Analyzing Performance and Playing Style Using Ball Event Data

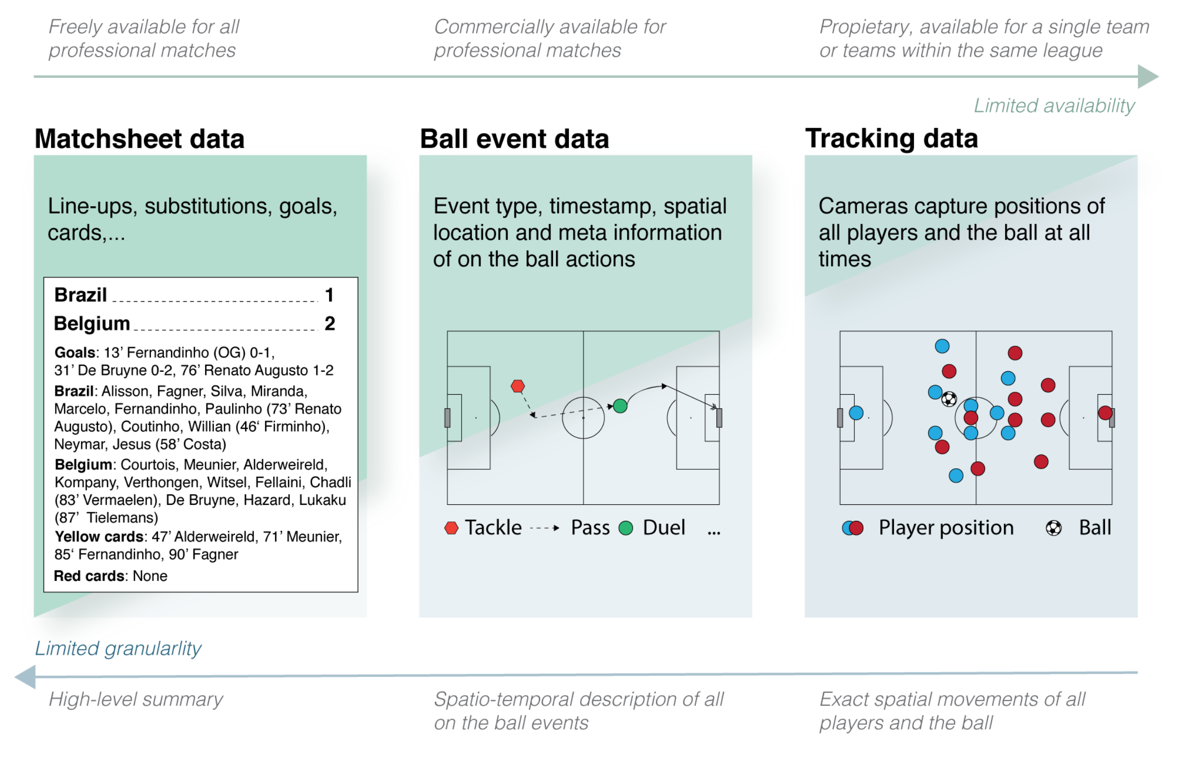
Jan Van Haaren, Pieter Robberechts, Tom Decroos, Lotte Bransen, Jesse Davis

The data that are collected during football matches can help clubs answer the multitude of questions that they are faced with each day. The ever increasing availability of ball event data, which describes the actions that the players perform with the ball, has led to a wide range of approaches to analyze the performance and playing style of players and teams in matches. This section provides an overview of the most important approaches in this area to date.

Introduction

Every day, football clubs are faced with a multitude of important questions that arise from a diverse set of groups within the club, ranging from the technical staff to the recruitment department. For example, a manager might want to know how many goals the striker should have scored given their goal-scoring opportunities, how many saves the goalkeeper should have made given the shots on target faced, or what strategy and tactics the team's upcoming opponent will use. Similarly, a scout might want to know what impact the actions of a potential signing would have on the team's performances, or whether a potential signing would fit the team's playing style.

