Beyond the Pass Completion Rate: Towards Analysing the Players' Passing Quality in Association Football

Gilbert Kotzbek * and Wolfgang Kainz

University of Vienna; Faculty of Earth Sciences, Geography and Astronomy; Department of Geography and Regional Research; 1010 Vienna, Austria;
* Correspondence: gilbert.kotzbek@univie.ac.at

Abstract: For approximately two decades, professional association football has been characterised by increased digitisation. This is associated with the capture of a vast amount of georeferenced data representing the players' and the ball's movement as well as game relevant events. In this context, the practicability of geographic information systems (GIS) in the course of the analysis of single players' passes is addressed in this article. The aim of this study is to relativise the players' pass completion rate's significance. As space is crucial in football, for each pass the analysis takes the game's current spatiotemporal structure into account in order to determine its inherent risk. This encompasses the opponents' positions first, between the passer and the receiver as well as second, in close proximity to both of them. Moreover, the calculation considers the passes' length and locations. In relation to the pass completion rate, the players' passing quality is derived. For this purpose, an ArcGIS custom tool was developed utilising Python. In addition to this, an expert survey was conducted interviewing professional game analysts. The results provide strong evidence that the tool's functionality is given although certain measures would definitely enhance the tool's applicability in contemporary football game analyses.

Keywords: GIS-based Football Game Analysis; GIS; sports analytics; association football; passing quality; ProzoneSports

1. Introduction

This article is one of a series of papers which are about GIS-based football game analysis. In this sense, association football (soccer) is meant. For instance, an approach for automated analysis of scoring attempt patterns utilising geographic information systems (GIS) was already presented [1]. Owing to the growing competition on both, national and international level, football clubs sought to find new ways of analysing and characterising the game in order to gain advantages over their opponents. As a consequence, professional teams invest heavily in sports analytics [2-4]. Furthermore, contemporary practices of football game analysis are highly subjective due to the dominance of video analysis [5]. In general, this is not recognised by professional game analysts, but more objective methods are advisable. Since a vast amount of georeferenced data is being captured for virtually all high level games [6], the application of GIS embodies such novel approaches as geo data provides the foundation for manifold spatiotemporal analyses [7-10]. The employed data can be described as football-specific geo data and comprise the players' and ball's movement as well as game relevant events [11]. In the course of the doctoral thesis GIS-gestützte Spielanalyse (= GIS-based Game Analysis) [5] at the University of Vienna, Austria, the practicability of GIS within the scope of football game analysis was examined. The thesis' main purpose, its contents, design and scope are briefly portrayed in [12,13]. Among others, the GIS-based analysis of the players' passing quality was addressed in [5]. Referring to that, this article presents the analysis' results and most significant findings.

As space is crucial in football, the analysis of each pass focuses on the current game’s spatiotemporal structure. The objective of this study is to relativise the significance of the pass completion rate and thus objectivise the interpretation of the players’ passing. The pass completion
rate is an expression for the players’ passing accuracy [14] and is calculated as displayed in Equation (1).

\[ Q_p(t) = \frac{P_s(t)}{P(t)} \times 100 \]  

(1)

It corresponds to the ratio between successful passes \( P_s(t) \) and all passes \( P(t) \) played in a specific period of time \( t \) such as half times, games or entire seasons. In this article’s context a successful pass is given if the recorded Team ID of both the passer and the receiver are identical. Although this parameter’s legitimacy is given, its significance is limited and elusive. Its major deficiency is attributed to its ignorance towards the game’s spatiotemporal structure when passes are played. As a consequence, all passes are considered as qualitatively equal. For instance, short passes between two central defenders often lack risk of the opponent’s intervention, but are equally assessed as long and risky passes into the attacking zone. Apart from this, parameters should not be interpreted independently of the game [15] and also the players’ tactical position has to be taken into account [6]. Therefore, the mere indication of the pass completion rate must be considered as inappropriate for evaluating a pass’ quality on single player level.

Although the acquisition of football-specific geo data dates back to the mid-1990s, corresponding publications are still rare in the domain of cartography and GI science [13]. The reason for this is unknown, but might be a result of the lack of awareness of football as a potential field of application. Hence, spatial or spatiotemporal analysis of footballers’ passing are generally limited to GI applications, which according to Bartelme [7] are not built upon generic GIS. However, it is almost impossible to draw a razor-sharp line between both approaches [7]. Furthermore, thematic overlaps to this article’s topic are scarce which is why solely two contributions are worth highlighting in this context.

First, two decades ago Taki et al. [16] developed the concept of the players’ Dominant Regions. These regions define areas on the pitch “where the player can arrive at earlier than all of the others.” For this, the so called Minimum Moving Time Pattern (MMT) was derived from the players’ trajectories. If passes are played within the intersection area of the passer’s and the receiver’s dominant regions, the pass cooperation quality is high. The principal objective of Taki et al. [16] was to evaluate the teamwork.

Second, Gudmundsson and Wolle [17] pursued a comparable approach to Taki et al. [16] and developed a pass analysis tool. Instead of dominant regions, the players’ passable and reachable areas were determined based upon kinetic information of the players’ and the ball’s trajectories. With this, the players’ passing abilities were evaluated and the passes’ difficulty was rated. Moreover, they compared the passes to passing alternatives. Although interesting in general, this paper lacks a profound description of both, the employed data and the applied analysis concept. Therefore, the results’ accomplishment is difficult to comprehend.

