# THE ANALYSIS OF TEAM TACTICAL BEHAVIOUR IN FOOTBALL USING GNSS POSITIONAL DATA

*GUANGZE ZHANG*

A thesis submitted in partial fulfilment of the requirements of Liverpool John Moores University for the degree of Master of Philosophy

This research programme was carried out in collaboration with the Bolton Wanderers Football Club

October 2022

# ***MPhil Project Supervisory Team***


***Dr. Sigrid Olthof***

*(Lead supervisor, Liverpool John Moores University)*

***Dr. Matthias Kempe***

*(Co-supervisor, University of Groningen)*

***Dr. Allistair McRobert***

*(Co-supervisor, Liverpool John Moores University)*

Acknowledgements

I am extremely grateful to my lead supervisor Sigrid Olthof, who has encouraged me to explore ideas and put them into place, and has guided me through the past year. Sigrid has been an invaluable mentor with massive passion for the research and the most importantly, for the life.

I would like to thank my supervisor Matthias Kempe whose rigor and knowledge in transdisciplinary research was the key to provoke those thoughts in this programme. Matthias is a role model as the sports engineer.

I am also grateful to my supervisor Allistair McRobert who introduced me to Sigrid one year ago. This team would have not been made up without him.

I am also thankful to Andrea Riboli, and sport scientists and analysts from Bolton Wanderers Football Club, for their data sharing and collaboration in this project.

And finally, thank you to my family who have been giving me their best support mentally and financially.

# Abstract

Tactical analysis in football is an emerging field focused on assessing the collective movement of teams. Advanced player tracking technology systems facilitate the data collection for tactical analysis. GNSS tracking systems is currently the most popular player tracking technology in football application and is mainly used in physical monitoring. It also captures players positional information as geographic coordinates (i.e., latitude and longitude coordinates) which requires extra data pre-processing for tactical analysis as opposed to Cartesian coordinates (i.e., X, Y coordinates). Given the lack of a comprehensive workflow on pre-processing raw GNSS positional data for calculating tactical measures in previous publications, this thesis aimed to present a workflow that provides exemplar data, processing steps, potential issues, and corresponding solutions. With the presented workflow, not only sport scientists but also practitioners are able to engage in tactical analysis using GNSS tracking systems and bring in their own understanding and perspective. In other words, GNSS tracking systems could play an important role in both physical and tactical analysis in real-world application.

Collective movements and actions vary as the match progresses along. The second objective was to use GNSS positional data to compare team tactical behaviour in different phases of a competitive match. The presented workflow was applied in data pre-processing of this analysis as a proof of concept. Although team tactical behaviour in football has been widely studied in recent years, there is no previous study that analyses team tactical behaviour in phase of attack, defence, and transition, based on tactical measures measured by positional data. In this thesis, effective playing time of a professional football match was divided into phase of in possession (IP), attack-to-defence transition (ADT), out of possession (OOP), and defence-to-attack transition (DAT). Team length, width, length per width ratio (LpW ratio), surface area, stretch indices, and interpersonal distance were calculated and compared to explore the difference of team tactical behaviour between phases. The findings showed that the team tactical behaviour during each phase was in line with the offensive and defensive tactical principles. The team presented a more dispersed and wider formation while in possession than other phases. The difference of all team tactical behaviour between IP and DAT indicated the potentiality of distinguishing defence-to-attack transition from in possession when analysing offensive tactical behaviour. Moreover, there was no significant difference across all tactical measures between defence-to-attack transition and defence, which implied that a short period of time was required for the team to switch to attack mode. In the future, the difference between transitions, attack, and defence should be valued in tactical analysis. Combining multi-type data with multi-disciplinary knowledge could inform stakeholders of dynamic team moving pattern and benefit decision making process. However, data quality (e.g., positional data and synchronisation of positional data and event data) should be prioritised in this type of study.

## *TABLE OF CONTENTS*

# Chapter 1 - General Introduction

## 1.1 Match analysis

The provision of performance analysis, feedback and future planning are important considerations in the football coaching process (Olsen and Larsen, 1997; Jayal et al., 2018). In invasion sports like football, behavioural events (i.e., on-ball and off-ball actions) are recorded by subjective game observation and objective data during a match for the purpose of performance analysis. Key information is then assessed to support coaching planning and practice, with the ultimate goal to improve athlete performance. This procedure was defined by Carling et al. (2005) as match analysis that is widely used in football practice and science. Match analysis is a tool that facilitates communication between coaches and athletes and supports decision-making processes of managers (Groom et al., 2011). The analysis can include aspects of technique, physiology, tactics, and psychology to have a comprehensive idea of players' performance (O'Donoghue, 2015).

Initially, coaches assessed player performance through subjective observation and post-game recall, which could potentially be unreliable and inaccurate due to personal perceptions, biases, and memory (Franks and Miller, 1991). To improve the efficiency and accuracy, a simple tally sheet for recording frequency counts on key events was devised and used to notate events live known as hand notational analysis (Carling et al., 2005). Notational analysis has now evolved to computerised notational systems (e.g., Hudl Sportscode), whereby analysts can manually tag match events with a computer. This development enabled subjective evaluation to be combined with objective information to provide insights into the strength and weakness of the own team and opponent team in offensive and defensive phases of game. Notational analysis could also reveal key performance indicators and guide the training process. For example, Castellano et al. (2012) analysed match statistics and identified that ball possession and shots on target best discriminated between match outcomes (i.e., wining, drawing, losing). Accordingly, ball control and effectiveness of attacking could be emphasised in training and match preparation.

From the 1990s, the physiological assessment developed a more prominent role in match analysis (Reilly et al., 1993). Physical monitoring emphasises the movement of individuals during activities, quantifying training and match performance (Castellano et al., 2011; Sampaio et al., 2014). It was achieved by hand notation at the beginning and by advanced technology nowadays. To monitor players' work rate, a coded map of the playing pitch with measuring cues alongside each side line was used to estimate moving distance of players (Carling et al., 2005). In recent years, modern player tracking technologies (e.g., wearable tracking systems and optical tracking systems) are widely applied to analyse individual contribution and team effort, which allows automatic player tracking and relieves laborious manual coding processes (Carling et al., 2008). With these advanced tracking technologies, physical monitoring is now employed in detecting physical demands in match (Lago-Peñas et al., 2012; Mallo et al., 2015), quantifying fatigue (Buchheit et al., 2013; Carling et al., 2018), and team preparation (Strudwick, 2013; Morgans et al., 2014).

Player tracking technology systems capture players' position and trajectories for quantifying physical performance, but are also extended to studying collective tactical behaviour (Low et al., 2020). Implementing and evaluating appropriate tactical performance is fundamental to match preparation to enhance the chances of winning the match. Tactical analysis used to be based on individual observations by domain experts (e.g., coach, scout) according to personal experiences and judgements (Mackenzie and Cushion, 2013). However, the interactions between team members and opposing teams are highly dynamic, and contextual circumstance in football match changes constantly. Subjective observation alone is not easy to have a holistic evaluation of the individual or collective performance. The lack of supportive data and time-consuming observational methods also limited the application of tactical analysis in the past (Rein and Memmert, 2016). From the beginning of 2010s, driven by novel developments in sensor technology and the blooming of big data (Gandomi and Haider, 2015; Memmert and Rein, 2018) and by the introduction of ecological dynamics in sports (Araujo and Davids, 2009; Vilar et al., 2012; Gréhaigne and Godbout, 2013), in-depth dynamic tactical assessments through quantifying collective behaviour became a trend in match analysis. This type of analysis requires players' positional data captured by player tracking technology systems.

## 1.2 Player tracking technology

Sports organizations have been investing financial and staff resources in technology that can quantify training and match performance. It is expected to improve preparation for competition by providing monitoring information on load-performance relationship (Buchheit et al., 2013; Akenhead and Nassis, 2016). Tracking systems in sports refer to those capturing player movement to quantify load-performance relationship within sports activity (Torres-Ronda et al., 2022b). Inside some single tracking device is the combination of multiple sensors that can provide information for different uses such as physical monitoring and tactical analysis (Torres-Ronda et al., 2022a). For example, heart rate sensor for monitoring internal load, inertial measurement units (IMU) and GNSS positioning for quantifying external load are all embedded in Catapult tracking devices (Catapult Innovations, South Melbourne, VIC, Australia). In the following, three types of player tracking systems (Figure 1.1) that can provide position information of moving objects will be discussed. Because these positional data are relevant for tactical analysis. IMU systems which only detect body movement but cannot provide position information, are therefore not included.

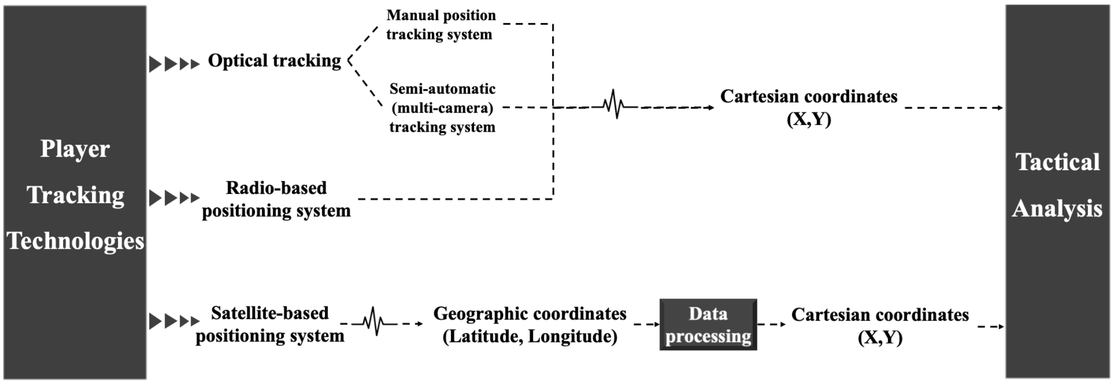



**Figure 1.1** Mainstream player tracking technology systems that provide position information of moving objects, and types of generated raw data. Geographic coordinates require extra data processing prior to tactical analysis.

Optical tracking systems are based on video image analysis to determine two-dimensional Cartesian coordinates (i.e., x and y coordinates). This tracking system can be manually or semi-automated operated. For example, a tracking software TACTO (Tool for Applied and Contextual Time-series Observation) requires manual operation on video footage (Fernandes et al., 2010). Users need to follow with a computer mouse the middle point between both feet on the ground to retrieve the virtual coordinates (in pixels) of the player (Duarte et al., 2010). The system is calibrated by retrieving virtual coordinates on the screen and actual coordinates of reference points on the pitch in advance. Direct Linear Transformation (DLT) is then applied to reconstruct the virtual coordinates of player movement into the real coordinates in meters (Serrano et al., 2014). Fernandes et al. (2010) demonstrated the high reliability and accuracy of the software. However, manual work is involved in the data collection, and it requires eye-hand coordination, visual sharpness, and concentration of the observer, which potentially could influence the data quality (Serrano et al., 2014). In recent years, machine learning and computer vision have been utilised into techniques that can automatically recognise and track trajectories of each player based on video footage (Yang and Wang, 2022). Modern optical tracking systems have the appeal of non-invasiveness (i.e., athletes carry no equipment), ball tracking, and indoor and outdoor use. Furthermore, the data from optical tracking systems are two-dimensional Cartesian coordinates which can be directly used for tactical analysis without extra data processing. However, high costs, infrastructure requirements for installation, and lack of portability (potentially having no access to away match data) limit its application (Torres-Ronda et al., 2022b).

Radio-based local positioning systems (LPS) work based on electronic signal transmitting to provide location information. In a sports setting, anchors that emit radio waves are placed around the playing field. Athletes are fitted with a vest carrying a lightweight transmitter (i.e., receiver) that returns the radio signal from the anchors. The reception time between anchors and receivers is synchronised and used to triangulate the position (Carling et al., 2008; Torres-Ronda et al., 2022b). In addition to indoor and outdoor application, radio-based positioning systems also share the trait of advances in accuracy and processing efficiency over other types of tracking technology (Carling et al., 2005). However, Angrisani et al. (2017) suggested that radio-based tracking systems is most reliable when it is fixed permanently. Accordingly, it lacks portability and requires a costly installation. Raw positional data are also captured as Cartesian coordinates that can be directly used for tactical variable processing.

Satellite-based positioning systems utilise the satellite network to provide information on the location. Global positioning system (GPS) used to be the overarching term for all tracking systems based on satellite networks. In fact, GPS is a branch of Global Navigation Satellite System (GNSS) that also includes GLONASS, Galileo, and BeiDou (Jiao et al., 2019; Shergill et al., 2021). To use consistent terminology throughout the thesis, GNSS will be used for the satellite-based tracking technology. The GNSS tracking device (receiver) constantly receives signals that are transmitted at the speed of light and contain precise timing information from the atomic clock in satellites. By decoding and synchronising received signals from at least four satellites, the difference in time (i.e., travel time) between each satellite and the receiver is determined, which is then used to calculate distance from the satellite to the receiver. The exact position is retrieved by distances between satellites and the receiver using trigonometry, captured as latitude, longitude, and altitude coordinates (Torres-Ronda et al., 2022b). In the football application, given the size relative to the Earth, the pitch is considered as a plane. Therefore, positional data from GNSS tracking systems only include latitude and longitude coordinates. However, those data require processing to be prepared for team tactical analysis. The limitations include no access to ball tracking and to opposition data, susceptibility to external interference such as weather condition and constructions surrounding the pitch (Shergill et al., 2021). Linke et al. (2018) demonstrated that GNSS (GPSports, Pro X, Canberra, Australia) showed lower (but acceptable) validity for measuring player position than optical tracking and LPS. The GNSS signal quality depends on the number of connected satellites. The ability of acquiring satellite signal may vary across tracking systems by different manufactures. To improve data reliability and validity, manufacturers for example Catapult (Catapult Innovations, South Melbourne, VIC, Australia), integrated GPS and GLONASS to enable more available satellites (Jackson et al., 2018).

The portability is one of the advantages of GNSS tracking systems over optical tracking and LPS. The GNSS receiver can be conveniently put into use without any extra installation of equipment that is required by the other two tracking technologies. This enables that teams travel with the GNSS tracking devices and record team performance continuously wherever the training or match take place. Besides, modern GNSS tracking systems combine multiple sensors within a single device, such as satellite positioning, accelerometer, magnetometer, gyroscope, and heart rate sensor (Torres-Ronda et al., 2022a). Multi-type data are automatically processed by algorithms to provide wider range of physical and technical measures than optical tracking and LPS, such as the running imbalance indicating the load difference between left and right legs (CatapultSports, 2022). Its high cost-effectiveness enables the wide application in a majority of football teams across the highest playing levels and age groups. Given the popularity of the system, this thesis will concentrate on the application of this system in football teams.

Positional data from three types of player tracking system above can be used to quantify team tactical behaviour. Because positional data from GNSS tracking systems are measured as latitude and longitude coordinates rather than Cartesian coordinates, extra data processing is necessary prior to calculating tactical measures. A workflow to outline processing steps of preparing raw GNSS positional data for tactical analysis is presented in chapter two.

## 1.3 Team tactical behaviour

Team tactical performance in football is defined as how the team manages space and time through individual (e.g., one-on-one attacking and defending events with or without the ball) and group actions (e.g., the cooperation within and between subgroup units) in attacking, defending, and transition situations (Carling et al., 2005; Garganta, 2009; Grunz et al., 2012). The speed and direction of team movement depends on the ball possession which continuously switches between opposing teams along the game. With ball possession, to build up and create scoring opportunity, teams tend to adopt different strategies to create spatial and numerical advantage over the opposition. Without ball possession, teams have to move with the opposition but concentrate on remaining defensive positioning and compactness. Welch et al. (2021) reported that the collective movement within the defensive game phase was more ordered, compact (i.e., low surface area) and faster moving compared to the attacking phase. This demonstrated that collective behaviour of a team varies among different match phases (e.g., attacking, defending, transitions). Team tactical behaviour can be determined using positional data from player tracking technology and is captured by tactical measures. Those measures provide information on position relation within the team and between teams, which can be used to identify offensive/defensive game features to create benchmarks for coaching and match preparation.

In a football game, two teams play within an enclosed area and limited time, with the aim of scoring goals while not conceding goals. Individual and team actions are influenced by players skills, playing condition, team cooperation and the organisation of the opposition (Jayal et al., 2018). Those constraints increase the degree of freedom and variability of football game (Garganta, 2009). Interactions within a team, between teams, and between player and environment exist and influence collective behaviour (Vilar et al., 2012; Duarte et al., 2012). In sport science literature, collective behaviour is also described as team tactical behaviour. The ecological approach, such as the dynamical system theory, enables capturing and identifying these intra-team (i.e., between teammates) cooperation, inter-team (i.e., between teams) coordination and competition (Araújo et al., 2015; Seifert et al., 2017).

Team tactical behaviour can be quantified by tactical measures calculated from the coordinates of players’ positions. Those coordinates are retrieved by player tracking technology systems with a sampling frequency of 5-25 Hz (Low et al., 2020). The process of constructing features (tactical measures) from raw data (coordinates) to capture collective behaviour is referred to as spatial aggregation which increases the interpretability of positional data (Goes et al., 2020). The aggregation allows tactical measures to focus on dynamic relation at various interpretable levels of play: 1) the team level, represented by competition between teams; 2) the sub-group level, represented by coordination of a small group of players (e.g., defensive line); and 3) the individual level, represented by dyadic interaction between two players (Grehaigne et al., 1997; Gréhaigne and Godbout, 2013). Tactical measures themselves provide information on position relation between players, sub-groups, and teams on the pitch. With these measures, coaches and sport scientists can identify weakness and strength on team tactics, which benefits planning for future competition.

Proposed team tactical measures as shown in Figure 1.2, describe players collective movements (i.e., intra-team coordination and inter-team competition) on the pitch (Memmert et al., 2016). Intra-team measures reflect the dispersion of players by variables such as length, width, and length-per-width radio (LpW ratio). These measures provide insights to the shape of the formation. The team centroid represents the mean longitudinal and lateral positions of outfield players. Stretch indices measured by the mean distance of all outfield players from a team to the centroid, along with LpW ratio reflect the extent of team spreading. The surface area refers to the convex hull enclosed by outfield players from a team. Inter-team measures reflect the pressure of two teams, displayed by the absolute distance between team centroids (Frencken et al., 2012). Space control is quantified by the Voronoi diagram to represent the nearer region to one player than others, also considered as the dominant region of the player (Taki and Hasegawa, 2000; Fujimura and Sugihara, 2005). Positional data of all players on the pitch are required to calculate this measure. These spatial features enable researchers to explore the team's tactical behaviour during match-play, but also the influence of manipulating independent factors such as playing formation (Baptista et al., 2020), pitch size (Clemente et al., 2018), quality of opposition (Folgado et al., 2014), playing (defending) strategy (Low et al., 2021), number of players on a similar relative pitch area (Olthof et al., 2019) on tactical behaviour.

Tactical measures have previously been analysed in different perspectives of time. To extract and compare these spatial features, a given time-window is needed to aggregate the data, also termed as temporal aggregation (Goes et al., 2020). Driven by various research questions and hypotheses, time windows are customised by researchers to explore details of collective movements under different match contexts. As a result, there are mainly three existing methods to set the time window: 1) fixed-size time window; 2) fixed-size time window linked to contextual information; 3) flexible-size time window (Goes et al., 2020). Firstly, with a fixed-size window spatial features have been aggregated over the course of fifteen-minute duration (Duarte et al., 2013b), a half match (Ric et al., 2016), or a bout of small-sided games (Gonçalves et al., 2016). This type of aggregation regularly averages tactical measures within a time window and provides insights into the specific period. Whilst spatial features are aggregated over a match period, periods of stoppage which do not belong to match tactics should be excluded from analysis (Clemente et al., 2014a; Clemente et al., 2014b). Secondly, a window of fixed size linked with match events for example, Frencken et al. (2012) used three-seconds moving window to identify variability of longitudinal and lateral inter-team distances. This 3-second window here was chosen based on the opinion of football coaches which would reflect the time that the team needs for reorganisation. Lastly, spatial features can also be aggregated according to the flexible window of oriented events such as playing sequences in ball possession (Silva et al., 2014), a sequence of passes (Rein et al., 2017), attacking and defending phases (Castellano et al., 2016). The main difference between time windows above is the types of information involved (i.e., only duration or combining with constraints). It is also important to note that when linking the spatial features with match events, the timelines of positional data and event data needs to be synchronized. Additionally, temporal analyses on those spatial features can identify team synchronisation and regularity of synchronisation. Entropies measuring the disorder of a system have been used to quantify the regularity of team synchronisation (Sampaio and Macas, 2012). Duarte et al. (2013a) calculated the entropy with a-half time window and found more regularity of collective movement in second half than first half. In this thesis, the flexible time window will be combined with different phases of match to analyse team tactical performance.