To better answer the aforementioned questions, football clubs are collecting increasing amounts of data during matches. As shown in Figure 1, the technical data that is currently collected during matches can be broadly divided into three types: matchsheet data, ball event data and tracking data. Matchsheet data provide high-level information about matches such as line-ups, substitutions, goals and cards. Ball event data describe the actions that players perform with the ball such as passes, dribbles, interceptions, tackles and shots. Tracking data provide the exact spatial locations of the players and the ball at all times.


Freely available for all professional matches
Commercially available for professional matches
Propietary, available for a single team or teams within the same league
Limited availability
Matchsheet data
Ball event data
Tracking data
Line-ups, substitutions, goals, cards,...
Event type, timestamp, spatial location and meta information of on the ball actions
Cameras capture positions of all players and the ball at all times
Tackle
Pass
Duel
...
Player position
Ball
Limited granularlity
High-level summary
Spatio-temporal description of all on the ball events
Exact spatial movements of all players and the ball

Brazil 1
Belgium 2

Goals: 13’ Fernandinho (OG) 0-1, 31’ De Bruyne 0-2, 76’ Renato Augusto 1-2

Brazil: Alisson, Fagner, Silva, Miranda, Marcelo, Fernandinho, Paulinho (73’ Renato Augusto), Coutinho, Willian (46‘ Firminho), Neymar, Jesus (58’ Costa)

Belgium: Courtois, Meunier, Alderweireld, Kompany, Verthongen, Witsel, Fellaini, Chadli (83’ Vermaelen), De Bruyne, Hazard, Lukaku (87’ Tielemans)

Yellow cards: 47’ Alderweireld, 71’ Meunier, 85‘ Fernandinho, 90’ Fagner

Red cards: None

Figure 1. The data collected about football games can be broadly divided into three types: matchsheet data, ball event data and tracking data. Ball event data strikes a balance between matchsheet data and tracking data in terms of granularity and availability.

The three types of football data differ not only in their granularity but also in their availability. Matchsheet data, which provide only limited high-level summaries about what happened in a football match, are available for virtually all professional and semi-professional football matches in the world. In contrast, tracking data, which provide the highest level of detail possible, are available only for a restricted number of competitions, mostly top divisions of the better-ranked European countries. Interestingly, ball event data, which attempts to strike a balance between the limited matchsheet data and the detailed tracking data, has become widely available in recent years.

Ball event data has become an increasingly popular data source for developing football analytics tools for three reasons. First, ball event data is easier to process and analyze than tracking data due to its smaller volume and easier structure. Second, ball event data is extremely interesting for player recruitment purposes due to its wide coverage of players in smaller competitions as well as important youth competitions. Furthermore, ball event data can be purchased from specialized companies, whereas tracking data is typically only available to the teams in the league. Third, ball event data is becoming increasingly information rich. In addition to describing the actions with the ball, data collection companies have recently started registering the locations of the relevant players at the time of key events such as shots.

In the remainder of this section, we will first describe the properties of ball event data in more detail and then discuss a number of metrics for performance analysis and playing style analysis that leverage ball event data.

Ball Event Data

Ball event data describe the most conspicuous events that occur in a football match. On one hand, the data describe the actions that the players perform with the ball, including passes, dribbles, crosses, interceptions, tackles, and shots. On the other hand, the data describe other important events such as fouls and cards. Ball event data are typically collected by humans who watch video feeds of matches through special annotation software. For each noteworthy event on the pitch, the annotator collects a timestamp, the location on the pitch (i.e., a (x,y) position), a type (e.g., pass, cross, or foul), and the players who are involved in the event. Depending on the type of the event, the annotator collects additional information such as the end location for a pass and the outcome of a tackle.