Besides these two examples, the analysis of passing networks, such as addressed in [18-20] are distantly related to this article’s topic. However, network analyses primarily focus on teamwork or collective team parameters instead of the assessment of the single players’ passing. In addition to this, analysis of passing sequences can be found in [21-23].

In the course of this study an ArcGIS custom tool named Analyse Passes was developed based upon Python in order to bridge the gap to analyses of the players’ passing quality. Its input parameters correspond to tracking and event data provided by ProzoneSports, a company which is market-leading in Europe [24] and recently joined forces with STATS [25]. Primarily the tool quantifies the passes’ risk based upon the current players’ positions on the pitch at the time of each pass. Afterwards, this supplementary parameter is related to the passes’ outcome (successful / not successful) to derive the players’ passing quality. Whilst the formula of the pass risk and the passing quality is predefined, their factors and thresholds can optionally be configured according to one’s own conception. This is reasonable, since the majority of game analysts complain about the opaqueness of contemporary analysis tools’ functionality [5]. For the evaluation of the tool’s practicability, the results were presented to professional game analysts, before the interviews were conducted. The experts’ feedback provides strong evidence that the tool’s functionality is given in
the course of football game analysis. However, there are still opportunities left to enhance the tool’s applicability, which are summarised in both Section 4.2 and Section 5 of this article.

Moreover, it is important to emphasise that although ArcGIS was applied and hence this article relates to a certain software, other GIS are considered as equally suitable for our purpose. This statement is applicable as the tool’s code described in Section 3 can be modified easily and quickly so that it is executable on other platforms too. Therefore, this article’s relevance arises from the presented concept of both the pass risk and the passing quality as well as their determination based upon the analysis of the game’s spatiotemporal structure which can be considered as novel. Despite the limitation to certain aspects, these two parameters provide deeper insights into the individual play and hence enrich the single player performance analysis.

2. Materials and Methodology

2.1. Data

In the following the applied data is briefly outlined. Additional information about the data capturing process can be found in [11,5]. As mentioned above, the data base was provided by ProzoneSports and is composed of tracking and event data representing an entire game. Whilst the former mirror the players’ and the ball’s trajectories, the latter contain information about both, ball relevant events such as shots, tackles or passes as well as other events as offside positions, substitutions or cards for example. In order not to be tempted to subjective interpretations a request for anonymized data was made and granted. Hence, the players are distinguished by a six-digit player ID. As a consequence, the teams’ philosophy as well as the coaches’ tactical specifications are unknown. Moreover, football is influenced by a series of external factors which cannot be taken into account due to lack of appropriate data. In this context not only internal factors such as mental, physical or medical conditions but also external factors such as temperature, humidity, altitude, ground conditions or the atmosphere within the stadium are beyond any game analysis [11, 5]. Therefore, the single player pass analysis introduced in this article is based upon spatial data exclusively.

According to Mitchell [9] as well as Zeiler and Murphy [10] both data sets correspond to point features whose spatial appearance as X/Y coordinates is time-dependent. Since each point is equipped with a consecutive time stamp and frame, tracking and event data can be attributively joined. This is essential for analysing the players’ passing. Although identical from a technical perspective, there is a semantic difference between both data types. Whilst event data are independent phenomena, tracking data can be described as continuous paths in form of Moving Point Objects (MPO) [26]. These are characterised as point clouds, and its elements are related to each other and are chronologically ordered [27]. Based on the players’ movements, assertions about their behaviour can be drawn [27].

It has to be taken into account, that in general these are influenced by environmental factors as well as intrinsic characteristics, spatial constraints and other objects’ impacts [26]. In this context, the trajectories reveal not only the individual, but also the collective behaviour [28,29]. Furthermore, according to Kang et al. [28] the players’ motion has to be distinguished from their movement. Whilst the former is associated with the objects’ spatial variance, the latter corresponds to its actions.

Without exception ProzoneSports’ data spatially refer to the international standardised pitch size of 105 m x 68 m, as displayed in Figure 1. However, according to the FIFA’s (Fédération Internationale de Football Association) laws of the game [30] the pitch size can vary within a predefined tolerance. As a consequence, the data might feature minor spatial inaccuracies due to scaling. In regard of the data’s validity and accuracy, Carling et al. [4] criticise that no standardised tests exists. Notwithstanding this, there is evidence that the spatial accuracy of both, tracking and event data is sufficient for scientific purposes and game analyses [30-34]. For instance, di Salvo et al. [31] conducted several test runs which were different in the paths’ course, distance, and velocity. Across all measurements, deviations between 1 cm and 23 cm were observed. Moreover, di Salvo et al. confirmed the speed data’s high accuracy based on match data in another study [32]. In addition to this, the tracking data’s validity was approved by O’Donoghue and Robinson [33], who examined the trajectories with regard to pitch areas’ transitions and path changes.
In contrast to tracking data which are recorded with 10 fps applying high speed cameras, event data are acquired manually after the game. Hence, their reliability is questionable. On this, Bradley et al. [34] came to the conclusion, that the observed mean absolute error of 0.007 s and 3.6 m for event time and position respectively, are acceptable. Nevertheless, in this context it is advisable to attributively join tracking and event data in order to solely employ the trajectories more accurate spatial information.