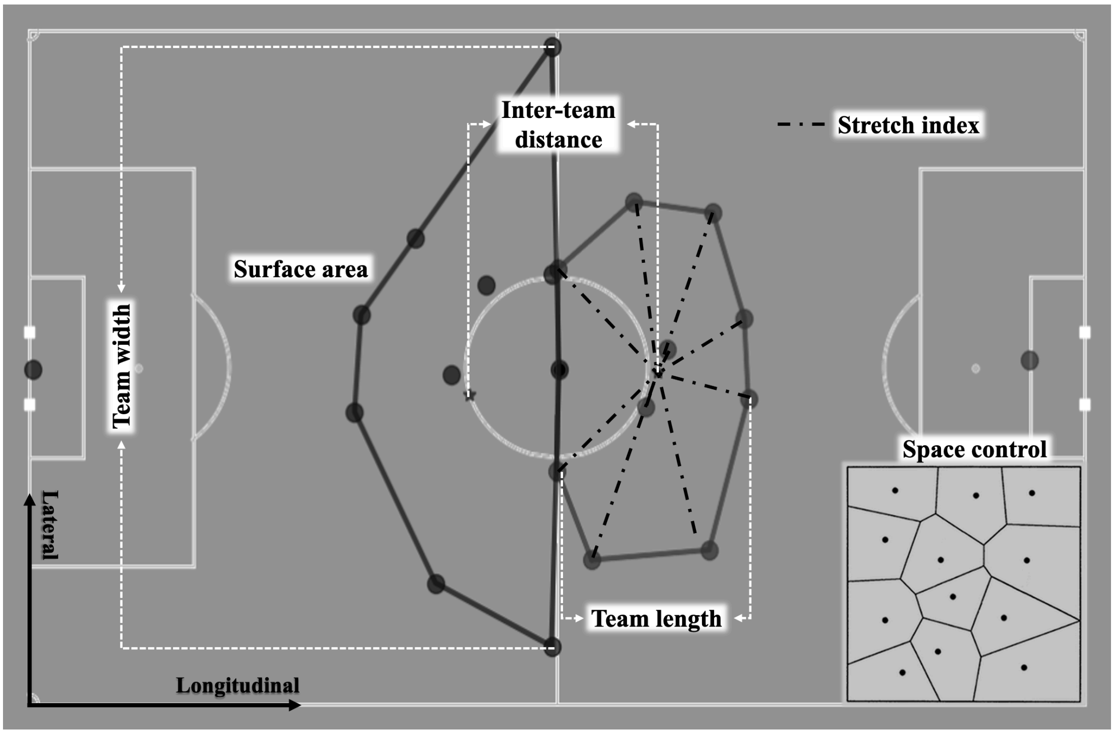



**Figure 1.2** Formation of opposing teams with tactical variables for intra-team and inter-team coordination. LpW ratio is calculated as the team length by the team width. A Voronoi diagram is given as an example of space control. Longitudinal and lateral direction are considered as the direction of pitch length and width respectively.

## 1.4 Match phases

Possession over the ball determines the match phases for a team, which has become an important perspective to analyse team performance in particular (Praça et al., 2022). During a football match, a team will cycle through several general phases: attacking (in possession), defending (out of possession), and transitions between attacking and defending. Tactical behaviour of a team changes as a response to match situations (e.g., phases, location) based on playing styles and strategies of the own team and opposition (Rico-Gonzalez et al., 2022). Ball possession influenced the strategy and tactical behaviour of team players (Clemente et al., 2013). Accordingly, teams develop and adopt different strategies and tactics for their own offensive and defensive actions across the phases as a response to the opposition and game status. This can be described by those aforementioned tactical measures. Analysing tactical behaviour in different phases facilitates identifying teams' (i.e., the own team and opposition) tactical property, whereby coaches can make comprehensive tactical plans for match preparation.

During an attack, a team might face against 1) high-press defence that the opposition is aggressively pressing in the defensive third of the team; 2) mid-block defence in the midfield third; 3) low-block defence in the attacking third (Teoldo da Costa et al., 2010). When the opposition is lured out of their defensive third, massive space is left behind the defensive line. A long-range delivery or a sequence of quick passes could extensively progress the ball forwards and even to the back of the defensive line, disorganising the defence of the opposing team. During the defence, similar defensive tactics above can be used against the opposition having ball possession. In between, as the team regains the ball possession, transition begins from defence to attack. Transition from attack to defence starts as the opposition regains ball possession (Gonzalez-Rodenas et al., 2016). As a result of players' different moving patterns as the response to match situations, team tactical behaviour varies across these phases. Data-driven approaches using event data (Worville, 2019) and positional data (Llana et al., 2020) have been proposed to divide the match into possession-oriented phases. Clemente et al. (2013) have demonstrated the differences in tactical behaviour between in possession and out of possession. Teams in attack tend to create space for penetrative actions (e.g., pass or dribbling). Welch et al. (2021) suggested that reducing the width and depth in defence can increase collective movement speeds and the level of collective order to close passing opportunities of the opposition. These findings evidence the different strategies and tactics execution in offensive and defensive phases. Nevertheless, because of distinct objectives, collective performance in transition from attack to defence and from defence to attack, differs from when the opposition and own team controlling ball possession, respectively. Previous studies only decomposed match-play into offensive and defensive phases based on ball possession. Dividing the offensive phase into defence-to-attack transition and in possession, and the defensive phase into attack-to-defence transition and out of possession, is expected to provide further insights on team tactical behaviour.

## 1.5 Thesis outline

In summary, the player tracking technology, especially GNSS tracking technology, has been extensively applied in match analysis. Positional data have been employed to quantify team tactical behaviour to uncover in-depth collective movement details. Based on tactical analysis, the weakness and strength of team tactics execution can be improved and built on respectively. Tactical behaviour in different match phases can be reflected with quantification of tactical measures. Strategy can then be adjusted for the following competition according to the properties of the team and the opposition. However, using positional data from GNSS tracking technology to quantify team tactical behaviour requires steps that are rather complex and time consuming. Although Folgado et al. (2014) suggested the necessary information required for tactical analysis using GNSS positional data, a depiction of raw positional data and referential procedures has yet to be reported in previous studies. A workflow that provides a detailed description of data-processing steps and solutions to potential issues could be informative for practitioners. This enables all users of GNSS tracking systems to explore their interests in team tactical performance.

The objectives of this thesis are two-fold and related. To enable practitioners and scientists using GNSS tracking systems to analyse tactical behaviour without time-consuming trial and error, firstly, it is aimed to determine a sensible and universal workflow for using GNSS positional data for tactical analysis, and to explore the challenges during data processing and corresponding solutions. Given that teams tend to adopt different tactics across match phases rather than keep same one over the match, the second objective is to use GNSS positional data to compare team tactical behaviour in four basic match phases (i.e., attacking, defending, transition from attacking to defending, and transition from defending to attacking). The proposed workflow is then put into practice.

In Chapter 2, an exploratory study is conducted to bridge the gap between GNSS positional data and data that is ready for tactical analysis. The detailed description of data processing, exemplar dataset and Python routine can facilitate practitioners to easily proceed with tactical analysis and to overcome the most common issues when working with this type of data.

In Chapter 3, GNSS positional data of a professional football team from a competitive match are used to compare team tactical behaviour across basic match phases: in possession, attack-to-defence transition, out of possession, and defence-to-attack transition. It is hypothesised that the team plays with different formations and tactical execution across these phases of match.

This thesis concludes with a general discussion of relevant results, strengths and limitations of the studies, and implications for soccer practice and science that can guide future study.

## References


Akenhead, R. & Nassis, G.P. 2016. Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *International Journal of Sports Physiology and Performance,* 11**,** 587-93.

Angrisani, L., Arpaia, P. & Gatti, D. Analysis of localization technologies for indoor environment. 2017 IEEE International Workshop on Measurement and Networking (M&N), 27-29 Sept. 2017 2017. 1-5.

Araujo, D. & Davids, K. 2009. Ecological approaches to cognition and action in sport and exercise: Ask not only what you do, but where you do it. *International Journal of Sport Psychology,* 40**,** 5.

Araújo, D., Passos, P., Esteves, P., Duarte, R., Lopes, J., Hristovski, R. & Davids, K. 2015. The micro-macro link in understanding sport tactical behaviours: Integrating information and action at different levels of system analysis in sport. *Movement & Sport Sciences - Science & Motricité***,** 53-63.

Baptista, J., Travassos, B., Goncalves, B., Mourao, P., Viana, J.L. & Sampaio, J. 2020. Exploring the Effects of Playing Formations on Tactical Behavior and External Workload During Football Small-Sided Games. *Journal of Strength and Conditioning Research,* 34**,** 2024-2030.

Buchheit, M., Racinais, S., Bilsborough, J.C., Bourdon, P.C., Voss, S.C., Hocking, J., Cordy, J., Mendez-Villanueva, A. & Coutts, A.J. 2013. Monitoring fitness, fatigue and running performance during a pre-season training camp in elite football players. *J Sci Med Sport,* 16**,** 550-5.

Carling, C., Bloomfield, J., Nelson, L. & Reilly, T. 2008. The role of motion analysis in elite soccer: contemporary performance measurements techniques and work rate data. *Sports Medicine,* 38**,** 24.

Carling, C., Lacome, M., McCall, A., Dupont, G., Le Gall, F., Simpson, B. & Buchheit, M. 2018. Monitoring of post-match fatigue in professional soccer: welcome to the real-world. *Sports Medicine,* 48**,** 8.

Carling, C., Williams, A.M. & Reilly, T. 2005. *Handbook of Soccer Match Analysis a Systematic Approach to Improving Performance,* London ;, Routledge.

Castellano, J., Blanco-Villasenor, A. & Alvarez, D. 2011. Contextual variables and time-motion analysis in soccer. *International Journal of Sports Medicine,* 32**,** 415-21.

Castellano, J., Casamichana, D. & Lago, C. 2012. The Use of Match Statistics that Discriminate Between Successful and Unsuccessful Soccer Teams. *Journal of Human Kinetics,* 31**,** 139-47.

Castellano, J., Silva, P., Usabiaga, O. & Barreira, D. 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**,** 153-163.

CatapultSports. 2022. *How to Detect Running Symmetry Metrics | Catapult Sports* [Online]. CatapultSports. Available: https://support.catapultsports.com/hc/en-us/articles/360001044935-How-to-Detect-Running-Symmetry-Metrics#h_69274c9b-eabf-4e66-ba90-76247db8dd89 [Accessed 7 Oct 2022].

Clemente, F., Couceiro, M., Martins, F., Mendes, R. & Figueiredo, A.n. 2013. Measuring Tactical Behaviour Using Technological Metrics: Case Study of a Football Game. *International Journal of Sports Science & Coaching,* 8.

Clemente, F.M., Santos Couceiro, M., Lourenço Martins, F.M., Sousa Mendes, R. & Figueiredo, A.J. 2014a. Intelligent systems for analyzing soccer games: The weighted centroid. *Ingeniería e Investigación,* 34**,** 70-75.

Clemente, F.M., Sequeiros, J.B., Correia, A., Sierra-Olivares, J., González-Víllora, S., Silva, F. & Martins, F.M.L. 2018. How dots behave in two different pitch sizes? Analysis of tactical behavior based on position data in two soccer field sizes. [¿Cómo se comportan los puntos en dos campos diferentes? Análisis del comportamiento táctico basado en los datos de posición en dos tamaños de campo de fútbol]. *RICYDE. Revista internacional de ciencias del deporte,* 14**,** 16-28.

Clemente, M.F., Martins, F.M.L., Couceiro, S.M., Mendes, S.R. & Figueiredo, A.J. 2014b. Inspecting teammates' coverage during attacking plays in a football game: A case study. *International Journal of Performance Analysis in Sport,* 14**,** 384-400.

Duarte, R., Araujo, D., Correia, V. & Davids, K. 2012. Sports Teams as Superorganisms. *Sports Medicine,* 42.

Duarte, R., Araújo, D., Correia, V., Davids, K., Marques, P. & Richardson, M.J. 2013a. Competing together: Assessing the dynamics of team–team and player–team synchrony in professional association football. *Human Movement Science,* 32**,** 555-566.

Duarte, R., Araújo, D., Folgado, H., Esteves, P., Marques, P. & Davids, K. 2013b. Capturing complex, non-linear team behaviours during competitive football performance. *Journal of Systems Science and Complexity,* 26**,** 62-72.

Duarte, R., Ferreira, R., Folgado, H. & Fernandes, O. 2010. Interpersonal dynamics in team sports: the role of TACTO and its applications. *International journal of sport psychology,* 41**,** 62-63.

Fernandes, O., Folgado, H., Duarte, R. & Malta, P. 2010. Validation of the Tool for Applied and Contextual Time-Series Observation. *International Journal of Sport Psychology,* 41**,** 2.

Folgado, H., Duarte, R., Fernandes, O. & Sampaio, J. 2014. Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches. *PLoS One,* 9**,** e97145.

Franks, I.M. & Miller, G. 1991. Training coaches to observe and remember. *Journal of sports sciences,* 9**,** 285-297.

Frencken, W., Poel, H.d., Visscher, C. & Lemmink, K. 2012. Variability of inter-team distances associated with match events in elite-standard soccer. *Journal of Sports Sciences,* 30**,** 1207-1213.

Fujimura, A. & Sugihara, K. 2005. Geometric analysis and quantitative evaluation of sport teamwork. *Systems and Computers in Japan,* 36**,** 49-58.

Gandomi, A. & Haider, M. 2015. Beyond the hype: Big data concepts, methods, and

analytics. *International Journal of Information Management,* 35**,** 137-144.
Garganta, J. 2009. Trends of tactical performance analysis in team sports : bridging the gap between research, training and competition. *Revista Portuguesa de Ciências do Desporto,* 9.
Goes, F.R., Meerhoff, L.A., Bueno, M.J.O., Rodrigues, D.M., Moura, F.A., Brink, M.S., Elferink-Gemser, M.T., Knobbe, A.J., Cunha, S.A., Torres, R.S. & Lemmink, K.A.P.M. 2020. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science,* 21**,** 481-496.
Gonçalves, B., Marcelino, R., Torres-Ronda, L., Torrents, C. & Sampaio, J. 2016. Effects of emphasising opposition and cooperation on collective movement behaviour during football small-sided games. *Journal of Sports Sciences,* 34**,** 1346–1354.
Gonzalez-Rodenas, J., Lopez-Bondia, I., Calabuig, F., Pérez-Turpin, J.A. & Aranda, R. 2016. Association between playing tactics and creating scoring opportunities in counterattacks from United States Major League Soccer games. *International Journal of Performance Analysis in Sport,* 16**,** 737-752.
Gréhaigne, J.-F. & Godbout, P. 2013. Collective Variables for Analysing Performance in Team Sports. *Routledge Handbook of Sports Performance Analysis.* Routledge.
Grehaigne, J.F., Bouthier, D. & David, B. 1997. Dynamic-system analysis of opponent relationships in collective actions in soccer. *Journal of Sports Sciences,* 15**,** 137-49.
Groom, R., Cushion, C. & Nelson, L. 2011. The delivery of video-based performance analysis by England youth soccer coaches: towards a grounded theory. *Journal of applied sport psychology,* 23**,** 16-32.
Grunz, A., Memmert, D. & Perl, J. 2012. Tactical pattern recognition in soccer games by means of special self-organizing maps. *Human Movement Science,* 31**,** 334-343.
Jackson, B.M., Polglaze, T., Dawson, B., King, T. & Peeling, P. 2018. Comparing Global Positioning System and Global Navigation Satellite System Measures of Team-Sport Movements. *International Journal of Sports Physiology and Performance,* 13**,** 1005-1010.
Jayal, A., McRobert, A., Oatley, G. & O'Donoghue, P. 2018. *Sports analytics: Analysis, visualisation and decision making in sports performance*, Routledge.
Jiao, G., Song, S., Ge, Y., Su, K. & Liu, Y. 2019. Assessment of BeiDou-3 and Multi-GNSS Precise Point Positioning Performance. *Sensors,* 19.
Lago-Peñas, C., Rey, E. & Lago-Ballesteros, J. 2012. The influence of effective playing time on physical demands of elite soccer players. *The Open Sports Sciences Journal,* 5.
Linke, D., Link, D. & Lames, M. 2018. Validation of electronic performance and tracking systems EPTS under field conditions. *PLoS One,* 13**,** e0199519.
Llana, S., Madrero, P., Fernández, J. & Barcelona, F. The right place at the right time: Advanced off-ball metrics for exploiting an opponent's spatial weaknesses in soccer. Proceedings of the 14th MIT Sloan Sports Analytics Conference, 2020.
Low, B., Coutinho, D., Goncalves, B., Rein, R., Memmert, D. & Sampaio, J. 2020. A

Systematic Review of Collective Tactical Behaviours in Football Using Positional Data. *Sports Medicine,* 50**,** 343-385.

Low, B., Rein, R., Raabe, D., Schwab, S. & Memmert, D. 2021. The porous high-press? An experimental approach investigating tactical behaviours from two pressing strategies in football. *Journal of Sports Sciences,* 39**,** 2199-2210.

Mackenzie, R. & Cushion, C. 2013. Performance analysis in football: a critical review and implications for future research. *J Sports Sci,* 31**,** 639-76.

Mallo, J., Mena, E., Nevado, F. & Paredes, V. 2015. Physical demands of top-class soccer friendly matches in relation to a playing position using global positioning system technology. *Journal of human kinetics,* 47**,** 179.

Memmert, D., Lemmink, K.A.P.M. & Sampaio, J. 2016. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine,* 47**,** 1-10.

Memmert, D. & Rein, R. 2018. Match analysis, Big Data and tactics: current trends in elite soccer. *Deutsche Zeitschrift für Sportmedizin,* 2018**,** 65-72.

Morgans, R., Orme, P., Anderson, L. & Drust, B. 2014. Principles and practices of training for soccer. *Journal of Sport and Health Science,* 3**,** 251-257.

O'Donoghue, P. 2015. *An Introduction to Performance Analysis of Sport*, Routledge.

Olsen, E. & Larsen, O. 1997. Use of match analysis by coaches. *Science and football III,* 3**,** 209-220.

Olthof, S.B.H., Frencken, W.G.P. & Lemmink, K. 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**,** 523-530.

Praça, G.M., Moreira, P.E.D., de Andrade, A.G.P., Clemente, F.M., de Oliveira, W.B. & Demétrio, G. 2022. Integrating notational and positional analysis to investigate tactical behavior in offensive and defensive phases of football matches. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*.

Reilly, T., Clarys, J. & Stibbe, A. 1993. *Science and Football II*, Spon.

Rein, R. & Memmert, D. 2016. Big data and tactical analysis in elite soccer: future challenges and opportunities for sports science. *SpringerPlus,* 5**,** 1410.

Rein, R., Raabe, D. & Memmert, D. 2017. "Which pass is better?" Novel approaches to assess passing effectiveness in elite soccer. *Human Movement Science,* 55**,** 172-181.

Ric, A., Torrents, C., Goncalves, B., Sampaio, J. & Hristovski, R. 2016. Soft-assembled Multilevel Dynamics of Tactical Behaviors in Soccer. *Frontiers in Psychology,* 7**,** 1513.

Rico-Gonzalez, M., Pino-Ortega, J., Castellano, J., Oliva-Lozano, J.M. & Los Arcos, A. 2022. Reference values for collective tactical behaviours based on positional data in professional football matches: a systematic review. *Biology of Sport,* 39**,** 110-114.

Sampaio, J. & Macas, V. 2012. Measuring tactical behaviour in football. *International Journal of Sports Medicine,* 33**,** 395-401.

Sampaio, J.E., Lago, C., Goncalves, B., Macas, V.M. & Leite, N. 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 and Medicine in Sport,* 17**,** 229-33.

Seifert, L., Araujo, D., Komar, J. & Davids, K. 2017. Understanding constraints on sport performance from the complexity sciences paradigm: An ecological dynamics framework. *Human Movement Science,* 56**,** 178-180.

Serrano, J., Shahidian, S. & Fernandes, O. 2014. Validation of a Manual Position Tracking Software (TACTO) to Quantify the Football Movements. *International Journal of Sports Science,* 4**,** 7.

Shergill, A.S., Twist, C. & Highton, J. 2021. Importance of GNSS data quality assessment with novel control criteria in professional soccer match-play. *International Journal of Performance Analysis in Sport,* 21**,** 820-830.

Silva, P., Duarte, R., Sampaio, J., Aguiar, P., Davids, K., Araujo, D. & Garganta, J. 2014. Field dimension and skill level constrain team tactical behaviours in small-sided and conditioned games in football. *Journal of Sports Sciences,* 32**,** 1888-1896.

Strudwick, A. 2013. Contemporary issues in the physical preparation of elite players. *Science and Soccer.* Routledge.

Taki, T. & Hasegawa, J. Visualization of dominant region in team games and its application to teamwork analysis. Proceedings Computer Graphics International 2000, 19-24 June 2000 2000. 227-235.