Figure 2 shows a visualization of thirteen consecutive events that led to Lionel Messi's goal for Barcelona in their LaLiga win over Getafe on January 6th, 2019. The sequence starts with Getafe’s Vitorino Antunes throwing the ball to his teammate Dakonam Djené (event 1) whose subsequent pass is intercepted (event 2) with Barcelona’s Clément Lenglet heading the ball to his teammate Ousmane Dembélé (event 3). This marks the start of a string of passes by Barcelona with Dembélé and Jordi Alba (events 4 and 5) advancing the ball with a give-and-go after which Dembélé passes to Arthur Melo (event 6). The Brazilian's subsequent high pass (event 7) is cleared by Getafe's Damián Suárez (event 8). However, Lionel Messi (event 9) recovers the ball and shoots, but Getafe goalkeeper David Soria (event 11) manages to parry the ball. This is followed by Messi immediately recovering the ball (event 12) and scoring with a tap-in (event 13).

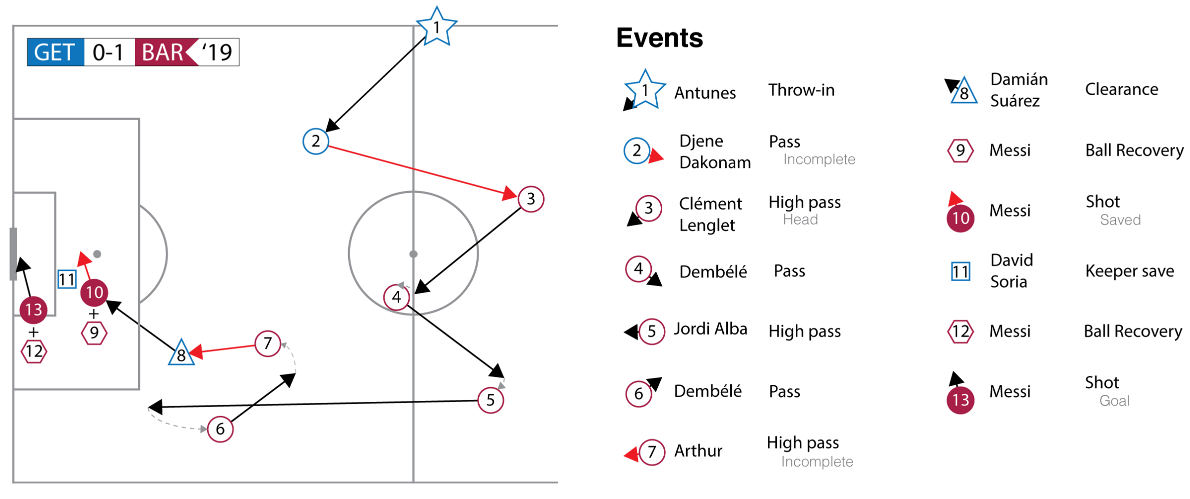


Figure 2. Visualisation of the events preceding Lionel Messi’s goal in the game against Getafe on January 6, 2019.. The dotted lines represent movements of the ball that are not explicitly modelled in the event stream.

Although the process of collecting ball event data is error-prone and time-consuming, the quality and information richness of the data has improved considerably in recent years (Bialik, 2014; Liu, Hopkins, Gómez, & Molinuevo, 2013). Data collection companies such as Opta, InStat, Wyscout, StatsBomb and Sportlogiq are increasingly annotating more types of events as well as more information about each event. For instance, some providers nowadays annotate separate events for putting pressure on the player in possession of the ball. Furthermore, some providers also annotate the locations of the relevant players at the time of key events such as shots. These additions provide a better understanding of what happens on the pitch and thus enable a more in-depth technical and tactical analysis of the game.

Performance Analysis

The low-scoring nature of football poses significant challenges for analyzing and measuring of teams' and players' performances. Sports media typically report basic performance statistics such as distance covered, number of assists, number of saves, number of goal attempts, and number of completed passes. While these statistics provide some insights into the performances of teams and players, they largely fail to account for the circumstances under which the players performed their actions. For example, successfully completing a forward pass deep into the opponent's half is both more difficult and more valuable than performing a backward pass in your own half without any pressure from the opponent whatsoever.

