- ggplot(data, aes(x, y)) +
  annotate_pitch(dimensions = pitch_custom, fill = "palegreen3", colour =
"white") +
  geom_point(aes(color = team, size = team)) +
  scale_color_manual(values=c("goldenrod", "navyblue", "grey1")) +
  scale_size_manual(values=c(5, 5, 2)) +
  geom_encircle(data = length, aes(x = DTGK, y = edge, fill = team),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
FALSE) +
  theme(legend.position = 'none')



anim = p + transition_manual(frames = time) +
  labs(title = "space in behind")



animate(anim, nframes = 1250, fps = 25, renderer = av_renderer())
```

a) b)

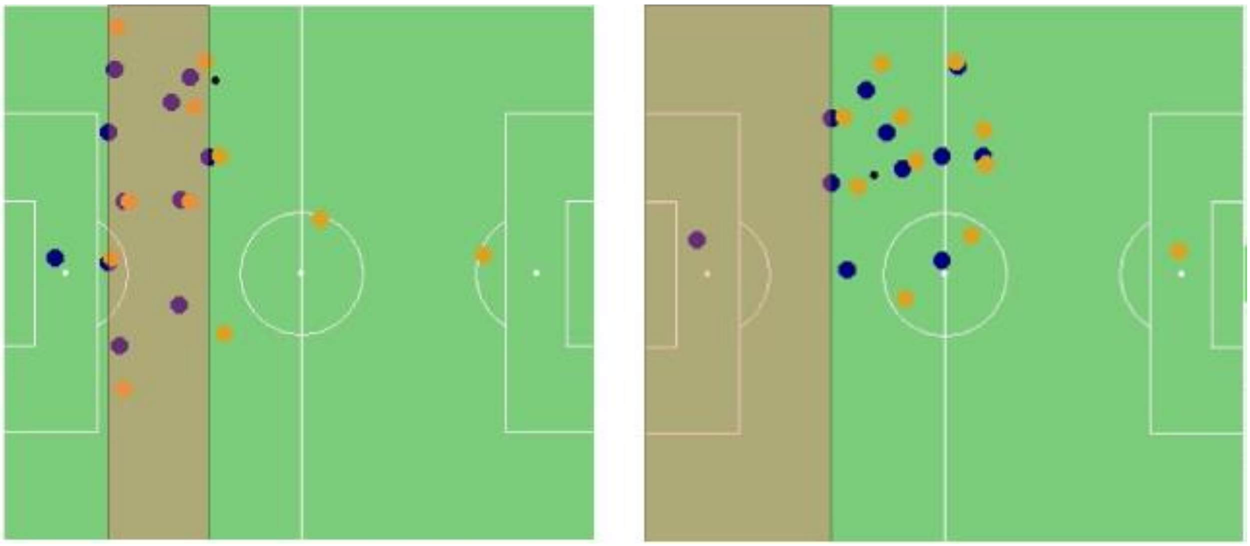


Figure 7.2: a) Team length described by distance between deepest defender and furthest attacker. B) Space behind defensive line described by the distance between the deepest defender and the goal line.

7.3.1.2 Team width

Similarly, to team length and depth, team width examines the positioning of the players across the pitch and is scrutinised during attacking phases of play. The difference between the positions of the furthest right (*max*) and left (*min*) players are calculated below.

```
width <- matchdata %>% filter(player.id != 1 & player.id != 12) # remove g
oalkeepers
width <- width %>% group_by(time,team) %>% mutate(width = max(y) - min(y))
```

for the visualisation, the team length and width included and are calculated as below.

```
width <- matchdata %>% group_by(time,team) %>%   mutate(length = max(x)) %
>% mutate(depth = min(x)) %>%
  mutate(right = min(y)) %>% mutate(left = max(y))
```

One team is isolated for the visualisation, and the data frame is split into 4 instances per time period. a reference of 1, 2, 3 or 4 is added to each time point and the x and y coordinates to describe the width are placed into new columns

```
width <- width %>% filter(team == 1) %>% filter(player.id != 1)

width <- width[rep(seq(1,nrow(width),11), each = 4),] # get only 4 data po
ints per time
width$ref = rep(seq(1,4,1), nrow(width)/4) # create reference vaues for da
ta

#adjust data so for visualisation - make calculation of points for x and y
of spaces
width <- width %>% mutate(x.points = case_when(ref == 1 | ref == 2 ~ lengt
h,
                                              T ~ depth)) %>%
  mutate(y.points = case_when(ref == 1 | ref == 3 ~ right,
                              T ~ left))
```

A time period is selected for analysis. *annotate_pitch* creates the pitch diagram on the plot, *geom_point* draws the players and the balls as dots and *geom_encircle* creates a shape that is defined by the x points and y points that were created in previous lines of code.

```
width <- width %>% filter(time >= 1415 & time < 1427) # data for period of
play analysed
data <- longdata %>% filter(time >= 1415 & time < 1427)

#visualisation of team length
p <- ggplot(data, aes(x, y)) +
  annotate_pitch(dimensions = pitch_custom, fill = "palegreen3", colour =
"white") +
  geom_point(aes(color = team, size = team)) +
  scale_color_manual(values=c("goldenrod", "navyblue", "grey1")) +
  scale_size_manual(values=c(5, 5, 2)) +
  geom_encircle(data = width, aes(x = x.points, y = y.points, fill = team)
,
              s_shape = 1, expand = 0,
              alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
FALSE) +
  theme(legend.position = 'none')



anim = p + transition_manual(frames = time) +
  labs(title = "team length")



animate(anim, nframes = 300, fps = 25, renderer = av_renderer())
```

Figure 7.3: The width of a team described by the distance between the distance between the furthest right and furthest left players.

7.3.2 Distance

7.3.2.2 Distance between defenders

The distance between defenders is representative of the discipline principle, encouraging players to remain equidistant from each other and acting as a unit. The following code outlines how to calculate the distances and create the visualisation for the metric. This section uses only the home team data and extracts the players identified in the defensive line only. Figure 7.4 provides a visualisation of the distances on the pitch

```
####### Discipline & patience #######

#create dataframe with only defenders
defence <- datahome %>% filter(player.id == 2 | player.id == 3 | player.id
== 5 | player.id == 6)
```

The data frame is then ordered to arrange the players by time, and then from the player furthest right to furthest left in each instant.

```
#order players depending on y axis position
defence <- defence[
  with(defence, order(time, y)),
]

order <- rep(seq(1,4,1),nrow(defence)/4)

defence <- cbind(defence,order)
```

The distances are then calculated for the defensive line. *dist1* is the distance between the first defender to the second defender, *dist2* is the distance between the second defender to the third defender and *dist3* is the distance between the third defender and the fourth defender.

```
defence <- defence %>% group_by(time) %>%
  mutate(dist1 = sqrt((x[which(order == 1)] - x[which(order == 2)])^2 +
           (y[which(order == 1)] - y[which(order == 2)])^2)) %>%
  mutate(dist2 = sqrt((x[which(order == 2)] - x[which(order == 3)])^2 +
                        (y[which(order == 2)] - y[which(order == 3)])^2))
%>%
  mutate(dist3 = sqrt((x[which(order == 3)] - x[which(order == 4)])^2 +
                        (y[which(order == 3)] - y[which(order == 4)])^2))
```

The section of data analysed is extracted for the visualisation, including player and ball data. Using *annotate_pitch* creates the custom football pitch described earlier. *geom_point* represents the players and ball as points and are customised to their shirt colour and size with the ball being smaller. *geom_line* identifies the lines between the defensive line and as oriented so that it goes along the y axis.

```
data <- longdata %>% filter(time >= 1333.56 & time <= 1383.56)


p <- ggplot(data, aes(x, y)) +
```

```
  annotate_pitch(dimensions = pitch_custom, fill = "palegreen3", colour =
"white") +
  geom_point(aes(color = team, size = team)) +
  scale_color_manual(values=c("goldenrod", "navyblue", "grey1")) +
  scale_size_manual(values=c(5, 5, 2)) +
  theme(legend.position = 'none')+
  geom_line(data = data[data$team == 1 & (data$player.id == 2 | data$playe
r.id == 3 | data$player.id == 5 | data$player.id == 6),], orientation = "y
")



anim = p + transition_manual(frames = time) +
  labs(title = "distance ebtween defenders")

animate(anim, nframes = 1250, fps = 25, renderer = av_renderer())
```

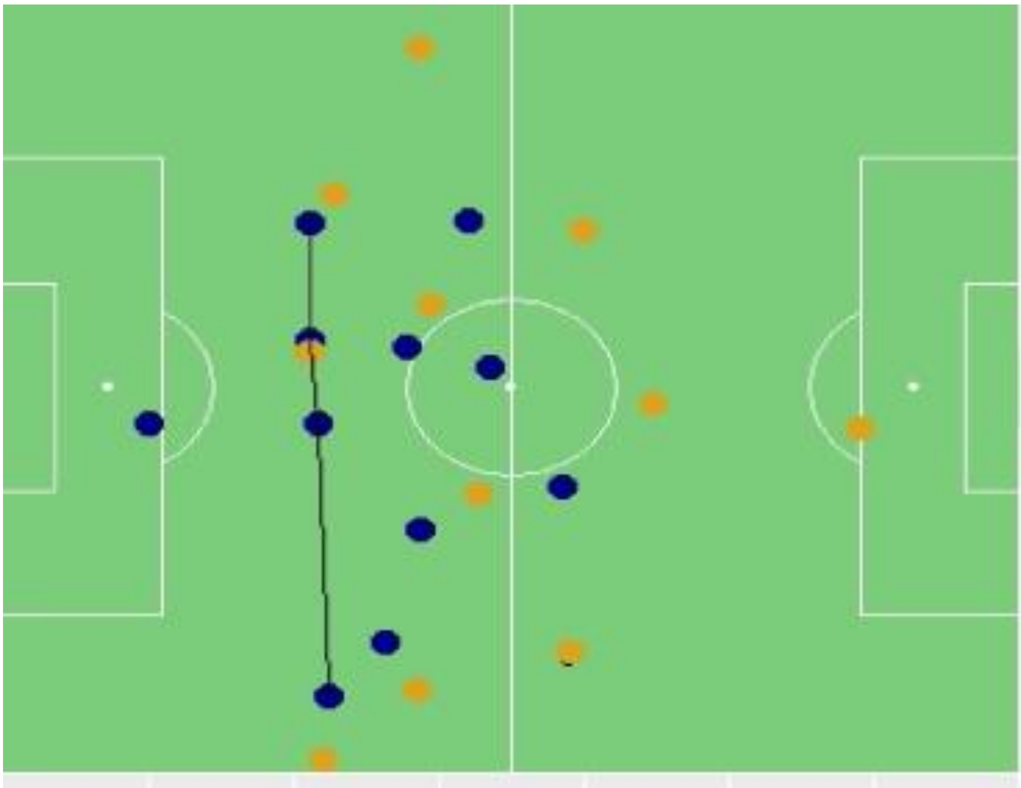


Figure 7.4: Distance between the defenders across the defensive line

7.3.3 Space

7.3.3.1 Surface Area

The surface area relates to the principle of compactness, measuring how concentrated the team shape is. This is calculated as the space inside the shape described by the peripheral outfield players as shown in figure 7.5. The following code outlines how to calculate the metric and create the visualisation for surface area. This section of the code extracts all player data and removes the goalkeepers

```
# create home and away dataframes for outfield players only
homeout1 <-  datahome %>% filter(player.id != 1)
awayout1 <- dataaway %>% filter(player.id != 12)
```

A function is then created to calculate the surface area of a team at each instance

```
SArea = function(g,h,data){
  SAreaCollect = c(NULL) #create empty matrix to fill with surface a
rea data
  for(i in 1:length(g)){ # loop through each instance of 10 players
using deldir
    dataf = data[g[i]:h[i],]
    convhull = deldir(dataf$x, dataf$y)
    SAreaCollect[i] = sum(convhull$summary[,4]) # collect sum of del
dir      column 4 (surface area)

  }
  return(SAreaCollect)
}
```

The Surface area function is applied to both data sets across a sequence based on the length of each data set.

```
homeout1sa <- SArea(seq(1,709090,10),seq(10,709100,10),homeout1)
awayout1sa <- SArea(seq(1,709090,10),seq(10,709100,10),awayout1)
```

each surface area value is replicated 23 times to fit into the longdata for visualisation and is joined using the *cbind* function.

```
homeSA <- (rep(homeout1sa, each = 23))
awaySA <- (rep(awayout1sa, each = 23))

SA <- cbind(homeSA, awaySA)
colnames(SA) <- c("home.SA", "away.SA")
surface.area <- cbind(longdata,SA)
```

The threshold of 600m^2 is selected based on data from 2018 world cup and a categorical variable is added to the dataframe, identifying when the surface area is above or below this value.

```
surface.area <- surface.area %>% mutate(compact = case_when(home.SA
< 600 ~ "Yes",
                                                          T ~ "No"
))
```

The time selected for analysis is extracted and the visualisation is created. The *annotate_pitch* creates a pitch, the *geom_point* uses player and ball data to represent them as dots on the pitch and *geom_encircle* creates the surface area diagram, creating a shape around the peripheral outfield players. the fill aspect adds the contextual identifier, if the team have a surface area < 600m^2 then the shape will be filled in a green colour, otherwise it is coloured in red as shown in figure 7.5.

```
data <- surface.area %>% filter(time >= 508.56 & time <= 538.56)


p <- ggplot(data, aes(x, y)) +
  annotate_pitch(dimensions = pitch_custom, fill = "palegreen3", col
our = "white") +
  geom_point(aes(color = team, size = team)) +
  scale_color_manual(values=c("goldenrod", "navyblue", "grey1")) +
  scale_size_manual(values=c(5, 5, 2)) +
  geom_encircle(data = data[data$team == 1 & (data$player.id != 1),]
, aes(fill = compact),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.leg
end = FALSE) +
  scale_fill_manual(values = c("red1", "forestgreen")) +
  theme(legend.position = 'none')
```

```
anim = p + transition_manual(frames = time) +
  labs(title = "area")


animate(anim, nframes = 750, fps = 25, renderer = av_renderer())
```

a) b)

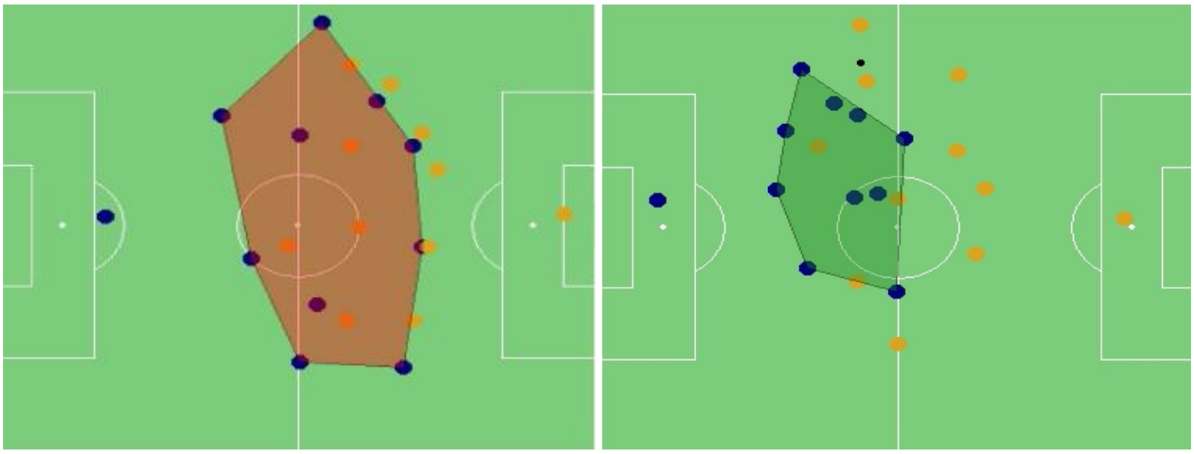


Figure 7.5: Surface area of a team described by the convex hull of the outfield players with areas a) above 600m^2 and b) below 600m^2

7.3.3.2 Space control

Space control, sometimes referred to as pitch control, is a method for identifying the areas closest to a player. In this example, space control is mapped and visualised as a method for predicting pass success. The model accounts for the movement of each player in the instance and considers the time it would take each player to reach each square meter of the pitch. A parsimonious model is used to describe this but can be adapted in multiples ways (Spearman and Basye 2017; Brefeld et al, 2019). This is used to identify the success likelihood of passes into areas. A function is created to calculate the pitch control. Firstly, a dataframe is created of 7140 grid references across the pitch marking the centre of a $1m^2$ tile. this list of x and y tiles is then repeated depending on the time period being investigated. The player data for this time period is then isolated and the x and y speed of each player on the pitch is calculated based on their time at an instant and their time 0.2s prior. Their position in 0.7 seconds is estimated by assuming they continue at their current speed, for that time. The position data is then repeated 7140 times to match up with the tile grid. The distance to each point on the grid by each player in their position they are predicted to be in based on their current speed is calculated. The model assumes all players will travel at the same speed towards each point and the quickest player from each team is used to identify who is likely to control the ball. A sigmoid function is used to calculate the likelihood that one or the other team will receive the ball with only a small number of the tiles, placed directly between the opponents where the model is less confident.

```
#pitch control
pitchcontrol <- function(data,time1,time2) {
  #get all x y coordinates, creating a grid on the pitch of 7140 points
  x <- rep(rep(0.5:104.5, 68),((time2-time1)*25+1))
  y <- rep(rep(0.5:67.5, each = 105),((time2-time1)*25+1))

  #calculate position of players in 0.7s after continuing at their current
speed
  data <- data %>% filter(time >= time1 & time <= time2) #extract time of
data to perform pitch control model
  xspeed <- as.data.frame(data[,seq(4,ncol(data),2)] %>% map(~(. - lag(.,
k = 5))/0.2)) # calculate speed in x axis
  yspeed <- as.data.frame(data[,seq(5,ncol(data),2)] %>% map(~(. - lag(.,
k = 5))/0.2)) # calculate speed in y axis
  xpos <- as.data.frame(data[,seq(4,ncol(data),2)] + xspeed*0.7) # calcula
te x coordinate after 0.7s
  ypos <- as.data.frame(data[,seq(5,ncol(data),2)] + yspeed*0.7) # calcula
te y coordinate after 0.7s
  xpos[1,] <- data[1,seq(4,ncol(data),2)] # interpolate 1st row x coordina
te
```

```
  ypos[1,] <- data[1,seq(5,ncol(data),2)] # interpolate 1st row y coordinate
  data[,seq(4,ncol(data),2)] <- xpos # replace data with new positions
  data[,seq(5,ncol(data),2)] <- ypos

  #repeat data 7140 times to merge with x and y grid for calculation
  data <- data[rep(seq_len(nrow(data)), each = 7140), ]
  data <- cbind(x,y,data) # combine 1x1 meter grid with player data

  #calculate x and y distance of player to all points on grid at each instance
  ydist <- as.data.frame(data[,seq(7,ncol(data),2)] %>% map(~(. - data$y)))
  xdist <- as.data.frame(data[,seq(6,ncol(data),2)] %>% map(~(. - data$x)))

  rdist <- sqrt((xdist^2)+(ydist^2)) # calculate radial distance from x and y distance
  rtime <- as.data.frame(rdist) %>%  map(~( .-1.79)/5 + 0.7 + (5/7)) # estimate time to get to grid point based on assumptions
  rtime <- as.data.frame(rtime) # make data into a dataframe
  rtime <- rtime %>% select(starts_with("h"), starts_with("a")) #extract all home and away data points

  htime <- rtime %>% select(starts_with("h")) # time it takes all home players to reach all points on grid
  atime <- rtime %>% select(starts_with("a")) # time it takes all away players to reach all points on grid
  htime <- apply(htime, 1, FUN=min, na.rm = T) # minimum time it takes all home players to reach all points on grid
  atime <- apply(atime, 1, FUN=min, na.rm = T) # minimum time it takes all away players to reach all points on grid



  #calculate value between 1 and 0 to indicate who is likely to win the ball at each location
  hchance <- exp(4.3*(htime-atime))
  achance <- exp(4.3*(atime-htime))
  pitchcontrol <- hchance/(achance+hchance)
  pitchcontrol <- cbind(x,y,pitchcontrol) # bind x and y grid to chance of team winning the ball

  return(pitchcontrol)
}
```

The pitch control function is run, along with time selected for the analysis. This visualisation can be computer intensive and requires large memory.