Teoldo da Costa, I., Garganta, J., Greco, P., Isabel, M. & Afonso, J. 2010. Assessment of tactical principles in youth soccer players of different age groups. *Revista Portuguesa de Ciências do Desporto,* 10**,** 147-157.

Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A. & Clubb, J. 2022a. Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis. *Sports Medicine - Open,* 8**,** 15.

Torres-Ronda, L., Clubb, J. & Beanland, E. 2022b. Tracking Systems in Team Sports: Back to Basics. *Sport Performance & Science Reports*.

Vilar, L., Araujo, D., Davids, K. & Button, C. 2012. The role of ecological dynamics in analysing performance in team sports. *Sports Medicine,* 42**,** 1-10.

Welch, M., Schaerf, T.M. & Murphy, A. 2021. Collective states and their transitions in football. *PLoS One,* 16**,** e0251970.

Worville, T. 2019. *Phases of play – An Introduction.* [Online]. Stats Perform. Available: https://www.statsperform.com/resource/phases-of-play-an-introduction/ [Accessed August 29 2022].

Yang, X. & Wang, H. 2022. Design and Implementation of Intelligent Analysis Technology in Sports Video Target and Trajectory Tracking Algorithm. *Wireless Communications and Mobile Computing,* 2022.

# Chapter 2: Bridging the gap between GNSS positional data and tactical analysis

## 2.1 Introduction

Player tracking technology is widely used in soccer, but the International Football Association Board (IFAB) did not allow the use of wearable technology during official competition until 2016 (IFAB, 2016). Teams use this technology mainly for capturing players' movements and monitoring physical conditioning. Physical measures used to quantify external load are presently based on three categories: speed, accelerations/ decelerations, and composite variables, such as dynamic stress load and repeated high intensity (Silva et al., 2018; Rago et al., 2020). Total distance, peak speed and distance covered by different speed zones attained are widely used indicators of training volume in professional football (soccer) (Akenhead and Nassis, 2016; Bourdon et al., 2017; Brink and Frencken, 2018). In addition to physical monitoring, positional data collected by player tracking technology can contribute to analysing tactical behaviour (Sampaio and Macas, 2012). Although tactical analysis based on player tracking technology is well-studied by sport scientists, this is a relatively unexplored area by sports teams, as processing the raw positional data from player tracking technology is complex.

As the definition in chapter 1, tactical behaviour in football is described as individual and collective actions of a team to best use player skills and manage spatial positioning over time to beat the opposition by scoring more goals. To investigate team positioning and distribution, several tactical variables have been conceptualized to capture this tactical behaviour, which are related to positions, distances, spaces, and numerical relations of players. As shown in Figure 1.2 (chapter 1), these measures include but are not limited to, centroid, width, length, length per width ratio (LpW ratio), surface area and space control of the formation, inter-team distance, and stretch indices (Frencken et al., 2011; Folgado et al., 2012; Fonseca et al., 2012; Frencken et al., 2012; Olthof et al., 2015; Memmert et al., 2016). These variables have previously been used to determine team collective behaviour in different football populations (e.g., age, gender, and playing level) (Castellano et al., 2017; Silva et al., 2014), playing formats (i.e., official match and small-sided games) (Castellano et al., 2016; Olthof et al., 2019), and phases of the match (Welch et al., 2021). Additionally, with the knowledge of dynamics, positional data are capable of providing insights into team position synchronization (Folgado et al., 2018) and regularity of players' moving patterns (Sampaio and Macas, 2012). These previous studies were all based on player tracking technology systems. Each of them has their own technical mechanism, application environment, and advantages and limitations.

### 2.1.1 Player tracking technology systems

Three mainstream types of tracking systems have been widely applied in football (Low et al., 2020; Goes et al., 2020): radio-based local positioning system, optical tracking system, and satellite-based positioning system. To use consistent terminology throughout the study, satellite-based positioning systems will be called as Global Navigation Satellite System (GNSS), a collective term for all satellite navigation systems providing geospatial positioning with global coverage (Jackson et al., 2018). In sport science, these systems are also referred to as Global Positioning Systems (GPS). All systems track player's positioning on the pitch and record players' positional data, but in different formats. Local positioning system (LPS) and optical tracking system share the trait that positional data are collected and saved as two-dimensional Cartesian coordinates (i.e., x and y coordinates). These coordinates enable calculating tactical variables without pre-processing. Nevertheless, the lack of portability and high expense limit the wide application of LPS and optical tracking in football teams (Torres-Ronda et al., 2022). By contrast, mobile, lower-expensive, and wider application make GNSS tracking systems popular in this sport. However, GNSS tracking systems record player position as geographic coordinates (i.e., latitude and longitude coordinates). Geographic coordinates are less suitable for determining tactical performance variables and require data processing steps to convert them to Cartesian coordinates.

In the Geographical Coordinate System (GCS), a tuple of three geographic coordinates (i.e., latitude, longitude, and altitude) is considered as a unique identifier of a precise geographic location (Maling, 1992). Accordingly, tracking systems in sports capture latitude and longitude coordinates of players at each timestamp, but without altitude or depth because of negligibly relative size of a pitch (compared to the Earth). The measurements of latitude and longitude are angles (Figure 2.1) and are not on a planar surface. Scientists have been using these coordinates of players to determine tactical behaviour, but this method requires steps to convert them into Cartesian coordinates (Folgado et al., 2014).

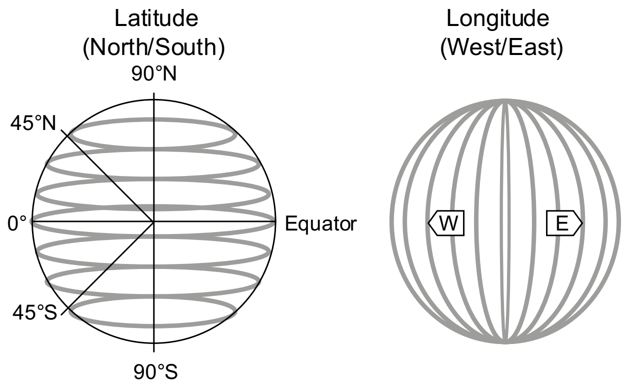



**Figure 2.1** Latitude ($\Phi$) and longitude ($\lambda$) measurements in Geographical Coordinate System. Latitude varies from 0° at the Equator to 90° (North and South) at the poles. Longitude varies from 0° at Greenwich to 180° East and West.

GNSS refers to a constellation of satellites that conveys signals providing positioning and timing data from space to receivers on earth. However, in discussion of tracking systems in sports, GPS is commonly considered as the overarching name of a range of tracking technologies based on satellites. In addition to GPS, GNSS network also includes GLONASS, Galileo and BeiDou systems (Table 2.1). Latest tracking systems in sports utilise multiple systems for enhanced accuracy and reliability of information. For example, inside most Catapult GNSS tracking devices (Catapult Innovations, South Melbourne, VIC, Australia) is a tracking engine that is accessing satellites from GPS and GLONASS (Jackson et al., 2018). For different versions and brands of tracking systems in sports, their effective sampling frequency of positional data varies from 1 Hz to 25 Hz (Low et al., 2020).

**Table 2.1** Overview of different types of Global Navigation Satellite Systems (GNSS).

| Global Navigation Satellite System (GNSS) | Owner | Active Satellite (*n*) | Precision |
|---|---|---|---|
| Global Positioning System (GPS) | United States | 32 | 0.3 - 5 m |
| GLONASS | Russia | 24 | 2 - 4 m |
| Galileo | European Union | 24 | 1 m (Public)<br>0.01 m (Encrypted) |
| BeiDou-3 | China | 35 | 3.6 m (Public)<br>0.1 m (Encrypted) |

The integral atomic clock in satellite provides precise timing information in the signal. GNSS receivers allow the simultaneous reception of signals (encoded time and date stamp) from several satellites, and then decode and synchronize the transmitted and received signals. By calculating the time that signals take from four or more satellites to the receiver on earth, and determining the distances (i.e., time multiplied by the speed of light), precise positioning information is triangulated and captured. A limitation in this process is that the GNSS signal consistency is susceptible to covers that could surround the pitch or weather conditions (e.g., constructions and overcast). Consequently, data instability (Shergill et al., 2021) combining with aforementioned data processing (i.e., converting geographic coordinates into Cartesian coordinates) limit the GNSS application on team tactical analysis.

Altogether, there are several boundaries that limit the use of raw GNSS positional data for team tactical analysis. The trial and error can be a time-consuming process. Therefore, this chapter aims to explore a workflow of processing raw GNSS positional data to overcome the limitations of this system. With more teams using GNSS tracking technology, this workflow may give practitioners and scientists access to tactical analysis. However, there is currently no scientific consensus on processing positional data for team tactical analysis. This chapter will present methodological guidelines for team tactical analysis. Furthermore, practitioners using GNSS tracking technology may also benefit from these guidelines for quantifying and analysing tactical behaviour in addition to physical monitoring.

## 2.2 Materials and methods

To track players' outdoor activity with GNSS technology, players are fitted with a specific harness that carries GNSS tracking devices during training sessions and matches. GNSS tracking systems are not capable of receiving signal in indoor settings. Positional data from GNSS devices can be used for tactical analysis, but this requires a workflow to make the data useful for calculating tactical variables. A general workflow has been described before by Folgado et al. (2014), but this lacks a detailed description of required information for data processing (e.g., how to retrieve pitch location coordinates) or the solution to potential data loss. Generally, to process raw data for tactical analysis, there are three parts of information required: 1) session information that gives the knowledge of start time, end time and outfield players; 2) coordinates of pitch location; and 3) raw positional data. As the intertwined flows show in Figure 2.2, they provide different ingredients which contribute interdependently to data processing. In the following, this process will be described step by step. All the processing was conducted in Python 3.8. Customised Python routines and exemplar datasets can be accessed via GitHub (https://github.com/kehanabcd/GNSS-positional-data-to-Tactical-analysis).

An exemplar dataset to illustrate the processing steps was used in this chapter. Players from a Spanish academy team (under-18) played small-sided games (SSG) and were wearing Catapult tracking system (10 Hz, Optimeye S5, Catapult). The dataset includes three types of SSG: 4 vs. 4 + goalkeepers (GKs), 5 vs. 5 + GKs, and 6 vs. 6 + GKs. Each SSG was played for 5 minutes.

### 2.2.1 Raw GNSS positional data

Raw positional data of each player were exported and extracted from the GNSS tracking system. Generally, it comprises columns of timestamp, and latitude and longitude coordinates of a player at each timestamp (e.g., 5 or 10 Hz) during the activity. At this point, it is important that timestamps from the positional data and the start and end timestamps correspond and are of a similar format.

### 2.2.2 Session information: start and end timestamps

Session information provides start and end timestamps of match halves, match phases, and/or the training session. The timestamp may present as: 1) Unix timestamp, the running total of seconds that have elapsed since the Unix Epoch; or 2) playing time, a combination of hour, minute and second, starting from 00:00:00. Exported timestamp information can be directly used to subset player's positional data without the noise of activities outside match-play or unnecessary training activities. This would be outlined in the following Step I.

In positional data, occasional missing data may occur. To solve the issue of potential missing timestamps in the positional dataset, the start and end timestamp were then used to create a synchronous timeline to later fill in those missing data through interpolation, ensuring that the final team positional dataset is aligned with the specific sampling frequency (e.g., 10 Hz). This process is outlined in the Step VI.

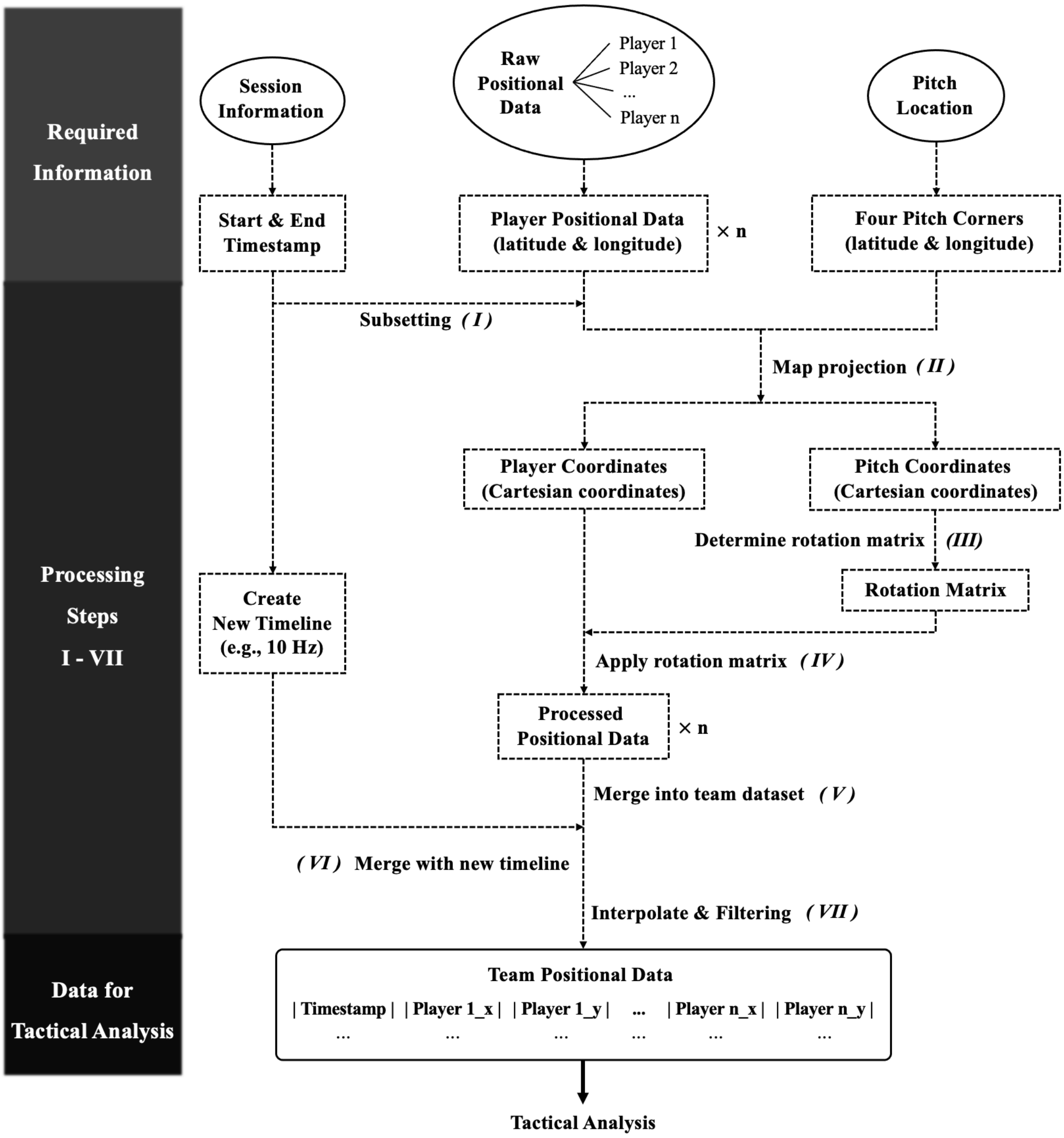



**Figure 2.2.** Workflow of data processing for tactical analysis.

### 2.2.3 Pitch location: four pitch corners

Match and training sessions can be played on different locations which have unique geographic coordinates. The location of each pitch is required for the calculation of the rotation matrix (Folgado et al., 2014). Although Folgado et al. (2014) provided an outline for the rotation matrix, no adequate method to determine the pitch location has been described. In practice, there are two adoptable approaches to retrieve pitch location. The first approach is using a web mapping platform, where the pitch location is being obtained through an online platform that provides satellite imagery such as Google Maps. Four corners of the pitch are visually determined, and latitude and longitude coordinates are manually obtained by mouse clicking on those corners which will reveal the coordinates. An alternative approach is using GNSS tracking devices to acquire coordinates of each pitch corner. There are also two applicable protocols, either placing one device at each corner in turn or placing four devices simultaneously at four corners. The latter costs less time and controls external factors (e.g., overcast) that potentially influence signal quality. The mean location over that collection period can then be used as the location of the pitch. In addition, walking along one of side lines (e.g., the length of the pitch) with a GNSS tracking device is also an option to establish the pitch location.

The methods from web mapping and GNSS tracking devices were further outlined in the Appendix. Web mapping demonstrated stable outcomes from intra-observer and inter-observer reliability tests. It is also a relatively easy method to perform and will not require a data processing step, as would be necessary with using the device method. Lastly, data from GNSS tracking devices showed relatively large variability over the course of the protocol. The web mapping protocol is therefore the recommended protocol to retrieve coordinates for the pitch rotation step.

Based on the outcome of this protocol, the pitch location retrieved from web mapping was used for further analysis in this chapter. As shown in Figure 2.2, these geographic coordinates were then converted to Cartesian coordinates (i.e., x- and y-coordinates) ahead of calculating the rotation matrix.

### 2.2.4 Preparing the data for tactical analysis

With the start and end timestamps, coordinates of pitch location, and the positional data in hand, the main part of data processing is ready for proceeding. The data processing steps outlined in Figure 2.2 will be step-by-step described below.

**Step I. Subsetting for useful data**

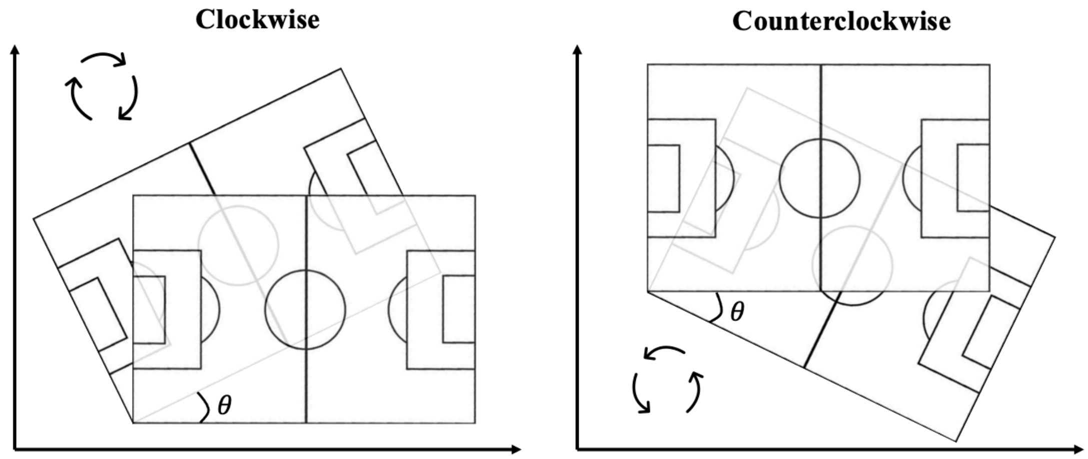
In most cases, a match or training session usually starts after activation of GPS units and stops before deactivation of GPS units, which means player's positional data include noise of activities that are not directly related to match-play or training formats. Therefore, start and end timestamps from session information play a role in excluding these data. By matching the start and end timestamps of the session information with timestamps in positional data, useful positional data can be extracted from the original dataset and then be used as a subset. Generally, the timestamp in the Unix-format consists of nine or ten decimals representing date and time of the activity. Given the 10 Hz data sampling, a minimum of six decimals were required to subset the positional data for each data sample, as the exemplar timestamp in Figure 2.4a. This step has been outlined in the code (*line 73 to 103, 277 to 280*).

**Step II. Convert geographic coordinates into Cartesian coordinates (Map projection)**

Map projections include sets of mathematical models, which are used to transform geographic coordinates to Cartesian coordinates. Methods of map projection include, but are not limited to, 1) Transverse Mercator Projection, also known as Gauss-Krüger projection (Rod Deakin et al., 2010); 2) Universal Transverse Mercator projection (Karney, 2011); 3) Lambert conical projection (Weisstein, 2009); and 4) Stereographic double projection. Mathematical formulas and equations are presented and elaborated in those cited references above. The accuracy of each method depends on the region that is being mapped. Universal Transverse Mercator projection which has been used as a standard projection for decades, was applied in this chapter. This projection has been outlined in the code (*line 11 to 47*).

