FOOTBALL GAME ANALYSIS: A NEW APPLICATION AREA FOR CARTOGRAPHERS AND GI-SCIENTISTS?

Gilbert Kotzbek, Wolfgang Kainz

Gilbert Kotzbek, BSc MSc
Department of Geography and Regional Research
University of Vienna
Universitätsstraße 7, 1010 Vienna, Austria
Phone: +43 (1) 4277-48627; e-mail: gilbert.kotzbek@univie.ac.at

Univ.-Prof. Dipl.-Ing. Dr. Wolfgang Kainz
Department of Geography and Regional Research
University of Vienna
Universitätsstraße 7, 1010 Vienna, Austria
Phone: +43 (1) 4277-48640; e-mail: wolfgang.kainz@univie.ac.at

Abstract
Game analyses in association football are increasingly valuable for professional teams to remain competitive. Unfortunately conventional methods of post-game analyses often fail to analyse and visualise the complex spatio-temporal patterns of the football sufficiently. These circumstances lead to a rising demand for rich geospatial sports data. Actually, vast amounts of geo data are captured. In this article, the two basic kinds of football specific geo data are described. This paper also presents the possibility to utilise these special datasets for spatio-temporal analyses, conducted with a Geographical Information System (GIS). Furthermore, the modelling of a football field using a GIS is portrayed. Additionally, the implementation of football specific geo data is demonstrated. Moreover, our central future research project is introduced. Since football can be thought as a union of space and time, game analyses cannot ignore its spatio-temporal nature. The suitability of GIS applying football specific geo data is also discussed.

Keywords: football game analysis; GIS; Geographical Information Systems; geovisualisation; spatio-temporal analysis; football specific geo data; tracking data; event data; soccer; association football

INTRODUCTION
Association football (soccer) is much more than a game, played by two teams with eleven players each, who attempt to score at least one more goal than the opposition within 90 minutes to win the game. It is one of the most popular sports in the world, which attracts millions of spectators into the stadiums every weekend worldwide. Researchers from different scientific disciplines e.g. sport science, medicine, sociology and psychology as well as economics and many more, attempt to analyse different facets of the game. Although post match analysis is only one aspect within the entire field of possible football related applications it is certainly a crucial one. As in other sports disciplines, game analyses are very important for game preparation. Not only do football teams need to evaluate teamwork and each player’s individual performance, but also tactics, performance analysis and scouting are potential fields of application [1] [7] [10] [11]. Game analysis is providing professional teams with another useful planning and tactical tool, in order to maximise their performance. In order to maintain a competitive advantage in professional football teams are now investing heavily in sports analytics [9]. Unfortunately, statistical methods of summarising matches often fail to answer the given specific questions about the gameplay because many of them ignore the spatio-temporal characteristics of football matches [7].

The purpose of this article is to demonstrate that Geographical Information Systems (GIS) can be applied in the field of football game analysis in an appropriate manner. We assume that this could open the way to more detailed questions and satisfy answers concerning the gameplay by taking the spatio-temporal nature of football into account in order to support tactical preparation, recovery and in match decision making. As football is a spatial and time-limited sport, it is of fundamental importance to understand its spatial-temporal patterns. It is crucial to understand space and time relationships in sports in order to convey players with adequate tactical instructions. Our intent is to specify and highlight the main advantages of using GIS technology in the course of football game analysis. Further, we aim to
provide an overview of our upcoming research. With this article, we hope to ignite further conversation and to encourage cartographers and GI-scientists to use GIS technology for game analysis in football.

The rest of this paper is organized as follows: First, we will describe the game of football as a spatio-temporal framework, including its objects and restrictions. Second, football specific geo data will be scrutinised. Afterwards we will depict how to model a football field as an analysis-framework using ArcMap 10.1. Then we will present a brief introduction to our research project to convey an overview on our future investigations. Finally, this paper will be concluded in its last section.