```
#perform pitch control function on data
data.vis <- pitchcontrol(widedata, 1000, 1005)
```

the time for the longdata is isolated for the visualisation and a time column is created and added to the pitch control data frame.

```
data <- longdata %>% filter(time >= 1000 & time <= 1005) # filter long dat
a so pitch control and player data time matches

data.vis <- as.data.frame(data.vis) # change pitch control data into a dat
aframe
time <- rep(seq(1000,1005,0.04), each = 7140) # get time for pitch control
data
data.vis <- cbind(data.vis,time) # bind time with pitch control data
```

The diagram is plotted using *ggplot*, *annotate_pitch* is used to create the pitch dimensions and shape, *geom_tile*, fills in the 7140 $1m^2$ grid and *geom_point* adds in the players and the ball as dots. the plot is then animated and a still image is shown in figure 7.6.

```
#create image of data
p <- ggplot(data.vis, aes(x,y)) + #use pitch control data
  annotate_pitch(dimensions = pitch_custom, fill = "palegreen3", colour =
"white") + #use custom pitch design (line 103) and colour
  geom_tile(aes(fill=pitchcontrol, alpha = 0.5)) + # use tile plot to colo
ur each square on grid correct colour
  scale_fill_gradient2(midpoint = 0.5, low = "blue", mid = "white", high =
"gold1", space = "lab") + # colour grid specific colours
  geom_point(data = data, aes(x,y, color = team, size = team)) + #create d
ot plot for players in match
  scale_color_manual(values=c("goldenrod", "navyblue", "grey1")) + # colou
r dots as team colours
  scale_size_manual(values=c(4, 4, 2)) + # give ball and players specific
dot sizes
  theme(legend.position = 'none') # remove legend

## Warning: Non Lab interpolation is deprecated

#animate the ggplot diagram
anim <- p + transition_time(time) +
  labs(title = "time: {frame_time}") +

  view_follow(fixed_y = TRUE)


animate(anim, nframes = 125, fps = 25, renderer = av_renderer())
```

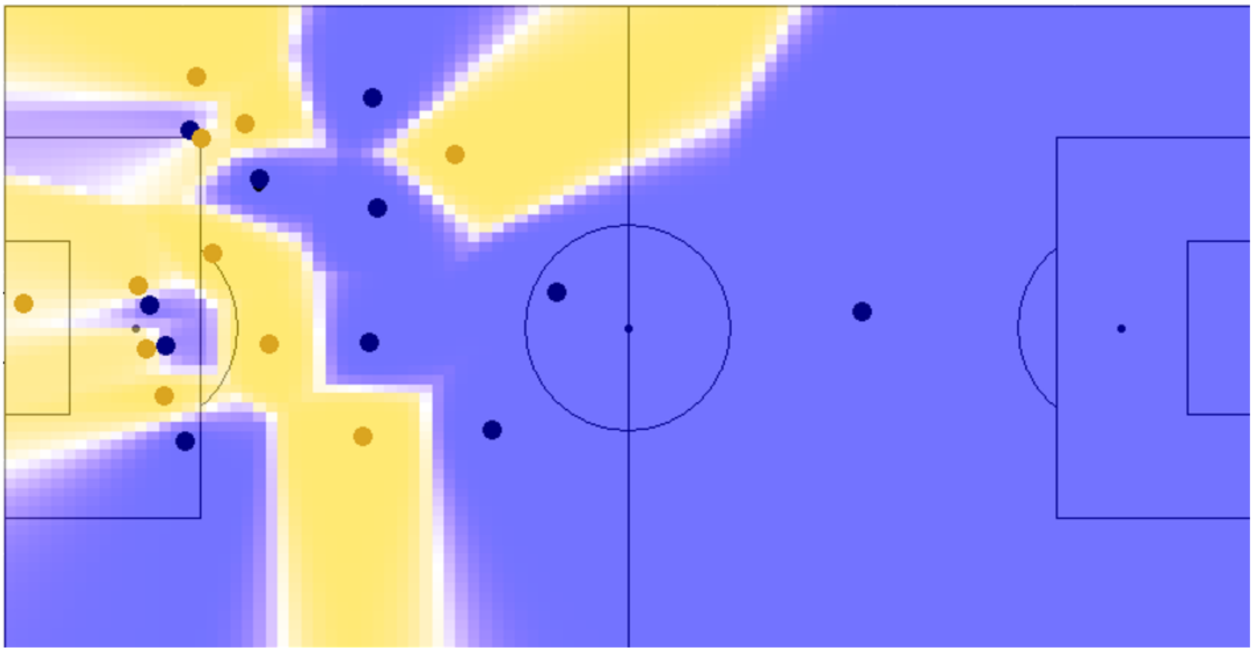


Figure 7.6: Pitch control, showing areas where each team are likely to control the ball depending on where it is passed.

7.3.4 Numerical relations

7.3.4.1 Number of outplayed opponents

This metric identifies how many defenders the ball has passed in each attack and can be combined with the previously mentioned pitch control model to identify and analyse passes that successfully penetrated the opponent. This metric is calculated by identifying the difference between the number of opponents closer to their own goal line than the ball at each touch. To do this requires the contact dataframe. The contact and contact number are extracted from the dataset and repeated row-wise 22 times and then bound with the *matchdata* dataframe.

```
contacts <- contact %>% select(contact, contact.num) #isolate contacts fro
m  ball data
```

```
contacts <- contacts[rep(seq_len(nrow(contacts)), each = 22), ] # repeat
contact data so it can be merged with match data
matchdata <- cbind(matchdata, contacts) #merge contacts data with matchdat
a
```

All instances that do not involve a ball contact are removed from the data set and the number of home defenders between the ball and their goal line are identified along with the number of away defenders between the ball and their own goal line.

```
#calculate number of opponents behind the ball for both sides
outplayed <- matchdata %>%  filter(contacts == 1) %>% # data of instances
only with contacts
  mutate(hdef.between.goal = case_when((team == 1 & ballx > x) ~ 1, TRUE ~
0)) %>%  # identify if a home defender is behind the ball
  mutate(adef.between.goal = case_when(( team == 0 & ballx < x) ~ 1, TRUE
~ 0))  # identify if an away defender is behind the ball
```

The total number of players at each instance are summed together, to create two separate dataframes.

```
outplayed.h <- outplayed %>%
  group_by(time) %>%
  summarise(sum.h.def = sum(hdef.between.goal)) #sum of home players betwe
en ball

outplayed.a <- outplayed %>%
  group_by(time) %>%
  summarise(sum.a.def = sum(adef.between.goal)) #sum of away players betwe
en ball
```

The difference between the number of defenders is calculated using a lag function to subtract the defenders between the ball and the goal at each instance. The dataframes are combined, with the values identifying instances whereby teams are penetrated. There is no visualisation for this metric.

```
#calculate change in opponents behind the ball for both sides
```

```
outplayed.h <- cbind(outplayed.h, outplayed.a[,2]) #merge both opponents
behind ball
outplayed.h <- outplayed.h %>%
  mutate(HNOO = lag(sum.a.def, 1) - sum.a.def) %>%
  mutate(Anoo = lag(sum.h.def, 1) - sum.h.def)
```

7.3.4.2 Overloads

Zones on the pitch where teams had numerical superiority are considerations for coaches. In situations where an attacking team have more players in an area than the defending team, this is commonly termed as an overload and relates to the balance of the competing teams. The following code outlines how to calculate the metric and create the visualisation for overloaded areas. The zones created are 7 dynamic areas selected based on the current position of outfield players on the pitch adapted from Vilar et al, (2013). Figure 7.7 demonstrates how this is calculated. This section of the code extracts all player data and removes the goalkeeper's data.

```
#make full player data set without GKS
outfielders <- playerdata1 %>% filter(player.id != 1 & player.id != 12)
```

The four reference points that encompass the whole match highest and lowest x and y coordinates are identified. Following this, the 5 mid points that are identified based on a percentage across the team length and team width are calculated

```
#mutate columns with total length and width zones of all outfield players
outfielders <- outfielders %>% group_by(time) %>% mutate(top = max(x)) %>%
mutate(bottom = min(x)) %>%
  mutate(left = max(y)) %>% mutate(right = min(y)) %>% ungroup
```

```r
#mutate attacking, defensive, half way and channel zones
outfielders <- outfielders %>% group_by(time) %>%
  mutate(attackx = top-(0.25*(top-bottom))) %>%
  mutate(halfx = top-(0.5*(top-bottom))) %>%
  mutate(defence = top-(0.75*(top-bottom))) %>%
  mutate(left.channel = left-(0.25*(left-right))) %>%
  mutate(right.channel = left-(0.75*(left-right)))
```

A *case_when* is applied to identify when players are in each zone where LW = left wing, LD = left defence, RW = right wing, RD = right defence, DEF = central defence, ATT = central attack and MID = central midfield

```r
#create column to classify what zone each player is in
outfielders <- outfielders %>%
  mutate(zone = case_when((x >= halfx & y >= left.channel) ~ "LW",
                          (x < halfx & y >= left.channel) ~ "LD",
                          (x >= halfx & y <= right.channel) ~ "RW",
                          (x < halfx & y <= right.channel) ~ "RD",
                          (x < defence) ~ "DEF",
                          (x > attackx) ~ "ATT",
                          T ~ "MID"))
```

After this step, 2 separate data frames are created to count how many players are in each zone for each team.

```r
#split up data for home team and away team
homeout <- outfielders %>% filter(team == 1)
awayout <- outfielders %>% filter(team == 0)

#for home team count the number of players with each zone getting a column
homeout <- homeout %>% group_by(time) %>% mutate(LW = sum(zone == "LW")) %
>%
  mutate(LD = sum(zone == "LD")) %>%  mutate(RW = sum(zone == "RW")) %>%
  mutate(RD = sum(zone == "RD")) %>%  mutate(DEF = sum(zone == "DEF")) %>%
  mutate(ATT = sum(zone == "ATT")) %>%  mutate(MID = sum(zone == "MID"))

#for away team, count the number of players with each zone getting a colum
n
awayout <- awayout %>% group_by(time) %>% mutate(LW = sum(zone == "LW")) %
>%
  mutate(LD = sum(zone == "LD")) %>%  mutate(RW = sum(zone == "RW")) %>%
  mutate(RD = sum(zone == "RD")) %>%  mutate(DEF = sum(zone == "DEF")) %>%
  mutate(ATT = sum(zone == "ATT")) %>%  mutate(MID = sum(zone == "MID"))
```

The difference between the home team value and the away team value is calculated to identify which team is overloading each section.

```
numerical.adv <- homeout[,17:23] - awayout[,17:23] # take home count per z
one away from away count
numerical.adv <- numerical.adv[seq(1,nrow(numerical.adv),10),] # remove du
plicated rows so 1 per instance
time <- seq(0,nrow(numerical.adv)/25-0.04,0.04) # create time

numerical.adv <- cbind(time, numerical.adv) # combine time and numerical a
dvantage df
```

to set up the overload visualisation from the outfielder data, the first 4 rows must be extracted from each time point and given a reference value 1 to 4.

```
num.adv.vis <- outfielders[rep(seq(1,nrow(outfielders),20), each = 4),] #
get only 4 data points per time
num.adv.vis$ref = rep(seq(1,4,1), nrow(num.adv.vis)/4) # create reference
values for data
```

Following this, 5 columns are created using x coordinates and 3 columns using y coordinates that are used to identify the 7 unique zones.

```
#adjust data so for visualisation - make calculation of points for x and y
of spaces
num.adv.vis <- num.adv.vis %>% mutate(x1 = case_when(ref == 1 | ref == 2 ~
halfx,
                                             T ~ top)) %>%
  mutate(x2 = case_when(ref == 1 | ref == 2 ~ halfx,
                          T ~ bottom)) %>%
  mutate(x3 = case_when(ref == 1 | ref == 2 ~ bottom,
                          T ~ defence)) %>%
  mutate(x4 = case_when(ref == 1 | ref == 2 ~ attackx,
                        T ~ defence)) %>%
  mutate(x5 = case_when(ref == 1 | ref == 2 ~ attackx,
                        T ~ top)) %>%
  mutate(y1 = case_when(ref == 1 | ref == 3 ~ left,
                        T ~ left.channel)) %>%
  mutate(y2 = case_when(ref == 1 | ref == 3 ~ right,
                        T ~ right.channel)) %>%
  mutate(y3 = case_when(ref == 1 | ref == 3 ~ left.channel,
                        T ~ right.channel))
```

The original numerical advantage dataframe is replicated 4 times to it can be joined to the visualisation dataframe using *cbind*.

```
#repeat rows so columns join easily
num.adv <- numerical.adv[rep(seq_len(nrow(numerical.adv)), each = 4), 2:8]

num.adv.vis <- cbind(num.adv.vis, num.adv)
```

The desired time values for analysis are extracted from the longdata and overload data for visualisation. The following code describes how to create the animation. The *annotate_pitch* function creates a pitch for the visualisation, *geom_point* uses player and ball data to create points on the pitch, *geom_encircle* creates the zones and is defined by the previously described columns created *x1* - *x5* and *y1* - *y3*. *scale_fill_gradient2* colours each zone to the team colour and the intensity identifies how strong an overload each team has in that zone, a white zone indicates a numerically balanced zone. The final 2 lines animate the measure over time.

```
nadv <- num.adv.vis %>% filter(time >= 886.56 & time <= 916.56)
data <- longdata %>% filter(time >= 886.56 & time <= 916.56)

p <- ggplot(data, aes(x, y)) +
  annotate_pitch(dimensions = pitch_custom, fill = "palegreen3", colour =
"white") +
  geom_point(aes(color = team, size = team)) +
  scale_color_manual(values=c("goldenrod", "navyblue", "grey1")) +
  scale_size_manual(values=c(5, 5, 2)) +
  geom_encircle(data = nadv, aes(x = x1, y = y1, fill = LW),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
FALSE) +
  geom_encircle(data = nadv, aes(x = x2, y = y1, fill = LD),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
FALSE) +
  geom_encircle(data = nadv, aes(x = x1, y = y2, fill = RW),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
FALSE) +
  geom_encircle(data = nadv, aes(x = x2, y = y2, fill = RD),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
```

```
FALSE) +
  geom_encircle(data = nadv, aes(x = x3, y = y3, fill = DEF),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
FALSE) +
  geom_encircle(data = nadv, aes(x = x4, y = y3, fill = MID),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
FALSE) +
  geom_encircle(data = nadv, aes(x = x5, y = y3, fill = ATT),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.legend =
FALSE) +
  scale_fill_gradient2(low = "goldenrod", mid = "white", high = "navyblue"
, midpoint = 0) +
  theme(legend.position = 'none')




anim = p + transition_manual(frames = time) +
  labs(title = "numerical superioirty")




animate(anim, nframes = 750, fps = 25, renderer = av_renderer())
```

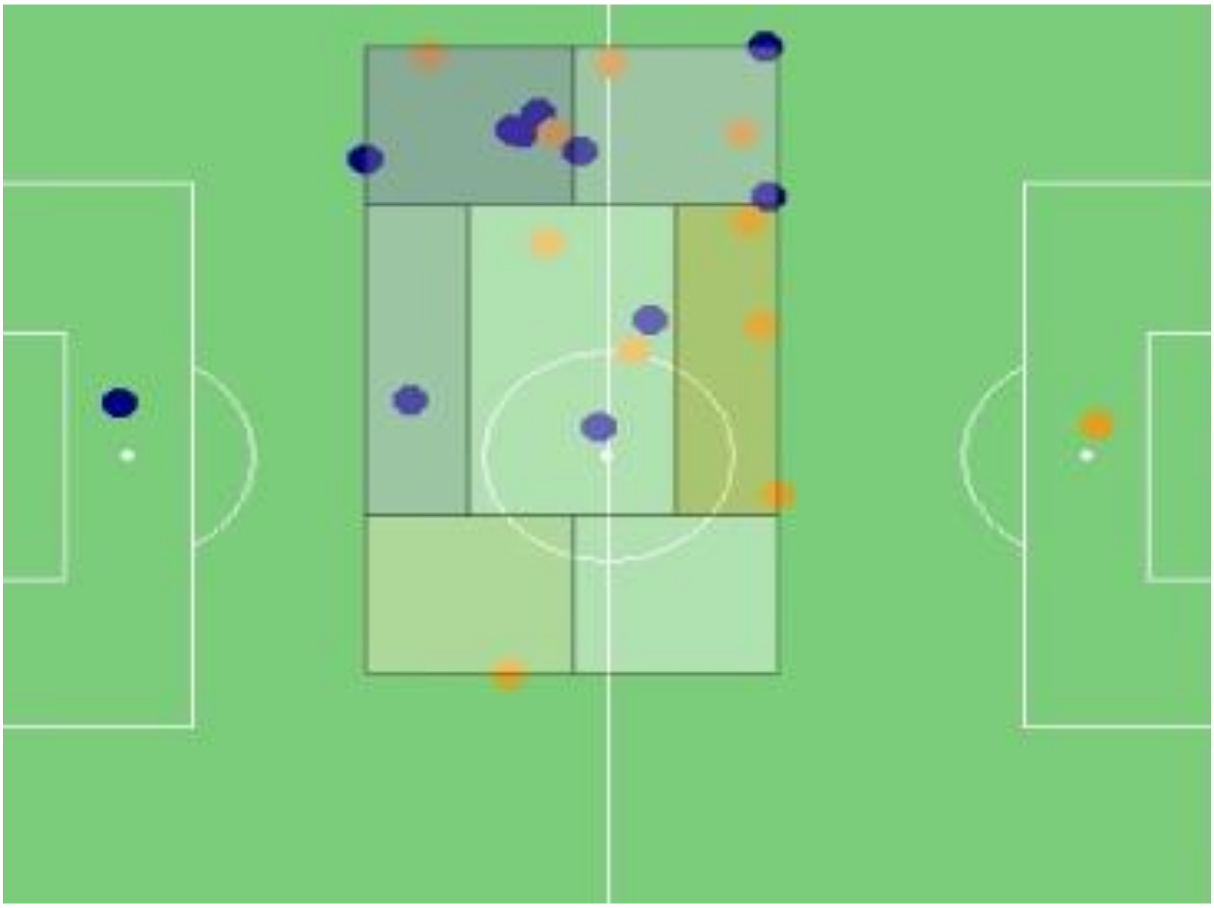


Figure 7.7: Visualisation of 7 zones and number of players in each zone, where colour represents which team has an overload in the zone.

7.3.5 Passing Networks

7.3.5.1 Network Intensity

The measure of network intensity identifies how fast each pass is being completed when a team is in possession. Coaches will generally seek to increase the rate at which their team passes, relating to the mobility principle. This metric requires the previously mentioned `possession` and `passing` dataframes. Firstly, the time of each possession for both teams and the time the ball is out of play is extracted.

```
possession.start <- possession %>% filter(Category == "team2 pos" |
Category == "team1 pos" | Category == "out of play")
```

The time is in the format MM:SS:MS and is changed into seconds, to do this, each component of the possession start time is extracted separately, changed into a numeric format and bound into a new dataframe and combined into seconds.

```
minutes = substr(possession.start$Start,1,2) #extract minutes as the
two numbers
seconds = substr(possession.start$Start,4,5) # extract seconds as th
e two numbers
milliseconds = substr(possession.start$Start,7,8) #extract milliseco
nds as the 2 numbers

minutes <- as.numeric(minutes) #change all too numeric
seconds <- as.numeric(seconds)
milliseconds <- as.numeric(milliseconds)

time <- as.data.frame(cbind(minutes,seconds,milliseconds)) # create
dataframe of time

time <- time %>%
  mutate(time = minutes*60 + seconds + milliseconds/100)
```

The *time* data is combined with the *possession.start* data.