Fueled by the availability of more extensive ball event data, football analytics researchers and enthusiasts have proposed several metrics for valuing shots, saves, passes and other types of actions. First, people looked at expected-goals (xG) metrics to objectify shot quality and goalkeeper skills. More recently, the sports analytics community has extended the concept of expected goals to quantify passes and the ability to carry the ball into dangerous areas of the pitch. These metrics assign values to actions other than shots and allow for the comparison of the attacking contribution of all players. We now discuss each of these in more detail. Although we focus on assessing the actions of individual players, the same metrics could be used to analyze the performance of both players and teams in matches by aggregating the data.

Shots

Motivated by the fact that some shots are better than others, the expected-goals (xG) metric, which has received significant attention, attempts to quantify the quality of a goal-scoring opportunity. To do so, an expected-goals model assigns a value between zero and one to each shot which represents the probability that the shot will result in a goal. Building an expected-goals model requires addressing three key points:

1. What game situations (e.g., penalty, shot from open play) should the expected-goals model distinguish among?
2. How can relevant characteristics of the shot be defined from the event data?

3. How can the expected-goals model be learned from the data?

The game situation in which the shot takes place affects the chance that it will result in a goal. For example, a penalty is more likely to result in a goal than a shot arising from a counter attack. Some approaches ignore the game situation (Decroos, Dzyuba, Van Haaren, & Davis, 2017). However, it is far more common to distinguish among shots arising from a small number of distinct game situations and build one expected-goals model for each one. Distinguishing among different situations better captures the context of the opportunity, which in turn results in more accurate models. Caley (2015) considers six different shot types: Shots from direct free kicks, shots following a dribble past the keeper, headed shots assisted by crosses, headed shots not assisted by crosses, non-headed shots assisted by crosses, and non-headed shots not assisted by crosses. In contrast, IJtsma (2015) considers ten different shot types: open-play shots, open-play headers, penalties, shots from direct free kicks, shots following an indirect free kick, shots arising from a corner, shots arising from a throw-in, shots following a rebound from a save, shots following a rebound from the woodwork, and shots during counters.

Defining features about the shot that are indicative of its quality is the key step in building an expected-goals model. Moreover, this step is where domain-knowledge about football comes into play. One category of features captures characteristics of the shot itself such as the location on the field, the distance to the goal, and the shot angle. Another category of features is based on considering the actions that preceded the shot. One such feature is whether the previous action was a completed dribble. By definition, this entails removing the on-the-ball defender from the play, which should make the shot easier. Similarly, whether the previous action was a through-ball also increases the quality of the shot. A huge number of features have been considered; both Caley's and IJtsma's blogposts provide detailed descriptions of possible features.

Ultimately, an expected-goals model needs to assign a real-valued number between zero and one to each goal scoring opportunity. Typically, this is done by applying a data-driven machine learning technique to a large historical data set of goal scoring opportunities that contains the observed features for each one, as well as the true result (i.e., goal or no goal). In a nutshell, machine learning techniques analyze historical data in order to learn the predictive relationship between different observed values for the features and the observed outcome of the goal opportunity. From a technical perspective, any model that returns a probability is suitable, but typically either a logistic regression model or gradient boosted tree model is used in practice. The key difference between a logistic regression model and a gradient boosted tree model is that the latter model can represent more fine-grained differences between goal scoring opportunities (Decroos, 2019/2019). The positive examples are the shots that resulted in a goal and negative examples are those that missed (e.g., off target, blocked, saved).

Saves

Viewing shots from the goalkeeper's perspective, the expected saves (xS) metric attempts to quantify the probability that a goalkeeper will prevent a shot on target from resulting in a goal. To do so, an expected-saves model assigns a value between zero and one to each shot on target which represents the probability that the shot will *not* result in a goal. Both expected-saves models and expected-goals models typically use a similar set of features to describe the characteristics of a goal-scoring opportunity. The key difference is that expected-saves models only consider shots on target, while expected-goals models consider both shots on target and shots off target.