**Step III. Determine rotation matrix**

Because matches are played on different locations (i.e., home and away matches), the location of the pitch also varies in the visualized coordinate system. To analyse the tactical behaviour in different matches, rotating the pitch as a method of normalisation was suggested by Folgado et al. (2014). In this chapter, to make pitch length and width parallel with coordinates, clockwise or counterclockwise rotation varied for different pitches (Figure 2.3). The corresponding rotation matrix would be saved and then applied to the player's position data in the following processing step (Step IV). Steps of calculating rotation matrix are as follows: (1) Set an origin for pitch coordinate system, taking the lower left vertex of pitch as an example; (2) Determine the other vertex which also should be on the pitch length that should be parallel with the x-axis after rotation; (3) Determine the angle between the pitch length and x-axis (Folgado et al., 2014); (4) Calculate the rotation matrix (RM) using the angle, as shown in Figure 2.3. These steps have been outlined in the code (*line 202 to 265*).




$$RM = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad RM' = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

**Figure 2.3** Two types of rotation and corresponding rotation matrices.

**Step IV. Apply the rotation matrix to positional data**

The pitch coordinates are ready for rotation as geographic coordinates have been projected to Cartesian coordinates in Step II. Accordingly, positions of players moving on the pitch were then rotated with the same rotation matrix (Folgado et al., 2014). Player position at each timestamp is represented as a one-dimensional array of x-coordinate and y-coordinate (Equation 1). The rotation matrix is a two-dimensional array (Equation 2). A dot product of two arrays is the result of the rotated position in which the first element is the processed x-coordinate and the second is the processed y-coordinate, as shown in Equation 3. This step has been outlined in the code (*line 147 to 150*).

$$Player\ position = \begin{bmatrix} x_1 & y_1 \end{bmatrix} \tag{1}$$

$$Rotation\ matrix = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \tag{2}$$

$$Rotated\ position = \begin{bmatrix} x_1 & y_1 \end{bmatrix} \cdot \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} x_1 a + x_1 c & y_1 b + y_1 d \end{bmatrix} \tag{3}$$

**Step V. Merge into team positional dataset**

After applying the rotation matrix to each player's positional data, the rotated Cartesian coordinates of each player are ready to be merged into a team positional dataset for the purpose of team tactical analysis. In practice, while players in a team should have the same start and end timestamp in a session and therefore also a similar data volume, there may have different numbers of rows (data points) in their respective dataset due to signal instability. During merging, "timestamp" was set as the column to full outer join on, preserving union of keys from all frames (Figure 2.4a). This enables all original data points are kept in the merged dataset. The merging step can also be considered as the synchronisation of players' data. This has been outlined in the code *(line 166 to 170)*.

**Step VI. Merge with created timeline**

In the merged team positional dataset, missing data points (i.e., NaN or null value) could be found within several rows (Figure 2.4b). This is due to these timestamps existing in some players' positional data but not in that of other players. Data loss could also be easily identified by the difference between the total rows in the dataset and the number of rows that it should contain. For example, in the scenario that positional data is collected using 10-Hz GPS devices for 60 seconds, there are supposed to be 600 rows in the dataset (i.e., 60 secs x 10 Hz = 600 data points). However, there were less than 600 rows in the exemplar team dataset (Figure 2.4b) even with the merging of last step. This suggested that data loss was present. Therefore, a newly created timeline based on the start and end timestamps is required to ensure to work with a dataset of known sampling frequency. After merging the timeline with the team positional dataset, a dataset that contained all timestamps (rows) was then presented (Figure 2.4c). Creating and merging steps have been outlined in *line 294 to 298*, and *343 to 344* of the code, respectively.

**Step VII. Interpolation and filtering**

As a result of the previous step, more null values might have emerged in the dataset. There were rows where null values exist in a part of the columns, which indicated that those timestamps were previously lost in those corresponding players' data (i.e., partial data loss). In addition, they also existed in all columns at some rows (i.e., timestamps), which indicated that those timestamps were previously lost in data of all players (i.e., complete data loss).

In the exemplar data for this chapter, approximately 40% of all rows contained at least a null value for one or more players (i.e., partial data loss). A maximum of five continuous rows with null values occurred in the dataset as partial data loss. That corresponds to a consecutive data loss of 0.5-second period for this player. Furthermore, 13.6% of timestamps were lost simultaneously for the whole team (i.e., complete data loss). No continuous missing timestamps (complete data loss) were found in the exemplar dataset. Although data loss should be remained to a minimum, this will not hinder further data processing and the data analysis. Data can be interpolated as a solution. The code for checking data loss has been outlined in *line 300 to 340*.

Mathematical interpolation is a type of estimation that can be used to construct and fill in the null value based on those known data points (Steffensen, 2006). The missing x-coordinates and y-coordinates in this analysed dataset can be interpolated respectively. Linear interpolation is for data points in one spatial dimension, such as x-axis and y-axis. Therefore, a Python routine of linear interpolation was used in this chapter (Figure 2.4d). For example, to estimate these n-1 continuous missing data points in Equation 4, the $N_{th}$ missing data point $X_N$ can be retrieved by Equation 5 and then filled into the data sequence. Accordingly, the fewer continuous missing data points there are, the more reliable the interpolation is.

$$Dataset = \{x_1, NaN, \ldots, NaN, x_n\} \quad (4)$$

$$x_N = x_1 + \frac{x_n - x_1}{n-1} \times N \quad (5)$$

The accuracy of the positional data from the GNSS tracking technology is susceptible to external factors. To increase data accuracy, Folgado et al. (2014) used a Butterworth low-pass filter to smooth positional data. A study on GNSS positional data by Sandru et al. (2016) suggested that a combination of Butterworth and Kalman filter showed an increased performance on reducing position error. In addition, Savitzky-Golay filter is a type of data smoothing that is easy to apply and has been utilised for football tactical analysis in the practical world (Shaw, 2021). In this chapter, a python routine of Savitzky-Golay filter created by The SciPy community (2015) was applied for data smoothing. Interpolation and filtering steps have been outlined in the code (*line 346 to 351*).

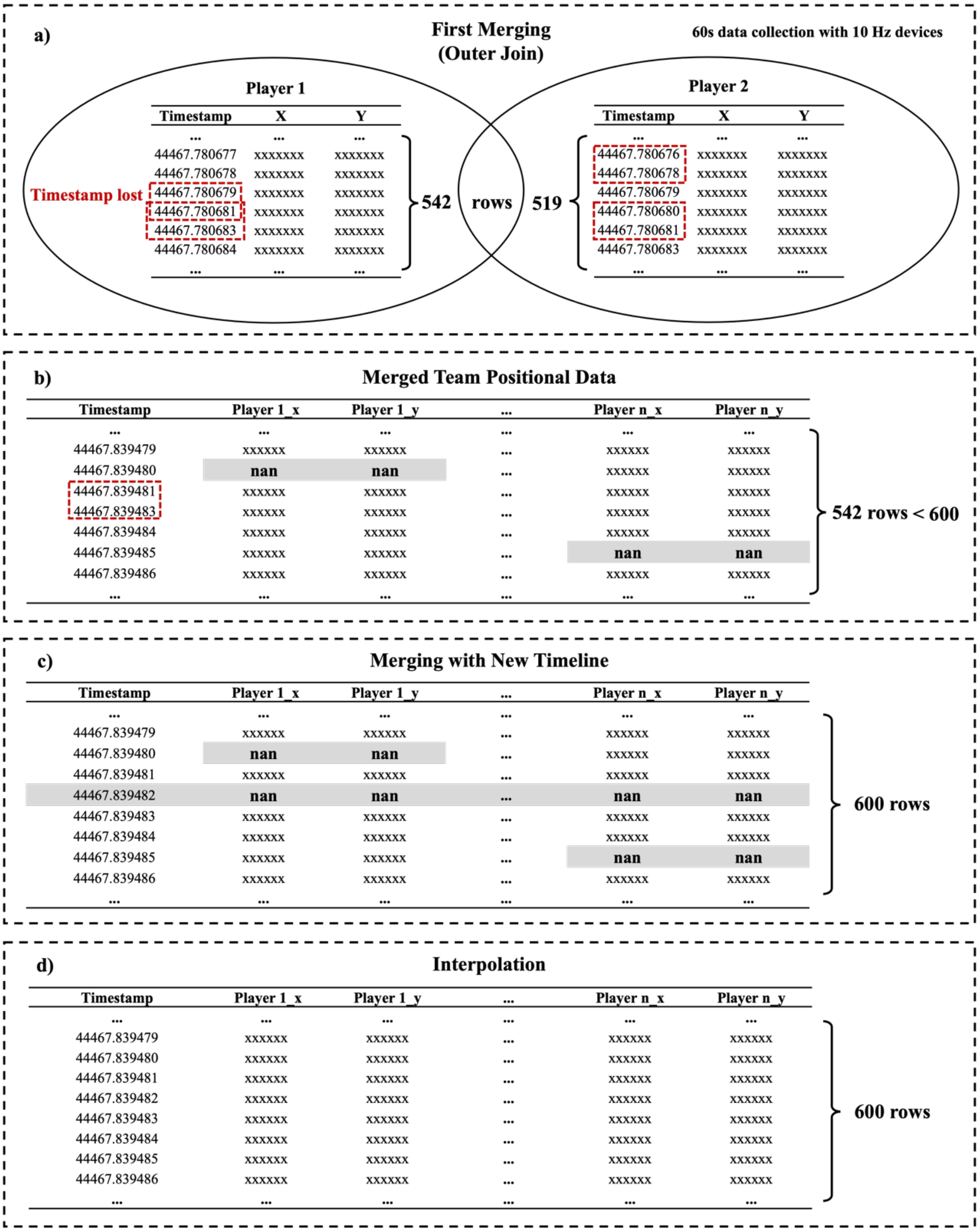



**Figure 2.4.** Details of the data processing steps from player positional data to a team dataset.

### 2.2.5 Basic team tactical measures

Raw GNSS positional data are now processed to accessible data for tactical analysis. Combining the knowledge of geometry, proposed tactical measures (Figure 2.5) are then easily to be retrieved. Calculation for these tactical measures have been outlined in previous literature (Folgado et al., 2012; Memmert et al., 2016). They include, but are not limited, to the following metrics. Surface area refers to the area covered by outfield players (Figure 2.5a). Team length is defined as the distance between the most backward and the most forward player (attacking direction). Width is defined as the distance between the most lateral players on either side of the pitch (Figure 2.5b). The centroid of the team is calculated as the mean position ($\bar{x}, \bar{y}$) of all outfield players (Figure 2.5c). Space control defined by Voronoi cell describes players' and teams' dominant regions (Figure 2.5d). These measures are extensively used to quantify collective movement patterns.

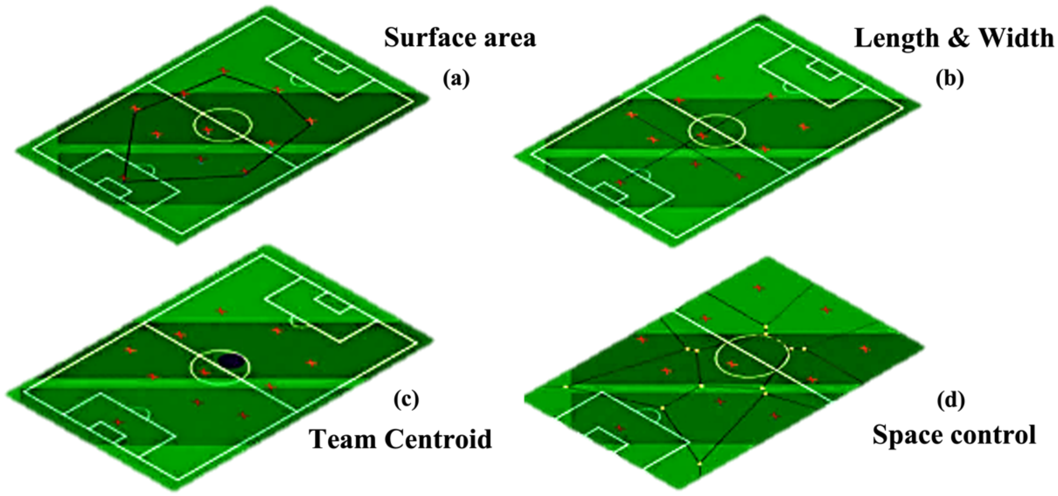



**Figure 2.5.** Example of basic team tactical measures.

## 2.3 Discussion

GNSS tracking systems capture player's position information as geographic coordinates which cannot be directly used for tactical analysis. The gap between the raw GNSS positional data and tactical analysis may limit the number of people working in this area. This chapter presented a workflow to use raw GNSS positional data for team tactical analysis in football. In comparison with the methodology briefed by Folgado et al. (2014), exemplar data, data processing steps, and potential issues and corresponding solutions which were overlooked in previous studies, were presented and detailed in this chapter. A referential customised Python routine was also attached. It is also recommended to have some basic understanding and skills of data management systems, to adjust the presented code to the user's conditions. This chapter as a toolbox is expected to facilitate researchers and practitioners of football teams using GNSS tracking systems to analyse team tactical behaviour. As a result, player tracking technology systems are now useful for both physical and tactical purpose. Software packages from most GNSS tracking systems usually facilitate the physical monitoring by calculating several physical variables, but not tactical variables. It therefore has limited the tactical analysis with these systems. Additionally, partial or complete data loss may be present due to data instability caused by external factors, such as atmospheric condition (Shergill et al., 2021). The workflow in this chapter described the solution to the most common issues that are likely to arise in data collection and processing in football practice: 1) the optimal way for retrieving coordinates of pitch location, 2) synchronizing the starting and end timestamp of football activities with raw positional data, and 3) addressing data loss in GNSS positional data.

Pitch location coordinates are an indispensable ingredient for the tactical analysis using GNSS positional data, especially for rotation matrix (Folgado et al., 2014). But the approach to retrieving coordinates of pitch location has never been reported in previous studies. This chapter proposed and compared two pragmatic methods to retrieve geographic coordinates of pitch corners (details can be found in the Appendix). Web mapping platform features low time cost, high consistency, and accessibility. Results of intra-observer and inter-observer analysis in the Appendix highlighted this method as the most reliable way to retrieve those coordinates. However, the satellite imagery of a pitch that generates on mapping platform might not be latest updated, which implies it is possible to collect incorrect coordinates if there was a renovation of the pitch or stadium recently. In addition, because most high-resolution imagery on the web mapping platform is aerial photography taken by flying aircrafts, a part of the pitch may be invisible if the stadium roof or high building surrounds the pitch. To remedy the situation, the visible information needs to be fully exploited to determine at least a corner, pitch size and the most importantly rotation matrix. Alternative to this method, the GNSS tracking device approach is also applicable, while this requires longer collection time and extra data processing steps to determine coordinates of a pitch.

After rotating the player's coordinates, positional data of each player were merged into a team dataset. Due to internal issues of tracking devices and external conditions (e.g., weather), that dataset contained missing data points and timestamps. In other words, players trajectories were missing within this period, affecting following calculation of tactical measures. A customised timeline based on start and end timestamps was created to fill in the missing timestamps and to ensure working with data of a consistent sampling frequency (i.e., 10 Hz). This chapter used a dataset from a training session, with approximately 40% of partial data loss and 13.5% of complete data loss respectively. More than 95% of data loss were two consecutively missing data points. Previous studies compared data validity of different GNSS tracking systems (Jackson et al., 2018), but they did not report data loss or acceptable levels of missing data. In practice, the maximum consecutive data loss was 5 samples (or 0.5 seconds). Although data loss should be kept to a minimum, this duration is acceptable to be interpolated. Players will move around on the pitch in this time window, but their movement trajectories can still be tracked by interpolation. Whilst the issue of data loss has been solved in the latest version of GNSS tracking device (Vector S7, Catapult Innovations, South Melbourne, VIC, Australia), examining and reporting data quality are supposed to become a routine in the study of this area. In future studies, to what extent and what type (i.e., partial and complete) of data loss is acceptable should be discussed to reach a consensus, whereby researchers can maintain standards of original data and reliability of study result.

Future tactical analysis could combine positional data with match event data, explore interactions between opposing teams, and involve more contextual information. However, GNSS tracking systems compromise ball tracking and the access to opponent data, which accordingly constrain the analysis with respect to these aspects. Combining those tactical variables with match video can be considered as a way of compensation. For example, video data and positional data can be combined and employed to explore tactical behaviour in different match situations (e.g., phases and conditions). Driven by varying objectives along the match, teams tend to adopt different tactics in phases of play, as a response to actions of the opposition. Analysis on team tactical behaviour in these match phases can provide insights into tactics execution, benefiting the future competition.

## 2.4 Conclusion

To conclude, this chapter presented a workflow to prepare raw GNSS positional data for tactical analysis in football. Detailed processing steps were presented and elaborated to use positional data from GNSS player tracking technology for team tactical analysis. The gap between raw GNSS positional data and tactical analysis is bridged. This workflow facilitates practitioners and scientists to analyse team tactical behaviour using common player tracking technology systems. GNSS tracking technology can be now fully used for tactical and physical insights.

## References


Akenhead, R. & Nassis, G.P. 2016. Training Load and Player Monitoring in High-Level Football: Current Practice and Perceptions. *International Journal of Sports Physiology and Performance,* 11**,** 587-93.

Bourdon, P.C., Cardinale, M., Murray, A., Gastin, P., Kellmann, M., Varley, M.C., Gabbett, T.J., Coutts, A.J., Burgess, D.J., Gregson, W. & Cable, N.T. 2017. Monitoring Athlete Training Loads: Consensus Statement. *International Journal of Sports Physiology and Performance,* 12**,** S2161-S2170.

Brink, M.S. & Frencken, W.G.P. 2018. Formative feedback for the coach reduces mismatch between coach and players' perceptions of exertion. *Science and Medicine in Football,* 2**,** 255-260.

Castellano, J., Fernández, E., Echeazarra, I., Barreira, D. & Garganta, J. 2017. Influence of pitch length on inter- and intra-team behaviors in youth soccer. *Anales de Psicología,* 33.

Castellano, J., Silva, P., Usabiaga, O. & Barreira, D. 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**,** 153-163.

Folgado, H., Duarte, R., Fernandes, O. & Sampaio, J. 2014. Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches. *PLoS One,* 9**,** e97145.

Folgado, H., Goncalves, B. & Sampaio, J. 2018. Positional synchronization affects physical and physiological responses to preseason in professional football (soccer). *Research in Sports Medicine,* 26**,** 51-63.

Folgado, H., Lemmink, K.A.P.M., Frencken, W. & Sampaio, J. 2012. Length, width and centroid distance as measures of teams tactical performance in youth football. *European Journal of Sport Science,* 14**,** S487-S492.

Fonseca, S., Milho, J., Travassos, B. & Araujo, D. 2012. Spatial dynamics of team sports exposed by Voronoi diagrams. *Human Movement Science,* 31**,** 1652-9.

Frencken, W., Lemmink, K., Delleman, N. & Visscher, C. 2011. Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science,* 11**,** 215-223.

Frencken, W., Poel, H.d., Visscher, C. & Lemmink, K. 2012. Variability of inter-team distances associated with match events in elite-standard soccer. *Journal of Sports Sciences,* 30**,** 1207-1213.

Goes, F.R., Meerhoff, L.A., Bueno, M.J.O., Rodrigues, D.M., Moura, F.A., Brink, M.S., Elferink-Gemser, M.T., Knobbe, A.J., Cunha, S.A., Torres, R.S. & Lemmink, K.A.P.M. 2020. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science,* 21**,** 481-496.

IFAB 2016. Laws of the Game 2016/17. *The Player's Equipment.* Switzerland: The International Football Association Board.

Jackson, B.M., Polglaze, T., Dawson, B., King, T. & Peeling, P. 2018. Comparing Global Positioning System and Global Navigation Satellite System Measures of Team-Sport Movements. *Int J Sports Physiol Perform,* 13**,** 1005-1010.

Karney, C.F. 2011. Transverse Mercator with an accuracy of a few nanometers. *Journal of Geodesy,* 85**,** 475-485.

Low, B., Coutinho, D., Goncalves, B., Rein, R., Memmert, D. & Sampaio, J. 2020. A Systematic Review of Collective Tactical Behaviours in Football Using Positional Data. *Sports Med,* 50**,** 343-385.

Maling, D.H. 1992. CHAPTER 4 - The geometry of the spheroid. *Coordinate Systems and Map Projections.* New York.

Memmert, D., Lemmink, K.A.P.M. & Sampaio, J. 2016. Current Approaches to Tactical Performance Analyses in Soccer Using Position Data. *Sports Medicine,* 47**,** 1-10.

Olthof, S.B., Frencken, W.G. & Lemmink, K.A. 2015. The older, the wider: On-field tactical behavior of elite-standard youth soccer players in small-sided games. *Human Movement Science,* 41**,** 92-102.

Olthof, S.B.H., Frencken, W.G.P. & Lemmink, K. 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**,** 523-530.

Rago, V., Brito, J., Figueiredo, P., Costa, J., Barreira, D., Krustrup, P. & Rebelo, A. 2020. Methods to collect and interpret external training load using microtechnology incorporating GPS in professional football: a systematic review. *Research in Sports Medicine,* 28**,** 437-458.

Rod Deakin, M. N. Hunter & Karney, C. 2010. THE GAUSS–KRÜGER PROJECTION. *Victorian Regional Survey Conference.* Warrnambool.

Sampaio, J. & Macas, V. 2012. Measuring tactical behaviour in football. *International Journal of Sports Medicine,* 33**,** 395-401.

Sandru, F.D., Nanu, S., Silea, I. & Miclea, R.C. Kalman and Butterworth filtering for GNSS/INS data. 12th IEEE International Symposium on Electronics and Telecommunications (ISETC), 27-28 Oct. 2016. 257-260.