```
possession.start <- cbind(possession.start, time)
```

this full step is repeated but for the possession end times, so that for each attack, a start and end time are established in seconds

```
#get possession end
possession.end <- possession %>% filter(Category == "team2 pos" | Ca
tegory == "team1 pos" |
                                      Category == "out of play")

minutes = substr(possession.end$End,1,2) #extract minutes as the two
```

```
numbers
seconds = substr(possession.end$End,4,5) # extract seconds as the tw
o numbers
milliseconds = substr(possession.end$End,7,8) #extract milliseconds
as the 2 numbers

minutes <- as.numeric(minutes) #change all too numeric
seconds <- as.numeric(seconds)
milliseconds <- as.numeric(milliseconds)

time <- as.data.frame(cbind(minutes,seconds,milliseconds)) # create
dataframe of time


time <- time %>%
  mutate(time = minutes*60 + seconds + milliseconds/100) # create fo
rmula for time in seconds

possession.end <- cbind(possession.end, time)
```

the categories used will also add on an end to the string, to differentiate the start and the end time for each row

```
possession.end$Category <- paste("end", possession.end$Category, sep
=".")
```

Again, a similar process will happen for the `passing` dataframe, isolating the time components and outputting them in seconds.

```
#extract passes in same format

minutes = substr(passing$Start,1,2) #extract minutes as the two numb
ers
seconds = substr(passing$Start,4,5) # extract seconds as the two num
bers
milliseconds = substr(passing$Start,7,8) #extract milliseconds as th
e 2 numbers

minutes <- as.numeric(minutes) #change all too numeric
seconds <- as.numeric(seconds)
milliseconds <- as.numeric(milliseconds)

time <- as.data.frame(cbind(minutes,seconds,milliseconds)) # create
dataframe of time


time <- time %>%
  mutate(time = minutes*60 + seconds + milliseconds/100) # create fo
```

```
rmula for time in seconds

passing <- cbind(passing, time)
```

The *possession.start*, *possession.end*, and *passing* data are bound by row and then ordered chronologically,

```
network <- rbind(possession.start,passing, possession.end)

network <- network[
  with(network, order(time, Category)),
]
```

The home and away data are separated for independent analysis

```
home.network <- network %>% filter(Category == "team1 pass received"
| Category == "end.team1 pos" |
                                    Category == "team1 pos")

away.network <- network %>% filter(Category == "team2 pass received"
|
                                    Category == "team2 pos" |
                                    Category == "end.team2 pos")
```

when the home team passes, a reference will be made as a 1 in the *ref* column, additionally, for each time the team become in possession of the ball *ref2* will indicate this with a 1 in the column. The *cumsum* function is used to create a *possession* column, allowing to group the data by attack.

```
home.network <- home.network %>% mutate(ref = case_when(Category ==
"team1 pass received" ~ 1,
                                                      T ~ 0)) %>%
  mutate(ref2 = case_when(Category == "end.team1 pos" ~ 1,
                          T ~ 0)) %>%
  mutate(possession = cumsum(ref2)-ref2)
```

The *group_by* function, groups the dataframe by each attack and allows for use of the mutate function to calculate the required variables. Passes calculates the number of passes in each attack, *pos.time* calculates the time of the attack in seconds and intensity calculates the number of seconds between each pass. No visualisation is presented for this metric.

```
home.network <- home.network %>% group_by(possession) %>%
  mutate(passes = sum(ref)) %>%
  mutate(pos.time = max(time) - min(time)) %>%
  mutate(intensity = pos.time/passes)
```

7.3.6 Combination metrics

7.3.6.1 Triangles

Players naturally form triangles with their teammates when playing. These triangles can be measured through the distance and angle between each side of the triangle and the area of the triangle. The shapes described by the players can be important to the coaches' principles of play and how players are able to support their teammates in possession. This section will provide an example of how to measure and visualise the triangle formed by three players in the central midfield area. The wide data form is used for this analysis, isolating players 6, 8 and 10 from the home team. First, distances between the 3 are calculated as follows.

```
triangles <- widedata %>%
  mutate(dist.6.8 = sqrt((h6.x - h8.x)^2 + (h6.y - h8.y)^2)) %>%  #
calculate distance between player 6 and 8
  mutate(dist.6.10 = sqrt((h6.x - h10.x)^2 + (h6.y - h10.y)^2)) %>%
#calculate distance between player 6 and 10
  mutate(dist.8.10 = sqrt((h8.x - h10.x)^2 + (h8.y - h10.y)^2)) # ca
lculate  distance between player 8 and 10
```

the angles for each corner of the triangle are calculated through the following code, 180/pi calculates the angle in degrees as the original output is in radians.

```
triangles <- triangles %>%
  mutate(angle6 = 180/pi*(acos((dist.6.8^2 + dist.6.10^2 - dist.8.10
^2)/
       (2*dist.6.8*dist.6.10)))) %>% # calculate angle at player 6
  mutate(angle8 = 180/pi*(acos((dist.6.8^2 + dist.8.10^2 - dist.6.10
^2)/
       (2*dist.6.8*dist.8.10)))) %>%   # calculate angle at player
8
  mutate(angle10 = 180/pi*(acos((dist.8.10^2 + dist.6.10^2 - dist.6.
8^2)/
       (2*dist.8.10*dist.6.10)))) # calculate angle at player 10
```

Heron's formula is used to calculate the area as shown in equation 1. This requires an s value, which is found by dividing the perimeter of the triangle, a, b and c by 2. The perimeter of the triangle is identified as the sum of each side a, b and c.

$$Area = \sqrt{s(s-a)(s-b)(s-c)} \qquad [1]$$

```
triangles <- triangles %>%
 mutate(s.triangle1 = (dist.6.8 + dist.6.10+ dist.8.10)/2) %>% # cal
culate s value for area calculation
  mutate(area = sqrt((s.triangle1*(s.triangle1-dist.6.8)*(s.triangle
1-dist.6.10)*(s.triangle1-dist.8.10))))
```

Data is isolated for analysis and visualisation. A plot is created using the annotate pitch to display the pitch and *geom_point* draws the players and the ball on the plot as dots. *geom_encircle* creates a triangle of the 3 selected players for analysis, the output is shown in figure 7.8.

```
data <- longdata %>% filter(time >= 391 & time <= 403)

#triangles visualisation

p <- ggplot(data, aes(x, y)) +
  annotate_pitch(dimensions = pitch_custom, fill = "palegreen3", col
our = "white") +
  geom_point(aes(color = team, size = team)) +
```

```
  scale_color_manual(values=c("goldenrod", "navyblue", "grey1")) +
  scale_size_manual(values=c(5, 5, 2)) +
  geom_encircle(data = data[data$team == 1 & (data$player.id == 6 |
data$player.id == 8 | data$player.id == 10),], aes(fill = team),
                s_shape = 1, expand = 0,
                alpha = 0.4, color = "black", na.rm = TRUE, show.leg
end = FALSE) +
  theme(legend.position = 'none')

#animate graph
anim <- p + transition_time(time) +
  labs(title = "time: {frame_time}") +
  view_follow(fixed_y = TRUE)

#create animation
animate(anim, nframes = 300, fps = 25)
```

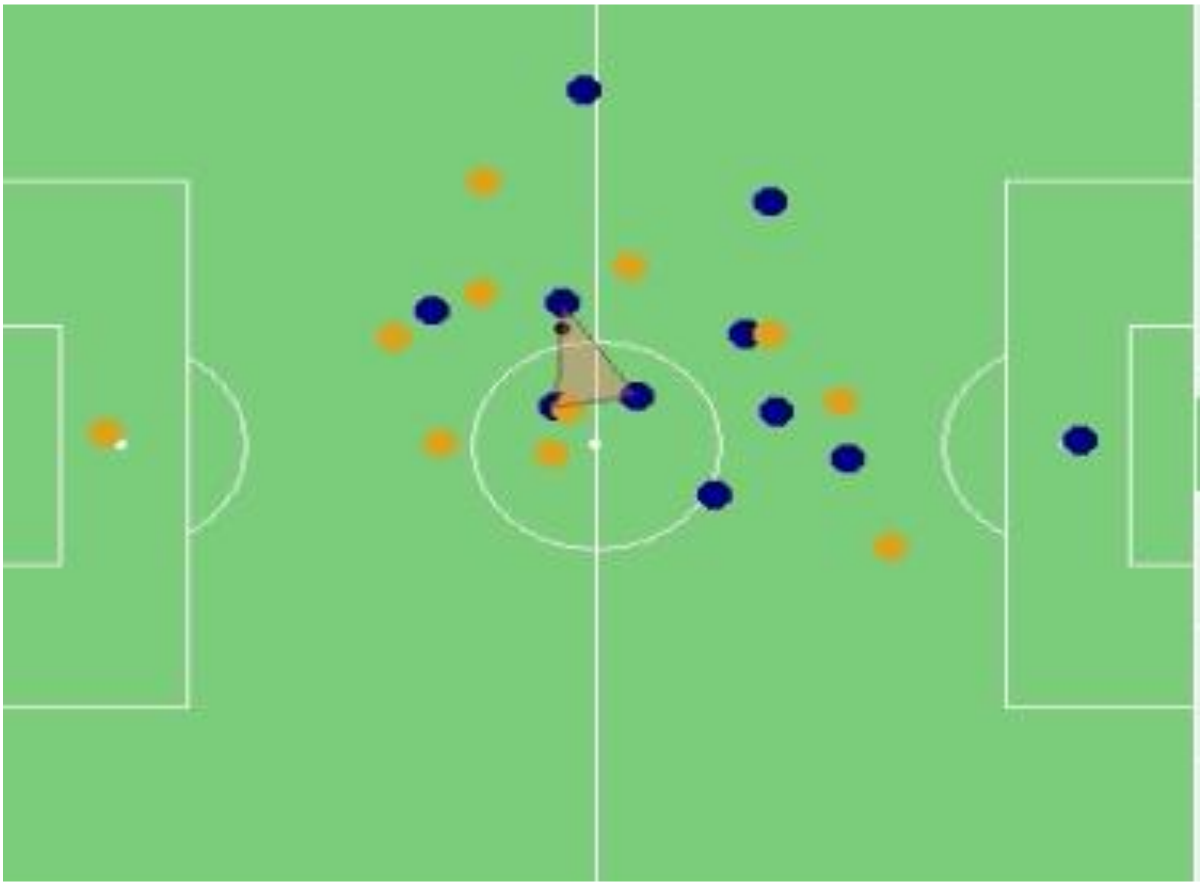


Figure 7.8: Visualisation of the triangle created by three pre-determined players

7.3.6.2 Total Pressure

This metric calculates how much pressure the defending team are applying to the player on the ball. Three aspects are considered when identifying the pressure that the player is under. These are space pressure, time pressure and decision pressure. these concepts are combined to estimate the total pressure on the player in possession. Pressure is an important part of delaying the opponent and relates to the discipline of the defending team. The first section looks at space pressure. And attempts to use the following model, adapted from Link et al (2016) demonstrated in figure 7.9. Players in the highlighted radii increase the pressure on the player in possession. Pressure is further increased if their position is between the player and the goal, and if the player is closer to the player on the ball. If the player is within a 1-meter radius of the attacker then the space pressure value is 1. If no players are within the illustrated radius, then the pressure is 0.

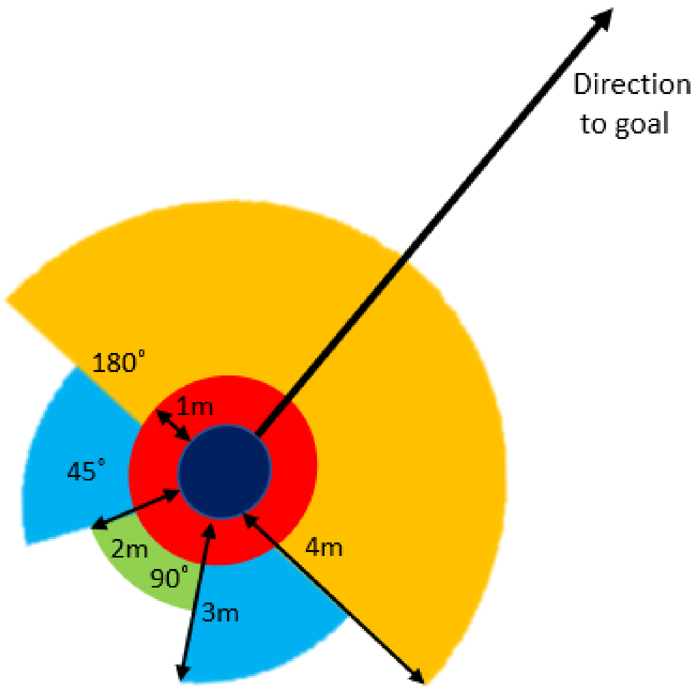


Figure 7.9: Space pressure model adapted from link

The following code outlines how to calculate space pressure. Using the *Longdata* dataframe, the team column is changed into a numeric representation, 0 = away, 1 = home and 4 = ball. This data is then ordered by time, then by team, then by *player.id* to make sure there is a consistent order. An id column for each individual is created, repeating values 1 to 23 and added to the dataframe

```
####### Delay #######

#create data with unique id for each player - order data first, then
assign  unique ids
pressure <- longdata
pressure$team <- as.numeric(pressure$team) # change team to 1(away)
2(home)  and 3(ball)

#reorder data so ball is last and players recur in an order
pressure <- pressure[
  with(pressure, order(time, team, player.id)),
]

#create unique id for each player and ball
id <- rep(seq(1,23,1),nrow(pressure)/23)
pressure <- cbind(pressure,id)
```

The following functions, *distance*, *distance2* and *distance3* are created to calculate the distance between each id at every time instance. The distances are output in a list and so are extracted, transformed into a dataframe, and then bound to the list after naming each column the distance to the individual reference id.

```
#### calculate distance between all points
distance = function(x1,x2,y1,y2) sqrt(((x2-x1)^2)+((y2-y1)^2)) #dist
ance
distance2 = function(x,y,.pred) distance(x, x[.pred], y, y[.pred])
distance3 = function(x, y, id){
  dists = map(1:23, ~distance2(x,y, which(id == .x)))
}

#use distance formula
P2 <- pressure %>%
  group_by(time) %>%
```

```
  mutate(distances=distance3(x, y, id))


distances <- P2$distances # extract distance list
distances <- do.call(rbind.data.frame, distances) # change list to d
ataframe
colnames(distances) <- c(paste0("dist", 1:23)) # change column names
pressure <- cbind(pressure,distances) # merge dataframes
```

The row representing the ball is removed, and then new columns are created identifying which reference id is closest to the ball to identify who and what team is in possession. The x and y coordinates of the player closest to the ball are also extracted and the distance between each player and the closest player to the ball are calculated in a new column.

```
pressure <- pressure %>%  filter(team == 1 | team == 2) #remove ball
data    from data as no longer needed)

#new columns with id and group as closest to position of id 23
pressure <- pressure %>% group_by(time) %>% mutate(closest = id[whic
h.min(dist23)]) %>%
  mutate(team.pos = team[which.min(dist23)]) %>% ungroup

#create columns for x and y coords player closest to ball (or player
in       possession)
pressure <- pressure %>%  group_by(time) %>% mutate(Bx = x[which(id=
=closest)]) %>%
  mutate(By = y[which(id==closest)]) %>%  ungroup()

#calculate distance between ball carrier and all players
pressure <- pressure  %>% mutate(distance = sqrt(((Bx-x)^2) + (By-y)
^2))
```

The distance of each player to their own goal is calculated, along with the distance of each player to the opposition goal. Using this data, the angle of each defending player relative to the ball carrier and the goal they are defending are calculated. This is multiplied by 180/pi to get the value in degrees as the original output is in radians

```
#calculate distance between players and their own goal
pressure <- pressure %>% mutate(dist.own.goal = case_when((team == 1
) ~ sqrt(((105-x)^2)+((34-y)^2)),
```

```r
    T ~ sqrt(((x)^2)+((34-y)^2))))

#calculate distance between attacking players and their opponents go
al
pressure <- pressure %>% mutate(dist.goal = case_when((team == 1) ~
sqrt(((105-Bx)^2)+((34-By)^2)),
    T ~ sqrt(((Bx)^2)+((34-By)^2))))



#calculate angle at ball carrier with opponent goal, the defender
pressure <- pressure %>% mutate(angle = case_when((team.pos == team)
~ 0,
          T ~ 180/pi*(acos((dist.goal^2 + distance^2 - dist.own.goal
^2)/(2*dist.goal*distance)))))
```

The following code works through the pressure applied by a defender if they appear in the zones described in figure 7.9. For each instance, a calculation based on the sum of the opponent's space pressure is calculated. the value k used as a weighting factor for the space pressure calculation can be altered, in this instance the value -2.3 is selected. The current calculation identifies a value between 1, representing the player is under extreme space pressure and 0, where the player is under very little pressure.

```r
pressure <- pressure %>%
  mutate(space.pressure = case_when(distance <= 1 & team.pos != team
~ 1,
        (angle <= 90 & distance < 4 & team.pos != team) ~ (4-distanc
e)*0.25,
        (angle <= 135 & distance < 3 & team.pos != team) ~ (3-distan
ce)*0.25,
        (angle > 135 & distance < 2 & team.pos != team) ~ (2-distanc
e)*0.25,
                                   TRUE ~ 0))

pressure <- pressure %>% group_by(time) %>%  mutate(sp.pressure = 1
- exp(-2.3*sum(space.pressure)))
```

After space pressure has been calculated, the time pressure is calculated independently. This is determined through the time to contact measure identifying the expected time the player has before the closest defender reaches them. This process uses the pressure dataframe used in the space pressure calculation. The speed of each player at an instance is calculated as the difference between the players position and the players position 0.24s back in time. A column is created with lagged x and y values and the formula $Speed = \frac{distance}{time}$ is applied.

```
#time to contact


# Create shifted x and y and rename columns
x.24 <- as.data.frame(lag(pressure$x, 132))
y.24 <- as.data.frame(lag(pressure$y, 132))
colnames(x.24) <- c("x.24")
colnames(y.24) <- c("y.24")

#combine in df
TTC <- cbind(pressure,x.24,y.24)

#calculate distance change in player over 0.24 seconds
TTC <- TTC %>% mutate(dist.change = sqrt(((x-x.24)^2 + (y-y.24)^2)))

#calculate speed of player at instance
TTC <- TTC %>% mutate(speed = dist.change/0.24)
```

The distances between players of each team are calculated in separate grouped data frames, one for each team, and arranged for data analysis. *min_col1* and *min_col2* identifies which player is closest to that player in each team. if the player is in the same team, the *reference id* will be the same player and will be removed in a later step.

```
#extract line distances for each team individually
team1 <- TTC[,8:18]
team2 <- TTC[,19:29]

#calculate minimum distance to line for each group
```

```
team1$min_col1 = colnames(team1)[apply(team1, 1, which.min)]
team2$min_col2 = colnames(team2)[apply(team2, 1, which.min)]

#make data frame
min_col1 <- as.data.frame(team1$min_col1)
min_col2 <- as.data.frame(team2$min_col2)

#change column names
colnames(min_col1) <- c("min.col1")
colnames(min_col2) <- c("min.col2")
```

The data are bound and arranged so the reference min columns are identified. A new column is created to specify the closest player in the opposition team. The values in this new column titled ref contains the player reference with the string *dist* in front of the *id* value. The *dist* string is removed using the *gsub* function and the column is changed to numeric for processing.

```
#merge closest players with passing.lane df
TTC <- cbind(TTC,min_col1,min_col2)

#when team 1 possesses. find closest team 2 players to each line
#when team 2 possesses find closest team 1 player to each line
team.seperateness <- TTC %>% mutate(ref = case_when((team == 1) ~ mi
n.col2,
                                                      (team == 2)
~ min.col1,
                                                      T ~ "dist0")
)


#remove "d" from column to match up with id column and make numeric
team.seperateness$ref <- gsub("dist","",as.character(team.seperatene
ss$ref))
team.seperateness$ref <- as.numeric(team.seperateness$ref)
```

The x and y values of the closest opposition players are identified and denoted as *Cx* and *Cy* as well as creating a new column for the speed that the player is traveling, denoted as *Cs*.