Yam (2019a) introduces a framework to evaluate a goalkeeper's shot-stopping skills by estimating how many goals above or below average a goalkeeper prevented or conceded. To do so, the framework compares the actual number of goals conceded with the expected number of goals conceded. To compute the expected number of goals conceded, the framework leverages a post-shot expected-goals model. Unlike traditional expected-goals models, the model only considers shots on target since blocked shots and shots off target can never result in a goal. Furthermore, it accounts for the characteristics of the actual shot as well as the locations of the defenders.

Passes

The fact that expected-goals models and expected-saves models only assign values to shots is problematic for two reasons. First, shots arise very infrequently, particularly compared to other types of actions such as passes, crosses or dribbles. Second, focusing just on the shot ignores the contributions that other players made to creating the attempt. These observations have motivated attempts to measure the impact of actions other than shots on the scoreline as well. These models are sometimes called non-shot expected-goals models.

Since passes constitute the lion's share of the actions that happen in a match, passes have been a topic of particular interest in recent years. An important limitation of traditional pass-based statistics is that they fail to appropriately account for the circumstances under which a player performs a pass. For example, the percentage of successfully completed passes does not distinguish between a pass between two central defenders in their own half and a pass by an attacking midfielder who tries to reach a forward in the opponent's penalty area. The latter pass is at the same time both more valuable and more likely to fail. With data about passes becoming more widely available, several methods for evaluating passes have been proposed in recent years.

Most methods for valuing passes attempt to quantify a player's involvement in creating goal-scoring chances. Traditional, pass-based metrics usually only reward passes that result in goals. However, very few passes can be valued according to this criteria. Therefore, like for shots, methods for valuing passes typically resort to measuring their *expected* impact on the scoreline based on historical observations. The traditional approach to address this task is to compute the difference between the value of possessing the ball in the location before the pass and the value of possessing the ball in the location after the pass. The differences between methods arise in how they determine the value of possessing the ball in a particular location.

At a high level, three approaches to determine the value of a pitch location have been proposed. The first approach assigns each pitch location an expected-goals value that is the xG value if a shot would have been attempted from the location (Michalczyk, 2018). This approach is particularly appropriate for passes close to the opponent's goal. However, passes further away from the goal will all have expected-goals values close to zero. The second approach is to determine the proportion of play outs from a pitch location that result into a goal within a given number of actions or seconds (Bransen, 2017; Bransen & Van Haaren, 2018; Gyarmati & Stanojevic, 2016). The challenges are to determine the optimal number of actions or seconds to look ahead and to measure the similarity between play outs. For instance, a pass during a slow build up will likely have a different value than a pass during a fast counter attack. The third approach is to distribute the reward of a possession sequence (e.g., a goal) starting in a given pitch location to its constituent passes (Brooks, Kerr, & Guttag, 2016). The challenge is to decide on the optimal weighting scheme to distribute the credit across all the sequence's passes. Typically, passes at the end of the possession sequence receive more credit than passes at the start of the sequence.

Actions

Naturally, the interest in assessing passes led to the desire to evaluate more actions. While a number of different approaches have been proposed that evaluate a large number of actions, at a high-level they all function in the same way. When the team possesses the ball, each action within that possession is undertaken with the high-level objective of helping the player's team, either by increasing the chance that his team scores or decreasing the chance that the opposing team scores. The practical result of each (successful) on-the-ball action such as a pass or dribble is a change in the ball's location on the pitch. Intuitively, certain locations on the pitch are more valuable than others as they more readily lend themselves to generating goal scoring opportunities. For example, possessing the ball near the sideline close to the midfield line is generally not as threatening as possessing the ball in the centre of the pitch just outside the opponent's penalty box. This suggests that the value of an action can be derived by simply taking the difference between the value of the ball's new location (i.e., where the ball ends up as a result of the action) and the ball's original location. The value of a location can be thought of as the probability of team scoring during its current possession given the location where the ball is currently possessed.