Shaw, L. 2021. *LaurieOnTracking* [Online]. Available: https://github.com/Friends-of-Tracking-Data-FoTD/LaurieOnTracking [Accessed 01 May 2022].

Shergill, A.S., Twist, C. & Highton, J. 2021. Importance of GNSS data quality assessment with novel control criteria in professional soccer match-play. *International Journal of Performance Analysis in Sport,* 21**,** 820-830.

Silva, P., Duarte, R., Sampaio, J., Aguiar, P., Davids, K., Araujo, D. & Garganta, J. 2014. Field dimension and skill level constrain team tactical behaviours in small-sided and conditioned games in football. *Journal of Sports Sciences,* 32**,** 1888-1896.

Silva, P., Santos, E.D., Grishin, M. & Rocha, J.M. 2018. Validity of Heart Rate-Based Indices to Measure Training Load and Intensity in Elite Football Players. *Journal of Strength and Conditioning Research,* 32**,** 2340-2347.

Steffensen, J.F. 2006. *Interpolation*, Mineola, N.Y.

TheScipycommunity. 2015. *Signal processing (scipy.signal)* [Online]. The Scipy community. Available: https://docs.scipy.org/doc/scipy-

0.15.1/reference/generated/scipy.signal.savgol_filter.html [Accessed 04.08 2022].

Torres-Ronda, L., Clubb, J. & Beanland, E. 2022. Tracking Systems in Team Sports: Back to Basics. *Sport Performance & Science Reports*.

Weisstein, E.W. 2009. Lambert Conformal Conic Projection. MathWorld--A Wolfram Web Resource.

Welch, M., Schaerf, T.M. & Murphy, A. 2021. Collective states and their transitions in football. *PLoS One,* 16**,** e0251970.

# Chapter 3: Team Tactical Behaviour in Different Phases of Football Match

## 3.1 Introduction

Football as a team sport is a complex system (Duarte et al., 2013; Seifert et al., 2017). Players interact with other components during a game, including teammates, opponents, the ball, coaches, and even the referee and audience. To improve players' performance and coaches' insights into competition, physical, technical, mental, and tactical aspects of performance are analysed (Carling et al., 2005). Modern player tracking technologies have also developed and benefited the quantification of physical and team tactical performance (Buchheit et al., 2014; Torres-Ronda et al., 2022). Positional data from player tracking systems have been extensively utilised to capture and quantify team tactical behaviour (Folgado et al., 2012; Sampaio and Macas, 2012; Olthof et al., 2015).

As aforementioned (chapter 1 and 2), team tactical behaviour is described as the positioning, interaction, and collective movement of a team. In a football match, interactions occur within a team (coordination), between opposing teams (competition), and between player and environment (Gréhaigne et al., 2004; Duarte et al., 2012; Vilar et al., 2012; Araújo et al., 2015). These intra-team and inter-team moving patterns have been quantified as tactical measures (e.g., team centroids, surface area, stretch indices and inter-team distance) (Memmert and Rein, 2018; Goes et al., 2020b). These tactical measures provide information about the central position of the team, the dispersion of players, and how players collectively act in response to the opposition. To illustrate, players move collectively towards the goal of the opposition during attack (Olthof et al., 2019b). However, the speed and direction of movement vary across phases of match and depend on the action of the opposition. Low et al. (2021) compared the dispersion of the defending team and inter-team distance between opposing teams in different defending strategies. Clemente et al. (2013) demonstrated that without ball possession teams tended to be closer to the defensive zone and to reduce dispersion, compared to with ball possession.

Match phases are related to ball possession that repeatedly switches between opposing teams, cycling throughout the match (i.e., attack, defence, transition from attack to defence, and transition from defence to attack). Specific objectives within each phase lead to different tactical plans and executions of a team in the match. Both in attack and defence, a paramount tactical principle is to create numerical superiority around the ball (Silva et al., 2014b). During attack, teams control ball possession and collectively move the ball towards the opposition's goal. To achieve this, the player with the ball attempts to penetrate opposing defence by incisive pass or dribbling. Offensive support comes from teammates with specific missions in different positions. Movement of the striker can lure the opposing defensive line into higher position and therefore create an opening behind the defensive line. Movement to the side line can extend the effective playing space. Moving forwards of defenders also benefits the offensive organization by offering passing options to the player with the ball, which in the meantime narrows the gap between defenders and midfielders. Coordinated movement of goalkeepers in attack is also emphasized in modern football competition to ensure the offensive unity (Teoldo da Costa et al., 2010). Although players in different positions have their own responsibility in attack, the team moving forwards collectively as a unit with the same goal – creating scoring opportunities.

During defence, teams aim to slow down and intercept the opposing team's attempts to move forward with the ball. Compactness (i.e., keeping the close distance between team members) of defensive formation restricts the space that the opposition can utilise for offensive organisation. Pressing as a way of building defence is described as the collective behaviour of a team to win back ball possession from the opponents (Low et al., 2021). Defending tactics is adopted according to the attacking style of the opposition, physical condition of the team, and match status (i.e., goal difference). The ball location also influences the tactical action of the defending team. High pressing at the attacking third requires intensively coordinated collective actions of the team. Few players delay the ball progress by approaching the attacker with the ball. Positioning of other off-ball defenders is also critical to provide defensive support by occupying vital spaces and closing passing opportunity. High pressing can also be considered as delaying attack for defensive organisation. Mid-block defence at middle third and low-block defence at defensive third requires high concentration and unity of collective movement against the opponent attack. Similar to the offensive unity, players in different positions exercise respective responsibility to contribute into defensive unity, working as a team.

In addition, transitions in between include the attack-to-defence transition and defence-to-attack transition. Whenever the team regains or loses ball possession, they require a short period to reshape and reorganise positioning for attack and defence. Accordingly, team tactical behaviour in transitions could differentiate from the actions while the team (or the opposition) controls ball possession. These moments of the game are extremely important as they are executed at a higher speed within a shorter period than other match phases (Barbu and Stoica, 2020). Transition from defence to attack starts as the team regains ball possession and is also termed as counterattack (Olsen and Larsen, 1997; Gonzalez-Rodenas et al., 2015). Players tend to spread and utilise the momentary opportunities to break through defence before the defence reorganization. On the other hand, as the opposition regains ball possession, the team transits from attack to defence. To delay the opponent's counterattack, aggressive (e.g., immediate counter pressure) and cautious (e.g., sitting off and retaining compactness) tactics can be adopted by different teams against different oppositions. The transition phase is generally shorter than phases of attack and defence, players are required to make decision and react within a short period. This may lead to a temporary chaos of team tactical behaviour and influence the team unity.

In summary, collective moving patterns vary across these phases because of different tactical principles in attack, defence, and transitions (i.e., from attack to defence, and vice versa). In offence, it is essential to create space and advance to the opposition's goal, which is reflected by a bigger surface area and average interpersonal distance. In contrast, in defence to protect space and retain compactness, the team formation tends to present a smaller surface area and team dispersion. Driven by these differences, previous studies explored team tactical behaviour in offensive and defensive phases. Clemente et al. (2013) compared team tactical movement in attack and defence. They confirmed that the teams made use of the width of the pitch, and moving as a unit in both defence and attack. Moura et al. (2012) found that in defence teams spread a greater area when suffering shots than when performing tackles. Welch et al. (2021) compared team tactical behaviour in phases of attack, defence and out of play. Their findings indicated that the collective movement was more compact and faster moving in defence than attack. Goes et al. (2020a) found lower a longitudinal inter- and intra-team synchrony of team interactions during successful attacks than unsuccessful attacks. Team tactical behaviour in different phases of the official match demonstrated a greater spread in tactical positioning characterises in offence than defence (Praça et al., 2022). These previous studies confirmed the difference of team tactical behaviour across match phases and the value to analyse tactical behaviour at the phase level rather than the full-match level.

Those studies based on tactical measures from positional data explored team tactical behaviour in offensive and defensive phases. However, they did not consider the difference of team tactical behaviour between controlling ball possession and the phase of transition. Therefore, the main objective of this chapter is to explore and compare the tactical behaviour in four basic match phases: attack, defence, transition from attack to defence, and transition from defence to attack. Besides, this chapter also aims to provide a proof of concept for the workflow (proposed in chapter 2) being able to be applied in real tactical analysis. It is hypothesised that team tactical behaviours differ across match phases and scientifically demonstrate common tactical principles.

## 3.2 Methods

### 3.2.1 Players

The dataset contained the positional data of 13 professional football players during one competitive match, including 10 starting outfield player and 3 substitutes (mean$\pm$*SD*: age = 26.3$\pm$2.4 years; professional playing experience = 4.7$\pm$1.5 years). The goalkeeper was excluded from the analysis. All players are from the first team competing in the English Championship during the 2020/2021 season. De-identified data from all players were compiled into a data repository, and the Research Ethics Committee at Liverpool John Moores University approved secondary data analyses.

### 3.2.2 Data collection

**Table 3.1** Definition of match phases, home team (HT) is considered as the analysed team.

| | Phases | Definition | Sub-phases | Definition |
|---|---|---|---|---|
| **With the ball** | **In Possession** | HT controls ball possession, originating from transitions or restarts. | vs High Press | OT are pressing to influence play in OT Attacking Third. |
| | | | vs Mid Block | OT defensive line above OT Defensive Third & most OT players in Middle Third, defending actively or passively. |
| | | | vs Low Block | HT controls possession in HT Attacking Third. Most OT players in OT Defensive Third. |
| | | | IP Transition | Originating from HT possession, play fast progresses through at least two thirds in few passes. Accordingly, OT defensive line has to drop quickly (in front or behind the ball). |
| | **DA Transition** | HT regains ball possession. | vs Counter press | HT are attempting to consolidate possession against counterpressure. |
| | | | Counter Attack | HT are attempting to progress towards the opponent goal in a quick and incisive manner. |
| **Without the ball** | **Out of Possession** | OT controls ball possession, originating from transitions or restarts. | High Press | HT are pressing to influence play in HT Attacking Third. |
| | | | Mid Block | HT defensive line above HT Defensive Third & most HT players in Middle Third, defending actively or passively. |
| | | | Low Block | OT controls possession in OT Attacking Third. Most HT players in HT Defensive Third. |
| | | | OPP Transition | Originating from OT possession, play fast progresses through at least two thirds in few passes. Accordingly, HT defensive line has to drop quickly (in front or behind the ball). |
| | **AD Transition** | OT regains ball possession. | Counter press | HT are applying either indirect or direct counter pressure immediately or soon after ball turnover. |
| | | | Defensive Recovery | In response to an opposition counter attack, HT are quickly moving towards their own goal. |
| **Others** | **IP Unstructured** | Originating from HT action, neither team has controlled possession with multiple ball turnovers in short period of time, and little attempt to consolidate possession. | \ | \ |
| | **OPP Unstructured** | Originating from OT action, neither team has controlled possession with multiple ball turnovers in short period of time, and little attempt to consolidate possession. | \ | \ |
| | **Set Piece A** | HT takes a set piece. | \ | \ |
| | **Set Piece D** | HT defends a set piece. | \ | \ |

*HT* home team, *OT* opposition team, *DA Transition* defending to attacking transition, *AD Transition* attacking to defending transition

Positional and video data from one English professional league match were analysed. The sampling rate of video data is 59.95 fps. The players' positional data were collected using Vector S7 tracking devices (Catapult Innovations, South Melbourne, VIC, Australia) at a sampling rate of 10 Hz. The reliability of the current device has been previously tested (Jackson et al., 2018). The analysed team played at home for 97 minutes including stoppage time (first half = 46 minutes; second half = 51 minutes). According to the league table prior to the match, the opposition was placed behind the analysed team.

Video footage was recorded from the top level of the stadium stand to have the full view over the full pitch. This recording was then used to annotate the possession of the ball to provide insights of match phases (i.e., in possession, out of possession, attacking-to-defending transition, and defending-to-attacking transition). This was achieved by professional analysts using video notation software (Hudl Sportscode, USA) to record the time point of each change in match phases, according to the definition of match phases (Table 3.1).

Some match phases consist of several subphases. To illustrate, when the team regains ball possession, the defence-to-attack transition (DA transition) starts. The first attempt is to consolidate possession against counterpressure (vs Counter press) and then move the ball fast and incisively forwards (Counter Attack) until a foul is given (Ball out of play). The DA Transition here comprises "vs Counter press" and "Counter Attack", ending up with "Ball out of play". In this chapter, four basic match phases, i.e., *in possession* (IP), *out of possession* (OOP), *defence-to-attack transition* (DAT), and *attack-to-defence transition* (ADT) were taken into account. *Ball Out of Play* (BOOP) is naturally a part of football match but not regarded as effective playing time, therefore not included in the definition and the analysis.

### 3.2.3 Data processing

Data processing comprises the following steps: 1) processing raw positional data (i.e., latitude and longitude coordinates) of each player into a team dataset (Cartesian coordinates); 2) synchronising positional data with event data; and 3) calculating tactical measures for the match phases.

#### From raw data to processed team data

The workflow of preparing raw positional data for tactical analysis has been detailed in chapter 2 and applied in this chapter. No data loss was found in the team dataset.

#### Synchronisation

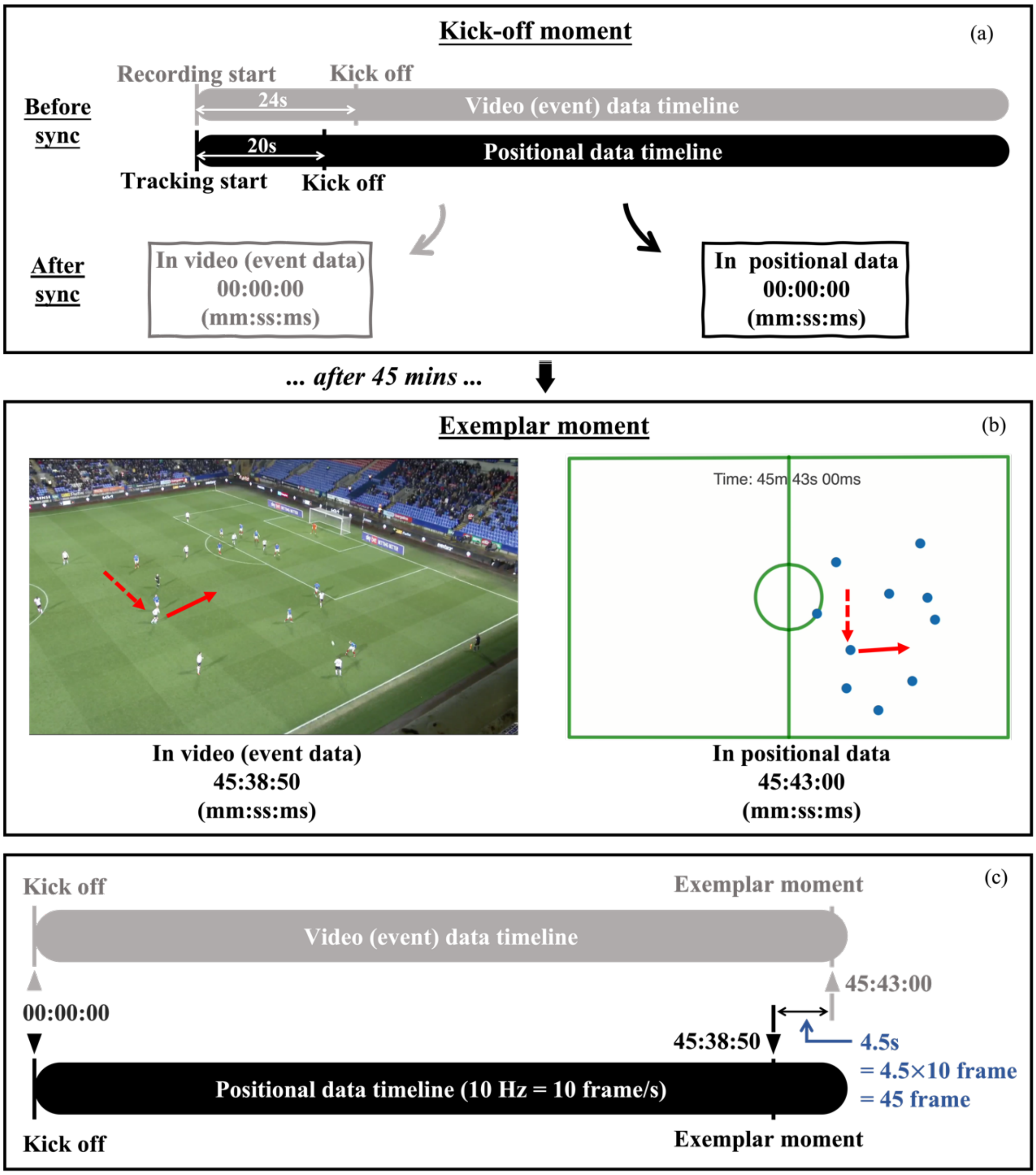
To capture collective movement in different match phases, it requires positional and phase information from positional data and event data, respectively. Synchronisation of two types of data is also indispensable to ensure that their timelines are aligned without a time lag. Otherwise, players' positioning and movement in an attacking moment might be considered and analysed as defensive behaviour due to a time lag. Consequently, the analysis of tactical behaviour in different phases will be meaningless.

When synchronising these two types of data, this leads to two potential issues: 1) human errors on event data timestamps because of manual notation; 2) systematic offsets caused by two systems (i.e., the video footage and positional data) having their own clock (Anzer and Bauer, 2021). The first type of time lag may occur between video footage and event data. Event data are manually notated based on watching the video footage and using computerised software by analysts. Consequently, the timestamp tagged by analysts may differ from the timestamp in video footage. In the current study, event timestamps were verified with the video footage and both timelines were aligned. This means that the manual notational errors did not occur in the dataset.

However, systematic offsets between event data and tracking data occurred in the dataset. Tracking start and end time in the positional data did not align with kick-off and end-whistle time in the event data, respectively. This was identified by visual inspection of video and positional data and is a common error when positional and event data are analysed (Olthof et al., 2019a; Anzer and Bauer, 2021). Player movements in both the video footage and animations generated from the positional data were observed and compared. An example of this process was outlined in Figure 3.1. In the current datasets, the recording of video footage and player tracking started earlier than the kick-off. In the half time, video recording was paused but the player tracking system was not paused. Therefore, the process of identifying correct start and end timestamps and synchronisation for the positional data with the event data (video footage) was achieved for each half independently. QuickTime Player (Apple Inc., version 10.5) was utilised to play video in millisecond and correct for inconsistencies in the synchronisation.

The criterion for resolving systematic offsets is that the start, end, and each timestamp in between are synchronised between the positional data and event data. The pitch (location coordinates) and timestamp information from positional data were also plotted in animations to help the visual inspection of the synchronisation (Figure 3.1b). First, by visual inspection, the identified start timestamps (first and second half) from the positional data were set as 0 second. All tracking timestamps were shifted by the time difference (i.e., the identified start timestamp minus 0). With respect to the event data, the kick-off timestamps were set into the 0 second. Same shifting method as for the positional data was applied for the event data. As a result, the kick-off time in the event data was aligned with the start time in positional data (Figure 3.1a-c).

Second, to retrieve end timestamps (first and second half) in positional data, exemplar moments with player drastic movement (e.g., changing of direction) in vicinity of the half end were selected and compared by visual inspection (Figure 3.1b). It was expected to ensure that end timestamps in two types of data corresponded with each other. However, there was a lag between the end timestamp in positional data and the end timestamp in the event data (Figure 3.1c). The lags were detected in both two halves and gradually accumulated along the match playing. In other words, the lags occurred for each timestamp between the start and end of the half and became greater towards the end of the half. A solution and equations to this lag are elaborated in Figure 3.1d. Because the same issue occurred in both halves, all timestamps in positional data were shifted by the equation 5 in Figure 3.1d. By selecting several moments along the match playing (i.e., at 15, 30, 45, 60, 75, 90 minute) to verify synchronisation, the timeline in positional data was aligned with that in event data along the entire match. Systematic offsets in the dataset were resolved (Olthof et al., 2019a; Anzer and Bauer, 2021).




(d)

(1) $V_1$ = playing speed of video footage

(2) $V_2$ = sampling frequency of positional data $= 10\ \mathrm{Hz} = 10\ \mathrm{frames}/_{s}$

(3) $t = 45 \times 60 + 43 = 2743\ s$

(4) $V_1 \cdot t = V_2 \cdot t + 45$ frame

$V_1 = 10.016\ \mathrm{frames}/_{s}$

*To modify the positional data timeline:*

$$(5)\quad Timestamp\ (i) = \frac{timestamp\ (i)}{V_2/V_1} = \frac{timestamp\ (i)}{10/10.016}$$

**Figure 3.1** After determining and synchronising (a) start timestamps, (b) exemplar moments were selected to compare end timestamps in two types of data. (c) The same lag was detected in both halves. It is hypothesised that this issue is caused by different playing speeds of video footage and positional data animation (10 frames per second, based on the timeline of positional data). (d) Equations were applied, and the result proved two timelines and synchronised.