2.2. Systems Applied

ArcGIS for Desktop 10.x – Background Geoprocessing (64 Bit) was utilised for the purposes of this study because it is a proven, functional comprehensive and well documented GIS. Since Python is Esri’s preferred programming language [35] it was reasonable to employ it for the development of the Analyse Passes custom tool. One of the advantages executing the script as a custom tool is that the tool’s target group do not have to be familiar with Python at all. However, minor experience with GIS in general and executing ArcGIS tools in particular are recommendable. Assuming that GIS is generally unknown to game analysts, the developed tool was designed as simple as possible. Therefore, as it is outlined in more detail in Section 3, the user must only select the input data and enter a player ID, since all other input parameters are optional.

As the configuration of the applied computer system is a critical factor in regard to the tool’s performance, its characteristics have to be stated. It is based on a Windows 7 64 Bit operating system and is composed of 8 GB RAM and an Intel® Core™ i5-3470 4 x 3.20 GHz processor.

2.3. Theoretical Conception of the Passes’ Risk and the Passing Quality

The passing quality is composed of the pass risk and the passes’ outcome. Whilst the latter is a mere distinction in successful and misdirected or so-called bad passes, the former is a parameter developed in the course of this study. Hence, this new parameter may have to be adopted in future if the results suggest the necessity of this action. In relation to the following remarks, the pass risk was calculated as displayed in Equation (2) and corresponds mathematically to a summarisation of five factors in total.

\[ P_{\text{Risk}} = f_L + f_{\text{OPP}} + f_{\text{OPR}} + f_{\text{OPC}} + f_Z \]  

As it is shown in Section 3, not only the tool’s default values and underlying thresholds, but also additional coefficients are freely selectable via the tool’s Tool Dialog Box (TDB). This also includes the five pass risk factors. The first determinant \( f_L \) classifies the passes’ length. Presumed that longer passes are more likely to be intercepted by the opponent, they feature a higher risk. Furthermore, it is taken into account if opponent players put pressure on the passer and receiver (\( f_{\text{OPP}}, f_{\text{OPR}} \)). According to the
assumption, that the more opponents in close proximity to those two players, the higher the pressure and thus the higher the pass risk. For this, the current opponents’ positions are spatially joined with buffers around the passer and receiver at the moment of each pass event. Moreover, those positions are also spatially joined with the pass corridor, which is defined as a buffer around the virtual pass line (fOPC). The basic concept here is, that the more opponents are positioned between the passer and the receiver, the more likely a ball’s interception is. In this context through balls are taken into account implicitly. For better illustration, the buffered areas are depicted in Figure 2. In this context, it is most important, that not only the buffers around the passer and the receiver are spatially erased from the pass corridor, but also that the corridor’s buffer distance is less or equal the radii of the players’ close proximity buffers. Otherwise miscalculations cannot be excluded. Beyond that, the pitch’s zone is also taken into account based on the idea that the closer a pass to the own goal line, the more dangerous an intercepted pass and thus the higher the pass risk respectively (fZ).

Figure 2. Schematic illustration of the passer’s and receiver’s close proximity as well as the pass corridor (modified from [5]).

The pass risk also might be an indicator for bad passes. Apart from this, in combination with the passes’ outcome it facilitates conclusions about the players’ passing quality. For instance, at high risk a successful pass mirror high quality passing, whereas a bad pass is expectably and thus being tolerable under certain circumstances. Vice versa, at low risk a bad pass is an expression of fairly poor passing quality, whereas a successful pass can be expected, which is why its passing quality is rated lower than at high risk. Based upon these deliberations the classification of the passing quality ranges descending from very good, good and expectable for successful passes to acceptable, bad and very bad for bad passes.

2.4. Data Preparation Process

Since the raw data is provided as xml files, they have to be prepared in a GIS-appropriate manner in advance. For this, the so called Match Data Preparation tool was developed. Its functionality is comprehensively described in [11]. As all other designed tools in the scope of [5], it was realised utilising Python and can be executed as an ArcGIS custom tool. The tool’s output is exemplarily illustrated in Figure 3.

Figure 3. GIS-appropriate prepared raw data for one half. The colours symbolise different tracking objects and events respectively: (a) Tracking data; (b) Event data (data source [25]).
After the tool’s execution, each single tracking object and event category respectively are stored as point Feature Classes (FC) within an Esri File Geodatabase (GDB) named match_data.gdb. In it the data are physically divided into Feature Datasets (FDS) not only by their fundamental tracking and event data characteristic, but also by the game’s half times (…/[first|second]_half_{tr|vl}). According to Zeiler and Murphy [10] as well as Nasser [36] this structure is preferred for managing geo data. Despite the concept’s strict thematic segregation of data, for each half time the tool creates FCs containing all events chronologically ordered (…/[Half_ID]_all_events).