Metrics such as Valuing Actions by Estimating Probabilities (VAEP) (Decroos, Bransen, Van Haaren, & Davis, 2019), xG Added (Mackay, 2017), xG Threat (Singh, 2019) and Attacking Contributions (Yam, 2019) all exploit this type of reasoning. The differences arise in the technical modelling choices made to value the different locations. One standard idea is to view the possession as a Markov model. This involves discretizing the pitch into zones. Moreover, the model also assumes that previous actions will have no effect on how the rest of the possession will play out. That is, it would not differentiate between receiving the ball in the centre of the pitch just outside the penalty area via a long through ball versus receiving the ball in the exact same location via a short, lateral pass. Another approach involves training a machine-learned model to predict the probability that the team possessing the ball after an action will score in the near future (e.g., the next five to ten actions). This enables reasoning about the characteristics of past actions.

Playing Style Analysis

A recurring concept when discussing football is the style of play, which is applicable on both the player and team level. On the player level, this refers to a player’s behaviour on the pitch. For example, both Messi and Ronaldo are great players, but each one approaches the game in a different way. On the team level, this manifests itself in terms of the tactics the team employs. Naturally, a player’s behaviour is inherently linked to his team’s tactics. There is substantial value in gaining a better understanding of playing style as this can be leveraged in areas such as player scouting and match preparation. Simple descriptive statistics such as pass percentage or shot count are usually insufficient to capture playing style. Hence, there has been an explosion of interest in applying automated techniques to try to glean insights into both player and team behaviours.

Player Behaviour

Analyzing player behaviour can add substantial value to three important processes at professional football clubs. The first process is scouting. There are a huge number of players to scout, and clubs have a finite amount of resources to devote to this task. Using data to intelligently winnowing the list of prospective targets by identifying players whose style fits the team’s ethos would be very valuable. The second process is monitoring player development. Coaches often want players to behave in a certain way on the pitch. By analyzing playing behaviour, the coach can monitor whether his players correctly execute his instructions and give illustrative examples of both good and bad behaviours. The third process is match preparation. Understanding the behaviour of your own players and those of the opposing team can offer certain tactical advantages. For example, a team’s defenders will wish to know if an opposing striker tends to move towards the near post or the far post when receiving a cross and taking a shot.

Analyzing player behaviour essentially boils down to summarizing the player’s playing style in a way that is both human-interpretable and suitable for data analysis. Typically, the goal is to construct a fingerprint of a player’s playing style which captures distinguishing characteristics of a player’s behaviour such as which types of actions a player tends to perform and where or what types of gameplay patterns he tends to participate in. There are currently two distinct ways to do this: location-based and interaction-based. Location-based approaches consider the locations and action type information in an event. Typically, they then attempt to summarize which locations a player prefers to occupy on the field and what actions he tends to perform in each of these areas (Decroos & Davis, 2019; Gyarmati, Kwak, & Rodriguez, 2014). Interaction-based approaches consider the players involved in an event. Then, they focus on detecting player interaction patterns (e.g., a one-two pass where player A passes the ball to player B who immediately returns the ball to player A) and then for each player count how often they are involved in various patterns (Bekkers & Dabadghao, 2017; Gyarmati & Anguera, 2015; Wang, Zhu, Hu, Shen, & Yao, 2015).

Team Tactics

As interesting as analyzing individual player behaviour is, most works tend to focus on team tactics, which can also add substantial value to the workflow of the tactical decision maker, i.e., the coach. Analyzing team tactics is in some ways easier than analyzing player behaviour because there is more ball event data available per team than per player, yet also technically harder as team tactics are more complex than player behaviour. Usually, the more complex the concept you are trying to infer from your data, the harder it is on a technical level to infer it successfully.