The time of each substitution was determined and recorded based on video footage. Then positional data of substitutes were inserted into the team positional dataset according to the time points of match resuming after the substitution being made. Thus far, the positional data were prepared to calculate team tactical measures.

**Team tactical measures**

The processed team positional dataset was used to calculate the following tactical behaviour measures: centroid, length, width, length per width (LpW) ratio, surface area, stretch indices (lateral and longitudinal), and interpersonal distance (ID) of team members. The equations for length, width, LpW ratio, and ID are shown in Figure 3.2. Mean lateral and longitudinal position of all outfield players was calculated as team centroid. Stretch indices (longitudinal and lateral) were determined as the average distance between each player and team centroid in longitudinal and lateral direction. The surface area was calculated by the convex hull enclosed by outfield players. All measures were calculated on a team level at each timestamp. Then measures of timestamps within one phase were averaged as the measures of the phase. All data processing and calculation were conducted in a customised script in in Python 3.8.

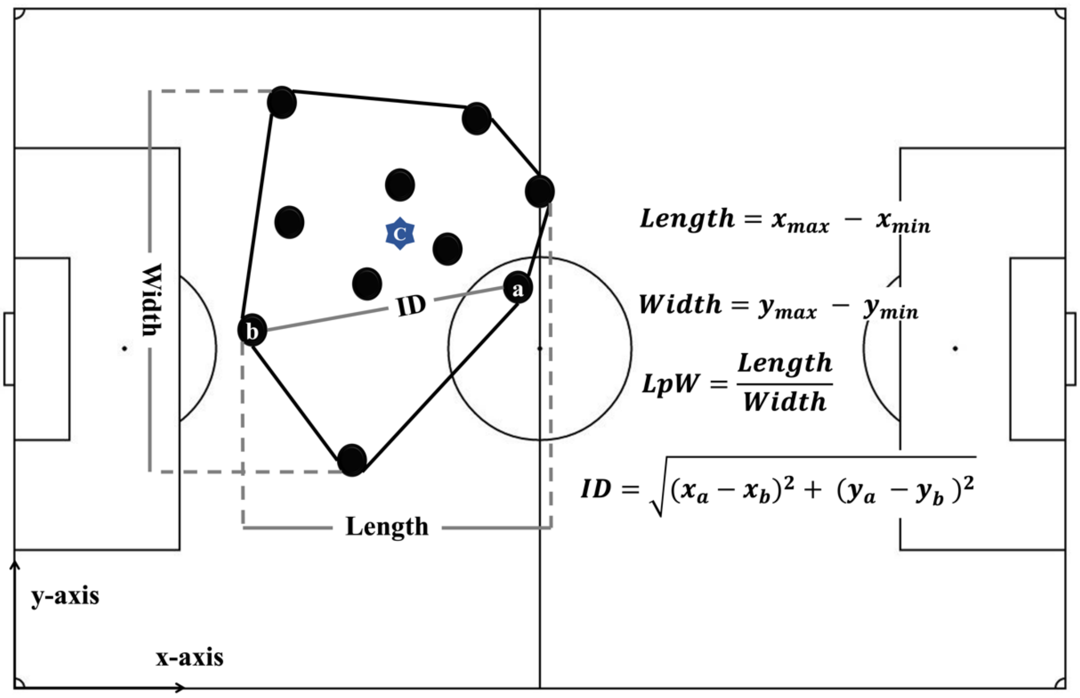


**Figure 3.2** Length, width, LpW and interpersonal distance (ID) of team formation, and corresponding equations. C is the team centroid.

### 3.2.4 Statistical analysis

Team tactical behaviour metrics were considered as dependent variables and compared according to four basic match phases (attack, defence, transitions from defence to attack, and transitions from attack to defence). A one-way ANOVA was used to compare team tactical measures in different phases. Because of unequal variance, results from Welch's test were used. Significant level was maintained at 5%. Eta squared ($\eta^2$) was calculated as effect size. For interpretation, magnitudes of effect size were considered as small ($\eta^2 < 0.06$), moderate ($0.06 \leq \eta^2 < 0.15$), or large ($\eta^2 \geq 0.15$) (Cohen, 1988). Pairwise comparison (Tamhane's post-hoc test) was conducted to determine which pairs of phases are significantly different, because of unequal variances (Field, 2013). Mean differences with 95% Confidential intervals (CI) are provided to assess to relationships between pairwise phases (Nakagawa and Cuthill, 2007). Cohen's *d* was calculated as effect size for pairwise comparison. All statistical calculation were conducted using IBM SPSS Statistics (version 26.0, IBM Corporation, Somers, New York, USA).

## 3.3 Results

Figure 3.3 provides an overview of the proportion of each phase accounting for the match. Effective playing time of 55 minutes and 57 seconds accounted for 57.7% of the total match duration. Total time spent in each phase was skewed, with the analysed team spending the most time to control ball possession (IP). Transitions accounted for approximately 30% of effective playing time, and 17.7% of match time. Two types of transitions shared similar proportion (total duration). As ball possession was constantly regained by two opposing teams, match phases might repeatedly switch between ADT and DAT, as shown in Figure 3.3. The short phases (≤3 seconds) of DAT (30.2%) and ADT (24.1%) accounted for larger proportion than IP (2.7%) and OOP (13.3%). In contrast, more long phases (≥20 seconds) were found in IP (31.5%) and OOP (22.7%) than DAT (2.3%) and ADT (3.6%).

Results from the ANOVA revealed significant differences ($p<0.001$) between phases across all tactical variables (Table 3.2). Effect sizes ($\eta2$) were large for surface area, width, lateral stretch index, as well as maximum and average interpersonal distance, were moderate for LpW ratio, longitudinal stretch index, minimum interpersonal distance, and were small for length.

The pairwise comparison revealed that all variables from IP (except for LpW ratio) showed significantly greater values than other three phases (Table 3.2 and Figure 3.4). The team played with longer and wider formation within IP than OOP (length: $p<0.001$, $d=0.78$; width: $p<0.001$, $d=1.86$), DAT (length: $p<0.01$, $d=0.54$; width: $p<0.001$,

$d$=1.52), and ADT (length: $p$<0.01, $d$=0.54; width: $p$<0.001, $d$=1.05). Furthermore, a significant lower LpW ratio was found for IP than OOP ($p$<0.001, $d$=-0.81), DAT ($p$<0.001, $d$=-0.83), ADT ($p$<0.05, $d$=-0.53). The team also played wider ($p$<0.01, $d$=0.60) within ADT than OOP.

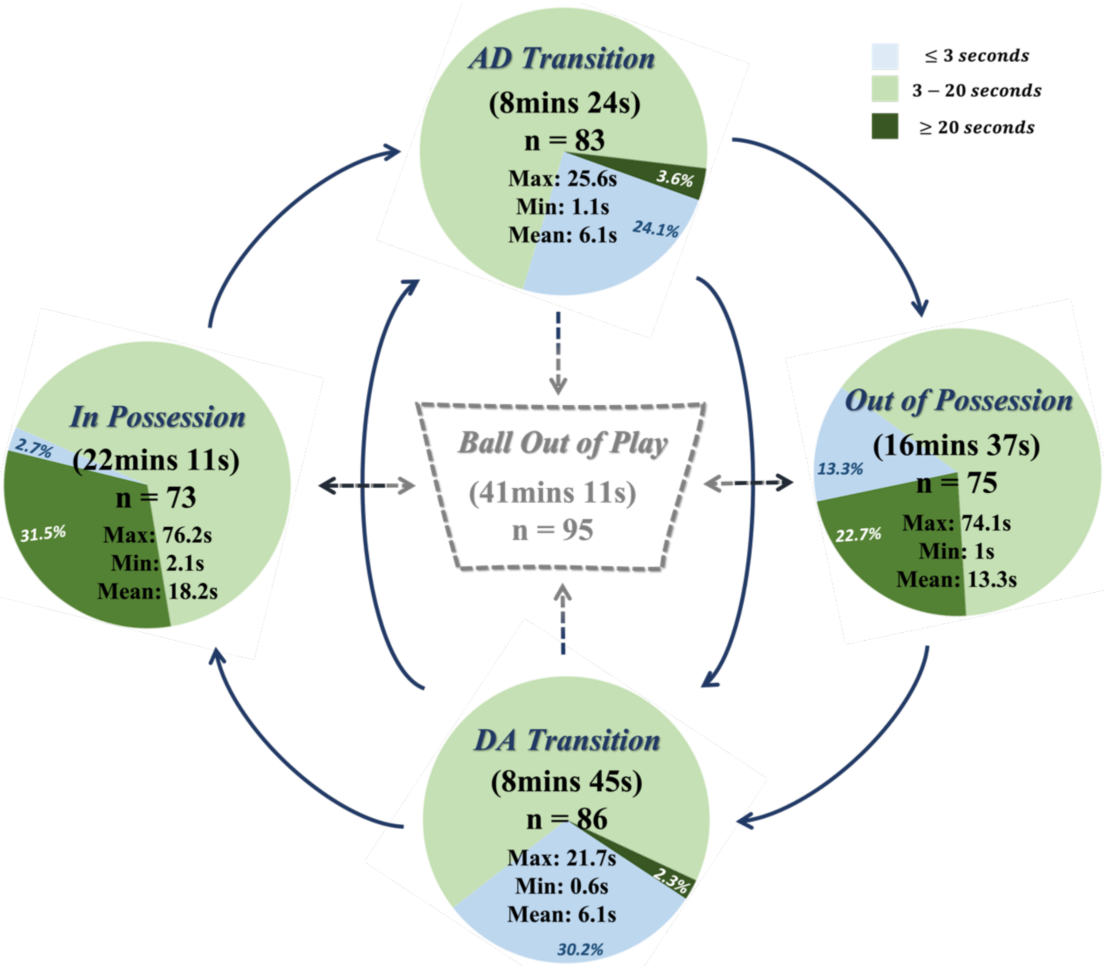



**Figure 3.3** Basic match phases, as well as corresponding counts, cumulative time, the maximum, minimum, and mean of each phase. Bar charts indicate the proportion of phases lasting less than 3 seconds and the rest. The edges indicate phase switching and the arrows indicate the direction of switching. All of four phases are likely to be followed by BOOP. But after BOOP, there only follows IP or OOP.

Pairwise analyses also revealed similar tendency in team dispersion behaviour (surface area, lateral stretch indices, maximum and average ID) between match phases. Larger area was covered in IP than OOP ($p$<0.001, $d$=1.94), DAT ($p$<0.001, $d$=1.61), ADT ($p$<0.001, $d$=1.25). In the direction of pitch width, stretch indices showed greater magnitude in IP than OOP ($p$<0.001, $d$=2.0), DAT ($p$<0.001, $d$=1.69), ADT ($p$<0.001, $d$=1.23). Longer average interpersonal distance within IP was found than OOP ($p$<0.001, $d$=1.98), DAT ($p$<0.001, $d$=1.66), ADT ($p$<0.001, $d$=1.33). In addition, players dispersed with larger surface area ($p$<0.001, $d$=0.65), lateral stretch indices ($p$<0.05, $d$=0.51), maximum ID ($p$<0.001, $d$=0.68), and average ID ($p$<0.01, $d$=0.60) within ADT than OOP. No significant difference in tactical behaviour was found either between DAT and ADT, or between DAT and OOP.

**Table 3.2** Mean and SD and statistics with the asymptotical $F$-value, $p$-value, effect size ($\eta^2$) of tactical variables.

| Tactical Variables | $^a$ In Possession | $^b$ Out of Possession | $^c$ AD Transition | $^d$ DA Transition | | | |
|---|---|---|---|---|---|---|---|
| | *Mean±SD* | | | | $F$ | $p$ | $\eta^2$ |
| Surface area ($m^2$) | 1418.0± 267.5 $^{bcd}$ | 911.0± 255.7 $^{ac}$ | 1081.1± 268.2 $^{ab}$ | 1010.5± 238.3 $^{a}$ | 39.125 | < 0.001 | 0.25 |
| Length (m) | 39.4±4.9 $^{bcd}$ | 35.2±5.8 $^{a}$ | 36.7±4.8 $^{a}$ | 36.7±5.1 $^{a}$ | 6.235 | < 0.001 | 0.04 |
| Width (m) | 52.0±7.1 $^{bcd}$ | 38.7±7.2 $^{ac}$ | 43.5±8.9 $^{ab}$ | 40.7±7.8 $^{a}$ | 37.33 | < 0.001 | 0.21 |
| LpW ratio (AU) | 0.78±0.15 $^{bcd}$ | 0.95±0.24 $^{a}$ | 0.88±0.23 $^{a}$ | 0.94±0.21$^{a}$ | 10.541 | < 0.001 | 0.06 |
| Stretch index longitudinal (m) | 11.7±1.5 $^{bcd}$ | 10.1±1.9 $^{a}$ | 10.8±1.7 $^{a}$ | 10.7±1.9 $^{a}$ | 9.647 | < 0.001 | 0.07 |
| Stretch index lateral (m) | 13.5±1.6 $^{bcd}$ | 10.1±1.7 $^{ac}$ | 11.1±2.2 $^{ab}$ | 10.5±1.9 $^{a}$ | 45.833 | < 0.001 | 0.23 |
| Max ID (m) | 56.0±6.1 $^{bcd}$ | 43.8±6.1 $^{ac}$ | 48.3±7.0 $^{ab}$ | 45.9±6.5 $^{a}$ | 41.915 | < 0.001 | 0.23 |
| Min ID (m) | 6.9±1.8 $^{bcd}$ | 5.8±2.0 $^{a}$ | 5.7±1.9 $^{a}$ | 5.6±1.8 $^{a}$ | 7.16 | < 0.001 | 0.06 |
| Mean ID (m) | 28.6±2.6 $^{bcd}$ | 22.8±3.2 $^{ac}$ | 24.7±3.2 $^{ab}$ | 23.9±3.0 $^{a}$ | 46.318 | < 0.001 | 0.22 |

* LpW = Length per Width, ID = Interpersonal distance. Superscripts a, b, c, d to indicate significant difference between corresponding two phases.

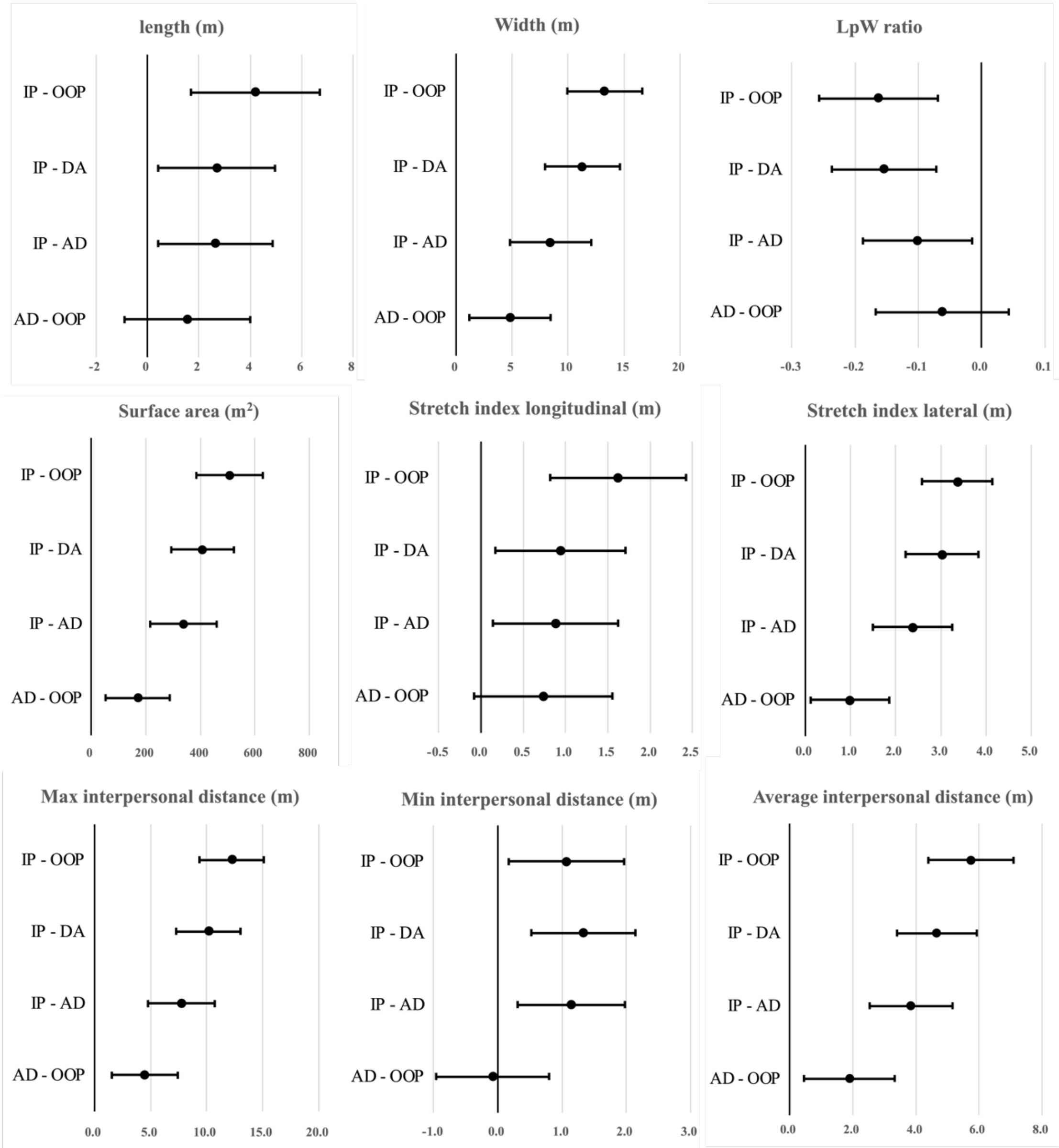



**Figure 3.4** Mean differences and 95% CI of pairwise differences between phases with significant difference for tactical behaviour. Pairwise phases (OOP and DA transition, DA transition and AD transition) without significant difference for any tactical measures are excluded.

## 3.4 Discussion

The football match, as a complex and dynamic team sport, includes coordination between teammates and competition between opposing teams. The match phases (i.e., attack, defence, and transition) influence the objective and decision making of a team, and then impact on collective movements and actions. Therefore, tactical analysis at phase level is expected to uncover more in-depth insights into team tactical behaviour. The aim of this chapter was to determine and compare team tactical behaviour of a professional football team in different phases of an official league match. The main findings were: 1) different proportion of short and long phase across four match phases; 2) significantly different team tactical behaviour between IP and each of other phases; 3) from OOP to IP, a larger increase in team width than length; 4) a wider and more dispersed team formation in AD transition than OOP; 5) the similar team tactical behaviour in DA transition and OOP. The workflow proposed in chapter 2 proved to be applicable in tactical analysis.

Larger proportions of short phase were found in two types of transition compared to the phase of in possession and out of possession. In contrast, phases of transition scarcely lasted for more than 20 seconds, but it is common for phases of in possession and out of possession. Longer phase means more time for decision making and action, and more players involved in the team coordination. This is a reason why team tactical behaviour varies across different match phases.

In possession, the team controlled the ball and tended to present longer and wider formation than during defence and transition phases. This demonstrated that when organising attack players spread out to control more area, which was also supported by the tendency of surface area across phases (Figure 3.4). These findings are in line with the greater spread team formation (especially team width) in attack to disorganise defence of the opposition (Clemente et al., 2013; Praça et al., 2022). Although significant differences of length (with small effect size) and width (with large effect size) were found between IP and each of other phases (i.e., IP-OOP, IP-ADT, and IP-DAT), the differences of team width (8 to 15 meters) between IP and other phases were larger than the differences of team length (less than 5 meters) between IP and other phases. This was also reflected by LpW ratios in different phases (Table 3.2). The dynamically lateral and longitudinal contraction and expansion corresponded to the "team unity" and "width and length" of offensive strategy (Teoldo da Costa et al., 2010). In the direction of attack, although team length increased in possession, the defensive line remained in proximity to offensive line to support offensive actions of teammates. Attacking as a unit to some extent can also benefit defence. Close distance between the defensive and offensive line (longitudinal compactness) reduces the space that opponents can use when the opposition regains ball possession and starts a counterattack immediately. However, in the direction of pitch width, players (especially right/ left wings/ midfielders/ fullbacks) tended to extensively make use of the space in flank to render the opposing team to widen their formation. The wider the formation, the bigger the gap between defenders. Accordingly, the team has more chances to make the penetrative pass or dribbling to move the ball towards the opponent goal and create scoring opportunity.

