Lotte Bransen*, Jan Van Haaren and Michel van de Velden

Measuring soccer players' contributions to chance creation by valuing their passes

https://doi.org/10.1515/jqas-2018-0020

Abstract: Scouting departments at soccer clubs aim to discover players having a positive influence on the outcomes of matches. Since passes are the most frequently occurring on-the-ball actions on the pitch, a natural way to achieve this objective is by identifying players who are effective in setting up chances. Unfortunately, traditional statistics such as number of assists fail to reveal players excelling in this area. To overcome this limitation, this paper introduces a novel metric that measures the players' involvement in setting up chances by valuing the effectiveness of their passes. Our proposed metric identifies Arsenal player Mesut Özil as the most impactful player in terms of passes during the 2017/2018 season and proposes Ajax player Frenkie de Jong as a suitable replacement for Andrés Iniesta at FC Barcelona.

Keywords: machine learning; pass evaluation; player performance; soccer analytics.

1 Introduction

In soccer clubs' quests for better results, the player recruitment and retention process is of vital importance. On one hand, soccer clubs aim to improve the level of the players who are already on their teams. On the other hand, they attempt to bring in better players from other clubs, which often forces them to pay large transfer fees. For example, during the summer transfer window for the 2017/2018 season, the twenty English Premier League clubs alone spent 1.8 billion euro on player transfers (Barnard et al. 2018).

Soccer clubs' scouting departments typically aim to discover players whom they expect to positively influence the outcomes of their teams' matches. Identifying players who are often involved in setting up chances by performing effective passes is a natural way to achieve this objective since passes are the most frequently occurring actions during soccer matches (Power et al. 2017). In the present work, we use a dataset covering 10,846,885 on-the-ball player actions of which 69% are passes.

Although vast amounts of data are collected during matches and training sessions, scouts are often still restricted to traditional pass-based statistics such as the number of assists (i.e. passes immediately prior to goals) and key passes (i.e. passes immediately prior to shots) or the percentage of successfully completed passes. Soccer clubs often lack experience with and knowledge of sophisticated statistical tools to implement a more data-driven approach to their player recruitment processes by analyzing the large quantities of valuable data they have at their disposal.

The main limitation of traditional pass-based statistics is that they fail to appropriately account for the circumstances under which the passes are performed. One example is the percentage of successfully completed passes, which does not differentiate between a pass between two central defenders on their own half and a pass by an attacking midfielder trying to reach a forward in the opponent's penalty area. While the latter pass is clearly more valuable in terms of creating a possible goal-scoring opportunity, it is also more likely to fail at the same time. Another example is a player's number of assists. Since a pass is only considered an assist when the receiving player manages to score, a player's assist tally highly depends on the abilities of his teammates as well. Hence, if the receiving player is a poor finisher, valuable passes are not registered as such.

To alleviate the limitations of traditional statistics, we propose a novel metric named Expected Contribution to the Outcome of the Match (ECOM) to measure players' involvement in setting up goal-scoring chances by valuing the effectiveness of their passes. Intuitively, our metric values passes that are more likely to lead to a goal higher than passes that are less likely to do so. Our approach values passes by first retrieving similar passes from historical data using a distance-weighted k-nearest-neighbors search and then aggregating their labels. Our domain-specific distance function accounts for both the characteristics of the passes and the circumstances under which the passes were performed.

***Corresponding author: Lotte Bransen,** SciSports, Amersfoort, The Netherlands, e-mail: l.bransen@scisports.com
Jan Van Haaren: SciSports, Amersfoort, The Netherlands, e-mail: j.vanhaaren@scisports.com
Michel van de Velden: Erasmus Universiteit Rotterdam, Rotterdam, The Netherlands, e-mail: vandevelden@ese.eur.nl

An extensive empirical evaluation reveals that Arsenal playmaker Mesut Özil, Manchester City midfielder David Silva and FC Barcelona star Lionel Messi were the most effective passers in the 2017/2018 season and that Ajax player Frenkie de Jong would be a suitable replacement for Andrés Iniesta at FC Barcelona. Furthermore, our experiments show that our proposed ECOM metric outperforms three baseline metrics for predicting the outcomes of future matches and carries valuable information to estimate player market values.