There are roughly three different ways to analyze team tactics. The first way is to summarize a team's playing style in a number of features (usually simple counts such as event counts or location occurrence counts) and then cluster teams based on those features. The second way is to extract patterns from the data using a pattern mining algorithm. Often, the biggest challenge in this approach is getting the representation of the data right so that the patterns extracted by the pattern mining algorithm are informative, intuitive and make sense to the end user. The third way is to attempt to model the complete behaviour of the team in a network-based approach such as a passing network or a Markov network (Cintia, Pappalardo, & Rinzivillo, 2015; Peña, 2014; Wang et al., 2015).

An important parameter of all three approaches is what information from events to consider. There are three main categories of information per event: (1) the player(s) involved, (2) the location, and (3) the type of the event. Some approaches focus only on one of them. For example, Bekkers and Dabadghao (2017) mine patterns that focus exclusively on involved players and Peña (2014) summarizes team's playing style purely based on event types.

However, the insights that can be gained from a single category of information is limited. Hence, most approaches combine two categories. Combining these categories is often technically challenging and the biggest technical contribution is in the way the approach combines them. Wang et al. (2015) and Cintia et al. (2015) analyze both the involved players and the locations of events by applying a network-based approach to model the transitions between players and zones on the pitch. Bojinov et al. (2016), Van Haaren et al. (2015), and Kerr (2015) all attempt to analyze team tactics using both location and event type.

Van Haaren et al. (2016) and Decroos et al. (2018) are even more ambitious and attempt to detect patterns that capture all three categories of information at the same time. While interesting, their results are so far more proof-of-concepts than useful in practice, because as mentioned earlier in this section, the more complex the concept you are trying to infer from your data, the harder it is technically to infer it successfully.

Conclusion

The ever increasing availability of ball event data has led to a wide range of approaches to analyze the performance and playing style of players and teams in matches. This section discussed recent approaches to evaluate shots, saves, passes and other types of actions as well as approaches to analyze player behaviour and team tactics.

Bibliography

Bekkers, J., & Dabadghao, S. (2017). Flow Motifs in Soccer: What can passing behavior tell us? *Proceedings of the MIT Sloan Sports Analytics Conference 2017*. Presented at the MIT Sloan Sports Analytics Conference, Boston, USA.

Bialik, C. (2014, June 10). The People Tracking Every Touch, Pass And Tackle in the World Cup. Retrieved 17 September 2019, from FiveThirtyEight website: https://fivethirtyeight.com/features/the-people-tracking-every-touch-pass-and-tackle-in-the-world-cup/

Bojinov, I., & Bornn, L. (2016). The pressing game: Optimal defensive disruption in soccer. *Proceedings of the MIT Sloan Sports Analytics Conference 2016*. Presented at the MIT Sloan Sports Analytics Conference, Boston, USA.

Bransen, L. (2017). *Valuing Passes in Football Using Ball Event Data* (Erasmus University Rotterdam). Retrieved from https://thesis.eur.nl/pub/41346

Bransen, L., & Van Haaren, J. (2018). Measuring football players' on-the-ball contributions from passes during games. *Proceedings of the Machine Learning and Data Mining for Sports Analytics ECML/PKDD 2018 Workshop*.

Brooks, J., Kerr, M., & Guttag, J. V. (2016). Using machine learning to draw inferences from pass location data in soccer. *Statistical Analysis and Data Mining*, *9*(5), 338–349.

Caley, M. (2015, October 19). EPL projections and expected goals method: Spurs are good! Retrieved 19 June 2019, from Cartilage Free Captain website: https://cartilagefreecaptain.sbnation.com/2015/10/19/9295905/premier-league-projections-and-new-expected-goals

Cintia, P., Pappalardo, L., & Rinzivillo, S. (2015, September 11). *A network-based approach to evaluate the performance of football teams*. Presented at the Workshop on

Machine Learning and Data Mining for Sports Analytics, Porto, Portugal.