Furthermore, a fictional pitch based upon the dimensions illustrated in Figure 1 can be created optionally. Its point, line and polygon FCs are stored separately in a GDB named pitch.gdb. From a geographical perspective the pitch’s centre is located at the intersection point of the equator and the zero meridian line of Greenwich. Although practically in the middle of the Atlantic Ocean, there are no consequences in regard to the proposed analysis of the players’ passing quality according to Section 2.3.

For this, four FC are required. Besides the FC comprising all events a collectively tracking FC of the opponent team (…/[Half_ID]_[Team_ID]_Merge) is necessary. This consists of player-specific trajectories which are merged based upon their Team_ID ({A\|B}). The merging has to be conducted in advance by executing the custom tool Team Merge, which is described in [5]. Although primarily build upon the ArcGIS system tool Merge, the Team Merge tool’s advantage is that the merging process can be automatically accomplished for both teams as well as both half times at once.

Moreover, an analytical pitch zoning is required. Similar to the pitch’s FCs, in total five zoning alternatives, which are stored separately in another GDB (…/analysis.gdb/analysis), can be created by executing the Match Data Preparation tool. In this particular case a polygon FC is applied representing a pitch’s equal vertical fragmentation (…/thirds). With this the pitch can be distinguished in a defending, central and attacking zone in dependence upon the game direction.

3. Analyse Passes Tool

Due to limited space the tool’s Python code can neither be embedded within the text, nor attached as an appendix. Therefore, the tool’s code as well as a flowchart of its processes can be reviewed online on demand (Code S1, Figure S1). In relation to this, the tool’s functionality is described in Section 3.1 and 3.2 in as much detail as possible, whereby particular attention is paid to strict compliance with the defined labels among others of variables, FCs, lists and dictionaries within the code and flowchart respectively. Besides field names and classifications, these labels are highlighted in italic letters within the text.

The tool’s TDB is illustrated in Figure 4 and consists of 19 input parameters altogether. Whilst the tool’s first two parameters are mandatory, all others are optional. First, the path to the input GDB (match_data.gdb) has to be indicated. Second, the Player_ID representative for the player, whose passing should be analysed is requested. With the following parameters, the pass risk’s determinants stated in Equation 2 can be adjusted according to one’s own conception or based upon the coach’s specifications. The third and fourth input parameter correspond to break values for the passes’ length classification. By default, 10 m and 20 m are predefined. In addition to this, 3 m are preset as buffer distances around the passer’s and the receivers’ positions as well as around the virtual pass line. Furthermore, the passes’ length is weighted with predefined values of 0.33, 0.66 and 0.99 respectively for short, middle range and long passes. Those factors are also employed for one, two and three or more opponents within the selected passer’s and receivers’ proximity. Moreover, passes within the defending, central and attacking zone are weighted with the values 0.99, 0.66 and 0.33 respectively by default. Besides this, the pass risk determined by reference to the entered values can be classified in very high, high, low and very low, wherefore the thresholds 1.65 and 2.64 are predefined based upon first analysis results. Beyond that, the analysis process can be skipped for one of the two half times. Since all factors and break values are freely selectable, the Analyse Passes tool facilitates a wide range of configuration possibilities and thus complies with the interviewed experts’ request in this matter.

The tool’s execution duration primarily depends on the number of passes played by the selected player. Moreover, the utilised computer system’s configuration has to be taken into consideration.
With reference to the system outlined in Section 2.2 analysing 8 passes of a goalkeeper required 101 seconds for instance, whereas the evaluation of 39 passes played by a central midfielder took 5 minutes and 30 seconds. In this context it has to be noted, that according to the thesis’ main objective the verification of GIS’ functionality in the course of football game analyses was the main focus [5], whereas the performance of single analysis tools was of lesser importance in this particular context.

![Tool Dialog Box of the Analyse Passes tool](image)

**Figure 4.** Tool Dialog Box of the Analyse Passes tool [5].

### 3.1. Preliminary Measures

The Python script of the Analyse Passes tool requires the import of the following modules and sitepackages respectively. First, the sitepackage *arcpy* to gain access to the full functionality of ArcGIS. Whilst the module *os* is applied for miscellaneous tasks in regard of the operating system, the tool is aborted with the *exit* function of the module *sys* if necessary, for instance in the case of an incorrectly entered player ID. In addition to this, the modules *ntpath* and *csv* are imported for handling file paths and the creation of temporarily required *csv* files respectively.

In order to keep the tool’s configuration as simple as possible, solely the database and the player to be analysed have to be specified mandatorily. Hence, several measures are necessary before conducting the analysis. First, a workspace is defined which corresponds to the input GDB’s root directory. Inside this, a *dBase* file containing player specific information, which is an optional output of the Match Data Preparation tool, is located (*player_info_file*). This file is employed for an evaluation of the entered player ID. Afterwards, an output folder is created that stores written summaries of the analysis results as text files (*.../Passes*). Then, the input event data is located within the FDS according to the input parameters and their labels are stored in a list (*event_fc_list*). Furthermore, the data’s projection is retrieved in order to create an output FDS within a new output GDB (*.../passes.gdb/*[first|second]_half_passes*).
Moreover, the game directions for both teams and both half times are determined based upon the home team goalkeeper’s X-coordinate at the game’s very first frame. This approach is expedient, since the players are not allowed within their opponents half before the kick off. In consideration of Figure 1, the home team plays from right to left whereas the away team plays from left to right if the coordinate value is greater than 0 and vice versa. Since a change of sides is commonly scheduled after the half time break [30], the defined teams’ game directions are swapped for the second half. Beyond that, the selected player’s team and jersey number (number_term) are read out from the player_info_file.