The remainder of this paper is organized as follows. Section 2 discusses the most relevant related work and Section 3 describes the dataset. Section 4 introduces our approach for valuing passes and rating players. Section 5 presents an experimental evaluation where we compare our ECOM metric to three baseline metrics. Section 6 presents a few concrete applications of our ECOM metric. Section 7 provides a conclusion and discusses future work directions.

2 Related work

The performances of players are hard to measure due to the low-scoring and dynamic nature of soccer matches. Since players only earn rewards for scoring goals, actions that do not lead to goal-scoring opportunities are hard to quantify. As a result, the sports analytics community has mostly focused on developing metrics for measuring the quality of chances. A widely-adopted metric is the expected-goals value of a shot, which is often abbreviated as "xG". The expected-goals metric assigns a probability between zero and one to each shot, reflecting its likelihood of resulting in a goal (Lucey et al. 2014; Eggels et al. 2016).

The observation that shots only constitute a small fraction of the actions that occur during sports matches has inspired sports analytics researchers to develop metrics for quantifying other types of actions as well Decroos et al. (2018). present an algorithm for valuing on-the-ball player actions in soccer. Their proposed HATTRICS-OTB algorithm values each action by estimating the likelihood that a team will score and concede a goal for both the game state before the action and the game state after the action. Cervone et al. (2016) present an approach to measure the offensive impact of basketball players. They introduce a metric named Expected Possession Value, which estimates the number of points a team will earn from a possession at any given point in time. Our approach follows a similar line of thought by using the expected rewards of phases to value individual passes and players.

Since passes constitute around 70% of all on-the-ball actions in soccer matches, sports analytics researchers have explored dedicated approaches to valuing passes as well. Power et al. (2017) introduce a supervised approach using hand-crafted features to measure the risk and reward associated to a pass. Rein, Raabe, and Memmert (2017) propose a Voronoi-diagram approach to assess the effectiveness of a pass by evaluating the attacking space dominance and the number of defenders between the ball carrier and the goal. Chawla et al. (2017) introduce a supervised approach to label passes as *good*, *ok* or *bad* based on features derived from the trajectories of the players and play-by-play action data. To obtain the ground-truth labels, two human observers watched video footage and rated the passes on a six-point Likert scale. Gyarmati and Stanojevic (2016) present an approach named QPass to measure the intrinsic value of a pass. Their approach divides the pitch into zones, estimates the value of having the ball in each zone, and rates each pass by computing the difference between the values of the destination zone and the origin zone. QPass is the approach that comes closest to our proposed approach. Unlike QPass, however, our approach also accounts for the circumstances under which the passes were performed.

In addition, sports analytics researchers have investigated the passing interactions between players as well as the passing behavior of teams (Beetz et al. 2009; (Duch, Waitzman, and Nunes Amaral, 2010); Grund 2012; Van Haaren et al. 2015). Furthermore, Gudmundsson and Horton (2017) provide an extensive overview of sports analytics approaches that operate on spatio-temporal match data.

3 Dataset

We use play-by-play action data as well as match sheet data provided by Wyscout[1] for 9061 matches played in the 2014/2015 through 2017/2018 seasons in the following leagues: the English Premier League, the Spanish Primera División, the German 1. Bundesliga, the Italian Serie A, the French Ligue 1, the Belgian Pro League and the Dutch Eredivisie. The play-by-play data describe the actions that happen during the course of a match, whereas the match sheet data provide the teams' line-ups, tactical formations (i.e. 4-4-2, 4-3-3, et cetera) and substitutions in each of the matches. Our dataset includes 7,447,548 passes, 203,309 goal attempts and 21,483 goals.

1 https://wyscout.com

Table 1: The representation of a play consisting of four actions in a match between FC Barcelona and Real Madrid, which starts with a throw-in from the sideline.