```
#identify x,y and speed value for closest defender to each player
team.seperateness <- team.seperateness %>%
```

```
  group_by(time) %>%
  mutate(Cx = map_dbl(ref, function(ref) if (ref > 0) .data$x[.data$
id == ref] else 0)) %>%
  mutate(Cy = map_dbl(ref, function(ref) if (ref > 0) .data$y[.data$
id == ref] else 0)) %>%
  mutate(Cs = map_dbl(ref, function(ref) if (ref > 0) .data$speed[.d
ata$id == ref] else 0)) %>%
  ungroup()
```

The distance between each player and their respective closest opponent is calculated in the column titled *TS* for team separateness. Following this, the data of the player closest to the ball is isolated for the time to contact calculation to be specific for the player on the ball. The time to contact is calculated in the final line of the code below.

```
#calculate distance between player and closest player of opposite te
am
team.seperateness <- team.seperateness %>%
  mutate(TS = sqrt(((x-Cx)^2) + (y-Cy)^2))

# keep only data of player closest to the ball for time to contact
TTC <- team.seperateness %>% filter(id == closest)

#calculate time to contact
TTC <- TTC %>% mutate(TTC = TS/Cs)
```

this value is used in an exponential function to generate a value between 1 and 0, where a time pressure value approaching 1 represents extreme time pressure and a value approaching 0 represents very little time pressure. an exponential equation is applied to be more representative of time pressure. in this situation, a 1 second difference in time to contact can yield a large difference (0.1s to 1.1s) or small difference (10s to 11s)

```
#calculate measurement based on
TTC <- TTC %>% mutate(time.pressure = 1/exp(TTC))
```

The final metric is decision pressure. this is calculated based on the number of teammates who are available for a simple pass and acts as a method for calculating a team's ability to force the team in possession to move the ball predictably. This is measured using the passing lane metric (Steiner et al, 2019) as demonstrated in figure 7.10. The passing lane is measured as the angle between the line between two teammates and the closest opponent to the line. An intuitive threshold for this is set at 10° where if the passing lane is below this value, a simple pass is not open for this player, if it is above this value, then a simple pass is open for this player.

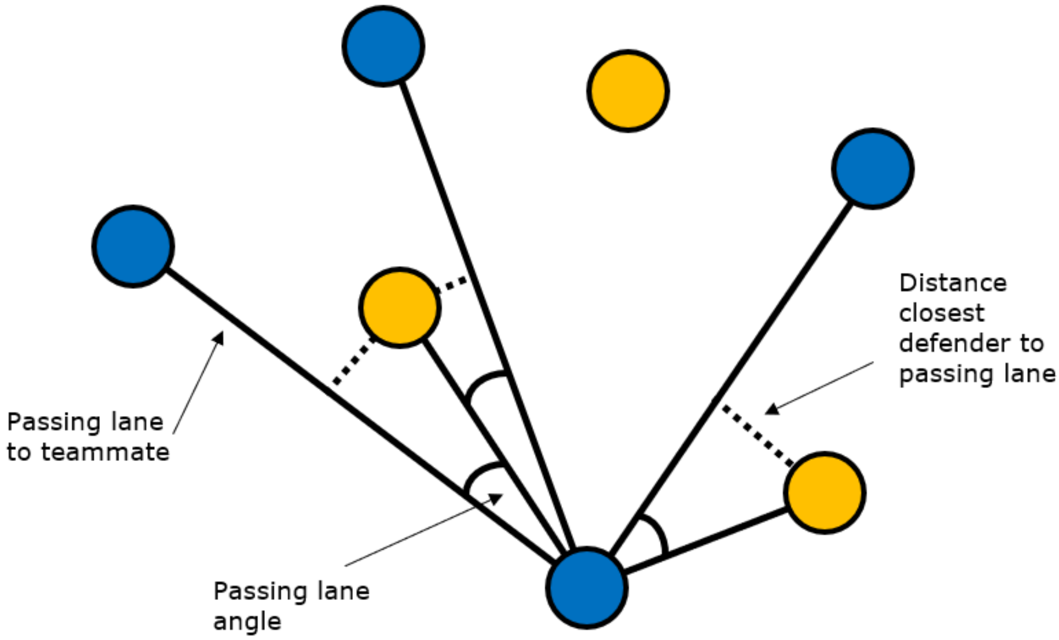


Figure 7.10: Passing lane angle calculation

The initial phases of the passing lane calculation require the data to be organised in a similar manner to the space pressure and are repeated in the following code.

```
# passing lane

#create data with unique id for each player - order data first, then
assign  unique ids
passing.lane <- longdata
passing.lane$team <- as.numeric(passing.lane$team) # change team to
1(away) 2(home) and 3(ball)

#reorder data so ball is last and players recur in an order
passing.lane <- passing.lane[
  with(passing.lane, order(time, team, player.id)),
]

#create unique id for each player and ball
id <- rep(seq(1,23,1),nrow(passing.lane)/23)
passing.lane <- cbind(passing.lane,id)

#### calculate distance between all points
distance = function(x1,x2,y1,y2) sqrt(((x2-x1)^2)+((y2-y1)^2)) #dist
ance function
distance2 = function(x,y,.pred) distance(x, x[.pred], y, y[.pred]) #
distance3 = function(x, y, id){
  dists = map(1:23, ~distance2(x,y, which(id == .x)))
}

#use distance formula
PL2 <- passing.lane %>%
  group_by(time) %>%
  mutate(distances=distance3(x, y, id))


distances <- PL2$distances # extract distance list
distances <- do.call(rbind.data.frame, distances) # change list to d
ataframe
colnames(distances) <- c(paste0("dist", 1:23)) # change column names
passing.lane <- cbind(passing.lane,distances) # merge dataframes


group3 <- passing.lane %>% filter(team == 3) #extract ball data in c
ase it   needs to be added later
passing.lane <- passing.lane %>%  filter(team == 1 | team == 2) #rem
ove ball from data

#new columns with id and group as closest to position of id 23
passing.lane <- passing.lane %>% group_by(time) %>% mutate(closest =
id[which.min(dist23)]) %>%
  mutate(team.pos = team[which.min(dist23)]) %>% ungroup
```

```
#create columns for x and y coords player closest to ball (or player
in possession)
passing.lane <- passing.lane %>%  group_by(time) %>% mutate(Bx = x[w
hich(id==closest)]) %>%
  mutate(By = y[which(id==closest)])
```

The distance between each player and each passing lane are computed using the following code. This is repeated for all player ids on the pitch.

```
#repeating function that calculates distance of each player to line
between  player in possession
#and all other players on the field
passing.lane <- passing.lane %>%  group_by(time) %>%
  mutate(d1 = sqrt(((Bx-x[which(id==1)])*(y-By)-(By-y[which(id==1)])
*(x-Bx))^2)/sqrt(((Bx-x[which(id==1)])^2) + (By-y[which(id==1)])^2))
%>%
  mutate(d2 = sqrt(((Bx-x[which(id==2)])*(y-By)-(By-y[which(id==2)])
*(x-Bx))^2)/sqrt(((Bx-x[which(id==2)])^2) + (By-y[which(id==2)])^2))
%>%
  mutate(d3 = sqrt(((Bx-x[which(id==3)])*(y-By)-(By-y[which(id==3)])
*(x-Bx))^2)/sqrt(((Bx-x[which(id==3)])^2) + (By-y[which(id==3)])^2))
%>%
  …
  mutate(d22 = sqrt(((Bx-x[which(id==22)])*(y-By)-(By-y[which(id==22
)])*(x-Bx))^2)/sqrt(((Bx-x[which(id==22)])^2) + (By-y[which(id==22)]
)^2))
```

each team is extracted individually, and the closest defender to each passing lane is identified for each instance depending on what team has possession.

```
#extract line distances for each team individually
team1 <- passing.lane[,35:45]
team2 <- passing.lane[,46:56]

#calculate minimum distance to line for each group
team1$min_col1 = colnames(team1)[apply(team1, 1, which.min)]
team2$min_col2 = colnames(team2)[apply(team2, 1, which.min)]

#make data frame
min_col1 <- as.data.frame(team1$min_col1)
min_col2 <- as.data.frame(team2$min_col2)

#change column names
```

```
colnames(min_col1) <- c("min.col1")
colnames(min_col2) <- c("min.col2")

#merge closest players with passing.lane df
passing.lane <- cbind(passing.lane,min_col1,min_col2)
```

Reference values are created to identify who is closest in each team. the *d* string is removed in front of the reference values using the *gsub* function and the values of players who are not defenders in that instance are changed to 0.

```
#when team 1 possesses. find closest team 2 players to each line
#when team 2 possesses find closest team 1 player to each line
passing.lane <- passing.lane %>% mutate(ref = case_when((team == 1 &
team.pos == 1) ~ min.col2,
                                                        (team == 2 & team.pos == 2) ~ mi
n.col1,
                                                        T ~ "d0"))

#remove "d" from column to match up with id column and make numeric
passing.lane$ref <- gsub("d","",as.character(passing.lane$ref))
passing.lane$ref <- as.numeric(passing.lane$ref)
#change to 0 in cases where closest player is not relevant
passing.lane$ref <- ifelse(passing.lane$ref == 0, passing.lane$id, p
assing.lane$ref)
```

New columns are created containing the x and y coordinates for each player identified as closest to the passing lane.

```
#identify x and y value for closest defender to each passing lane
passing.lane <- passing.lane %>%
  group_by(time) %>%
  mutate(Cx = map_dbl(ref, function(ref) if (ref > 0) .data$x[.data$
id == ref] else 0)) %>%
  mutate(Cy = map_dbl(ref, function(ref) if (ref > 0) .data$y[.data$
id == ref] else 0)) %>%
  ungroup()
```

the angle of the passing lane to the closest player is calculated and the GK is removed as a passing option

```
#calculate angle between a(receiver), b(passer) and c(inercepter)
passing.lane <- passing.lane %>% mutate(a = sqrt(((Bx-Cx)^2) + (By-C
y)^2)) %>%
  mutate(b = sqrt(((x-Cx)^2) + (y-Cy)^2)) %>%
```

```
  mutate(c = sqrt(((Bx-x)^2) + (By-y)^2)) %>%
  mutate(angle = 180/pi*(acos((a^2 + c^2 - b^2)/(2*a*c))))

## Warning in acos((a^2 + c^2 - b^2)/(2 * a * c)): NaNs produced

#remove GK data
passing.lane <- passing.lane %>% filter(player.id != 1 & player.id !
= 12)
```

The below code identifies if each passing lane is open if the angle is >10°. A count is performed to identify how many open passing lanes there are for the player in possession. If no passes are on, the decision pressure is 1, this decreases in 0.25 increments until there are 4 or more simple passes open. In such an event the player on the ball is perceived to have no decision pressure.

The next step then combines all the data and calculates a total pressure value

```
passing.lane <- passing.lane %>% mutate(open = case_when(angle > 10
~ 1,
                                                          T ~ 0)) %>%
  group_by(time) %>% mutate(options = sum(open)) %>%
  mutate(decision.pressure = case_when(options == 0 ~ 1,
                                       options == 1 ~ 0.75,
                                       options == 2 ~ 0.5,
                                       options == 3 ~ 0.25,
                                       T ~ 0))

#create only 1 row per isntance of passing lane
passing.lane <- passing.lane[seq(1,nrow(passing.lane),20),]
```

The space pressure, time pressure and decision pressure values are isolated and bound into a separate data frame along with a corresponding time value. From here, a weighting factor of 0.7 for space pressure, 0.2 for time pressure and 0.1 for decision pressure is used to calculate a total pressure, ranging from 1 to 0, where a value approaching 1 is high pressure and a value approaching 0 is low pressure.

```
#create df with only pressure values and time
press <- cbind(TTC[,c(3,40,53)],passing.lane[,68])
```

```r
#calcualte overall pressure with 0.7 weight for space, 0.2 for time and 0.1 for passing lanes
press <- press %>% mutate(pressure = 0.7*sp.pressure + 0.2*time.pressure + 0.1*decision.pressure)
```

The calculations are not meaningful when the ball is not currently under control as the closest to ball calculations are not relevant in this context. Therefore, a data frame containing when each touch occurred in the match is used to isolate moments when a player is touching the ball, so the pressure is more representative.

```r
timetouch <- read.csv("timetouch.csv")

contact <- timetouch[,2] # time of each touch extracted

touchpress <- subset(press, time %in% contact) # calculate the pressure on each touch
```

Time values of each touch are bound to the dataframe. Both dataframes are then joined together so pressure values where the ball is not being touched are also included but identified as NA. Pressure values are then interpolated using `na.approx` and then transformed into a dataframe.

```r
time <- as.data.frame(seq(0,nrow(press)/25-0.04,0.04))
colnames(time) <- c("time")

interpolate.press <- left_join(time,touchpress)

## Joining, by = "time"

pressure.df <- na.approx(interpolate.press)
pressure.df <- as.data.frame (pressure.df)
```

The dataframe is arranged for visualisation, moving it to three columns including *time value, pressure value* and the *pressure type*. categories include *space pressure, time pressure, decision pressure* and *total pressure*

```r
#make dataframe suitable for ggplot animation
```

```r
s <- pressure.df[,1:2]
t <- pressure.df[,c(1,3)]
d <- pressure.df[,c(1,4)]
p <- pressure.df[,c(1,5)]

#add in type column for groupings
s$type <- "space pressure"
t$type <- "time pressure"
d$type <- "decision pressure"
p$type <- "total pressure"

#rename columns so they are joinable
colnames(s) <- c("time", "pressure", "type")
colnames(t) <- c("time", "pressure", "type")
colnames(d) <- c("time", "pressure", "type")
colnames(p) <- c("time", "pressure", "type")

pressuredata <- rbind(s,t,d,p)

### make animated bar graph of pressure measures
```

The time window selected for analysis is selected across the dataframe. An animation of a bar graph is created that displays the pressure, providing colour insight where low, medium and high values in pressure are yellow, orange and red respectively.

```r
df <- pressuredata %>% filter(time > 1988.56 & time < 2001.12)





p <- ggplot(df, aes(x = type, y = pressure, fill = pressure)) +
    geom_col() +
    scale_fill_gradient2(low = "yellow",
      mid = "orange",
      high = "red",
      midpoint = 0.5,) +
  coord_cartesian(ylim=c(0,1)) +
  labs(title = "pressure")