3.2. Iterative Analysis Process

The iterative analysis process relates to the event_fc_list. As the Half_ID ({H1|H2}) is part of the FC name, the player’s game direction as well as the output FDS can be assigned. The path of the input FC (event_fc) is compiled of the read out path of the input FDS as well as the current loop variable. This FC is then copied to the output FDS (new_event_fc), where it is attributively joined with the player_info_file via the fields Actor_1 and Player_ID. From now new_event_fc contains all information necessary to distinguish passes from bad passes. For this, the Team_ID of both, the pass events and their subsequent events are compared. If the IDs coincide the pass is classified as successful, if not, it is regarded as a bad pass. Table 1 illustrates the required fields for analysing the passing play.

<table>
<thead>
<tr>
<th>Field name</th>
<th>Characteristic</th>
</tr>
</thead>
<tbody>
<tr>
<td>P Frame</td>
<td>Frame at the time of the pass</td>
</tr>
<tr>
<td>Passer</td>
<td>Player_ID of the passer</td>
</tr>
<tr>
<td>P Team</td>
<td>Team_ID of the Passer</td>
</tr>
<tr>
<td>Passer X</td>
<td>X-coordinate of the passer at the time of the pass</td>
</tr>
<tr>
<td>Passer Y</td>
<td>Y-coordinate of the passer at the time of the pass</td>
</tr>
<tr>
<td>Receiver X</td>
<td>X-coordinate of the receiver at the time of the pass reception</td>
</tr>
<tr>
<td>Receiver Y</td>
<td>Y-coordinate of the receiver at the time of the pass reception</td>
</tr>
<tr>
<td>Passing</td>
<td>Differentiation between pass and bad pass</td>
</tr>
<tr>
<td>Pass_ID</td>
<td>Consecutive ID to identify each pass event</td>
</tr>
</tbody>
</table>

These attributes are stored in a list (pass_event). While iterating through new_event_fc events named Pass or Cross are searched and if found, the associated attributes of the passers are retrieved. Besides this, the passes’ OBJECTID (pass_oid) is recorded. Then, still within the same loop new_event_fc once again is iterated, whereby the attributes of the event where OBJECTID = pass_oid + 1 relate to the receiver and thus are read out. After the pass’ classification its attributes are appended to the list pass_event. Once the iteration of new_event_fc is over, pass_event is examined if pass events occurred at all. Only then a temporary csv file (csv_file) is derived from this list which serves as an input table for the creation of fictional passes’ lines (passes_l). Moreover, csv_file is converted to a dbase file (dbf_file) which is attributively joined with passes_l via the field Pass_ID subsequently. Thereafter, both the csv_file as well as the dbf_file are no longer of use and thus are deleted.

Following this, passes_l is attributively extended with a new field named Length_Class, which is updated considering the field Shape_Length according to the input parameters. Within this all passes are categorised as short, middle or long. Then, the line features’ start and end points are derived as point features (points). Moreover, this new FC is attributively extended with a new field named Point_Type, in order to distinguish the passers’ and the receivers’ point data. This attribute is updated with the strings start and end based upon a counting. If the count value is odd the features correspond to starting points and vice versa end points.

Afterwards, buffers around both, the line (line_buffer) and points (points_buffer) are created applying the adjusted buffer distances. Furthermore, in dependence upon the selected player’s Team_ID the corresponding team-specific merged point FC (merge_fc) is located within the input
GDB. Before the spatial joins can be conducted, the creation of a list containing all passes' start frames is necessary (start_frame_list). This list is then iterated, whereby where clauses are derived from the current loop variables in order to limit the features in the course of the creation of feature layer from points_buffer (layer_1), merge_fc (layer_2) and line_buffer (layer_3). According to Figure 2, layer_1 is then erased from layer_3 to avoid spatial overlapping analysis areas (pass_corridor). Following this, both, layer_1 and pass_corridor are spatially joined with layer_2 to determine the number of opponents within these areas. Since the counting has to be inevitably conducted separately for each pass, the individual join outputs have to be merged subsequently. As a result, only two FC remain, and these represent the passer's and receivers' close proximity (merge_fc) as well as all pass corridors (merge_cor_fc). After cleaning up these two FC from unnecessary fields each single join output FC besides a number of other FC (new_event_fc, points_buffer, line_buffer, pass_corridor) are no longer of use and thus deleted immediately.

In order to create an attributively connection between merge_fc and points, a new field named Point_ID is added to both FC. It is a concatenation of the fields Pass_ID and Point_Type. Whilst Point_ID serves as a join field between points and merge_fc, Pass_ID is employed for an attributive join between points and merge_cor_fc. As a result of these twofold join process, points contains all information gathered so far, while merge_fc and merge_cor_fc can be deleted. Within points, the counting values for each feature are assigned to new fields named Near_Opponents and Near_Opponents_Cor. Then, thirds is located within the analysis.gdb and further spatially intersected with points (int_out_feat). Following the deleting of unnecessary fields of int_out_feat, the passes' initial zones are classified within the field zones applying the abbreviations own, mid and opp by taking the game direction into account. From this moment forth, points is of no use anymore and thus deleted.