FOOTBALL: A SPATIO-TEMPORAL FRAMEWORK

From a geographical perspective, football can be considered to be a spatio-temporal framework. Within this scheme spatio-temporal phenomena and objects are correlated. According to Kim et al. [7] these comprise the ball, the players and the football field including the goals. In addition, the referee and his/her assistants as well as the game time also have to be considered as spatio-temporal objects. Except for the football field all of them are dynamic, if the condition of the field during the game is disregarded. This fact causes the great complexity of this sport, which is also responsible for its unpredictability as well as its intrinsic high factor of excitement. Naturally, the pitch’s condition and the weather could change throughout the match. Therefore, it would be worth to consider this too. Unfortunately, there is a lack of pitch conditions’ geospatial data. Furthermore, different kinds of interactions such as passing, tackling, shooting as well as activities without the ball occur between the moving elements on the field. The two competitive teams can also be regarded as cooperative complex networks [1] [14]. Moreover, the activity and interaction between players of one team is directly affected by the interplay of the other team and vice versa [14].

This interdependence causes a certain spatio-temporal structure of a football game, which is defined by some fundamental restrictions. First, the field itself is limited to a predefined size. In association football the International Football Association Board (IFAB) standardised the rectangular field size as 105 x 68 metres. At this point, it has to be mentioned that not all fields are the same size. There is a tolerance range of several meters [22]. Within the field boundaries there are also a couple of defined areas, e.g. the penalty area, which are primarily relevant for the set of rules. In terms of time the game is limited to two half-times of 45 minutes each including a variable additional time depending on the course of the game.

Football is a union of space and time. Within this predefined framework the players’ and ball’s positions are unknown in advance. However, after the game almost every single action of each player is being recorded and stored. The result of this process is a vast amount of raw data, which is becoming increasingly valuable in the eyes of coaches, players and analysts. For this certain purpose knowledge of the ball’s and the players’ positions in course of the game is absolutely necessary. In the context of cartography and geoinformation science it is a matter of particular interest that a vast amount of football specific geo data is captured on a regular basis because geo data provides the backbone of a vast array of cartographic and geoinformation projects. In addition, these data are appropriate for all kinds of GIS-based analyses. Since cartographers and GI-scientists have been studying space-time relationships for a long time we assume that their experiences as well as technologies can help to better understand the dynamics of football matches.

FOOTBALL SPECIFIC GEO DATA AS AN INTERFACE TO GI-TECHNOLOGY

Within the scientific environment, different data capturing methods were applied in the last two decades in the field of football and comparable sports [4] [5] [8] [9] [10] [11] [12] [13]. Almost all of them are extracting the player profile and ball passes from video recordings. This seems reasonable because nearly every football game played in the course of an international competition or professional national league is shown on TV. Moreover video is a robust method, and seeing the game in replay is a method very close to reality.

In professional football video recordings are also the most used data source. These days, several companies provide data and offer various analysis systems [15] [16] [17] [18] [19] [20] [21]. In the case of football specific geo data, two different kinds of data have to be distinguished – tracking data and event data – which are provided in different data formats such as ‘.pos’ or ‘.xml’.

Tracking data are not only x/y-coordinates including a certain timestamp but they also provide speed details which is calculated using the x,y and time changing over time. Such data information is available for each player as well as the referee team and the ball [15] [21]. Furthermore the z-coordinate is provided for the ball’s position in order to understand the flight of the ball. The data is processed almost entirely automatically with the use of a couple of fixed cameras located around the field. Owing to the fact that these cameras have to be installed at a certain height, tracking is limited to fairly large stadiums. Whereas tracking in the German Bundesliga is possible in each stadium, Austria’s stadiums are too small in most cases. The number and position of the cameras can vary depending on the applied
system. The temporal resolution is extremely high and can reach up to 25 records per second [15], whereas the spatial resolution is about 30 cm, which can be considered as sufficient. Whigham [13] reported “…that it is not necessary to have events referenced with accuracy below approximately 1 metre, which is essentially the local neighbourhood of an individual.” Although his research was dealing with rugby, this statement is also valid in the case of football because of the similarities between these kinds of sports regarding its spatio-temporal framework. In fact, football was developed from rugby more than 150 years ago. Tracking data creates the possibility to observe and analyse the players’ movements during the whole game retrospectively.