  anim = p + transition_manual(frames = time) +
    labs(title = "pressure")
```

```
  animate(anim, nframes = 12.56*25, fps = 25, renderer = av_renderer
())
## nframes and fps adjusted to match transition
```

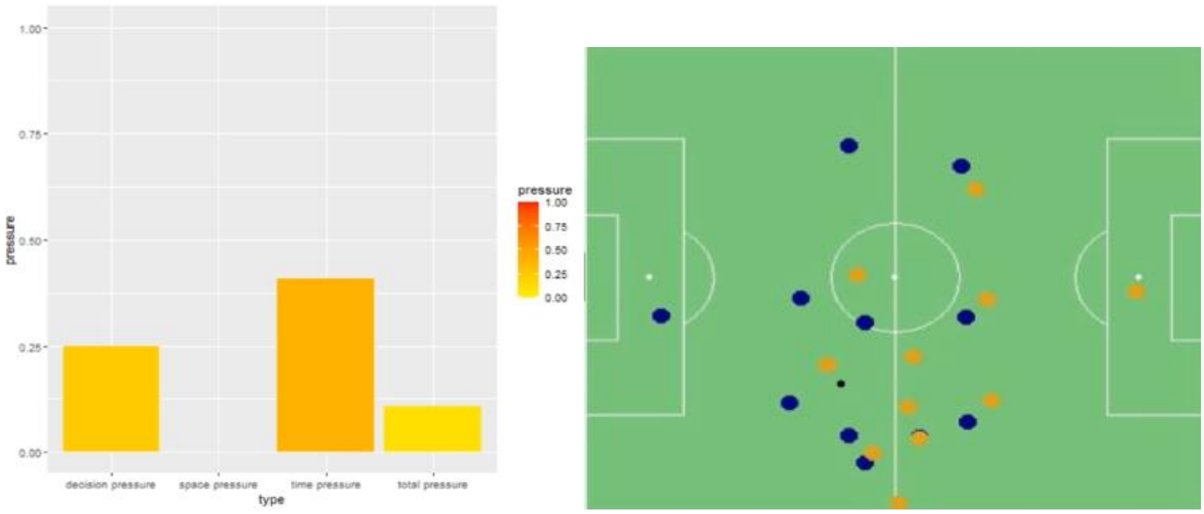


Figure 7.11: Pressure bars and top-down view of pitch output of visualisation.

The range of metrics covered in this chapter, and the accompanying code offers practitioners guidance in applying the methods used through this thesis. If required, further adaptations can be made to suit coach preferences depending on their own philosophy and principles of play. The accompanying visualisations allow for analysis processes to be understood more easily by coaches and players helping them relate mathematical concepts to their own beliefs and strategies in the game.

CHAPTER 8

Conclusion

This Chapter provides an overview of the thesis, highlighting the original contribution to the research base, the limitations of the work, as well as identify how the findings comprised herein can be applied in practical settings. Finally, considerations for future research will be made. Firstly, it is important to reflect on the research question that this body of work revolves around, namely, to evaluate whether novel collective behaviour metrics are capable of monitoring, influencing, and predicting the performance of football teams. Following the adapted model presented in the introduction, the research in this thesis identified important questions for novel approaches to be integrated into practice. Subsequently, the approaches were evaluated to be reliable, impactful and valid and were then developed further to assist coaching processes. Further to this, an assessment of these new approaches was required to be conducted to assist the application of evidence-based practice in sport.

At the outset of the research project, the area of study was still in relative infancy having only come under serious investigation in the previous 10 years. Naturally, there were still broad areas within the literature that required study. The two separate systematic reviews on spatial-temporal and social network metrics assisted in streamlining the areas of focus within the research field. Similarities in limitations were identified in the two systematic reviews including: 1) authors scarcely providing conceptual clarity relating to why metrics would be relevant to coaching principles; 2) authors often providing loose or broad practical applications where guidance on applying conclusions are vague; 3) nearly all the research examined male athletes; and 4) research being conducted with unjustified sample sizes. Additionally, research involving spatial-temporal analyses ignored or assumed appropriate reliability of measurement systems when calculating metrics as well as seldom providing clear limitations to the research. On the other hand, studies applying social network analysis appeared inconsistent in measuring network properties by applying different equations for the same metric. Additionally, there was a much greater focus on competitive adult matches and little examination of training and youth development.

Understanding these limitations through the systematic reviews provided a foundation for the next phase of the research investigating limitations within the literature that are critical for practical application. Firstly, the reliability of spatial-temporal metrics were analysed examining the impact on a range of variables through simulating expected error values of position measurement systems on top of real-world data. This highlighted that current systems with an accuracy of measuring players position to 0.5 m are likely to be reliable. Systems with less accuracy may still be reliable in measuring some spatial temporal metrics such as team centroid, however, practitioners should be cautious about implementing other metrics, such as distance between centroid or dominant region if systems being used demonstrate measurement errors as large as 2 m. Additionally, restricting analyses to short sections of play such as attacks which are often done in practice, may lead to low reliability for certain metrics.

The final investigation in this research explored coach perceptions of collective behaviour metrics applying spatial temporal or social network analysis. Through iterative interviews and presenting visualisations of metrics, clear links were created between a selection of metrics and important tactical principles of international coaches. This investigation proposed an iterative process for practitioners to create a system to assist and influence the development of youth players and preparation for matches of elite teams. Additionally, it identifies that individual coaches have a separate understanding and conceptualisation of tactical principles from one another. Consequently, practitioners must work alongside their own coaches to manifest and adapt their own specific set of metrics through repeating the processes presented in this thesis. This research clarifies one of the primary weaknesses of the literature base through anchoring measurements within concepts and principles that are understood by coaches and through the pedagogical chapter, offers instruction to practitioners wishing to replicate the analysis procedures used.

8.1 Limitations

This thesis has provided an original contribution to the literature base investigating collective behaviour metrics in football. However, there are limitations to the methodologies applied across the research project that have largely been covered in each individual chapter. Across the whole thesis, the biggest weakness is a lack of exploration of the novel metrics identified through the coaches in chapter 7. Coaches identified that certain tactical metrics were relevant and related to concepts they coached, however their ability to explain or predict performance or results was not investigated. Moreover, based on the feedback of coaches, further adaptations to the metrics could be made more suitable for analysis. Another limitation of the work is the minimal focus on non-linear analysis methods that measure concepts including predictability or synchronisation through entropy and relative phase among other approaches. These techniques additionally require understanding with regards to the reliability of their output and whether the concepts they measure are relevant to coaches. These techniques have the potential to offer an alternative to understand the complex behaviours of football teams. There are many reasons why the scope of the thesis is not as comprehensive as initially planned including limited access to data and time constraints.

8.1.1 Impact of Covid-19

The Coronavirus pandemic and subsequent protocols enforced worldwide had a major impact on the planned research involved in this thesis. Staff at the Scottish FA were furloughed and unable to work, consequently making it impossible at points to collaborate, including at a time when translating the metrics into concepts understood by coaches was the primary area of investigation. The shutdown of Scottish football for 5 months and postponement of international fixtures and the shifting of priorities for the Scottish FA in dealing with the challenges of the pandemic. Additional time could have allowed for further collaboration with the Scottish FA to perform an evaluation on the metrics identified in chapter 7. More generally speaking, the lockdown procedures in Scotland provided their own difficulties through not allowing the author access to a creative space that promotes efficient working. Moreover, the cancellation of conferences and other events severely limited the possibility of networking and creating partnerships that could lead to novel applications of data analysis.

8.2 Practical applications

The thesis provides an overview for practitioners on how to apply collective behaviour metrics. This includes a systematic overview of the primary metrics used in both spatial temporal and social network analysis. A system for evaluating reliability of spatial temporal metrics is also provided so practitioners can be confident the tools they are applying are trustworthy. Information regarding the impact of bio-banding and pitch size alteration on network structure in small-sided games are offered. Most importantly, a generalisable methodology has been presented where coaches can replicate the iterative interview method within their own teams and replicate the methods used to create specific metrics that are relevant to an individual coach’s tactics and beliefs. Additional support for replicating the analysis procedures in R studio are provided in the pedagogical chapter where practitioners can use and adapt the code provided to measure principles of play.

From the range of metrics explored in this work, there are several that resonated with coaches and can be recommended to practitioners. Included within this list is 1) network intensity, a measurement of the number of passes per time in possession, 2) distance between defenders, modelling the defensive unit as a line and the distance between each point in the line, 3) triangles, conceptualising 3 players as a unit and measuring the distance, angles and area constructed by the shape they adopt, and 4) depth measurements, described by the distance between the deepest defender and furthest attacker, and the distance between the deepest defender and the goal line. Similarly, other measurements including surface area, team width and pressure may also be relevant to informing coaching sessions. Moreover, the aforementioned metrics can be adjusted and personalised to suit an individual's philosophy. Conversely, coaches were hesitant of pitch control, a method for calculating likelihood of pass success rate through arranging areas of the pitch that players are likely to reach first. This method appears to have gained traction within research over the last 5 years, yet coaches were sceptical over its application to training sessions. the pitch control model may be useful for other practical tasks in football teams such as talent identification through measuring a player's ability to make difficult passes.

8.3 Future research

The thesis provides a solid foundation on the practical application of collective behaviour metrics applying spatial temporal and network analysis techniques. However, many avenues for further research are still required for this type of analysis to maximise its impact. One of the most critical and challenging areas of focus is understanding the signal of some collective behaviour metrics. This is less important in measurements such as network intensity where direction is clear in the sense that faster passing relates to a higher performance. However, for metrics such as triads, it is assumed there are sweet spots or goldilocks zones for the properties of triangles created. Examination of multiple various positional triads could be observed to identify the regular measurements of elite players. Further understanding of how training specific collective organisation can then be analysed through manipulating constraints in 3v3 games including pitch shape and dimensions, number of goals and factors such as maturation level for youth players. Data from this information can be gathered and inform training practices to encourage intelligent dynamics for individual players and these principles can be extrapolated to other variables explored in this thesis to inform coaches of the impact of their training practices.

Additionally, further research into several of the metrics introduced into this thesis are necessary before adoption into applied settings. Specifically, the constants selected for weighting the pressure metrics. The coaches agreed that the pressures made intuitive sense, however, further investigation into how the pressure impacts the decision making of players on the ball is critical in understanding how to train players to apply pressure to ball carriers, as well as guiding players to using the ball when under pressure. Other values that require further investigation include the angular threshold for decision pressure and the spatial threshold for surface area. A final recommendation for future research is further grounding of network metrics into clearer relationship with performance. This thesis has leaned towards spatial temporal metrics, partly due to measurements more easily being conceptualised by coaches due to their intrinsic link to space and time; two central constructs within football. Currently, many network structures largely ignore aspects of space and time within the passes made and consequently lose important information in the structures that are made. To maximise such a powerful analytical tool, integration of spatial temporal data into network metrics have the potential to provide greater impact and deeper insight.

8.4 Final remarks

This thesis has provided an original contribution to the development of collective behaviour metrics in football, providing insight into the current limitations of the literature base, and evaluation of the reliability of metrics. The thesis also succeeds in anchoring current and adapted collective behaviour metrics with coaching concepts and principles, and provides guidelines for practitioners to follow helping create specific and relevant metrics for their own team. Many challenges still lie in wait for this discipline to become commonplace among professional teams, however, some football clubs have begun to implement some of the related models explored within this project. It is an exciting time to be involved in the development collective behaviour metrics in football and it is the author's belief that they will become a foundational pillar in football performance analysis.

Appendices

Appendix I

This table outlines the tagged quotes made by coaches through the interviews in chapter 6, as well as identify what theme and main theme they have been attributed to.

sub themes & quotes	themes	main themes
diamonds and triangles	support	Penetration
"I do have a preference for play in a back four with a diamond shape in midfield" (1.1)		
"So going back to your point about the diamonds, I think in possession it gives it gives you obviously, almost like 3 different areas within the midfield" (3.1)		
"just going back to your point, in terms of the midfield and the triangles, that that absolutely is pivotal in terms of trying to control games and win games " (3.2)		
"whether the triangles match, because not sometimes it's just say, my team are playing two holding midfielders and the number 10. So, in my, the way, I see the game that's triangle up and the other team might be playing triangle up as well, which means there's not, it's not man for man" (3.2)		
"When you find a lot of the teams, the lesser team don't like that they want the triangles to match. So they will play they will play the triangle down to match up with the team that so basically to make man for man in midfield" (3.2)		
", I definitely think you could say the triangle between fullback wide player and midfielder. So Right back, right, midfielder on right, central midfielder. There's also triangle between right centre back, right, midfielder, and right back those kinds of		

progressions as the team builds through the thirds. And the you could also probably say the same between the wingers, the striker and the attacking midfielder if you've played with a ten" (3.2)	
passing options	
"control the middle of the pitch, which is, I think quite important in the game gives you know , good opportunities to try and get penetration." "And the teams understand that has to be good and if we're utilising let's say the nine as the target, what then happens after that, who's the expectation of support is it the seven is at the 11, we're probably asking the 10 and the eights ago or if the eights position isn't good, can the six go and the eight holds as we move up the pitch, but I think the important thing is that, and it happens to us a lot, you know. In the youth game that we struggled to get support to the ball quicker off against really good teams so it can get can come back on" "in possession, it gives you really good short passing options" (3.1) "But at the end of the day, it comes down to quick time decision making and execution. Do people actually… they may see on a screen there, but when they're on the pitch, and the environment is significantly different? Do they still see? You know what I mean? Those passes" (2.2) "I'm being cheeky, even though [team 2] are compact, there are still pass options through them available" (3.2) "So look how high the defenders are in relation to play. So they're creating… There are extra pass options" (3.2)	
angles	
" I think angles are important, because obviously, if you're looking for penetration, you have to try and play forward. I think playing forward in straight lines is much easier for the opposition to defend against as pardon me, much more difficult for the player particularly, it's going to receive the ball to be in a	

position where they can see where the ball is coming from and where you want to go" (1.1)

"in part why I like that formation because you get to get some natural angles occurring between you know, even the two central defenders and your deep or your base of the diamond you get good angles with the you know, obviously all four members of the diamond" (1.1)

"just shows, it helps players understand how they are creating those angles" (2.2)

teammate distances

"Yeah. And it's good to show as well, that, you know, sometimes I know that, in that game. Last week, where our nine was and where our back line was, there was probably only about maybe 20 or 22 yards between them" (2.1)

"So we're quite happy for the distance between the Center back and the fullback to be there, because we know that that central midfielder can drop in there as well" (2.2)

"think what's important for young players there is, is you can see that what each reaction, you know that you're trying to keep specific distances between that so there's a midfield balance, if that makes sense" (2.2)

"The distances are really important. But I also think it's the players that that need to sort understand that, you know, you don't just move to support the ball" (2.2)

coordination

"but I think real success comes from relationships and players being instinctive, and trusting them to be proactive" (2.1)

"it's about players identify and actually themselves, we have the spaces to play, and what the relationship or what key personnel they have around them to help manipulate that situation in their favour" (2.1)

“one player will run away from the ball and one player might go towards the ball, or they might even cross crossover, then one of the midfield players might run in behind, and one of like one of the strikers could drop in one of the wide players could come in” (3.1)	
overloads in wide areas	width
“can commit your own numbers forward and create overload so you hear that terminology used quite a bit, create an overload, especially in the wide area, you know, I've got a winger can and commit a midfield that and a fullback out there and create a three versus two and try and you know, take advantage of that numerical superiority in those areas” (1.1) **“I think you always look to create overloads obviously, in wide areas, that's probably the easiest place to do it. But also overloads in the middle of the pitch, can be quite significant dependent on the shape that your opponents playing” (2.1)** **“Normally, in the why there is I think it's very hard to do that central. There's a lot of times we asked the one of the midfielders to go out into the fullback area, like we call it false fallback movement.” (3.1)**	
creating space	
“then use the ball to find those spaces or opportunities to penetrate” (1.1) **“so if the if the winger receives the ball very wide on the touchline the space to actually make the forward run and to be inside” (1.1)** **“then if you can play through that you then go into the attacking half of the pitch where they might have less defenders and you've created you know, more space for you to go and attack” (1.1)** **“the opponent may be to have committed a lot of numbers forward so a big space with few opposition's “ (1.1)**	

“there's lots of space, and we've got the opportunity to get attackers into that space, and we can play the ball in that space. So that's a break” (1.1) **“they're able to play a pass forward. You know, I should be thinking to myself, well, can I make a forward run You know, to receive a pass beyond their midfielders or even beyond their back four, you know, in an area that might create some penetration, you know, can I make a run that might affect the defender and adjust make him have to adjust his position, that creates a space for my teammate on the ball to either dribble into or play a pass,” (1.1)**	
disrupting opponents	
“we've created some distance between opposition players“ (1.1) **“I think I t's, well, a lot of the players are still learning about that. And they can get dragged out of position. And the opposition are good enough to move you to create space for another man, another player” (3.1)** **“I think it's always really precarious when one of your central defenders is, you know, arguably outside the width of the penalty box” (3.2)**	
attacking shape	
“can we control possession and make sure that we make good use of the width of the pitch? “ (1.1) **“if you are playing against a team who likes to press high, you know, you might encourage your team to take up certain positions, almost to attract more players toward the ball” (1.1)** **“you want to attack with that group of six, you know, trying to be trying to get themselves in front of the ball, if they're not already there, and a group before just sitting behind the ball when you're not attacking part of the pitch. And that gives you sort of good balance.” (1.1)**	

"three passes to kind of, you know, secure possession, until you can actually go from more of a defensive shape into your attacking sort of shape, which is normally a little bit wider and bit more expansive" (1.1) **"for an expansive shape, I'm looking for centre to split fullbacks to go High, I'm looking for my three midfielders to play on three different lines, if that makes sense, I'm looking for my wide players to probably be wide, but not be in line with fullbacks and I'm looking for the nine to kind of stretch the game" (2.1)** **" I think in this in the scenario you're describing, I think it's very important to establish how you want to get started, whether how you're going how you're going to I think a lot of the times the there's a real need in that scenario you're describing for the central defenders to step out with the ball how you're going to do that is one of the midfielders going to drop in to make it a back three to make it slightly wider." (3.1)**	
passing speed	mobility
"how can we move the ball quickly , quite often from side to side to get the team opposition team shifting side to side and look for those opportunities to go between or you know" (1.1) **"you have to get the speed the game up" (3.1)** **"But a lot of teams, the performance, the speed of the performance dips, the longer the game goes they can't maintain the speed, the intensity. And in my opinion, good teams and good players can keep the intensity all the way to the final whistle" (3.1)** **"the factor for me in that scenario is how fast the ball is travelling to the player." (3.2)**	

“So the speed at which the ball is being received is also connected to the applied speed of pressure, in my opinion” (3.2) **“And I really believe in that and you know, if we're playing a pot six team or a pot five team, and you have to, you have to try to expose. they can't play at the same speed as what you can” (3.2)**	
contact time	