The determination of the passes' risk is related to int_out_feat, which for this is attributively extended by the fields Pass_Risk and Pass_Risk_Value. Since the assignment of new attribute values affects the starting points only, a list with the Pass_ID of those points is generated while iterating through int_out_feat. Based upon the previously gathered factor values the pass risk is calculated according to Equation (2), then transferred to Pass_Risk_Value and conclusively classified within Pass_Risk. Next, a new field named Pass_Quality is added to int_feat_out, which is populated with the predefined categories in dependence upon the distinction between passes and bad passes and the already classified Pass_Risk. Afterwards, int_feat_out and passes_l are attributively joined via the field Pass_ID in order to pass the fields zones, Pass_Risk_Value, Pass_Risk and Pass_Quality to the line features. While iterating through passes_l further information about the passes are collected with that the pass completion rate (pass_rate), the bad pass rate (bad_pass_rate) and the average pass risk (avg_risk_value) are calculated. Finally, the accumulated information is summarised which is not only displayed via the Results Window, but also stored as a text file within the designated output folder (.../[Half_ID]/[Team_ID]/[number_term]/passes.txt).

3.3. Evaluation of the Tool’s Functionality

The tool’s functionality was evaluated conducting an expert survey, followed by a multilevel qualitative content analysis. Detailed information about its design can be found in [5], whereas the evaluation’s fundamental concept is briefly outlined hereafter. It has to be mentioned that it is impossible to evaluate the presented method and its implementation as an ArcGIS custom tool quantitatively as its reasonability basically depend on the user. Hence, a qualitative method has to be employed in this particular case. For the survey’s purpose, the tool’s TDB and results were presented to eight professional game analysts from Austria and Germany. As among other professionals they also participated in a previous expert survey in late 2014 on contemporary football game analysis’ methods and techniques also conducted by the author in the course of his thesis [5]. Hence, this research project was not unknown to the experts. However, none of them had worked with GIS at the time of both surveys.

The second survey was conducted from 9 June 2016 to 4 July 2016. Whilst three experts were met personally, the others were interviewed online via Skype. Regardless of that, all conversations were
recorded acoustically not only to archive them, but also to produce interview protocols. According to Bogner et al. [37] these are content-related summaries and suitable for the evaluation of interpretative knowledge which is not restricted to individuals but can also be shared collectively. Furthermore, the conducted interviews can be described as partly standardised as well as explorative [37]. Their guideline ensures the experts’ feedback completeness and contains questions about the tool’s parameters, potential thematic adaptations and its benefit in the course of football game analysis.

The protocols were assessed with a qualitative content analysis according to Mayring [38]. Since this methodology is object related [38], adjustments in respect of the particular analysis purposes are necessary [37-39]. For Gläser and Laudel [39] the aim of this approach is to type and to configure the texts’ key messages. For this, relevant information is systematically extracted from the protocols and classified in terms of a textual structuring [38]. In addition to this, a frequency analysis was conducted concurrently in order to weight the experts’ collective statements. Although, a quantitative method, it is reasonable if generalisations are derived [38]. Based on this, the tool’s functionality in regard to the determination of the passing quality on single player level was finally evaluated.

4. Results

This section is divided into two main parts, beginning with a description of the tool’s outcome. This is followed by summarised results of the expert survey and the answer to this study’s central question about the tool’s functionality in the course of professional football game analyses.

4.1. Tool’s outcome

The tool’s output is exemplarily illustrated in Figure 5 representing the passes played by a wide midfielder on the right side during the game’s first half. Besides the passes’ length and spatial distribution, their outcome as well as their risk and quality are displayed.

![Diagram of passes played by a midfielder](data:image/png;base64,iVBORw0KGgoAAAANSUhEUgAAAIwAAAD2CAIAAAB8XbAaAAAABGdBTUEAALGPC/xhBQAAAgAElEQVR42m2Y3wBwDQAQfVEQyNnOkAA9gAAAAAElFTkSuQmCC)

**Figure 5.** A midfielder’s passes’ location and quality during the first half [5] (data source [25]).

The passes are depicted as arrows. Their colour indicates the passes’ outcome, utilising an associative colouration, whereby green and red signifies successful and bad passes respectively. Furthermore, the pass directions mirror the proportion of vertical and horizontal passes. Since the line features are created as Euclidian distances between the passer’s and the receivers’ positions,
those lines might not represent the ball’s curve exactly. This may in particular be the case for long and high passes such as crosses inside the penalty area or kick outs by the goalkeeper. However, this possible spatial fuzziness does not affect the analysis considerably. In order to identify each pass, they are labelled with their Pass_ID at the lines’ centre, which also serves as a link to further information contained within the FCs’ attribute table. Beyond that, the pass risk as well as the passing quality are illustrated at the passer’s positions as point symbols. Whilst the pass risk is depicted by the size of the arbitrarily graded points, the passing quality is indicated by colour.

Detailed information about the selected player’s passes can be either received interactively or readout from the created text file. In relation to Figure 5 its associated text file is illustrated in Figure 6, where the passing of player 345596 is summarised in written form.