Although the spatial resolution of tracking data gained with camera-based systems seems to be acceptable, we presume that other techniques and methods would lead to a more accurate outcome. In this case we would advocate investigations, which evaluate the advantages and disadvantages of dGPS as well as RTK-GPS systems for tracking the dynamic spatio-temporal objects in a football game. Indeed, several professional teams in Europe and in the US utilise GPS technology in their training sessions. Unfortunately, these kinds of techniques are not allowed under match conditions and are also disregarded in the rulebooks. Hence, analysts are only able to use the data captured from camera tracking technology.

Event data is captured manually from video by so called scouts, who gather the data while repeatedly watching every game scene. Event data can best be described as position based attributive data. Due to the fact that the data capture procedure depends on individuals’ subjective perceptual processes these data are susceptible to errors. To minimize the expected error, all records are checked by one other person. Data providing companies like Opta [16], Impire [20] or Amisco [17] as well as its subsidiaries Mastercoach [18] and Prozone [19] provide a huge amount of data for every game. Hence, 1,500 to 3,000 event data records per one single game are common. In addition, these data are also classified. Thus, up to more than 60 categories including manifold forms are available. For instance, the category ‘shot’ may be classified into ‘shot on target’, ‘shot missing the target’ as well as ‘deflected shot on goal’, etc. Furthermore, all data records are provided with spatio-temporal information. Therefore, an event data set for one single game is highly informative and provides a valuable dataset from which detailed analysis can be performed.

Whilst this data is being collected it requires quite a bit of pre-data processing before it can be imported into a GIS. As mentioned before, these kinds of geo data are provided in different data formats. Thus it cannot be ensured that it can be used in a GIS without any previous data preparation. Moreover, the local coordinate system that underlies the data can vary from one provider to another. These circumstances must be considered when transforming the data records into one’s own defined analysis environment reference system. Fortunately, our investigation showed that it is possible. We successfully managed to prepare test data for use in a GIS. The datasets used in this study are provided by Opta [16], Mastercoach [18] and Impire [20].

On the one side, a vast amount of football specific geo data is being captured and on the other side, there are cartographers and GI experts and analysts, who have the knowledge about GIS as well as the required skills to deal with geo data. At the moment, the vacuum between these two domains has not been filled yet. There exist a real opportunity for cartographer and GI analysts to bridge this gap to understand the dynamic nature of football games better, which will in turn improve the quality of football game analyses. Furthermore, we assume that proven cartographic guidelines are suitable to visualize the obtained analysis results. Our intention is corroborated by similar approaches in other sports such as described in [2] and [6].

MODELLING A FOOTBALL FIELD USING ARCMap 10.1

Analysing football specific geo data requires an appropriate analysis environment. We assume that ArcMap 10.1 provides a proven and robust GIS platform for which we conduct our research. As the provided data contains single coordinate (x,y) points only, it is appropriate to model a football field in a GIS (See: Figure 1(a)). The initial point dataset was a simple dBase table, in which essential points e.g. corner points, penalty area corner points, etc. of the standardized field with a size of 105 x 68 metres [22] were entered. Afterwards, the table was added in ArcMap 10.1, where the points were displayed by its x/y-coordinate information subsequently. On that account, it is also important that a projected coordinate system is used. From a geographical point of view, the field’s centre is located at the intersection point of the equator and the zero meridian line of Greenwich. Then, line and polygon features were manually created based on the displayed points. Finally, some graphical adaptations were made to support visual impressions such as 5 metre grass stripes. This is relevant for drawing fast conclusions on questions like “Was the ball out?” Regarding the football field’s visual appearance there are countless possibilities. Apart from the visible point, line and polygon features there is also the opportunity to create several invisible areas on the field. These can be crucial for the game analysis as certain questions may require subdividing the football field into smaller zones [1] [11]. However, the type of analysis is ultimately dictating how to represent the field.
As already mentioned in the previous section of this paper, conventional football specific geo data has to be transformed from a Cartesian to a projected coordinate system prior to its implementation in a GIS. Figure 1 (b), (c) and (d) shows that we successfully prepared the data by Opta [16], Mastercoach [18] and Impire [20]. Although each supplied data set initially had a different local coordinate system, it is possible to use them within one single analysis environment now. This creates numerous opportunities when combining the datasets to receive detailed information about the concerning football game. Naturally, all used data sets have to reflect the very same game.