Field	Action 1	Action 2	Action 3	Action 4
Match id	1256	1256	1256	1256
Player name	Jordi Alba	Lionel Messi	Luis Suárez	Sergio Ramos
Team name	FC Barcelona	FC Barcelona	FC Barcelona	Real Madrid
Action type	Throw-in	Pass	Cross	Clearance
$(x, y)_{start}$	(73.2, 0.0)	(75.6, 8.3)	(86.3, 11.4)	(15.5, 30.0)
$(x, y)_{end}$	(75.6, 8.3)	(86.3, 11.4)	(89.5, 38.0)	(23.4, 22.4)
Success	True	True	False	True
Time in seconds	2254	2258	2261	2262

For each action in each match, our dataset contains a reference to the player who performed the action, the type of the action (e.g. a pass or a shot), the start and end locations of the action (i.e. their (x, y)-coordinates), an indicator whether the action was successful or not and the timestamp of the action in the match. The dataset records the locations of the actions from the perspective of the team in possession of the ball, which is assumed to always play from the left side to the right side of the pitch. Since pitch dimensions vary from one venue to another, we standardize these locations to a pitch of 105 m long and 68 m wide, which is either the required or recommended pitch dimension in most international and domestic competitions (Union of European Football Associations 2018; Fédération Internationale de Football Association 2018; Deutscher Fussball-Bund 2017).

Table 1 shows an example of four consecutive actions in our dataset. The example describes a play in a match between FC Barcelona and Real Madrid, which starts with a throw-in from the sideline. Jordi Alba throws the ball to his teammate Lionel Messi (Action 1), who passes the ball to fellow attacker Luis Suárez (Action 2). Luis Suárez crosses the ball into the penalty area (Action 3), where Real Madrid defender Sergio Ramos clears the ball (Action 4).

4 Approach

Measuring a player's involvement in creating goal-scoring chances is challenging due to the low-scoring nature of the game. A soccer player only gets a few occassions during a match to earn reward from his passes, which is when his team scores a goal. Hence, our proposed ECOM (**E**xpected **C**ontribution to the **O**utcome of the **M**atch) metric resorts to computing the expected rewards from passes instead of distributing the actual rewards from goals across the preceding passes. Intuitively, our proposed ECOM metric reflects the number of goals that is expected to arise from a player's passes per 90 min of play.

We value each pass by estimating its expected added reward based on similar passes in historical play-by-play data. We consider both geometrical and contextual features of the passes to determine their similarity. In particular, we value each pass by computing the increase or decrease in likelihood of scoring a goal that arises from the pass. Hence, we positively value passes that increase the likelihood of scoring and negatively value passes that decrease that likelihood.

Our approach to computing the ECOM metric constitutes the following five steps. First, we split a match into possession sequences, which are sequences of actions where the same team remains in possession of the ball. Second, we label the possession sequences and their constituting passes using an expected-goals model. Third, we introduce a domain-specific distance function to measure the similarity between passes. Fourth, we value each pass by computing the expected added reward of the pass using a k-nearest-neighbors search leveraging our distance function. Fifth, we compute each player's ECOM rating by aggregating their pass values and normalizing them for 90 min of play.

4.1 Splitting matches into possession sequences

Our dataset represents each match as a sequence of consecutive actions. More formally, a match M is a sequence of actions $[a_1, \ldots, a_n]$, where n is the total number of actions in the match. To simplify the notation, we use variables p_i to represent actions a_i that are passes.

In order to value a pass $p_i \in M$, we view a match as a sequence of possession sequences, which are subsequences of M where the same team is in possession of the ball. More formally, our approach views a match M as a sequence of possession sequences $[S_1, \ldots, S_m]$,

where m is the total number of possession sequences in the match. Each possession sequence S_t is a sequence of actions $[a_{k_t+1}, \ldots, a_{k_t+l_t}]$, where l_t is the number of actions in possession sequence S_t and $k_t = \sum_{s=1}^{t-1} l_s$ is the total number of actions in the possession sequences preceding possession sequence S_t.