Decroos, T. (2019). *ML-KULeuven/socceraction* [Jupyter Notebook]. Retrieved from https://github.com/ML-KULeuven/socceraction (Original work published 2019)

Decroos, T., Bransen, L., Van Haaren, J., & Davis, J. (2019). Actions speak louder than goals: Valuing player actions in soccer. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Presented at the KDD 2019, Anchorage, Alaska - USA.

Decroos, T., & Davis, J. (2019, September). *Player Vectors: Characterizing Soccer Players' Playstyle from Match Event Streams*. Presented at the ECML/PKDD, Würzburg, Germany.

Decroos, T., Dzyuba, V., Van Haaren, J., & Davis, J. (2017). Predicting soccer highlights from spatio-temporal match event streams. *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, 1302–1308.

Decroos, T., Van Haaren, J., & Davis, J. (2018). Automatic Discovery of Tactics in Spatio-Temporal Soccer Match Data. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 223–232.

Gyarmati, L., & Anguera, X. (2015). Automatic Extraction of the Passing Strategies of Soccer Teams. *ArXiv Preprint ArXiv:1508.02171*.

Gyarmati, L., Kwak, H., & Rodriguez, P. (2014). Searching for a Unique Style in Soccer. *ArXiv Preprint ArXiv:1409.0308*.

Gyarmati, L., & Stanojevic, R. (2016). QPass: A Merit-based Evaluation of Soccer Passes. *ArXiv Preprint ArXiv:1608.03532*.

Ijtsma, S. (2015, August 14). A close look at my new Expected Goals Model. Retrieved 19 June 2019, from 11tegen11 website: http://11tegen11.net/2015/08/14/a-close-look-at-my-new-expected-goals-model/

Kerr, M. G. S. (2015). *Applying machine learning to event data in soccer*. Massachusetts

Institute of Technology.

Liu, H., Hopkins, W., Gómez, A. M., & Molinuevo, S. J. (2013). Inter-operator reliability of live football match statistics from OPTA Sportsdata. *International Journal of Performance Analysis in Sport*, *13*(3), 803–821. https://doi.org/10.1080/24748668.2013.11868690

Mackay, N. (2017, July). Improving my ‘xG added’ model – Mackay Analytics. Retrieved 21 June 2019, from Mackay Analytics website: https://mackayanalytics.nl/2017/07/28/improving-my-xg-added-model/

Michalczyk, K. (2018, November 18). An attempt to extend xG gain. Retrieved 19 June 2019, from Kuba Michalczyk website: https://kubamichalczyk.github.io//2018/11/18/An-attempt-to-extend-xG-gain.html

Peña, J. L. (2014). A Markovian model for association football possession and its outcomes. *ArXiv:1403.7993 [Math, Stat]*.

Singh, K. (2019, February). Introducing Expected Threat (xT) [Blog]. Retrieved 20 June 2019, from https://karun.in/blog/expected-threat.html

Van Haaren, J., Dzyuba, V., Hannosset, S., & Davis, J. (2015). Automatically Discovering Offensive Patterns in Soccer Match Data. *Advances in Intelligent Data Analysis XIV*, 286–297.

Van Haaren, J., Hannosset, S., & Davis, J. (2016). Strategy discovery in professional soccer match data. *Proceedings of the KDD ’16 Workshop on Large-Scale Sports Analytics*, 1–4.

Wang, Q., Zhu, H., Hu, W., Shen, Z., & Yao, Y. (2015). Discerning Tactical Patterns for Professional Soccer Teams: An Enhanced Topic Model with Applications. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2197–2206. https://doi.org/10.1145/2783258.2788577

Yam, D. (2019, February 21). Attacking Contributions: Markov Models for Football. Retrieved 21 June 2019, from StatsBomb website:

https://statsbomb.com/2019/02/attacking-contributions-markov-models-for-football/