Although text files fulfil the thesis’ particular purpose [5], alternative output formats are worth to consider in dependence upon the intended aim. For example, the tool’s script could be adopted so that the extracted and determined information are embedded in a comprehensive game analysis report automatically. For this, appropriate Python modules and libraries such as the Plotly Python Library for creating diagrams or Report Lab’s PDF Library and the sphc module for the report’s export as PDF- and HTML-file respectively are available [5]. In this context, the benefit of online solutions, perhaps applying ArcGIS Online facilitate an interactive access to the analysis results, provided that this approach is requested.

In order to speculate about the causes of bad passes, their parameters are itemised within the written summary. Furthermore, the pass completion rate, the average pass risk, the number of passes in each pitch’s third as well as in regard of their length are stated. Moreover, the percentage of passes played with a certain quality is listed. As stated in Figure 6 the pass completion rate of player 345596 was 81.82 % in the first half. Hence, more than four fifths of all his passes were successful. However, this and other common player specific values have to be interpreted in relation to the player’s tactical position [6]. This value is a sign of a fairly high pass accuracy, however, to derive statements about the passing quality from this might be elusive. Therefore, a closer look at the summary’s conclusion is advisable. According to the determinations outlined in Section 2.3 not even one pass of player 345596 was very good, though 22.22 % and 77.78 % were good and expectably successful respectively. In terms of the player’s bad passes, exactly half of them were bad and very bad respectively.

In contrast to the pass completion rate, the introduced passing quality corresponds to a player’s feature independent of the tactical position within the team and thus is more appropriate for comparing different players. For this purpose, the overall passing quality of a goalkeeper, a wide right defender, the previously analysed midfielder and a right forward are exemplarily illustrated in Figure 7.

Figure 6. Written summary of the midfielder’s passing during the first half [5] (data source [25]).
Figure 7. Overall passing quality of four players distinctive in regard of their tactical position (modified from [5], data source [25]).

The data refers to the game’s first half, whereby the numbers in brackets after the players’ ID’s indicate the total number of passes. Since the pass completion rate is also displayed implicitly, the bar diagram coincides with the findings of Carling et al. [6], whereupon single player parameters have to interpreted in dependence upon tactical positions. This approach is reasonable as forwards are usually marked more frequently for instance, especially within the attacking zone, which is why their passes are comparatively risky. Referring to Figure 7 the goalkeeper 348837 displays the best passing performance by far followed by the defender 347300. Although the pass completion rate of the forward 341918 is slightly lower than those of the midfielder 345596, his overall passing quality is marginally higher.

In relation to Figure 7, Figure 8 illustrates the associated spatial distributions of the four players’ passes in the first half.

Figure 8. Comparison between the passing of four players distinctive in regard of their tactical positions for the first half (modified from [5], data source [24]).
Across the four selected players, the forward’s passing was the most spatially variable in consideration of the tactical positions. In spite of that, the four depictions suggest that the deeper the passes are played in the opponents’ half, the more likely is a bad pass. This seems to be plausible, as it is obvious that zonal defence is more small-scale in proximity to the goals. However, the pass risk’s zonal factor $f_z$ is diametrically opposed to that, as it is based upon the assumption that the closer a pass is to the own goal line, the more dangerous a bad pass becomes. Hence, it is possible that this zonal factor may alleviate the tool’s result, which is probably why in reference to Figure 7 very good passes are scarce. In order to reassess this, an extensive survey analysing a bigger data base is suggested.

Beyond comparisons between the passing of different players, the tool is also suitable to contrast one player’s passing separated by half times. For this purpose, Figure 9 illustrates once again the spatial and quantitative features of player’s 345596 passing, though unlike Figure 5 for both playing periods. With regard to the passes’ spatial distribution, it is noticeable that the player’s passing was horizontally more dispersed in the second half. In this case, it might be of interest to categorise the passes according their directions. In this context, the creation of four circular sectors of 90° each around the players is feasible. With this a spatial join could be conducted with the ball’s position at the time of the frame subsequently to the pass events in order to determine the passes’ directions.

![Figure 9](image)

Figure 9. Comparison between midfielder’s passing in the first and second half (modified from [5], data source [25]).

Moreover, the player’s overall passing quality was far better in the second half, as the bar diagram displays in Figure 10.

![Figure 10](image)

Figure 10. A midfielder’s overall passing quality for both playing periods (modified from [5], data source [25]).

Although the majority of passes was expectably successful only, no pass was very bad as it was the case in almost 10% of all passes within the first half. Notwithstanding this, for key situations it is advisable to extract further information interactively and assessing the game’s spatial structure at the
time of each pass. This approach could be enhanced by animating the players’ and the ball’s movement for a certain period of time before and after selected pass events. For this purpose, the Playback Manager of the ArcGIS extension Tracking Analyst is most appropriate utilising the Frame as time field [5]. By contrast, applying the field Time_Sample juddering was observed which may be a consequence of the data field’s more complex structure.

In conclusion, this subsection demonstrates that compared to the commonly applied pass completion rate the tool provides a deeper insight into the passing of single players. However, the pass risk, the passing quality and the pass completion rate cover specific aspects of the players’ passing and thus are legitimate as single parameters. Therefore, all of them have to be taken into account for analysing the passing and its interpretation, considering their specific significance though.