“That's what the top players can do their contact time on the ball is absolute minimal because you're trying to pass the time on to the next player to find the space. So the ability for players to play at speed is crucial” (3.1) **“I think it it comes down to opportunity, you know if it's in that transition moment, if the opposition quickly can establish comfortable possession with with you know, it might just be one first time pass where where there's just no opportunity to win the ball back to tackle because the opposition the opponent has one First time passed the ball into a different area” (3.1)** **“so critical in terms of player development, team development, winning games, absolutely love that and, is something I really believe in speed of play, obviously, it's not just the speed of the pass, it's the contact time in between, you know, the amount of time it takes a player to control the ball and play the ball” (3.2)** **“I'd be really interested to know what your contact time is, at the very top level” (3.2)**	
movement	
“you try to focus on mobility, so, you know, how can we get runs that will, in a sense destabilise, the opposition's organisation” (1.1) **“what you require from the players is obviously you know, forward runs” (1.1)**	

"when you're obviously in more of a creation sort of phase closer to the opposition's goal, space and obviously, time in possession are mostly A lot more condensed. So you have to kind of, you know, move and think and play a little bit quicker." **"successful counter attack is obviously speed. You know, can you can you play forward quickly" (1.1)** **"so when we say speed, doing things quickly, it will be thing, you know, a short number of seconds, three, four or five seconds can progress quickly up the pitch, you know, if you've kept possession, but maybe been forced to play, you know, sideways or backwards, within a couple of passes," (1.1)** **"But the key thing is that application again, of one of the attacking principles, which is mobility, you know, anytime someone's in a position, we need players to try and make movements that will threaten the opposition" (1.1)** **"you've got your awareness, trying to move the ball away from pressure" (2.1)**	
risk	creativity
"I think you will find that it's it's ones that they've chosen the pass and it goes into that white or yellow area, there's a turnover, you know, why have we chosen that pass? " (1.1) **"I think it's it's having that element of control" (1.1)** **"regain the ball high up the pitch or in certain areas in the opposition's half when they are trying to build up. You can create good goalscoring opportunities very quickly from those situations" (1.1)** **"But some people do, you know, take less risks in the build up phase, because you give away possession in that area, you can lead to chances of being created against your team, or maybe taking more risks, or the risk of losing possession a bit higher up the pitch, because the reward is, you know, a pass that gets**	

between or around or over. Or a dribble that beats a fullback, for example, creates a chance" (1.1) **"I think the moments, you're kind of referring to the, the player has an immediate decision on whether there is an opportunity to expose the transition , and obviously, you know, are the opposition out of shape? Or? Or is it a moment to try to retain possession and not risk the ball? " (3.1)** **"the only the only thing that changed in that in my mind in that regard is risk, I'll take more risk against the bottom 6 team. In terms of the pot one teams, you know, that the track the defensive transition that you've already highlighted in your videos, the top teams will punish you on the transition" (3.1)** **"if I really want to reduce the space, then the forwards will engage on the halfway line, if I'm willing to take slightly more risk, because I want to engage further up the pitch," (3.1)**	
breaking lines	
"trying to penetrate opposition defensive line" (1.1) **"they can play between lines, cant they, they can break lines" (2.2)** **"think when something is creative it penetrates a backline or the end result as potentially maybe getting in behind or creating an overload situation" (2.2)** **"Creativity for me is all related to creating chances and goalscoring opportunities. And obviously, that usually means breaking the offside line" (3.2)**	
patterns of play	
"is it going around because we've, you know, move the ball in such a way that it's got the opposition concentrated but concentrated more to one side of the pitch and you can go around , or have you moved the ball in a certain way or made some movements to get in behind the opposition's defence	

when there's some space in behind you can actually play a ball over the top so it's, it's through around or over the top" (1.1) **"I think it can show up patterns of for example" (2.1)** **"they actually don't bed don't patterns, at Rangers, they actually just let them play because patterns are good, but they're predictable" (2.1)** **"So one example would be like, don't like the opposition to be able to play the ball inside our wide player. So I look I'm something I look for during the game" (3.1)**	
deception	
"a player can literally just go face to face with a defender at the top level unbeaten with it with a fake a turn some kind of trick" (3.2)	
1v1	
"I think sometimes that gets a lot of the attention of the ones that can come up with those moments of improvisation, you know, whether it be good dribbling skills, or you know, I guess a lot of people refer to it as a flick or a trick, you know, to try and get that that little moment of penetration" (1.1) **"you may try and isolate somebody in a one v one dribbling situation, you know, so if you've again, got the ball in one area of the pitch and you've attracted a few defenders to it, you may ask one of your own players is quite a talented dribbler, one v one skills or even someone with a lot of peace, to just remain, you know, in a wide position, you get the ball to them." (1.1)** **"It's a one v one and if that midfield, midfielder, goes to join its a 3 v 2 isn't it, see if they recover quicker than the three that we've got on that right hand side" (2.2)** **"But it also captures a really crucial detail in terms of a defender in the defender, who is one v one has to slow him down enough for those three players to become relevant and to get back" (3.2)**	

"the really, really top players get paid the big bucks, because they can create their own chances, they don't need somebody to create a chance for them, because they can beat their opponent one v one, which is obviously an incredible skill" (3.2)		
defensive shape	compactness	Delay
"the first thing I'd look forward is the moment that we lost possession, what was the organisation of the team" (1.1) **"the right balance of players in the right positions" (1.1)** **"compactness from side to side, so not much distance between" (1.1)** **"at the fundamentals, so obviously, at the end, compact, deny in space" (2.1)** **"other situation maybe, you know, if we're, if we're working across the pitch, sometimes the players will become extremely narrow, and be very calm, very fixated on what's happening across the right hand side of the pitch, or making them understand that you can only come as far as if the switch the ball and two passes, we can then move back across to defend on that side" (2.1)** **"how quickly you can get back in shape after you lose the ball. And that is something that we coach" (3.1)** **"the tactical instruction of the coach and the team, knowing whether on those transitions, whether it is what is the objective to get back into shape, and be compact as quickly as possible, like you're speaking about, or is it to try and win the ball back immediately and to actually counter press? " (3.2)**		
reaction		

“there's a transition, where's the importance? Or where's the risk that I maybe need to protect, you know, and step into?” (2.1) **“how they should then react” (2.1)** **“Reaction , a quick an immediate decision on whether to press the ball and win the ball back, or whether to get back in shape behind the ball” (3.1)** **“guess is identifying what is the awareness in transition? Isn't it? Are, we reacting to it” (2.2)**	
recovery	
“ideally, you'd like to have them back within two or three seconds, but that's not always possible.” (1.1) **“the difference between two seconds to get back in your shape could mean the difference between conceding a goal or not” (1.1)** **“Or, you know, you can recover and get all the 10 outfield players plus a goalkeeper in a goal side position” (1.1)** **“numbers either sometimes become even, or, you know, the opposition gets, if you've got four players committed the attack, and they get five back, well, then your opportunity to counter attack is kind of lost” (1.1)** **“we get back in to shape and if that happens, then we all do it together. And it has to be fast and we have to get back in and stop the opposition hurting us.” (3.1)** **“but so many of those recovering players, in my opinion, especially at the younger age groups, they don't realise the importance of just getting back. And actually the what if scenario, like, for me, that always is a what if scenario. And I think a lot of the younger players really struggle with that” (3.2)**	

controlling opponents decisions	
“if an opposition to a counter attack, but we're able to kind of funnel them into a wide position” (1.1) “work a lot with the national team at the moment of protecting the middle of the pitch. So it's about making sure that we try and, you know, keep numbers in there for a start and makes it really difficult for the opposition to play in that area And if we're applying pressure, can we show them into a wide position, so can we make them dribble or pass into a wider position” (1.1) “the angle that you go and approach the opposition player as well and get that pressure block passing lines , you'll hear that one quite a bit” “So if we apply pressure to the ball, and my the player that's applying pressure, you know, does it from obviously kind of one side, if you like, and stops a pass into a certain area, the defenders, then in the next line, you know, are aware that it's very unlikely that the opponent's going to play into that area, if the pressure is right.” (1.1) “And we look to force the opponent to play around the outside” (2.1) “pressure can also be about you know denying in space or channelling your opponent so that the space that we have to plan becomes smaller” (2.1) “I think it's very important to try and disrupt the opposition and make them play passes, they don't really want to play” (3.1) And once that starts, once it starts like that against a top team, they're like the red arrows. So if you can stop it at source then obviously helps you and you're almost forcing them to play passes, they don't want to play (3.1)	

"you haven't got wide players, so you are exposed in the wide areas, but the moment the ball goes wide and you were forcing the ball, wide So it takes out the randomness where we almost know where the ball is going," (3.1) **"And I think it's like that when we're making the play predictable. So our players know where the ball is going to go. And they know their jobs. Once it goes there, you take out the randomness. And I definitely don't want the opposition to play through the middle of us because I think that's very dangerous" (3.1)**	
anticipation	
"you like in your decision making. And almost being prepared to lose possession" (1.1) **"anticipating that loss of possession, I think once a possession has turned over, depending on the position of those players, where we then still able to have somebody apply the right kind of pressure to the opponent in possession to try and slow them down" (1.1)**	
length	depth
"ones where they kind of always have a constant distance from the deepest defender, you know, maybe one of your centre backs is behind the rest of the line" (1.1) **"it was very evident to me, but probably difficult for us to show to the players that there was a huge gap, when we were in possession and very, very good, controlled possession higher up the pitch, you couldn't actually see our backline. Now, yeah, that was one to do with the camera angle. But that was also to do with the fact that they had to squeeze the game" (2.1)** **"And our probably our biggest sort of reference to those thirds are, you know, if we're in possession, we would expect our backline to be on the border of the first and the second third to squeeze the game to prevent the counter and to prevent the big spaces" (2.1)** **"Its still an important thing and your players have to understand that, when you're out of possession, that there's a**	

lot there's a lot of times where it's important to have depth, width and length." (2.2)	
lines	
"Also looking for that compactness from defensive line to front players. And I quite like the visual that you showed on your software there before with the three lines" (1.1) **"the exact opposite of that we become compact and tight with smaller spaces between the lines and across the pitch" (2.1)** **"making sure that we're really compact across the pitch and between the lines is important" (2.1)** **"it's very important to the defensive line, get, get the line correct as well, in terms of the distance, not too deep, not too high" (3.1)** **"it affects the distances between our forwards and our midfielders, and our midfielders and our defensive line" (3.1)** **"but you don't want the opposition to have comfortable possession in the middle of the pitch in between your lines. So by the striker dropping in and reducing that depth, you're kind of trying to make the play predictable and just trying to control it a bit more." (3.1)**	
cover	Balance
"Do we need to slide right a little bit, etc. and making sure obviously, they've got that depth of that little bit cover" (1.1) **"looking for the closest centre back to drop and give cover and sometimes what happens, you know, with a young player, as the young player becomes fixated, the centre back will become fixated on the striker, where actually the first priority is to give cover if that player if that fullbacks beat in a one v one situation" (2.1)** **"one thing that we speak about a lot is kind of in midfield balance, you know, so making sure that defensively, we're not**	

flat. And we're looking to potentially apply pressure, but also cover space” (2.1) **“I think that has it's about it's about covering the space and, and being able to cover the movement of the opposition” (3.1)**	
overloads near the ball	
“And if we the end the centre of the pitch, do you need to step in? Or is it better for you to delay at that moment in time the progress of that player, and for me, it's better to delay because the amount of times that we've seen a player step in, and then the overload kills us centrally” (2.1) **“I think you've really captured a critical moment in the game there where, you know, the model showing that there's no overloads. But actually, the team are really in trouble” (3.2)**	
adjusting	Discipline
“we then have to go and we have to re-evaluate the situation again” (2.1) **“I agree view, they're very important. And it's one of my most common shouts to the team is to adjust” (3.1)** **“our players know how to adjust like the if the left hand side of the diamond goes to the ball, the top of the diamond drops down one and obviously the six at the bottom of the diamond adjusts over and the right hand side of the diamond can then adjust over because we all know where the ball is gone. And also the strikers have to react to that as well” (3.1)** **“So it I know, obviously related to pressure is it's not just one player pressuring you, is not just one player pressing one player, it's the whole team, there's a player behind the player, and the player behind him, there's a there's an adjustment happening, whether it's front to back or side side” (3.1)** **“if you're part of the Midfield three like that, and you're looking at balance, you need to react to the movements of the three of you that are playing in there together, you know” (2.2)**	

“So if they go back, and then out the other side, it takes longer and [team 1] can adjust” (3.2) **“I personally don't think the Scotland players aren’t adjusting enough. Like for me, they need to be adjusting more aggressively, especially in the right back .” (3.2)** **“But you have to be willing to run you have to be willing to, to, to read and then get back out and the whole team needs to get back out. It's, for me really, really important to get those adjustments. And always, you can't… you can't take risks. You always have to think what if” (3.2)** **“they need to be talking to each other and adjusting so that it has a knock on effect where that gap between the two centre halves gets closed” (3.2)**	
triggers	
“So getting them to understand the importance of who's where we start defender, from obviously the nine and what are the triggers” (2.1) **“our programme will go and high press most teams when they play on a week to week basis. So looking at, you know, there may be moments that we can high press, and what are the triggers?” (2.1)** **“we actually have possession of the ball, we move up the pitch the way the left player loses possession. Do we want to just drop back then? Or do we want to press No, we want to play it because we've got the team up the pitch. Whereas if we were looking at situations where it's the Dutch goalkeeper that has the ball, and they're expansive, are we going to set a high press? No, we're not. “ (2.1)** **“So it could be that we use the border between two and three. To… That becomes you know, the line of when the opponent comes in, that's when we trigger the press” (2.1)**	

“We speak about where we're going to engage with the opposition, whether it's at the top of the circle, whether it's the halfway line” (3.1)	
time	
“the speed at which you can approach the ball and get that get that pressure on” (1.1) **“on his touch and things that sort of tried to get the pressure on the opposition player as they received the ball, rather than getting there after they've already had a first and they’re in command of the ball then” (1.1)** **“I think first and foremost will be that we would ask, you know, the closest person to press the ball to then afford us time to get into shape to get into good shape” (2.1)** **“So we really forced Portugal to play out of their comfort zone, not giving them time on the ball, pressing them, really disrupting them, and not allowing them possession” (3.1)** **“usually means they're obviously closed down fast and their time on the ball is reduced” (3.1)** **“Well, usually in my opinion, it usually means, the person receiving the ball has more time, because the opposition have had less time to adjust. You know that you began this course speaking about the distances in terms of pressure. If you the distance, the difference between a first time ball or me controlling the ball, then playing it to my teammate usually affects how much time he's going to have when he receives the ball” (3.2)**	
distance to opponent	
“I think, yeah, distances is one key one” (1.1) **“things like get up to the ball, you know, so get close enough, so that the player in possession of the ball feels as if they're under**	

pressure. So it feels that the distance is quite tight feels probably that they're within risk of losing the ball" (1.1) **"you hear people say, like, you know, make them get their head down, you know, the closer an opponent gets to you, if you're in possession of the ball, the more you kind of have to pay attention to ball and get your head down, so you can't see the kind of longer penetrate and pass potentially" (1.1)**	
working as a team	
"Where's the ball? Where's my teammates, and where's the opposition player" (1.1) **"you'll have backline players that the ball was at the midfield and they just switch off. And they don't understand No, actually, my role. And my position needs to continue continually change in relation to can I be, can I be of support in attack" (2.1)** **"When we are out of possession, I suppose, the most important thing, and whether young players as the understand the role within the team's defensive unit, because like I said earlier, when I watch a lot of youth games, they very much play as individuals" (2.1)** **"you're looking at your sort of pressure, cover, balance situations . And then as you get deeper, something's understanding the difference between the delaying the opponent, and actually winning the ball and what is the priority at that moment in time" (2.1)** **"they're obviously trying to find space in between our players. And it's very important that our players don't get dragged if you go and watch, in my opinion, lesser players or junior players, they will just stay with their man they don't pass on, they don't there needs to be a defensive structure, an organisation where you don't get dragged around, and you have to do it efficiently in my opinion" (3.1)** **"Just because of the timing isn't it, it's the thought process of the individuals . So sometimes you know if the balls on one side**	

of the pitch, the fullback on the other side sometimes becomes kind of distracted by the wide player as opposed to actually be in cover and balance" (2.2)	
pressure	
"can apply pressure to, you know, the opponent and possession of the ball, high up the pitch as best you possibly" (1.1) **"okay what as your role as a first defender is to press and try and make the play predictable" (2.1)** **"it could be about physical distances. encouraging the, you know, if we can get called for an affair and colleges that are attacking players to put their head down and the decision making then maybe isn't as good or bad execution is stressed" (2.1)** **"the worst decision, in my opinion, is to try to get to the ball when you know you can't get to the ball, because then you just get passed around and you're out of position" (3.1)** **"it's the decision from the central defenders not to pressure once the transition happens, right there. Yeah, the defenders are protecting the space in behind them initially" (3.2)** **"They have a different objective first, which is to counter press, and to try and win the ball because they feel that that kind of flux, that moment of transition is their best chance to win the ball back" (3.2)** **"If you do force it back, like does the ball go all the way back to the halfway line? Like the moment it goes to the halfway line by the time he takes a touch there? The back line should be an obviously when it goes back to the goalie. The whole team has to you have to change your my opinion from protect to press . Like there's a there's a transition where the team is compact that you've you've done the objective has happened you've got what you wanted, you force them back on the team has to change collectively change from protect to press in my opinion" (3.2)**	

	player development	team performance
barriers to development		
“that creates a fantastic visual, you know, that I think, is very much clearer, and has less distractions” (2.1) **“So you're trying to keep things simple for them. But get them to understand it's not clear cut. So it's not always about the ball. And it's not always about the player that your marking” (2.1)**		
learning styles		
“there’s so many players nowadays that are visual learners and really struggle with, you know, if I'm on the pitch and I'm speaking about being compact or being expansive, they don't know how that transfers to visual so having a backup like that, you know, I think gives real clarity and all the time you know” (2.1)		
learning experiences		