Also, the team's shape was different between match phases. In possession, the team formation presented a rectangular shape in the lateral direction, which was significantly different from the transition phase from defence to attack. The LpW ratio was larger and approached a value of 1 during out of play and in transition from defence to attack. Olthof et al. (2019b) found that the team formation changed toward a nearly squared shape (LpW ratio close to 1) in official matches in comparison with a more rectangular shape (LpW ratio < 1) in 11-a-side training games. Praça et al. (2022) reported a similar result that LpW ratio neared a value of 1 in offensive phases but presented a smaller value in defensive phases. This current analysis indicates that the shape of team formation varies along the official match playing and is valuable to be studied at the subphase level (i.e., attack, defence, and transition phases). In addition, distance between players (from own and opposing team) and team dispersion also influence the space that can be used to organise attack. Frencken et al. (2011) suggested that the variation of tactical behaviour represented the interchanging attacking and defending behaviour, which is functional for attacking teams to explore defending weakness of the opposition. Olthof et al. (2015) reported that the game-to-game variability of lateral inter-team distance was higher than that of longitudinal inter-team distance. This current analysis revealed that interpersonal distance and stretch index (especially lateral) varied along the match and presented higher magnitude during in-possession phase than other phases.

This analysis also showed that several tactical measures (team width, surface area, lateral stretch index, as well as maximum and average interpersonal distance) in AD transition were significantly larger than those in OOP. Previous studies showed a greater surface area and stretch index in offence than defence (Clemente et al., 2013; Praça et al., 2022). The current study further confirms that the differences of these tactical measures between AD transition and OOP were similar to, but less than the differences between IP and OOP. This aligns with the fact that teams are reorganising defence from attacking formation in AD transition. The significant differences between DA transition and IP across all tactical behaviours suggest that when regaining ball possession, the team's counterattack characteristics differentiates from the attacking behaviour when they control the ball. The team covered less playing area and presented less dispersion when regaining the ball than when controlling and consolidating the possession. In practice, this informs stakeholders of the team tactical movement after the team regains ball possession. Coaches are then able to compare team's tactics execution in DA transition and their attacking strategy.

Nevertheless, there is no difference across all measures between OOP and DA transition. This means that the team presented the similar formation in defence and when transiting from defence to attack. A possible explanation for this behaviour is that the team in DA transition inherited the most tactical movement pattern from defence phase and reorganised attack slowly when regaining ball possession. This can be specifically built on if the coach desire higher tempo in transition from defence to attack. The length of DA transition period of the opposition could be also determined and exploited to explore options in attack. In future research, the short period after regaining (losing) ball possession is supposed to be distinguished from in-possession (out-of-possession) phase.

Frencken et al. (2013) proposed a 3-s time window for tactical analysis, based on expert football coaches' advice on the maximal time allowed for a team to respond to game events. This method of exclusion possibly emphasises the fact that shorter phases correspond to less players involved in the team action. However, approximately a quarter of transition phases lasted less than three seconds in the current analysis (Figure 3.3). If the 3-s window had been put into place, a large proportion of transition phases would have been excluded from the analysis. Consequently, the findings would be less comprehensive for understanding the team's performance over the match. The minimum response time allowed for a football team has yet to be suggested in previous study. The current analysis found the inheritance between OOP and DA transition, which could possibly be explained that the team required a short period to switch to offensive mode. This can be considered when building the consensus on exclusion criterion of phase length in team tactical analysis in the future.

Although the findings provided useful insights into team tactical behaviour across match phases, there are still limitations in this analysis. First, although there are more than 300 phases from the analysed match, the limited sample of match could compromise the validity of these findings. It requires more matches to be verified in the future. This analysis applied one match to explore team tactical behaviour in different match phases, which can be considered as the proof of concept. In addition, notational analysis was conducted by professional analysts from a football club, based on their own definition of match phases. Although the event data were visually inspected with the video footage, the validity of this analysis is dependent on the data quality and the definition of match phases by the football club. Lastly, this analysis had no access to position information of the opposition and the ball, and did not consider the movement of the goalkeeper. Interaction between two opposing teams is one of the missing parts of this analysis and could also be a limitation of the study using GNSS tracking data.

Those findings and limitations above also indicate the potential research direction in the future. For future study, given that up to date GNSS tracking systems cannot provide ball position information, notational analysis is a viable method to retrieve match information related to ball position (e.g., where the phase starts). Based on ball position, match phases can be further broken down and integrated with positional information of the event (Anzer and Bauer, 2021). Tenga et al. (2010) reported that 51.6% of the starting zone of the offensive phases occurs in the defensive half and 45.5% in the middle half. Tactical behaviour of phases can be linked with the starting zone information in the future. With respect to notational data, data validity is also important and commonly examined by the Intraclass Correlation Coefficient of inter-observer and intra-observer agreement. However, in a practical world and this chapter, it is impossible that a data analyst repeatedly notates a same match, nor that several data analysts repeat a same work. Analysts respecting the data quality and decision makers respecting analysts' suggestion is a pragmatic alternative in terms of data quality in practical application. Goalkeepers in modern football play an important role in not only defence but also offense in team unity. But this position was often not taken into consideration in football analysis (Knoop et al., 2013). Corrêa et al. (2014) reported a smaller defensive area as well as less variability when the goalkeeper acted as an outfield player in futsal. In the future, analysis into proximity of goalkeepers to defensive lines can inform coaches and players of the coordination between defenders and the goalkeeper.

Team performance is dependent upon how players are dynamically positioned according to the teams' overall space distribution principles and the dynamic functional constraints at the scale of the environment (Sampaio and Macas, 2012). To determine the regularity in players' moving pattern and team coordination, entropies (Duarte et al., 2013; Silva et al., 2014a) and relative phase (Folgado et al., 2014; Goncalves et al., 2014) have been respectively applied in previous studies. Welch et al. (2021) inspired by collective structure in biological systems, proposed a novel approach - density plot - to visualise collective behaviour in different phases of play in football match. In the future, synchronisation at team or group level can be compared across match phases, and knowledge from ecological systems can be extended into the application in football analysis. They are expected to meet the dynamic characteristics of football match and to inform the decision-making process of coaches and players.

## 3.5 Conclusion

This chapter aimed to compare professional football team's tactical behaviour in different match phases. Effective playing time was divided into four phases: *in possession*, *out of possession*, *transition from attack to defence*, and *transition from defence to attack*. While controlling the ball, the team presented a wider dispersion and less compactness than in transitions and defence. While losing ball possession and transiting from attack to defence, the team played with less dispersion and more compactness than *in possession*, and with wider formation compared to *out of play*. Team tactical behaviour in attack and defence was relatively reserved in the following transition. These findings are in line with offensive and defensive tactical principles and confirm the importance of breaking down offensive and defensive phases into subphases. This chapter also proves the potentiality and applicability of the workflow proposed in chapter 2. Integration of event data and positional data is a promising and informative approach for practitioners in this area. To ensure the tactical behaviour is captured and analysed for each phase, the data quality is a high priority in this type of analysis. Positional data loss, human error in notational analysis, and systematic offsets between event data and positional data are supposed to be resolved and reported prior to the interpretation of analysis.

## References


Anzer, G. & Bauer, P. 2021. A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Front Sports Act Living,* 3**,** 624475.

Araújo, D., Passos, P., Esteves, P., Duarte, R., Lopes, J., Hristovski, R. & Davids, K. 2015. The micro-macro link in understanding sport tactical behaviours: Integrating information and action at different levels of system analysis in sport. *Movement & Sport Sciences - Science & Motricité***,** 53-63.

Barbu, D. & Stoica, D. 2020. Increasing the execution speed of offensive and defensive tactical actions in the football game at the time of transition. *Journal of Sport and Kinetic Movement,* 36**,** 5-13.

Buchheit, M., Allen, A., Poon, T.K., Modonutti, M., Gregson, W. & Di Salvo, V. 2014. Integrating different tracking systems in football: multiple camera semi-automatic system, local position measurement and GPS technologies. *Journal of Sports Sciences,* 32**,** 1844-1857.

Carling, C., Williams, A.M. & Reilly, T. 2005. *Handbook of Soccer Match Analysis a Systematic Approach to Improving Performance,* London ;, Routledge.

Clemente, F., Couceiro, M., Martins, F., Mendes, R. & Figueiredo, A.n. 2013. Measuring Tactical Behaviour Using Technological Metrics: Case Study of a Football Game. *International Journal of Sports Science & Coaching,* 8.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences,* New York, Routledge.

Corrêa, U.C., Davids, K., Silva, S., Denardi, R.A. & Tani, G. 2014. The influence of a goalkeeper as an outfield player on defensive subsystems in futsal. *Advances in Physical Education,* 4**,** 84-92.

Duarte, R., Araújo, D., Folgado, H., Esteves, P., Marques, P. & Davids, K. 2013. Capturing complex, non-linear team behaviours during competitive football performance. *Journal of Systems Science and Complexity,* 26**,** 62-72.

Duarte, R., Araujo, D., Freire, L., Folgado, H., Fernandes, O. & Davids, K. 2012. Intra- and inter-group coordination patterns reveal collective behaviors of football players near the scoring zone. *Human Movement Science,* 31**,** 1639-51.

Folgado, H., Duarte, R., Fernandes, O. & Sampaio, J. 2014. Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches. *PLoS One,* 9**,** e97145.

Folgado, H., Lemmink, K.A.P.M., Frencken, W. & Sampaio, J. 2012. Length, width and centroid distance as measures of teams tactical performance in youth football. *European Journal of Sport Science,* 14**,** S487-S492.

Frencken, W., Lemmink, K., Delleman, N. & Visscher, C. 2011. Oscillations of centroid position and surface area of soccer teams in small-sided games. *European Journal of Sport Science,* 11**,** 215-223.

Frencken, W., Van Der Plaats, J., Visscher, C. & Lemmink, K. 2013. Size matters: Pitch dimensions constrain interactive team behaviour in soccer. *Journal of Systems*

*Science and Complexity,* 26**,** 85-93.

Goes, F.R., Brink, M.S., Elferink-Gemser, M.T., Kempe, M. & Lemmink, K. 2020a. The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data. *Journal of Sports Sciences***,** 1-10.

Goes, F.R., Meerhoff, L.A., Bueno, M.J.O., Rodrigues, D.M., Moura, F.A., Brink, M.S., Elferink-Gemser, M.T., Knobbe, A.J., Cunha, S.A., Torres, R.S. & Lemmink, K.A.P.M. 2020b. Unlocking the potential of big data to support tactical performance analysis in professional soccer: A systematic review. *European Journal of Sport Science,* 21**,** 481-496.

Goncalves, B.V., Figueira, B.E., Macas, V. & Sampaio, J. 2014. Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. *Journal of Sports Sciences,* 32**,** 191-9.

Gonzalez-Rodenas, J., Lopez-Bondia, I., Calabuig, F., Pérez-Turpin, J.A. & Aranda, R. 2015. The effects of playing tactics on creating scoring opportunities in random matches from US Major League Soccer. *International Journal of Performance Analysis in Sport,* 15**,** 851-872.

Gréhaigne, J.-F., Richard, J.-F. & Griffin, L.L. 2004. *Teaching and Learning Team Sports and Games,* New York, Routledge.

Jackson, B.M., Polglaze, T., Dawson, B., King, T. & Peeling, P. 2018. Comparing Global Positioning System and Global Navigation Satellite System Measures of Team-Sport Movements. *International Journal of Sports Physiology and Performance,* 13**,** 1005-1010.

Knoop, M., Fernandez-Fernandez, J. & Ferrauti, A. 2013. Evaluation of a Specific Reaction and Action Speed Test for the Soccer Goalkeeper. *The Journal of Strength & Conditioning Research,* 27**,** 2141-2148.

Low, B., Rein, R., Raabe, D., Schwab, S. & Memmert, D. 2021. The porous high-press? An experimental approach investigating tactical behaviours from two pressing strategies in football. *Journal of Sports Sciences,* 39**,** 2199-2210.

Memmert, D. & Rein, R. 2018. Match analysis, Big Data and tactics: current trends in elite soccer. *Deutsche Zeitschrift für Sportmedizin,* 2018**,** 65-72.

Moura, F.A., Martins, L.E., Anido Rde, O., de Barros, R.M. & Cunha, S.A. 2012. Quantitative analysis of Brazilian football players' organisation on the pitch. *Sports Biomech,* 11**,** 85-96.

Olsen, E. & Larsen, O. 1997. Use of match analysis by coaches. *Science and football III,* 3**,** 209-220.

Olthof, S.B., Frencken, W.G. & Lemmink, K.A. 2015. The older, the wider: On-field tactical behavior of elite-standard youth soccer players in small-sided games. *Human Movement Science,* 41**,** 92-102.

Olthof, S.B.H., Frencken, W.G.P. & Lemmink, K. 2019a. 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**,** 523-530.

Olthof, S.B.H., Frencken, W.G.P. & Lemmink, K. 2019b. When Something Is at Stake:

Differences in Soccer Performance in 11 vs. 11 During Official Matches and Training Games. *Journal of Strength and Conditioning Research,* 33**,** 167-173.

Praça, G.M., Moreira, P.E.D., de Andrade, A.G.P., Clemente, F.M., de Oliveira, W.B. & Demétrio, G. 2022. Integrating notational and positional analysis to investigate tactical behavior in offensive and defensive phases of football matches. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology*.

Sampaio, J. & Macas, V. 2012. Measuring tactical behaviour in football. *International Journal of Sports Medicine,* 33**,** 395-401.

Seifert, L., Araujo, D., Komar, J. & Davids, K. 2017. Understanding constraints on sport performance from the complexity sciences paradigm: An ecological dynamics framework. *Human Movement Science,* 56**,** 178-180.

Silva, P., Aguiar, P., Duarte, R., Davids, K., Araujo, D. & Garganta, J. 2014a. Effects of Pitch Size and Skill Level on Tactical Behaviours of Association Football Players During Small-Sided and Conditioned Games. *International Journal of Sports Science & Coaching,* 9**,** 993-1006.

Silva, P., Travassos, B., Vilar, L., Aguiar, P., Davids, K., Araujo, D. & Garganta, J. 2014b. Numerical relations and skill level constrain co-adaptive behaviors of agents in sports teams. *PLoS One,* 9**,** e107112.

Tenga, A., Ronglan, L.T. & Bahr, R. 2010. Measuring the effectiveness of offensive match-play in professional soccer. *European Journal of Sport Science,* 10**,** 269-277.

Teoldo da Costa, I., Garganta, J., Greco, P., Isabel, M. & Afonso, J. 2010. Assessment of tactical principles in youth soccer players of different age groups. *Revista Portuguesa de Ciências do Desporto,* 10**,** 147-157.

Torres-Ronda, L., Beanland, E., Whitehead, S., Sweeting, A. & Clubb, J. 2022. Tracking Systems in Team Sports: A Narrative Review of Applications of the Data and Sport Specific Analysis. *Sports Medicine - Open,* 8**,** 15.

Vilar, L., Araujo, D., Davids, K. & Button, C. 2012. The role of ecological dynamics in analysing performance in team sports. *Sports Medicine,* 42**,** 1-10.

Welch, M., Schaerf, T.M. & Murphy, A. 2021. Collective states and their transitions in football. *PLoS One,* 16**,** e0251970.

# Chapter 4: General Discussion

This chapter presents the conclusions of this thesis, emphasizing the contributions in application. Additionally, the limitations of this approach and ideas for future work are proposed. This thesis aims to present a toolbox (i.e., workflow, possible issues and corresponding solutions, and referential python codes) for team tactical analysis using GNSS positional data. Additionally, the other objective is to utilise the proposed workflow to compare team tactical behaviour in different phases of match.

## 4.1 Main findings

Team tactical strategy and execution vary along the match playing, influenced by the specific match condition (e.g., match phases). The primary motivation of this work is focused on exploring the difference of team tactical moving patterns between match phases. To achieve this, positional data from GNSS tracking systems require data pre-processing steps (as opposed to optical tracking and LPS) for tactical analysis. However, whilst massive studies previously used GNSS positional data to analyse tactical behaviour (Sampaio et al., 2014; Goncalves et al., 2017; Baptista et al., 2020), they scarcely reported the data processing procedure. Consequently, researchers and practitioners without the relevant background need time-consuming exploration to embark on tactical analysis using GNSS data. Besides, it is less trustworthy when comparing findings from studies with different standards of data processing. Therefore, this thesis presented a blueprint for processing raw GNSS positional data for tactical analysis. Based on the blueprint and primary motivation, team tactical behaviour was quantified and compared between phases of in possession, out of possession, transition from defence to attack, and transition from attack to defence. The following discussion will focus on these two parts of the thesis: 1) quantifying team tactical behaviour using raw GNSS positional data; and 2) integrating team tactical measures with event data to compare team tactical behaviour across match phases.

### 4.1.1 Using GNSS positional data for team tactical measures

The workflow in chapter 2 detailed on 1) data cleansing; 2) map projection; 3) calculating and applying rotation matrix; 4) merging individual data into a team dataset; 5) creating timeline when data loss occurring; 6) interpolation and filtering. Necessary information for each step and potential issues in data processing were clearly outlined, which can support practitioners understanding the procedure for using GNSS tracking systems for tactical analysis. Solutions to those issues and the referential Python code were provided as a toolbox that might reduce time-consuming exploration and facilitate the data processing for tactical analysis. Besides, two pragmatic approaches to retrieve coordinates of pitch location were also compared and presented in the Appendix. The first approach is placing GNSS tracking devices at four corners of the pitch for minutes. If collected coordinates drift in visualisation, average coordinates can then be used to determine the specific couple of coordinates (i.e., x and y coordinates) for each corner. Web mapping platforms are the other sensible tool. Observers click at each corner with the computer mouse to retrieve coordinates of the pitch. These coordinates are required to calculate the rotation matrix which is necessary for data processing. However, methods of retrieving coordinates of pitch location have not been reported in previous studies. Although Folgado et al. (2014) briefly reported a series of required steps for this type of data processing, there is still a lack of the handbook that can guide sport scientists and practitioners to prepare necessary ingredients for data processing. It can also be laborious to explore and solve emerged issues without guidelines. With the elaborate workflow in chapter 2, sport scientists and practitioners can easily embark on tactical analysis with GNSS positional data.

Additionally, a consensus on acceptable level of positional data loss is expected to be discussed and reached by researchers in the future. Occasional data loss from GNSS tracking systems is a common issue (Capaccio et al., 1997). It may not affect the primary use of physical analysis, but it will affect the team tactical analysis for each timestamp. However, the volume and frequency of data loss have not been mentioned in previous literature, but researchers should mention this alongside the measures that have been put in place to account for this data loss. In chapter 2, the dataset contained approximately 40% of partial data loss and 13.5% of complete data loss respectively. Two consecutive missing data accounted for 95% of data loss. Linear interpolation was used to fill in null values. Similar information on data quality and processing methods are supposed to be reported in future study. The working mechanism of GNSS tracking systems outlined in chapter 2 also supports seeking solutions if a new issue of data processing emerges in the future. By means of the workflow can GNSS tracking systems not only be employed for physical monitoring but also tactical analysis by sport practitioners on a training or match basis. Please note that the code might need refinement if new unexpected issues emerge during data processing, which possibly cost extra time.

### 4.1.2 Integration of tactical measures and match phases

Team tactical behaviour has been analysed in training and match-play previously (Frencken et al., 2012; Sampaio et al., 2014). Because team tactical behaviour varies along the match playing, analysing it on phase level are expected to meet the dynamics of football match and reveal more in-depth insights on tactics execution. The current analysis is the first attempt to decompose effective match playing to phases of attack, defence, and transition phases. Quantified tactical measures are combined with phase information from notational analysis. The findings from this thesis confirm the value of analysing tactical behaviour in phase level instead of full-match level. The team tended to play with a wider and more dispersed formation when in possession than other phases. Team tactical behaviour distinguished the in-possession phase from the defence-to-attack transition. Collective movement in a defence-to-attack transition was possibly inherited from the previous phase of out of play.

The reason of different behaviour in those phases can be explained by the relationship between task, person, and environment from constraint-led approach theory (Seifert et al., 2017). Players in possession are likely to consolidate possession or move forwards at high speed, which differentiates from their behaviour in defence. With respect to the constraint of person, while organising attack, centre backs mainly focus on remaining team unity and creating options for teammates, but strikers aim to create scoring opportunity for themselves or teammates. The playing style (e.g., aggressive high pressing or cautious sitting off) of the opposition also influences the team's tactical execution, from a perspective of the environment constraint. In summary, individual and collective behaviour depends on the specific objective under each situation. Splitting the match into different phases based on objectives benefits the understanding of team tactical behaviour, as this thesis showed.