4.2. Result of the Tool’s Evaluation

The experts’ feedback concerning the tool’s functionality was positive without exception. In the following the key findings are briefly outlined. The associated interview protocols on which the evaluation was based on can be found in the appendix of [5]. According to the consulted game analysts, the tool is a detailed and practical extension to common pass analyses on single player level. Furthermore, its outcome was described as graphically well prepared. Moreover, the possibility of examining single passes interactively in the context of the current game’s spatial structure was received favourably. Besides the profound observation of the players’ passing, the tool’s results also enable the analysts to characterise the game’s situation at the time of the passes but also in dependence upon the pass performing body part. In regard of the passes’ directions, the number of opponent players within the pass corridor can be considered as an indication concerning this matter according to the experts. Nonetheless, it is advisable to capture this information parametrically as it was suggested in the previous subsection. Moreover, the proposed utilisation of multiple ring buffer instead of single buffer for the determination of the performed pressure on the passer and the receiver, was unanimously approved of. This would enable the users to further grade the players’ close proximity and thus refine the results. However, for this purpose classifications based upon tactical positions are necessary. Beyond that, the majority of the experts also suggested to take the players’ speed and movement direction into account, since these two factors have a certain effect on the possibility of intercepting the ball especially within the pass corridor.

In addition to these recommendations, the acquisition of the players’ intention is most desirably but is also most challenging. Furthermore, it is of great interest not only what happens after the passes but also if alternative passes would have been better. Apart from this, the player specific passing quality may be extensible to a team specific parameter.

5. Discussion and Conclusions

Although recommendations for potential or necessary adaptations were obtained from the consulted experts, the tool’s evaluation provides strong evidence that its functionality is given in the context of professional football game analyses. However, it is obvious that further studies are inevitable to overcome remaining issues. Beyond that, one of the thesis’ main findings was that the success of GIS in the scope of football game analysis depends on the establishment of an interface between GIS-based analysis and the predominant video analysis. “As both approaches have their shortcomings, a combination of both would alleviate these.” [5] Since it is common to time tag and thematically
classify certain events within footages of football games an attributive connection to tracking and event data is practically existing. Hence, this issue is solely a subject of the technical implementation. Although this suggestion concerns GIS technology in general, it also pertains to all developed tools in the thesis’ scope, including the Analyse Passes tool. Furthermore, a rearrangement of the tools based upon free and open source GIS such as QGIS would probably be more attractive especially for financially weaker clubs. In this context it is also worth considering to create an independent graphical user interface comprising several tools in order to provide customised analysis solutions according to the respective analysts’ requirements.

In absolute agreement with the experts’ feedback, the validation of the parameters formula is supported. In this case, especially the impact of the zonal factor on the calculation of the pass risk is questionable due to the appearance of contradictorily assumptions in Section 4.1. Furthermore, instead of buffering the virtual pass lines between passer and receiver, the ball’s true curve would certainly depict the passes in greater detail. Moreover, this would prevent incorrect spatial join results in individual cases where the ball’s path is extremely curved. Independently, it is reasonable that not only the tool’s validation but also its further development should be conducted in close cooperation with the target group consisting of professional game analysts, coaches as well as sports scientists in a wider sense. The experts’ feedback does not contain any utopian recommendations in the context of the tool’s adaption. In terms of the parametrical determination and classification of the passes’ directions, a solution was already considered briefly in Section 4.1. This approach is theoretically also suitable for the detection of the players’ movement directions. Besides this, taking the players’ speed and body parts into account is possible without great efforts, because this information are already provided as attributes within the tracking and event data. Moreover, the application of multiple ring buffers instead of single buffers is only a question of choosing the applicable ArcGIS system tool.

To sum up, this article introduced a novel approach applying GIS technology for the assessment of passing quality on single player level in association football, which exceeds the often overrated pass completion rate’s significance. In this context the passes’ outcome is related to their risk, which is determined by a spatial analysis of the game’s current spatial structure at the time of each pass. The notation novel indicates that the Analyse Passes tool’s development is not concluded yet. Therefore, we will continue research not only in this particular case but also in general in the field of GIS-based football game analysis based on the results and findings presented in [5].

**Supplementary Materials:** Please request the author to forward the password for the download of the following online resources at [https://ucloud.univie.ac.at/index.php/s/izl5hERO8QwkB8v](https://ucloud.univie.ac.at/index.php/s/izl5hERO8QwkB8v), Figure S1: Flowchart of the Analyse Passes tool, Code S1: Analyse_Passes.py

**Acknowledgments:** We gratefully acknowledge the provision of the data base by ProzoneSports.

**References**

8. Longley P.A.; Goodchild M.F.; Maguire D.J.; Rhind D.W. Geographical Information Systems and Science, 2nd ed.; John Wiley & Sons: Chichester, UK, 2005
27. Sester M.; Feuerhake U.; Kuntzsch C.; Zourlidou S. Interpretation of Moving Point Trajectories. Photogrammetric Week 2015, 15, 329-341
39. Gläser J.; Laudel G. *Expertinterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierter Untersuchungen*; Verlag für Sozialwissenschaften: Wiesbaden, Germany, 2004