“but I would think even for players, is quite a strong visual” (1.1) **“I think that is a fantastic visual and learning tool for young players. Because I think, when you watch the game, you can be very much distracted with the ball” (2.1)** **“one of the things when you're working with players that are of the level that we are working with, and they dominate the the ball a lot is that they, they can be really expansive. And because of the best players on that at the best clubs, they don't always understand the requirement to squeeze the game or to become compact, because individually, they're physically equipped to go and win the ball back and a one to one situation if they lose it” (2.1)** **“it is different in youth football for me, because for me, it's about the long term athlete development” (2.1)** **“our results will be better. You know, they'll be more respectable, but in terms of the players learning, you're denying an opportunity to try and for them to try and play under pressure as well” (2.1)**		

“because they're not faced with those high pressure situations within the within the club game, because a lot of them don't play a competitive enough level, in my opinion. And that's probably been much more visible this year, because we've missed 18 months of educating players to play out the most competitive, challenging environment both in training and games” (2.1) **“it's really important that you don't overcomplicate it with players as well, because I think sometimes simple is better” (2.1)** **“I think you have to recreate the moments in in in the pathways of young players, you have to keep putting them in those positions, in game related practices” (3.1)**	
available coaching time	
“But I think the biggest challenge for us will be in attack. Finding ways to penetrate our opponent because we have very little time to spend on it to develop those relationships” (2.1) **“You're lucky if you've got a couple of training sessions before you play so I just have one day on in possession and one day out of possession and I have to try to really, really keep it simple on key points. So that when the team Play, they have that kind of clarity on what's expected defensively on what's expected offensively .” (3.1)**	
flexible tactics	match preparation
“I'm one that very much strives to stick to the principles of the game, you know, those are the constant strains that will happen within the game, be it you know, formations or, you know, defending styles, even attacking styles, but, you know, so it's, it's making sure that you stick to the principles of the game and everything becomes about the application of those principles” (1.1) **“working with the younger players is, are we pressing? Or are we not pressing? And it's with, in my opinion, you have to do both” (3.1)**	

pitch size	
“So the size of the pitch becomes a challenge but that at the end of the day that is the same for both teams” (2.1)	
opponent ability	
“And with the hope being, that we could play against a team that are better than us, whereby the focus would be on how we defend and how we identify and utilise counter attacking opportunities. As well as obviously setting up properly, being organised and disciplined, you know, I'd say, say please, as well. And then on the flip side of that, can you play a team that's at your level” (2.1) **“depending on the strengths of the opponent, and the rotations of the opponent, and how good I think the opponent is, will dictate whether we engage at the top of the circle, or the halfway line” (3.1)** **“It depends on the situation, it depends on the opponent, you know, if you're playing against a team that's maybe more lower ranked, potentially would leave somebody up higher because of that potential to then count or at least you've got a target to hit. But if you're playing against somebody that's maybe have a higher rank, and then you're probably getting everybody behind the ball, and that's why you become as flat as you potentially do.” (2.2)** **“I think it’s different from match to match. And it's dependent on the strengths of the opponent as well and the qualities that they have you know, you're going to be much less likely to play a high line if you're playing against maybe teams that switch the play well and have real pace in the frontline And then there'll be other situations where you might be quite comfortable taking more risk. So I think it's so it's all kind of your opponent specific really” (2.2)**	
game context	
“we try and replicate in training “ (2.1) **“They have to play what they see at the end of the day. The Of course, the preparation is very important. And you want players that are coachable. You want players that that are willing to do**	

as they've been asked. But for me it comes down to them making the right decision at the right time in the moment of the game in the context of the game" (3.1) "I think it's about training it coaching it, game related practices, so that the players recognise those moments and have the intelligence to make the right decisions in the games" (3.1) "But I'll also prepare the team to be able to do different things within the same game, depending on the context of the game" (3.1) "I guess it's very situational, isn't it? It depends on the way that the team plays" (2.2) "also this context of the game, you know, it's you can see there is 0-0, 33 minutes," (3.2) "the context of the game, the level of the opposition, all the things that come in and affect the players subconscious in terms of whether to… to how aggressive to be how much risk to take, in that particular moment of the game." (3.2) "when you show me a clip, the first thing I, the first place I look is to see what minute of the game is, what's the score, because the context of this, the scenario has to be realistic" (3.2)	
player strengths	
"I think you want to try and have as much variation as possible . But if you do have strengths and your team then you would obviously try and utilise them" (2.1) "So therefore, you prepare the team differently. And you, you obviously select the team related to that with the skills that you're going to have got to do that. And that's obviously shown in the shape of the team, the formation that you choose to play, and then the roles within that as well. So for me, again, just comes down to the individual game preparation" (3.1)	

“But he maybe is the one that we can release because the other players may be at a better ability to press with intensity, in those deeper positions do you know what I mean and then get up to support as well. So it's all specific, isn't it to personnel I suppose, And qualities.” (2.2) **“It all has to be aligned to the strengths and weaknesses of the players and the opposition” (3.2)**	

References

AGUIAR, M. et al., 2017. Regularity of interpersonal positioning discriminates short and long sequences of play in small-sided soccer games. *Science and Medicine in Football,* 1(3), pp. 258-264

AGUIAR, M. et al., 2015. Footballers' movement behaviour during 2-, 3-, 4- and 5-a-side small-sided games. *Journal of sports sciences,* 33(12), pp. 1259-1266

ALMEIDA, C.H., FERREIRA, A.P. and VOLOSSOVITCH, A., 2014. Effects of Match Location, Match Status and Quality of Opposition on Regaining Possession in UEFA Champions League. *Journal of Human Kinetics,* 41, pp. 203-214

AQUINO, R.L.Q.T. et al., 2016b. Biochemical, physical and tactical analysis of a simulated game in young soccer players. *Journal of Sports Medicine and Physical Fitness,* 56(12), pp. 1554-1561

AQUINO, R.L.Q.T. et al., 2016a. Periodization Training Focused on Technical-Tactical Ability in Young Soccer Players Positively Affects Biochemical Markers and Game Performance. *Journal of Strength and Conditioning Research,* 30(10), pp. 2723-2732

AQUINO, R. et al., 2018. Influence of Situational Variables, Team Formation, and Playing Position on Match Running Performance and Social Network Analysis in Brazilian Professional Soccer Players. *Journal of Strength & Conditioning Research,*

AQUINO, R. et al., 2017b. Influence of Match Location, Quality of Opponents, and Match Status on Movement Patterns in Brazilian Professional Football Players. *Journal of Strength & Conditioning Research,* 31(8), pp. 2155-2161

AQUINO, R. et al., 2017a. Skill-Related Performance in Soccer: a Systematic Review. *Human Movement,* 18(5), pp. 3-24

ARAUJO, D. et al., 2010. The role of ecological constraints on expertise development. *Talent Development and Excellence,* 2(2), pp.165-179

ASLAN, A., 2013. Cardiovascular Responses, Perceived Exertion and Technical Actions During Small-Sided Recreational Soccer: Effects of Pitch Size and Number of Players. *Journal of Human Kinetics,* 38, pp. 95-105

BAKEMAN, R. (2005). Recommended effect size statistics for repeated measures designs, Behavior Research Methods, 37(3), pp. 379-384.

BAPTISTA, J. et al., 2018. Exploring the effects of playing formations on tactical behaviour and external workload during football small-sided games. *Journal of strength and conditioning research,* BARTLETT, R. et al., 2012. Analysing Team Coordination Patterns from Player Movement Trajectories in Soccer: Methodological Considerations. *International Journal of Performance Analysis in Sport,* 12(2), pp. 1-1

BELTRAME, T. et al., 2020. Relationship between maximal aerobic power with aerobic fitness as a function of signal-to-noise ratio. *Journal of applied physiology,* 129(3), pp. 522-532

BERTOLLO, M., SALTARELLI, B. and ROBAZZA, C., 2009. Mental preparation strategies of elite modern pentathletes. *Psychology of Sport and Exercise,* 10, pp. 244 – 254.

BRADLEY, B. et al., 2019. Bio-banding in academy football: player's perceptions of a maturity matched tournament. *Annals of Human Biology,* 46(5), pp. 400-408

BRAUN, V. and CLARKE, V., 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health,* 11(4), pp.589-597.

BRECHOT, M. and FLEPP, R., 2020. Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals. *Journal of Sports Economics,* 21(4), pp. 335-362

BREFELD, U., LASEK, J. and MAIR, S., 2019. Probabilistic movement models and zones of control. *Machine Learning,* 108(1), pp.127-147.

BUCHHEIT, M. and SIMPSON, B.M., 2017. Player-Tracking Technology: Half-Full or Half-Empty Glass? *International Journal of Sports Physiology & Performance,* 12, pp. S2-35; S2-41

BULDU, J.M. et al., 2018. Using network science to analyse football passing networks: Dynamics, space, time, and the multilayer nature of the game. *Frontiers in psychology,* p.1900.

BUSH, M., ARCHER, D.T., BARNES, C., HOGG, B. and BRADLEY, P.S., 2017. Longitudinal match performance characteristics of UK and non-UK players in the English Premier League. *Science and Medicine in Football,* 1(1), pp.2-9.

CARPITA, M. et al., 2016. Discovering the drivers of football match outcomes with data mining. *Quality Technology and Quantitative Management,* 12(4), pp. 561-577

CASAMICHANA, D. and CASTELLANO, J., 2010. Time–motion, heart rate, perceptual and motor behaviour demands in small-sides soccer games: Effects of pitch size. *Journal of sports sciences,* 28(14), pp.1615-1623.

CASTELLANO J. and ECHEAZARRA, I., 2019. Network-based centrality measures and physical demands in football regarding player position: Is there a connection? A preliminary study. *Journal of sports sciences,* , pp. 1-8

CASTELLANO, J. et al., 2017. Influence of pitch length on inter- and intra-team behaviors in youth soccer. *Anales de Psicologia,* 33(3), pp. 486-496

CASTELLANO, J. et al., 2013. Identifying the effects from the quality of opposition in a Football team positioning strategy. *International Journal of Performance Analysis in Sport,* 13(3), pp. 21-21

CASTELLANO, J., ALVAREZ-PASTOR, D. and BRADLEY, P., 2014. Evaluation of Research Using Computerised Tracking Systems (Amisco and Prozone) to Analyse Physical Performance in Elite Soccer: A Systematic Review. *Sports Medicine,* 44(5), pp. 701-712

CASTELLANO, J. et al., 2016. The influence of scoring targets and outer-floaters on attacking and defending team dispersion, shape and creation of space during small-sided soccer games. *Journal of Human Kinetics,* 51(1), pp. 153-163

CHASSY, P., MALONE, J.J. and CLARK, D.P.A., 2018. A mathematical model of self-organisation in football. *International Journal of Performance Analysis in Sport,* 18(2), pp. 217-228

CHO, Y., YOON, J. and LEE, S., 2018. Using social network analysis and gradient boosting to develop a soccer win–lose prediction model. *Engineering Applications of Artificial Intelligence,* 72, pp. 228-240

CHUNG, D. et al., 2019. Number of players manipulation effect on space and concentration principles of the game representativeness during football small-sided and conditioned games. *Journal of Physical Education & Sport,* 19, pp. 381-386

CLEMENTE F.M. et al., 2016c. Physical and technical performances are not associated with tactical prominence in U14 soccer matches. *Research in sports medicine (Print),* 24(4), pp. 352-362

CLEMENTE, F. et al., 2014c. Intelligent systems for analyzing soccer games: The weighted centroid. *Ingenieria E Investigacion,* 34(3), pp. 70-75

CLEMENTE, F.M., COUCEIRO, M.S. and MARTINS, F.M.L., 2012. Towards a new method to analyze the soccer teams tactical behaviour: Measuring the effective area of play. *Indian Journal of Science and Technology,* 5(12), pp. 3792-3801

CLEMENTE, F.M. et al., 2013d. An online tactical metrics applied to football game. *Research Journal of Applied Sciences, Engineering and Technology,* 5(5), pp. 1700-1719

CLEMENTE, F.M. et al., 2014e. Using network metrics to investigate football team players' connections: A pilot study. *Motriz.Revista de Educacao Fisica,* 20(3), pp. 262-271

CLEMENTE, F.M. and MARTINS, F.M.L., 2017b. Network structure of UEFA Champions League teams: association with classical notational variables and variance between different levels of success. *International Journal of Computer Science in Sport (De Gruyter Open),* 16(1), pp. 39-50

CLEMENTE, F.M. et al., 2014b. Inspecting teammates' coverage during attacking plays in a football game: A case study. *International Journal of Performance Analysis in Sport,* 14(2), pp. 384-400

CLEMENTE, F.M. et al., 2014f. A network approach to characterize the teammates' interactions on football: A single match analysis. *Cuadernos de Psicologia del Deporte,* 14(3), pp. 141-148

CLEMENTE, F.M. et al., 2016a. Developing a tactical metric to estimate the defensive area of soccer teams: The defensive play area. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology,* 230(2), pp. 124-132

CLEMENTE, F.M. et al., 2015a. The social network analysis of Switzerland football team on FIFA world cup 2014. *Journal of Physical Education and Sport,* 15(1), pp. 136-141

CLEMENTE, F.M. et al., 2015b. General network analysis of national soccer teams in Fifa World Cup 2014. *International Journal of Performance Analysis in Sport,* 15(1), pp. 80-96

CLEMENTE, F.M. et al., 2014d. Using Collective Metrics to Inspect Spatio-Temporal Relationships between Football Players. *South African Journal for Research in Sport Physical Education and Recreation,* 36(2), pp. 47-59

CLEMENTE, F.M. et al., 2013b. Measuring Tactical Behaviour Using Technological Metrics: Case Study of a Football Game. *International Journal of Sports Science & Coaching,* 8(4), pp. 723-739

CLEMENTE, F.M. et al., 2013a. Measuring Collective Behaviour in Football Teams: Inspecting the impact of each half of the match on ball possession. *International Journal of Performance Analysis in Sport,* 13(3), pp. 10-10

CLEMENTE, F.M. et al., 2014a. Evaluating the Offensive Definition Zone in Football: a Case Study. *South African Journal for Research in Sport Physical Education and Recreation,* 36(3), pp. 25-37

CLEMENTE, F.M., 2018c. Performance outcomes and their associations with network measures during FIFA World Cup 2018. *International Journal of Performance Analysis in Sport,* 18(6), pp. 1010-1023

CLEMENTE, F.M. et al., 2013c. Interpersonal Dynamics: 1v1 Sub-Phase at Sub-18 Football Players. *Journal of Human Kinetics,* 37, pp. 181-191

CLEMENTE, F.M. et al., 2015e. Using Network Metrics in Soccer: A Macro-Analysis. *Journal of Human Kinetics,* 45(1), pp. 123-134

CLEMENTE, F.M. et al., 2015d. Soccer team's tactical behaviour: Measuring territorial domain. *Proceedings of the Institution of Mechanical Engineers Part P-Journal of Sports Engineering and Technology,* 229(1), pp. 58-66

CLEMENTE, F.M. et al., 2014g. A systemic overview of football game: The principles behind the game. *Journal of Human Sport & Exercise,* 9(2), pp. 656-667

CLEMENTE, F.M. et al., 2015c. 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), pp. 704-722

CLEMENTE, F.M., MARTINS, F.M.L. and MENDES, R.S., 2015. Are the prominent players the most accurate and efficient? Study in football players from different competitive levels. *Journal of Physical Education & Sport,* 15(3), pp. 472-478

CLEMENTE, F.M. et al., 2018a. The effects of large-sided soccer training games and pitch size manipulation on time-motion profile, spatial exploration and surface area: Tactical opportunities. *Proceedings of the Institution of Mechanical Engineers Part P-Journal of Sports Engineering and Technology,* 232(2), pp. 160-165

CLEMENTE, F.M. et al., 2018b. How to Use the Dots to Analyze the Behavior and the Collective Organization. *Computational Metrics for Soccer Analysis: Connecting the Dots,* , pp. 7-13

CLEMENTE, F.M. et al., 2016b. Performance Analysis Tool for network analysis on team sports: A case study of FIFA Soccer World Cup 2014. *Proceedings of the Institution of Mechanical Engineers Part P-Journal of Sports Engineering and Technology,* 230(3), pp. 158-170

CLEMENTE, F. and SARMENTO, H., 2020. The effects of small-sided soccer games on technical actions and skills: A systematic review. Human Movement, *Journal of Physical Education.* 1(1), pp. 657-668

COTTA, C. et al., 2013. A network analysis of the 2010 FIFA world cup champion team play. *Journal of Systems Science and Complexity,* 26(1), pp. 21-42

COUTINHO, D. et al., 2018. Effects of the pitch configuration design on players' physical performance and movement behaviour during soccer small-sided games. *Research in Sports Medicine,* 27(3), pp. 298-313

COUTINHO, D. et al., 2019. Effects of pitch spatial references on players' positioning and physical performances during football small-sided games. *Journal of sports sciences,* 37(7), pp. 741-747

COUTTS, A.J., 2017. Challenges in developing evidence-based practice in high-performance sport. *International journal of sports physiology and performance,* 12(6), pp.717-718

CUMMING, S.P. et al., 2018. Premier League academy soccer players' experiences of competing in a tournament bio-banded for biological maturation. *Journal of sports sciences,* 36(7), pp. 757-765

CUMMINS, C. et al., 2013. Global positioning systems (GPS) and microtechnology sensors in team sports: a systematic review. *Sports Medicine,* 43(10), pp. 1025-1042

DALEN, T. et al., 2016. Player Load,acceleration, and Deceleration during Forty-Five Competitive Matches of Elite Soccer. *Journal of Strength & Conditioning Research,* 30(2), pp. 351-359

DAVIDS, K. et al., 2013. An Ecological Dynamics Approach to Skill Acquisition: Implications for Development of Talent in Sport. *Talent Development & Excellence,* 5(1), pp. 21-34

DE SOUZA, N.M. et al., 2018. Space configuration and numerical relationship during professional soccer matches: a proposal for small-sided games design. *Human Movement Special Issues,* 2018(5), pp. 121-128

DECROOS, T., BRANSEN, L., VAN HAAREN, J. and DAVIS, J., 2019, July. Actions speak louder than goals: Valuing player actions in soccer. *In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 1851-1861).

DELLAL, A. et al., 2011. Comparison of physical and technical performance in European soccer match-play: FA Premier League and La Liga. *European Journal of Sport Science,* 11(1), pp. 51-59

DETERDING, N.M. and WATERS, M.C., 2021. Flexible coding of in-depth interviews: A twenty-first-century approach. *Sociological methods & research,* 50(2), pp.708-739.

DONALDSON, A. et al., 2016. Scientific evidence is just the starting point: A generalizable process for developing sports injury prevention interventions. *Journal of Sport & Health Science,* 5(3), pp. 334-341

DUARTE, R. et al., 2013a. Capturing complex, non-linear team behaviours during competitive football performance. *Journal of Systems Science and Complexity,* 26(1), pp. 62-72

DUARTE, R. et al., 2012b. Sports Teams as Superorganisms: Implications of Sociobiological Models of Behaviour for Research and Practice in Team Sports Performance Analysis. *Sports Medicine,* 42(8), pp. 633-642

DUARTE, R. et al., 2013b. Competing together: Assessing the dynamics of team-team and player-team synchrony in professional association football. *Human Movement Science,* 32(4), pp. 555-566

DUARTE, R. et al., 2012a. Interpersonal coordination tendencies shape 1-vs-1 sub-phase performance outcomes in youth soccer. *Journal of sports sciences,* 30(9), pp. 871-877

DUARTE, R. et al., 2012c. Intra- and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. *Human Movement Science,* 31(6), pp. 1639-1651

DUTTMAZUMDER A. et al., 2011. Neural network modelling and dynamical system theory: are they relevant to study the governing dynamics of association football players?. *Sports medicine (Auckland, N.Z.),* 41(12), pp. 1003-1017

EAVES, S.J., 2015. A history of sports notational analysis: A journey into the nineteenth century. *International Journal of Performance Analysis in Sport,* 15(3), pp. 1160-1176

EHRMANN, F.E. et al., 2016. Gps and Injury Prevention in Professional Soccer. *Journal of Strength & Conditioning Research,* 30(2), pp. 360-367

FENNER, J.S.J., IGA, J. and UNNITHAN, V., 2016. The evaluation of small-sided games as a talent identification tool in highly trained prepubertal soccer players. *Journal of sports sciences,* 34(20), pp. 1983-1990

FERNANDEZ, J. and BORNN, L., 2018. Wide Open Spaces: A statistical technique for measuring space creation in professional soccer. *In Sloan sports analytics conference*

FERRARI, W.R., SARMENTO, H. and VAZ, V., 2019. Match Analysis in Handball: A Systematic Review. *Montenegrin Journal of Sports Science & Medicine,* 8(2), pp. 63-76

FIGUEIRA, B. et al., 2018. Exploring how playing football with different age groups affects tactical behaviour and physical performance. *Biology of Sport,* 35(2), pp. 145-153

FILETTI, C. et al., 2017. A Study of Relationships among Technical, Tactical, Physical Parameters and Final Outcomes in Elite Soccer Matches as Analyzed by a Semiautomatic Video Tracking System. *Perceptual & Motor Skills,* 124(3), pp. 601-620

FOLGADO, H. et al., 2018c. Towards the use of multidimensional performance indicators in football small-sided games: the effects of pitch orientation. *Journal of sports sciences,*

FOLGADO, H. et al., 2014a. Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches. *PLoS ONE,* 9(5),

FOLGADO, H. et al., 2018b. Exploring how movement synchronization is related to match outcome in elite professional football. *Science and Medicine in Football,* 2(2), pp. 101-107

FOLGADO, H. et al., 2015. The effects of congested fixtures period on tactical and physical performance in elite football. *Journal of sports sciences,* 33(12), pp. 1238-1247

FOLGADO, H., GONÇALVES, B. and SAMPAIO, J., 2018a. Positional synchronization affects physical and physiological responses to preseason in professional football (soccer). *Research in Sports Medicine,* 26(1), pp. 51-63

FOLGADO, H. et al., 2014b. Length, width and centroid distance as measures of teams tactical performance in youth football. *European Journal of Sport Science,* 14, pp. S487-S492

FONSECA, S. et al., 2012. Spatial dynamics of team sports exposed by Voronoi diagrams. *Human Movement Science,* 31(6), pp. 1652-1659

FRENCKEN, W.G.P., LEMMINK, K.A.P.M. and DELLEMAN, N.J., 2010. Soccer-specific accuracy and validity of the local position measurement (LPM) system. *Journal of Science & Medicine in Sport,* 13(6), pp. 641-645

FRENCKEN, W. et al., 2012. Variability of inter-team distances associated with match events in elite-standard soccer. *Journal of sports sciences,* 30(12), pp. 1207-1213

FRENCKEN, W. et al., 2011. Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science,* 11(4), pp. 215-223

FRENCKEN, W. et al., 2013. Size matters: Pitch dimensions constrain interactive team behaviour in soccer. *Journal of Systems Science & Complexity,* 26(1), pp. 85-93