As shown in chapter 3, the integration of positional data and event data can produce more insights than analysing one of them alone. Nevertheless, issues related to data collection and processing need to be solved first for positional data, event data, as well as synchronisation for two types of data. In positional data collection, human error (e.g., a tracking device not turning on) or equipment malfunction would impact on data quality. In data processing, it is necessary to confirm that positional data from each player are merged according to timestamps. Otherwise, findings from the analysis are meaningless. Issues from event data are generally caused by human error, in other words, the time shift of an event between the notation tag and the reality (Boyd et al., 2011; Anzer and Bauer, 2021). This is due to the combined influence of the understanding of event definition, concentration, and decision and reaction time of the analyst. Given that it is time costly to notate a match, repeating the notation work to check the quality of event data is impossible to be fulfilled in practical application. Therefore, establishing a series of benchmarks to verify whether this is a noticeable notation error is a promising direction of future study. This can be achieved by massive datasets and long-term analysis. For example, hypothetically, through considerable studies, it is concluded that during defence the distance between defensive line and midfielder line is generally less than a value. This can be considered as a threshold. In following study, if the value at a moment exceeds the threshold, it reminds analysts to double check the event data. This, to some extent, ensures the quality of event data and saves time. Furthermore, if the reminder occurs but no notation error is detected in event data, that could be interpreted as a potentially valuable moment which is worth special attention for analysis and informing stakeholders.

Synchronisation is a crucial step that should be highlighted when integrating positional data with event data, but is scarcely reported in previous studies. Human error from notation analysis and systematic offset of two systems (i.e., player tracking technology and camera) are both potentially responsible for the time lag between two types of data (Anzer and Bauer, 2021). The human error, as aforementioned, can be minimised and controlled by clear understanding of event definition and concentration. However, the player tracking system and camera used by teams cannot be of uniform devices, which implies that there is not a universal solution for systematic offsets in synchronisation (especially when it is randomly varied). If it is a regular offset (e.g., an accumulative time lag), the regularity can be determined by time-consuming verification, as elaborated in chapter 3. In practical world, the match report is required shortly after the match. The time consumed on synchronisation and verification depends on the degree of urgency. In academic research, the acceptable level of time lag is suggested to be discussed and reached a consensus, which guarantees the standard of data quality and that findings are reliable to be compared across studies.

## 4.2 The limitation and directions for future research

The limitation of this thesis is that only one match was analysed in the study. The thesis is a proof of concept that the presented workflow has the potential for real-world application, and that team tactical analysis is worthy to be conducted in match-phase level. High volume of data and long-term analysis are expected to facilitate the comprehensive understanding of general tactical movement pattern and of the specific team tactical execution. The workflow presented in this thesis provides a toolbox and a technical basis for sports scientists and practitioners for future work. This type of analysis will lead to several data issues (i.e., GNSS data loss and synchronisation). Seeking solution for these issues can be time-consuming and requires understanding of the data sources and data handling skills. Requirements for the knowledge and understanding of the technology and data, to some extent, limit the number of people working in this area. Product companies should facilitate this type of analysis in their software and improve data quality.

In the future, larger number of positional data and event data are expected to be integrated for tactical analysis. In addition to team level, the tactical behaviour at subgroup level, such as the defensive line, can also be compared across different match phases. In the study at subgroup level, dynamic positioning and circulation of players are supposed to be considered. For example, the left back in attack sometimes acts as an extra winger, overlapping the winger to create attack opportunity. When losing ball possession, the opposition launch counterattack immediately without giving the left back time to move back into position. In this context, one of midfielders is likely to swap position with the left back. The question is that whether the left back should be included in defensive line at this period. To simplify data processing, the analysis might be conducted with a fixed positioning (Goncalves et al., 2014). In contrast, Goes et al. (2020) used clustering method to automatically identify dynamic formation and subgroups. An alternative can be always considering the nearest four players (if it is a four-back formation) to the own goal line as the defensive line, and so on for other subgroups.

Additionally, the starting zone of each phase also influences players' decision and action (Tenga et al., 2010). Information of match phase and starting zone (e.g., Thirds of football pitch) can be combined to, for example, compare tactical behaviour in transitions starting from different zones. Given the nonlinearity of football match, team synchronisation in different phases is also a promising direction in future study. In conclusion, combining multi-type detailed and useful information is expected to provide in-depth insights. But meanwhile it is also important to keep the data and findings reliable during data processing. Open source of processing workflow and corresponding example is a possible pathway for reaching a consensus on the quality of data and studies.

## 4.3 Practical implications

This thesis provided a toolbox that can facilitate team tactical analysis using GNSS positional data. The analysis of team tactical behaviour in different match phases is the proof of concept that the proposed workflow and tactical analysis in match-phase level are valuable in practical world. This type of study requires the integration of positional data and event data. Although the workflow can reduce the time for tactical analysis and maintain the standard of data processing, manual notation for event data is still a laborious work and seems difficult to guarantee the data quality. Clear definitions of match events and concentration of analysts during notation are beneficial to the quality of event data.

On the match day, or after training, event data are supposed to be notated and used for following tactical analysis. Positional data are exported from the player tracking technology system. If the start and end timestamps of the training or match-play are known, by means of the presented workflow data processing prior to tactical analysis can be completed within 5 minutes. In this step, processing time depends on the volume of raw data (i.e., the number of players and the length of data collection). Data noise may incur longer time of data processing. Integrating information from event data with positional data leads to the synchronisation for two types of data, which possibly requires prolonged visual inspection and processing to ensure that those timelines are aligned, as shown in chapter 3.

## 4.4 Conclusion

This thesis revealed that team tactical behaviour varies across different phases of match. Previous studies divided football match into offensive and defensive phases in tactical analysis. However, this thesis confirmed that offensive phases should be decomposed to *defence-to-attack transition* and *in possession*. Similar team tactical behaviour in *out of play* and *defence-to-attack transition* implied that the team requires a short period to reorganise offensive positioning. These findings arise from the integration of team tactical measures and match phase information. Those team tactical measures in current analysis were processed based on the workflow of data pre-processing proposed in this thesis, which proves the potentiality of the blueprint in real-world application.

## References


Anzer, G. & Bauer, P. 2021. A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer). *Front Sports Act Living,* 3**,** 624475.

Baptista, J., Travassos, B., Goncalves, B., Mourao, P., Viana, J.L. & Sampaio, J. 2020. Exploring the Effects of Playing Formations on Tactical Behavior and External Workload During Football Small-Sided Games. *Journal of Strength and Conditioning Research,* 34**,** 2024-2030.

Boyd, L., Ball, K. & Aughey, R. 2011. The Reliability of MinimaxX Accelerometers for Measuring Physical Activity in Australian Football. *International journal of sports physiology and performance,* 6**,** 311-21.

Capaccio, S., Lowe, D., Walsh, D.M.A. & Daly, P. 1997. Real-time differential GPS/GLONASS trials in Europe using all-in-view 20-channel receivers. *Journal of Navigation,* 50**,** 193-208.

Folgado, H., Duarte, R., Fernandes, O. & Sampaio, J. 2014. Competing with lower level opponents decreases intra-team movement synchronization and time-motion demands during pre-season soccer matches. *PLoS One,* 9**,** e97145.

Frencken, W., Poel, H.d., Visscher, C. & Lemmink, K. 2012. Variability of inter-team distances associated with match events in elite-standard soccer. *Journal of Sports Sciences,* 30**,** 1207-1213.

Goes, F.R., Brink, M.S., Elferink-Gemser, M.T., Kempe, M. & Lemmink, K. 2020. The tactics of successful attacks in professional association football: large-scale spatiotemporal analysis of dynamic subgroups using position tracking data. *Journal of Sports Sciences***,** 1-10.

Goncalves, B., Coutinho, D., Santos, S., Lago-Penas, C., Jimenez, S. & Sampaio, J. 2017. Exploring Team Passing Networks and Player Movement Dynamics in Youth Association Football. *PLoS One,* 12**,** e0171156.

Goncalves, B.V., Figueira, B.E., Macas, V. & Sampaio, J. 2014. Effect of player position on movement behaviour, physical and physiological performances during an 11-a-side football game. *Journal of Sports Sciences,* 32**,** 191-9.

Sampaio, J.E., Lago, C., Goncalves, B., Macas, V.M. & Leite, N. 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 and Medicine in Sport,* 17**,** 229-33.

Seifert, L., Araujo, D., Komar, J. & Davids, K. 2017. Understanding constraints on sport performance from the complexity sciences paradigm: An ecological dynamics framework. *Human Movement Science,* 56**,** 178-180.

Tenga, A., Ronglan, L.T. & Bahr, R. 2010. Measuring the effectiveness of offensive match-play in professional soccer. *European Journal of Sport Science,* 10**,** 269-277.

# Appendix

## Methods for retrieving pitch coordinates

In data processing section, rotation matrix is calculated based on coordinates of four pitch vertices. Given that the rotation matrix is also applied to raw positional data, it will exert an impact on the validity and reliability of processed team positional data, calculated tactical variables and therefore analysis afterwards. As a result, pitch coordinates need to be reliable. This part of study aimed to explore and compare the difference between two approaches: 1) GNSS device; 2) web mapping platform (e.g., Google Maps), to collecting coordinates of a 11-a-side football pitch. GNSS devices used to collect player's positional data can also be applied in pitch coordinates collection. Under some conditions (e.g., away stadium) it is not easy to have access to pitch and collect data on it. Existing and reliable web mapping platforms are then a viable alternative to GNSS collection. A secondary aim was to determine the inter-unit and intra-unit difference of GNSS devices, with two versions of devices.

### GNSS device collection

#### Experimental approach

There two available scenarios for the GNSS-collecting approach. First, with one device, set it at each corner and collect data in turn (i.e., four samplings with one unit), which implies it is costly in time because of repetition of sampling. Regardless, it is a pragmatic method that helps when only one practitioner working with one device on this. Second, with four devices, set each one at each corner, sampling at the same time, which saves time and furthest ensures the same external environment (e.g., number of available satellites) for sampling of each corner.

In this study, pitch coordinates were collected at each corner of an 11-a-sided football pitch by turns. Two versions of GNSS devices (Catapult, Optimeye S5, 10 Hz; Catapult, Vector S7, 10 Hz) were used to determine the inter-device and intra-device difference on collected positional data. For this purpose of study, two Optimeye S5 (S5) devices and two Vector S7 (S7) devices were set at a corner at the same time (Figure A-1) for 7-minute data collection. Data of each corner was collected by turns. Weather condition on the testing day was decent (17℃, 77% relative humidity, 18km·h$^{-1}$ wind speed) with no visible cloud cover.

To quantify GNSS measurement quality, two measures provided by suppliers of GNSS hardware were exported as positioning quality information: 1) the horizontal dilution of precision (HDOP) which measures how much the geometric arrangement of satellites being measured will affect the precision of the result (Witte and Wilson, 2004), and 2) number of fixed satellites. A HDOP of 1 to 1.5 is consider as high-quality data (CatapultSports, 2017), representing a good distribution of satellites. In other words, fewer satellites positioned in the space directly above the receivers lead to high accuracy. A HDOP greater than 3 is considered as poor data quality (CatapultSports, 2017), which indicates that more than 3 satellites are spaced directly above the receivers, and therefore a non-ideal positional arrangement of satellites. The time between turning on device and activating connection to GNSS system was recorded for each sampling.


**Figure A-1.** Set of GNSS devices for data collection at a pitch corner. S5 indicates Optimeye S5. S7 indicates Vector S7.

**Collection results**

In all 16 data samplings (4 samplings at each corner), the maximum and minimum connection time was 23.67 seconds and 87.25 seconds, respectively. GNSS connection was active in average 55.81 seconds. According to the time that connections take, there is no certain conclusion of which version of device has the advantage in activating connection to satellites. Figure A-2 illustrates that S7 devices takes less time to connect with more satellites, and to reach high precision (less HDOP). Drafting data points of GNSS-device collection are as shown in Figure A-3.

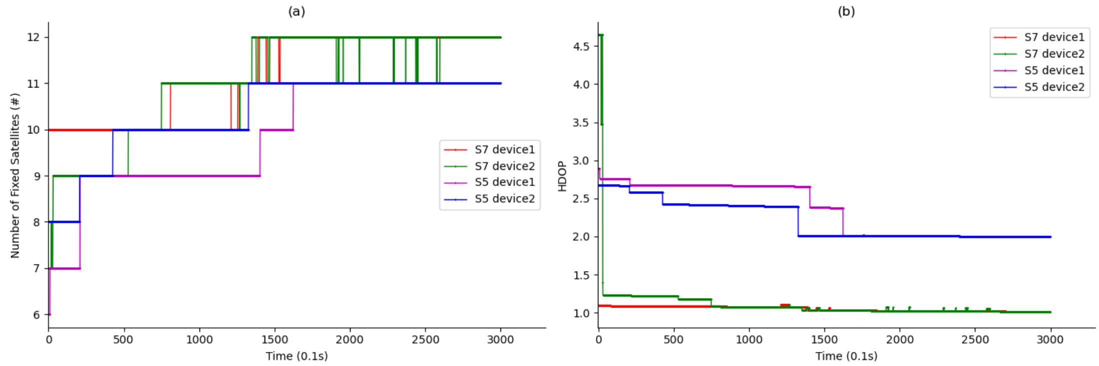



**Figure A-2.** (a) Number of fixed satellites and (b) HDOP of two S7 devices and two S5 devices set at the same corner for 5-minute data sampling.

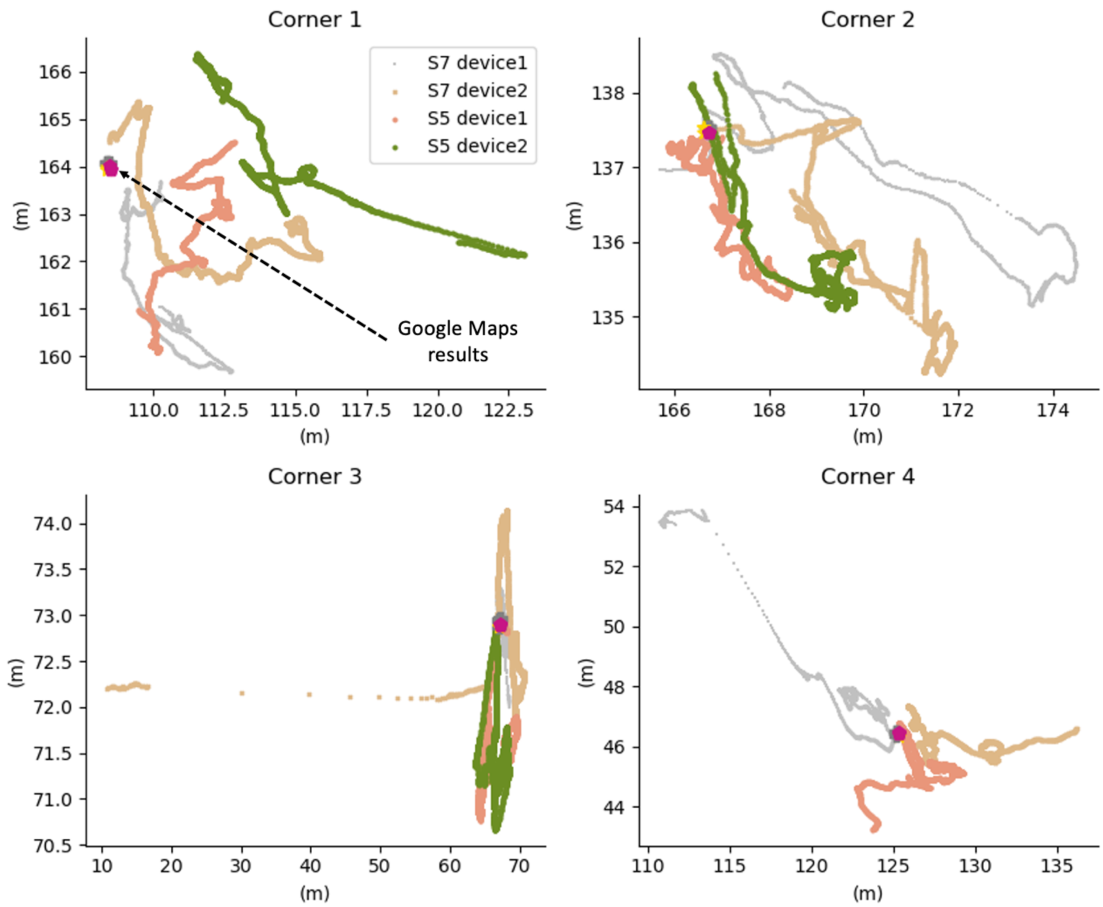



**Figure A-3.** Inter-observer Intra-device and inter-device results of GNSS collection, and results of Google Maps colletion. Two devices of each version of device (Optimeye S5, Vector S7) were included in GNSS collection.

## Web mapping platform

### Experimental approach

Web-mapping approach was operated on the Google Maps by three practitioners without relevant data collection experience. Operations by practitioners were conducted on different computers but with the same browser (Google Chrome) and browser display setting, and followed the same instruction. To simulate 1) operations by different observers; 2) operations by one observer at attempts, each practitioner was required to collect pitch coordinates by clicking on the pitch on Google Maps three times at different time (with random interval between each collection).

**Collection results**

For each observer, the variation between collections at one corner is maximumly 0.12 meter on x-axis and 0.07 meter on y-axis. Standard deviation of pitch corner locations is presented in Table A-1, indicating there is little variance between attempts and between observers. Average positions of four pitch corners by two approaches are presented in Figure A-4. Maximum Euclidian distance between results of two methods at one corner is 3.7 meters.

**Table A-1.** Standard deviation of x-coordinates and y-coordinates of three collections by each observer, as well as standard deviation of all collections.

| | Corner one | | Corner two | | Corner three | | Corner four | |
|---|---|---|---|---|---|---|---|---|
| | X | Y | X | Y | X | Y | X | Y |
| Observer 1 | 0.000 | 0.000 | 0.003 | 0.038 | 0.000 | 0.000 | 0.064 | 0.005 |
| Observer 2 | 0.063 | 0.041 | 0.003 | 0.038 | 0.003 | 0.038 | 0.003 | 0.038 |
| Observer 3 | 0.064 | 0.005 | 0.003 | 0.038 | 0.000 | 0.000 | 0.003 | 0.038 |
| All collections | 0.081 | 0.030 | 0.058 | 0.050 | 0.056 | 0.035 | 0.059 | 0.061 |

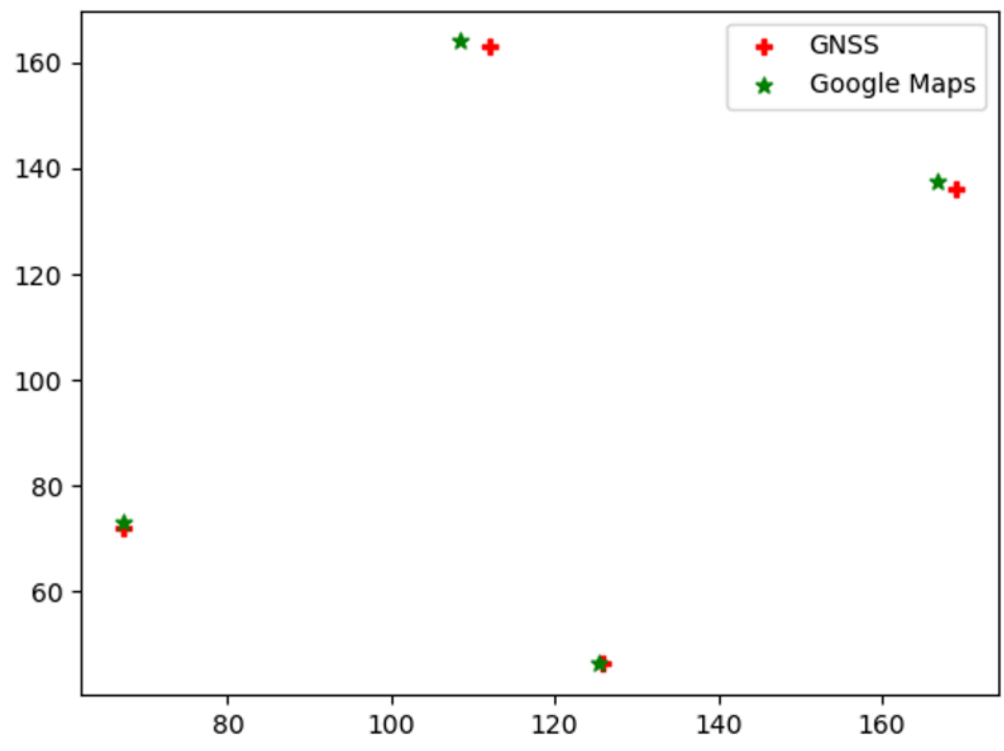


**Figure A-4.** Average positions of pitch corners without rotation, collected by GNSS devices and observers using Google Maps.

## Comparison of two approaches

Concerning precision of collected data, using web mapping platform is at an advantage over GNSS device. In term of time cost, two practitioners spent more than 30 minutes to complete this intra-device and inter-device collection. Generally, it takes at least 15 minutes to complete data collection using GNSS devices (without data exporting and processing). However, collection with web mapping platform took less than 10 minutes and presented four pitch corners that can be used directly for calculating rotation matrix. In practice, it is unpractical to use GNSS devices to collect coordinates of pitch corners in advance (e.g., no access to opponent home stadium).

In conclusion, web mapping platform is a pragmatic and effective method for collecting pitch coordinates in tactical analysis.