Figure 1: (a) Modelled football field with ArcMap 10.1; (b) Implementation of tracking data, kindly provided by Impire [20]; (c), (d) Implementation of event data, kindly provided by Mastercoach [18] in (c) and Opta [16] in (d)

GIS-BASED GAME ANALYSIS – AN OVERVIEW OF OUR UPCOMING RESEARCH PROJECT

“GIS-Based Game Analysis – Geovisualisation of Spatio-Temporal Phenomena Based on Football Specific Geo Data” is the translated and tentative working title of Mr Kotzbek’s doctoral thesis, which is planned to be finished in early 2017. In the course of this thesis several undertakings are intended. The basic research question is dealing with the benefits of qualitative GIS-based analysis of spatio-temporal phenomena in football. We assume that there are advantages compared to traditional game analysis methods. When analysing a football game it is of crucial importance to take the spatio-temporal nature of football into account. If space and time are ignored, misinterpretations are likely. This presumption can be regarded as the elementary research hypothesis. To verify it, it is necessary to approach several research domains, which are illustrated in Figure 2. The depicted research pyramid consists of four major research areas, while the sequence of investigation is determined from the bottom to the top. In addition, the respective objectives are listed in brackets for each area.

As a first step, it is essential to build up a solid scientific fundament to emphasise that there is a need for GIS-based game analysis, because of the existing shortcomings of conventional football game analyses. This intention requires a profound evaluation of conventional game analysis methods. The observed downsides are mainly caused by the lack of consideration of the spatio-temporal nature. This manifests itself in both, the analysis and the visualisation of football specific game analyses. Most of the time, single parameters e.g. ‘ball-winning ability’ or ‘pass rate’ etc. are presented separately without any hints about how these results were achieved. Furthermore, certain parameters are also often analysed isolated which leads to vague assumptions about the game [7] [9]. Therefore, conventional football game
analyses have to be scrutinised to determine all shortcomings. For that reason, the investigation focuses on team-
internal as well as team-external analysis methods. Regarding team-internal methods, expert interviews will have to be
conducted with the aim of finding out how professional football analysts are working within their teams. In order to
gain insight in this domain it is necessary to identify the key parameters as well as analysis and visualisation methods.
In the field of team-external methods, a couple of scientific investigations have already been done, especially in the area
of sports sciences. Moreover, analysis methods in the media such as newspapers and TV will have to be examined.
Afterwards, all gathered information about conventional football game analyses will also have to be confronted with
a set of criteria in the context of the spatio-temporal character of the football game to outline shortcomings from a
geospatial point of view. In general, if space and time are crucial in any sports, a mere representation of statistics,
without any relation to space and time, must be considered absolutely insufficient, for instance [2] [6].

The next step is concerned with single player analyses with the purpose of its optimisation. It is intended to achieve this
challenge by weighting of essential indicators with the assistance of a GIS by taking the spatio-temporal information
into account. This measure seems to be reasonable because the effect of a misdirected pass within the penalty area
cannot be ascribed the same impact on the game as one in the centre of the field, for example. The weighting is intended
to give the records a certain quality. The parameters are still given by the event data sets. But what is essential? To
figure this out, an online survey is envisaged. Therefore, coaches, players, media representatives and supporters are
going to be questioned. Nevertheless, we assume that the following parameters will play a key role in our work: ‘pass
rate’, ‘running distance’, ‘goal attempts’ and ‘ball-winning ability’, etc. By using appropriate GIS-tools it should be
possible to extract adequate weighting factors. In the context of the proposed GIS analysis itself, we unfortunately are
not able to point out any methods in detail yet, because the e-survey as well as the determination of the essential
indicators still needs to be carried out.