FRIAS, T. and DUARTE, R., 2014. Man-to-man or zone defense? Measuring team dispersion behaviors in small-sided soccer games. *Trends in Sport Sciences,* 21(3), pp. 135-144

FULLAGAR, H.H. et al., 2019. The translation of sport science research to the field: a current opinion and overview on the perceptions of practitioners, researchers and coaches. *Sports Medicine,* 49(12), pp.1817-1824

GAMA, J. et al., 2016a. Networks and centroid metrics for understanding football. *South African Journal for Research in Sport, Physical Education and Recreation,* 38(2), pp. 75-90

GAMA, J. et al., 2016b. Networks metrics and ball possession in professional football. *Complexity,* 21, pp. 342-354

GAMA, J. et al., 2014. Network analysis and intra-team activity in attacking phases of professional football. *International Journal of Performance Analysis in Sport,* 14(3), pp. 692-708

GAMA, J. et al., 2015. Small-World Networks in Professional Football: Conceptual Model and Data. *European Journal of Human Movement,* 35, pp. 85-113

GARRIDO, D.et al., 2020. Consistency and identifiability of football teams: A network science perspective. *Scientific reports,* 10(1), pp.1-10.

GLAZIER, P.S., 2010. Game, Set and Match? Substantive Issues and Future Directions in Performance Analysis. *Sports Medicine,* 40(8), pp. 625-634

GOES, F.R. et al., 2021. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science,* 21(4), pp. 481-496

GONÇALVES, B.V. et al., 2014. Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. *Journal of sports sciences,* 32(2), pp. 191-199

GONÇALVES, B. et al., 2019. Extracting spatial-temporal features that describe a team match demands when considering the effects of the quality of opposition in elite football. *PLoS ONE,* 14(8),

GONÇALVES, B. et al., 2017b. Exploring team passing networks and player movement dynamics in youth association football. *PLoS ONE,* 12(1),

GONÇALVES, B. et al., 2017a. Effects of Pitch Area-Restrictions on Tactical Behavior, Physical, and Physiological Performances in Soccer Large-Sided Games. *Journal of Strength and Conditioning Research,* 31(9), pp. 2398-2408

GONCALVES, B. et al., 2016. Effects of emphasising opposition and cooperation on collective movement behaviour during football small-sided games. *Journal of sports sciences,* 34(14), pp. 1346-1354

GONCALVES, B. et al., 2018b. Speed synchronization, physical workload and match-to-match performance variation of elite football players. *Plos One,* 13(7), pp. e0200019

GONCALVES, B. et al., 2018a. Changes in Effective Playing Space When Considering Sub-Groups of 3 to 10 Players in Professional Soccer Matches. *Journal of Human Kinetics,* 62(1), pp. 145-155

GOUMAS, C., 2017. Modelling home advantage for individual teams in UEFA Champions League football. *Journal of Sport & Health Science,* 6(3), pp. 321-326

GRÉHAIGNE, J.-. and GODBOUT, P., 2021. Opposition and dynamics at the core of game-play analysis in soccer. *International Journal of Physical Education,* 58(3), pp. 10-20

GRUND, T.U., 2012. Network structure and team performance: The case of English Premier League soccer teams. *Social Networks,* 34(4), pp. 682-690

GRUND, T.U., 2016. The Relational Value of Network Experience in Teams: Evidence From the English Premier League. *American Behavioral Scientist,* 60(10), pp. 1260-1280

GUDMUNDSSON, J. and WOLLE, T., 2014. Football analysis using spatio-temporal tools. *Computers, Environment and Urban Systems*, 47(1), pp.16-27.

HÄMMERER, D. et al., 2013. A lifespan comparison of the reliability, test-retest stability, and signal-to-noise ratio of event-related potentials assessed during performance monitoring. *Psychophysiology,* 50(1), pp. 111-123

HEADRICK, J. et al., 2012. Proximity-to-goal as a constraint on patterns of behaviour in attacker–defender dyads in team games. *Journal of sports sciences,* 30(3), pp. 247-253

HORVAT, T. and JOB, J., 2020. The use of machine learning in sport outcome prediction: A review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 10(5),

JOHNSTON, R.J. et al., 2014. VALIDITY AND INTERUNIT RELIABILITY OF 10 Hz AND 15 Hz GPS UNITS FOR ASSESSING ATHLETE MOVEMENT DEMANDS. *Journal of Strength & Conditioning Research,* 28(6), pp. 1649-1655

JOO, C.H., HWANG-BO, K. and JEE, H., 2016. Technical and physical activities of small-sided games in young Korean soccer players. *Journal of strength and conditioning research,* 30(8), pp.2164-2173.

KELLY, V.G., SCOTT, M.T.U. and SCOTT, T.J.U., 2014. The Validity and Reliability of Global Positioning Systems in Team Sport: a Brief Review. *Journal of Australian Strength & Conditioning,* 22(5), pp. 186-190

LAAKSO, T. et al., 2019. Interpersonal dynamics in 2-vs-1 contexts of football: The effects of field location and player roles. *Frontiers in Psychology,* 10(JUL),

LAAKSO, T. et al., 2017. Field location and player roles as constraints on emergent 1-vs-1 interpersonal patterns of play in football. *Human Movement Science,* 54, pp. 347-353

LACOME, M. et al., 2019. Can we use GPS for assessing sprinting performance in rugby sevens? A concurrent validity and between-device reliability study. *Biology of Sport,* 36(1), pp. 25-29

LESER, R. et al., 2019. Finding Efficient Strategies in 3-Versus-2 Smallsided Games of Youth Soccer Players. *Kinesiology,* 51(1), pp. 110-118

LESER, R. et al., 2015. Expert-oriented modelling of a 1vs1-situation in football. *International Journal of Performance Analysis in Sport,* 15(3), pp. 949-966

LINK, D., LANG, S. and SEIDENSCHWARZ, P., 2016. Real time quantification of dangerousity in football using spatiotemporal tracking data. *PloS one,* 11(12), p.e0168768.

LINKE, D., LINK, D. and LAMES, M., 2018. Validation of electronic performance and tracking systems EPTS under field conditions. *PLoS ONE [Electronic Resource],* 13(7), pp. e0199519

LINKE, D., LINK, D. and LAMES, M., 2020. Football-specific validity of TRACAB's optical video tracking systems. *PLoS One,* 15(3), p.e0230179.

LIU, H. et al., 2013. Inter-operator reliability of live football match statistics from OPTA Sportsdata. *International Journal of Performance Analysis in Sport,* 13(3), pp. 20-20

LIU, T. et al., 2019. Exploring home advantage and quality of opposition interactions in the Chinese Football Super League. *International Journal of Performance Analysis in Sport,* 19(3), pp. 289-301

LÓPEZ-FELIP, M.A., DAVIS, T.J. and FRANK, T.D., 2017. Synchronization Assessment for Collective Behavior. *CogSci.*

LORD, F. et al., 2020. Methods of performance analysis in team invasion sports: A systematic review. *Journal of sports sciences,* 38(20), pp. 2338-2349

LOW, B. et al., 2018. Exploring the effects of deep-defending vs high-press on footballers' tactical behaviour, physical and physiological performance: A pilot study. *Motriz.Revista de Educacao Fisica,* 24(2),

LOW, B. et al., 2019. A Systematic Review of Collective Tactical Behaviours in Football Using Positional Data. *Sports Medicine,*

MACHADO, J.C. et al., 2019. Enhancing learning in the context of street football: a case for nonlinear pedagogy. *Physical Education and Sport Pedagogy,* 24(2), pp.176-189.

MACKENZIE, R. and CUSHION, C., 2013. Performance analysis in football: a critical review and implications for future research. *Journal of sports sciences,* 31(6), pp. 639-676

MALONE, J.J. et al., 2017. Unpacking the Black Box: Applications and Considerations for Using GPS Devices in Sport. *International Journal of Sports Physiology & Performance,* 12, pp. S2-18; S2-26

MANUEL CLEMENTE, F., LOURENÇO MARTINS, F.M. and SOUSA MENDES, R., 2016. Analysis of Scored and Conceded Goals by a Football Team Throughout a Season: a Network Analysis. *Kinesiology,* 48(1), pp. 103-114

MANUEL CLEMENTE, ,FILIPE and SARMENTO, H., 2020. The Effects of Small-Sided Soccer Games on Technical Actions and Skills: a Systematic Review. *Human Movement,* 21(3), pp. 100-119

MARCOS, M.A., KOULLA, P.M. and ANTHOS, Z.I., 2018. Preseason Maximal Aerobic Power in Professional Soccer Players among Different Divisions. *Journal of Strength & Conditioning Research,* 32(2), pp. 356-363

MCGARRY, T., 2009. Applied and theoretical perspectives of performance analysis in sport: Scientific issues and challenges. *International Journal of Performance Analysis in Sport,* 9(1), pp. 128-140

MCHALE, I.G. and RELTON, S.D., 2018. Identifying key players in soccer teams using network analysis and pass difficulty. *European Journal of Operational Research,* 268(1), pp. 339-347

MCLEAN, S. et al., 2017a. What's in a game? A systems approach to enhancing performance analysis in football. *PloS one,* 12(2), p.e0172565

MCLEAN, S. et al., 2017b. Do inter-continental playing styles exist? Using social network analysis to compare goals from the 2016 EURO and COPA football tournaments knock-out stages. *Theoretical Issues in Ergonomics Science,* 18(4), pp. 370-383

MCLEAN, S. et al., 2018. A social network analysis of the goal scoring passing networks of the 2016 European Football Championships. *Human movement science,* 57, pp. 400-408

MCLEAN, S. et al., 2019. Integrating communication and passing networks in football using social network analysis. *Science & Medicine in Football,* 3(1), pp. 29-35

MEMMERT, D., LEMMINK, K.A.P.M. and SAMPAIO, J., 2017. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine,* 47(1), pp. 1-10

MEMMERT, D. et al., 2019. A tactical comparison of the 4-2-3-1 and 3-5-2 formation in soccer: A theory-oriented, experimental approach based on positional data in an 11 vs. 11 game set-up. *Plos One,* 14(1), pp. e0210191

MENDES, B., CLEMENTE, F.M. and MAURÍCIO, N., 2018. Variance in Prominence Levels and in Patterns of Passing Sequences in Elite and Youth Soccer Players: A Network Approach. *Journal of Human Kinetics,* 61(1), pp. 141-153

MENUCHI, M.R.T.P. et al., 2018. Effects of Spatiotemporal Constraints and Age on the Interactions of Soccer Players when Competing for Ball Possession. *Journal of Sports Science & Medicine,* 17(3), pp. 379-391

MOREIRA PRAÇA, G. et al., 2016. Influence of additional players on collective tactical behavior in small-sided soccer games. / Comportamento tático coletivo em Pequenos Jogos no Futebol: influência de jogadores adicionais. *Brazilian Journal of Kineanthropometry & Human Performance,* 18(1), pp. 62-71

MOURA, F.A. et al., 2012. Quantitative analysis of Brazilian football players' organisation on the pitch. *Sports Biomechanics,* 11(1), pp. 85-96

MOURA, F.A. et al., 2016. Coordination analysis of players' distribution in football using cross-correlation and vector coding techniques. *Journal of sports sciences,* 34(24), pp. 2224-2232

MOURA, F. et al., 2013. A spectral analysis of team dynamics and tactics in Brazilian football. *Journal of sports sciences,* 31(14), pp. 1568-1577

NAKAGAWA, S., & SCHIELZETH, H. (2012). The mean strikes back: mean-variance relationships and heteroscedasticity. *Trends in Ecology & Evolution,* 27(9), pp. 474-475.

NEVILL, A., ATKINSON, G. and HUGHES, M., 2008. Twenty-five years of sport performance research in the Journal of Sports Sciences. *Journal of sports sciences,* 26(4), pp. 413-426

NICHOLLS, S.B. et al., 2019. The implementation of performance analysis and feedback within Olympic sport: The performance analyst's perspective. *International Journal of Sports Science & Coaching,* 14(1), pp. 63-71

O'DONOGHUE, P. and MAYES, T. and O'DONOGHUE, P., 2012. Performance Analysis, Feedback and Communication in Coaches. In: T. MCGARRY, P. O'DONOGHUE and J. SAMPAIO, eds. *Routledge Handbook of Sports Performance Analysis*. London: Routledge. pp. 155-164

OGRIS, G. et al., 2012. Accuracy of the LPM tracking system considering dynamic position changes. *Journal of sports sciences,* 30(14), pp. 1503-1511

OLIVEIRA, P. and CLEMENTE, F.M., 2018. Network properties and performance variables and their relationships with distance covered during elite soccer games. *Journal of Physical Education and Sport,* 18, pp. 1045-1049

OLTHOF, S.B.H., FRENCKEN, W.G.P. and LEMMINK, K.A.P.M., 2019. A Match-Derived Relative Pitch Area Facilitates the Tactical Representativeness of Small-Sided Games for the Official Soccer Match. *Journal of strength and conditioning research,* 33(2), pp. 523-530

OLTHOF, S.B.H., FRENCKEN, W.G.P. and LEMMINK, K.A.P.M., 2015. The older, the wider: On-field tactical behavior of elite-standard youth soccer players in small-sided games. *Human Movement Science,* 41, pp. 92-102

OLTHOF, S.B.H., FRENCKEN, W.G.P. and LEMMINK, K.A.P.M., 2018. Match-derived relative pitch area changes the physical and team tactical performance of elite soccer players in small-sided soccer games. *Journal of sports sciences,* 36(14), pp. 1557-1563

OLTHOF, S.B.H., FRENCKEN, W.G.P. and LEMMINK, K.A.P.M., 2019. When Something is at Stake: Differences in Soccer Performance in 11 Vs. 11 during Official Matches and Training Games. *Journal of Strength & Conditioning Research (Lippincott Williams & Wilkins),* 33(1), pp. 167-173

PALUCCI VIEIRA, L.H. et al., 2018. Team Dynamics, Running, and Skill-Related Performances of Brazilian U11 to Professional Soccer Players During Official Matches. *Journal of Strength & Conditioning Research,* PAPPALARDO, L. et al., 2019. PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach. ACM Transactions on Intelligent Systems and Technology (TIST), 10(5), pp.1-27.

PASSOS, P. et al., 2006. Interpersonal dynamics in sport: The role of artificial neural networks and 3-D analysis. *Behavior Research Methods,* 38(4), pp. 683-691

PEIXOTO, D. et al., 2017. Comparison of network processes between successful and unsuccessful offensive sequences in elite soccer. *Human Movement,* 18(5), pp. 48-54

PINA, T.J., PAULO, A. and ARAÚJO, D., 2017. Network characteristics of successful performance in association football. A study on the UEFA champions league. *Frontiers in Psychology,* 8(JUL),

PODLOG, L. and EKLUND. C. R., 2009. High-level athletes' perceptions of success in returning to sport following injury. *Psychology of Sport and Exercise,* 10, pp. 535-544.

POLLARD, R. and GÓMEZ, M., 2009. Home advantage in football in South-West Europe: Long-term trends, regional variation, and team differences. *European Journal of Sport Science,* 9(6), pp. 341-352

PRAÇA, G.M. et al., 2018a. Analysis of network properties and tactical behavior of U-17 soccer athletes with different tactical skills. *Journal of Physical Education (Maringa),* 29(1),

PRAÇA, G.M. et al., 2018b. Defensive interactions in soccer small-sided games: An integrated approach between the fundamental tactical principles and the social network analysis. *Revista Brasileira de Cineantropometria e Desempenho Humano,* 20(5), pp. 422-431

PRAÇA, G.M. et al., 2016. Influence of additional players on collective tactical behavior in small-sided soccer games. *Revista Brasileira de Cineantropometria e Desempenho Humano,* 18(1), pp. 62-71

PRACA, G.M. et al., 2017. Network Analysis in Small-Sided and Conditioned Soccer Games: the Influence of Additional Players and Playing Position. *Kinesiology,* 49(2), pp. 185-193

PRACA, G.M., SOUSA, R.B.E. and GRECO, P.J., 2019. Influence of Aerobic Power on Youth Players' Tactical Behavior and Network Properties during Football Small-Sided Games. *Sports,* 7(3),

PRICKETT, P., 2021. *Football's Principles of Play.* London: Hawksmoor Publishing.

RAMOS, J. et al., 2017. Hypernetworks reveal compound variables that capture cooperative and competitive interactions in a soccer match. *Frontiers in Psychology,* 8(AUG), RAMOS, J., LOPES, R.J. and ARAUJO, D., 2018. What's Next in Complex Networks? Capturing the Concept of Attacking Play in Invasive Team Sports. *Sports Medicine,* 48(1), pp. 17-28

RATHKE, A., 2017. An examination of expected goals and shot efficiency in soccer. *Journal of Human Sport & Exercise,* 12, pp. S514-S529

REEP, C. and BENJAMIN, B., 1968. Skill and Chance in Association Football. *Journal of the Royal Statistical Society.Series A (General),* 131(4), pp. 581-585

REILLY, T., 1976. A motion analysis of work-rate in different positional roles in professional football match-play. *J Human Movement Studies,* 2, pp. 87-97

REIN, R., RAABE, D. and MEMMERT, D., 2017. "Which pass is better?" Novel approaches to assess passing effectiveness in elite soccer. *Human Movement Science,* 55, pp. 172-181

REIN, R. et al., 2016. Evaluation of changes in space control due to passing behavior in elite soccer using Voronoi-cells. *Proceedings of the 10th International Symposium on Computer Science in Sports (Iscss),* 392, pp. 179-183

RIBEIRO, J. et al., 2017. Team Sports Performance Analysed Through the Lens of Social Network Theory: Implications for Research and Practice. *Sports Medicine,* 47(9), pp. 1689-1696

RIC, A. et al., 2016. Soft-assembled Multilevel Dynamics of Tactical Behaviors in Soccer. *Frontiers in Psychology,* 7, pp. 1513

RYAN, S. et al., 2020. Measurement Characteristics of Athlete Monitoring Tools in Professional Australian Football. *International Journal of Sports Physiology & Performance,* 15(4), pp. 457-463

SAMPAIO, J. and MAÇÃS, V., 2012. Measuring Tactical Behaviour in Football. *International Journal of Sports Medicine,* 33(5), pp. 395-401

SAMPAIO, J., MCGARRY, T. and O'DONOGHUE, P., 2013. Introduction. In: T. MCGARRY, P. O'DONOGHUE and J. SAMPAIO, eds. *Routledge Handbook of Sports Performance Analysis*. London: Routledge. pp. 1-2

SAMPAIO, J.E. et al., 2014. Effects of pacing, status and unbalance in time motion variables, heart rate and tactical behaviour when playing 5-a-side football small-sided games. *Journal of Science & Medicine in Sport,* 17(2), pp. 229-233

SANTOS, P.M. and LAGO-PENAS, C., 2019. Defensive Positioning on the Pitch in Relation with Situational Variables of a Professional Football Team during Regaining Possession. *Human Movement,* 20(2), pp. 50-56

SANTOS, P., LAGO-PENAS, C. and GARCIA-GARCIA, O., 2017. The influence of situational variables on defensive positioning in professional soccer. *International Journal of Performance Analysis in Sport,* 17(3), pp. 212-219

SANTOS, S. et al., 2018. Differential Learning as a Key Training Approach to Improve Creative and Tactical Behavior in Soccer. *Research Quarterly for Exercise & Sport,* 89(1), pp. 11-24

SARMENTO, H. et al., 2018a. What Performance Analysts Need to Know About Research Trends in Association Football (2012-2016): A Systematic Review. *Sports Medicine,* 48(4), pp. 799-836

SARMENTO, H. et al., 2018b. Small sided games in soccer - a systematic review. *International Journal of Performance Analysis in Sport,* 18(5), pp. 693-749

SARMENTO, H. et al., 2014. Match analysis in football: a systematic review. *Journal of sports sciences,* 32(20), pp. 1831-1843

SCHUTZ, Y. and CHAMBAZ, A., 1997. Could a satellite-based navigation system (GPS) be used to assess the physical activity of individuals on earth?. *European journal of clinical nutrition,* 51(5), pp.338-339.

SERRA-OLIVARES, J., GARCÍA LÓPEZ, L.M. and GONÇALVES, B., 2019. Effects of the players' level and age group category on positional tactical behaviour during 7- and 8-a-side football youth games. *International Journal of Performance Analysis in Sport,* 19(2), pp. 236-247

SHAFIZADEH, M. et al., 2016. Informational constraints on interceptive actions of elite football goalkeepers in 1v1 dyads during competitive performance. *Journal of sports sciences,* 34(17), pp. 1596-1601

SIEGLE, M. and LAMES, M., 2013. Modeling soccer by means of relative phase. *Journal of Systems Science & Complexity,* 26(1), pp. 14-20

SIEGLE, M., STEVENS, T. and LAMES, M., 2013. Design of an accuracy study for position detection in football. *Journal of sports sciences,* 31(2), pp. 166-172

SILVA, P. et al., 2014b. Numerical relations and skill level constrain co-adaptive behaviors of agents in sports teams. *PLoS ONE,* 9(9),

SILVA, P. et al., 2016a. Sports teams as complex adaptive systems: manipulating player numbers shapes behaviours during football small-sided games. *SpringerPlus,* 5(1), pp. 1-10

SILVA, P. et al., 2016b. Practice effects on intra-team synergies in football teams. *Human Movement Science,* 46, pp. 39-51

SILVA, P. et al., 2014a. Field dimension and skill level constrain team tactical behaviours in small-sided and conditioned games in football. *Journal of sports sciences,* 32(20), pp. 1888-1896

SPEARMAN, W., BASYE, A., DICK, G., HOTOVY, R. and POP, P., 2017, March. Physics-based modeling of pass probabilities in soccer. *In Proceeding of the 11th MIT Sloan Sports Analytics Conference*.

STEINER, S. et al., 2019. Outplaying opponents—a differential perspective on passes using position data. *German Journal of Exercise and Sport Research,* 49(2), pp. 140-149

TAYLOR, S.J. and BOGDAN, R., 1989. On accepting relationships between people with mental retardation and non-disabled people: Towards an understanding of acceptance. *Disability, Handicap & Society*, 4(1), pp. 21-36.

TENGA, A. et al., 2015. Explorative Study on Patterns of Game Structure in Male and Female Matches from Elite Spanish Soccer. *International Journal of Performance Analysis in Sport,* 15(1), pp. 411-423

TENGA, A. and SIGMUNDSTAD, E., 2011. Characteristics of goal-scoring possessions in open play: Comparing the top, in-between and bottom teams from professional soccer league. *International Journal of Performance Analysis in Sport,* 11(3), pp. 545-552

THOMAS, C.H., OLIVER, J., KELLY, A. and KNAPMAN, H., 2017. A pilot study of the demands of chronological age group and bio-banded match play in elite youth soccer. *Graduate Journal of Sport, Exercise & Physical Education Research*, *5*(S1), p.211.

TOWLSON, C. et al., 2021. The effect of bio-banding on physical and psychological indicators of talent identification in academy soccer players. *Science & Medicine in Football,* 5(4), pp. 280-292

TRAVASSOS, B. et al., 2010. Eco-dynamics approach to the study of team sports performance. *Open Sports Sciences Journal,* 3(1), pp. 56-57

TRAVASSOS, B. et al., 2014a. How perceiving additional targets modifies teams' tactical behavior during football small-sided games. *Human Movement Science,* 38, pp. 241-250

TRAVASSOS, B. et al., 2014b. Tactical performance changes with equal vs unequal numbers of players in small-Sided football games. *International Journal of Performance Analysis in Sport,* 14(2), pp. 594-605

TREQUATTRINI, R., LOMBARDI, R. and BATTISTA, M., 2015. Network analysis and football team performance: a first application. *Team Performance Management,* 21(1-2), pp. 85-110

VAISMORADI, M. et al., 2013. Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing and Health Sciences,* 15(3).

VILAR, L. et al., 2013. Science of winning soccer: Emergent pattern-forming dynamics in association football. *Journal of Systems Science and Complexity,* 26(1), pp. 73-84

WIIG, A.S. et al., 2019. Analyzing passing networks in association football based on the difficulty, risk, and potential of passes. *International Journal of Computer Science in Sport (De Gruyter Open),* 18(3), pp. 44-68

WOODS, C.T., MCKEOWN, I., O'SULLIVAN, M., ROBERTSON, S. and DAVIDS, K., 2020. Theory to practice: performance preparation models in contemporary high-level sport guided by an ecological dynamics framework. *Sports medicine,* 6(1), pp.1-11.

WRIGHT, C., ATKINS, S., JONES, B. and TODD, J., 2013. The role of performance analysts within the coaching process: Performance Analysts Survey 'The role of performance analysts in elite football club settings.'. *International journal of performance analysis in sport,* 13(1), pp.240-261.

ZHOU, C., GÓMEZ, M. and LORENZO, A., 2020. The evolution of physical and technical performance parameters in the Chinese Soccer Super League. *Biology of Sport,* 37(2), pp. 139-145

ZLOJUTRO, N. et al., 2020. The Analysis of High-Intensity Running Performances in Bht Premier League of Bosnia and Herzegovina Matches. *Homo Sporticus,* (1), pp. 40-44

ZUBILLAGA, A. et al., 2013. Influence of Ball Position on Playing Space in Spanish Elite Women's Football Match-Play. *International Journal of Sports Science & Coaching,* 8(4), pp. 713-722