In general, GIS technology is assumed to be appropriate to increase the objectivity of the analyses. Since human beings
have to interpret the analysis results based on their individual perception combined with their experience, it is not
possible to eliminate the subjective factors. Nevertheless, GIS-based game analyses are supposed to support decision
making processes more efficiently than traditional methods. Just to mention one example: An average professional
player covers a distance of about 11 km per game [3]. Consider a player who ran less. The covered distance is often the
only information conventional game analyses provide. If the distance is below-average, some people might suggest that
this particular player performed weakly. Of course this could be true but what about other opportunities? Perhaps the
player did not need to run longer distances because his/her positional play was simply good. When observing and
recording the players’ position during the entire game, it should be possible to scrutinise different opportunities for
different situations using a GIS.

Figure 2: Research pyramid

303
As soon as the single player analysis has been concluded, the collective team level is scheduled by the research project’s timeline. Primarily, knowledge at the individual level has to be gained in order to understand the emerged complex spatio-temporal patterns between teammates as well as the competing teams as a whole. Both, the individual as well as the collective level is important and complementary. Furthermore, both are interrelated and overlapping [10] [11]. On that account, a clear distinction between these two levels cannot be drawn.

The third research domain will be defined thematically. We set out to focus on four specific problem areas, namely: ‘network of passing’, ‘spatial distribution and the use of space’, ‘teamwork’ and ‘play style’. Thereby, we attempt to analyse collective spatio-temporal phenomena in football to reveal its complex patterns and behaviours. The main goal is to provide detailed insights into the gameplay. In fact, similar analysing objects have been investigated in the field of other scientific disciplines, as described in [1], [10], [11], [12] and [14]. However, a GIS approach seems to be novel.

Regarding networks of passing sequences, GIS-based network analysing tools will be appropriate. In that case, players are represented as nodes and passes as edges. Additionally, analysing the length of the passing sequence, the time a player controls the ball, the passing accuracy as well as the players’ centrality, etc. are important in this context. Concerning spatial distribution and the use of space, GIS-based cluster analyses as well as proximity analyses are important in understanding the complex spatiotemporal network during football matches. Because space is so crucial in football, overlapping areas and connections between different analysing objects are inevitable. Hence, networks of passes have a certain impact on the spatial distribution of the players and vice versa and potentially affect the results of matches. Football can be thought as a dynamic mosaic. When focussing on one specific aspect of it, the entire picture must be considered. Otherwise it will not be possible to understand the game in its details. Based on several analyses of different spatio-temporal phenomena, particular game patterns of movement and behaviour should be detected. Combining the single results is presumed to be crucial to make assertions about the effectiveness of teamwork. In this particular case the players’ movement with the ball as well as their movements without the ball are important for the analysis process. Using a GIS should enable the analysts to evaluate one teams’ teamwork in terms of its space management and cooperative movement. A similar approach was employed by TAKI et al. [10]. Linked analyses are also supposed to be the key to determine the play style of one team and to find out whether a team favours for instance a short passing game or quick counter attacks. For this kind of analyses, knowledge of the passing networks is necessary. For instance, differences between ‘Tiki-Taka’ and ‘Kick & Rush’ can be revealed when comparing several indicators such as the length of the passing sequence as well as the time from the attack’s initial point to its outcome, e.g. a final shot on target. Moreover, the effectiveness of one specific play style can be determined. In order to do that, the outcome of each counter attack has to be analysed referred to the total amount of counter attacks conducted by one team during the game, for example.

Furthermore, buffering is assumed to be applicable for many issues, e.g. when analysing spheres of action. Through interpolations of point data clouds, representing the players, polygons of the two opposing teams can be created. In further consequence, these polygons are appropriate for intersection analyses, whereby spatial dominance of the opponent within defending lines is detectable, for instance. Additionally, requests of information as well as measuring tools can be applied as supplementary implements. Regarding football game analyses, GIS technology seems to provide numerous convenient analysing tools.

However, analysing something is only one side of the coin. The most convincing analysis results are worthless if they are not visualised appropriately. Geographic information lends itself to being mapped and visualised. If it is done correctly it can lead to further visual analysis and a greater understanding of the results. Therefore, aspects of geovisualisation are to be given highest importance. As football is a spatial phenomenon it is obvious to visualise its analysis outcomes in the form of a map. This approach takes the footballs’ spatio-temporal nature into account. The main objective is to provide adequate graphics, which allows the user to elicit information efficiently. Finally, an e-survey is envisaged again to evaluate the maps legibility. This can be seen as an iterative process because often the illustrations raise further questions and encourage further conversation. As respondents we intend to assemble the same groups as for the first questionnaire in order to validate if the produced maps are adequate or not.

Summarizing the results of each research domain is necessary to determine the benefits of GIS-based game analysis compared to conventional game analysis in football.

**SUMMARY AND CONCLUSION**

The relevance of space and time in sports is increasingly acknowledged by the public as well as coaches and team managers. Without continual improvement of their performance, professional football teams would be uncompetitive soon. Therefore, the teams make considerable efforts in game analyses which necessitate the capturing of high quality
data. Nowadays several companies provide these data. Regarding the data itself, it is of particular interest that the records can be considered as geo data sets, because all of them provide location information. Furthermore, a defined timestamp is also comprised. Owing to this fact, this special kind of data is applicable for analyses using Geographical Information Systems (GIS). The dynamic of field actions of players are driven by an equally dynamic team of coaches and analysts from opposing team, each try and out play the other using game plans and team systems. Teams are continually refining their game plans and systems in order to gain a competitive edge over their opposition and to not become predictable in their match play.

In this paper, we first described the game of football from a geographical point of view. It can be thought as a restricted spatio-temporal framework including several objects such as the ball, the referees’ team as well as the players of two competing teams and additionally the football field and the game time. Based on these elements the data are gathered. Concerning the football specific geo data, two different types have to be distinguished. These are: tracking data and event data. If analysing these data using a GIS is intended, first of all a football field has to be modelled. Hence, we also described our modelling process. Afterwards, we showed that it is possible to prepare football specific geo data in a GIS-appropriate manner. Moreover, we managed to implement these data in our analysis environment. The records were kindly provided by Opta [16], Mastercoach [18] and Impire [20].

Following this, we introduced our current research project, which is mainly based on Mr. Kotzbeks’ doctoral thesis. At the time of writing this paper his work is at the very beginning. Therefore no results can be provided yet. Instead of this, the main issues of the project were presented to share our ideas with the intention to foster further research. At this point, it has to be noted that the planned GIS based analyses discussed above, only cover some specific parts of the football game. To think that one doctoral thesis could cover and explain all aspects of the football game in detail would definitely be overconfident. In fact, this work can only be considered a very first step into a new application area of GI-Science. Therefore it is of great importance to highlight the possibilities of GIS-based game analysis as well as its advantages over traditional and conventional methods respectively. Furthermore, this research project is related to single games. In a broader sense, long-term analyses of certain teams for the duration of a whole season as well as beyond that also be of great interest. Cartographers and GI-Scientists are now at the entrance to this new field of application. The door is wide open for specialist researches of all kinds. We are very confident of the inherent high potential of this new application area, which is going to be revealed when football specific geo data are being analysed using GIS technology. This could pave the way to a better understanding of football, which is of great interest in the case of football game analyses.

REFERENCES


BIOGRAPHY

Gilbert Kotzbek received the BSc degree in Geography as well as the MSc degree in Cartography and Geoinformation with distinction both at the University of Vienna in 2010 and 2013 respectively. Since October 2013 he has been undertaking the doctoral program in the field of Geography at the University of Vienna. Currently he is also employed as a university assistant at the Department of Geography and Regional Research at the University of Vienna. His doctoral thesis focusses on the analysis and visualization of football specific geo data in a spatio-temporal context.

Prof. Kainz holds a graduate degree in technical mathematics and computer science and a PhD from the Graz University of Technology, Austria. Since 2002 he is a full professor of Cartography and Geoinformation Science at the Department of Geography and Regional Research of the University of Vienna, Austria. Previously he held various positions as researcher, assistant professor and full professor at research and academic institutions in Austria, United States of America and the Faculty of Geo-Information Science and Earth Observation of the University of Twente (ITC), The Netherlands. Prof. Kainz is the editor-in-chief of the ISPRS International Journal of Geoinformation. His research interests are in GIS theory, spatial databases, topology, and fuzzy logic.