We start a new possession sequence each time a team gains possession of the ball, which happens in the following situations: at the start of a half, when the team intercepts the ball, and after the opponent performs a shot, commits a foul followed by a freekick, or last touches the ball before it goes out of play.

4.2 Labeling possession sequences and passes

We assign to each possession sequence S_t a label $L(S_t)$ that represents its outcome. If the possession sequence S_t *does not* lead to a goal attempt, we set the label $L(S_t)$ to 0. If the possession sequence S_t *does* lead to a goal attempt, we set the label $L(S_t)$ to the probability of the goal attempt yielding a goal, regardless of its actual outcome. As explained in Section 2, this approach corresponds to computing the expected-goals values for the goal attempts as is commonly done in the soccer analytics community. We compute the expected values of goal attempts as goal attempts occur about ten times more often than goals. In our dataset, a match yields 2.4 goals and 22.4 goal attempts on average.

Furthermore, we assign to each pass the label of its constituting possession sequence. In particular, we set the label $L(p_i)$ for each pass $p_i \in S_t$ to $L(S_t)$. Thus, passes belonging to the same possession sequence receive the same label.

4.3 Computing similarities between passes

To measure the similarity between passes, we introduce a domain-specific distance function that incorporates the characteristics of the passes as well as the circumstances under which the passes were performed. Only considering passes that were performed under comparable circumstances leads to more accurate expected values for the passes. For example, if the ball was at the opposite side of the pitch a few seconds prior to a pass, it is likely that the pass was performed during a counter-attack when players typically have more time on the ball.

Our domain-specific distance function considers the following six components to compute the distance between a pass p_i and a pass p_j:

1. The difference in length of the passes (Δ_{ij}^1);
2. The Euclidean distance between the origins of the passes (Δ_{ij}^2);
3. The Euclidean distance between the destinations of the passes (Δ_{ij}^3);
4. The Euclidean distance between the locations of the ball 5 s prior to the passes (Δ_{ij}^4);
5. The Euclidean distance between the locations of the ball 10 s prior to the passes (Δ_{ij}^5);
6. The Euclidean distance between the locations of the ball 15 s prior to the passes (Δ_{ij}^6).

Hence, we obtain the following distance function: $d(p_i, p_j) = w_1\Delta_{ij}^1 + \cdots + w_6\Delta_{ij}^6$, where each w_b is a weight denoting the importance of the corresponding component. We define the similarity between a pass p_i and a pass p_j as $s(p_i, p_j) = \frac{1}{d(p_i,p_j)}$.

In our experimental evaluation, we investigate the design decisions for our distance function in further detail (Section 5.2.1) and automatically learn the optimal weights for the components from the available data (Section 5.2.3).

4.4 Valuing passes

We compute the expected added reward for a pass p_i based on the labels of similar passes, where more similar passes contribute more to the expected added reward than less similar passes. Given a particular pass p_i, we compute its expected added reward $V(p_i)$ as follows:

$$V(p_i) = V_e(p_i) - V_s(p_i),$$

where $V_e(p_i)$ reflects the expected end reward of the pass and $V_s(p_i)$ reflects the expected start reward of the pass.

We compute the expected start reward $V_s(p_i)$ by computing the average end label for the passes that end in the location where the pass originates from. Since passes hardly ever end in the exact same location, we divide the pitch into a grid of cells and assign each pass to the corresponding cell based on its end location. Based on the experimental evaluation in Section 5, we use cells of 15 by 17 m, which lead to robust expected start rewards on our dataset.

Given the l passes p_j in the cell that pass p_i originates from, we compute the expected start reward for a given pass p_i as follows:

$$V_s(p_i) = \frac{\sum_{j=1}^{l} L(p_j)}{l}.$$

Like Gyarmati and Stanojevic (2016), we compute the expected end reward $V_e(p_i)$ differently for successful and

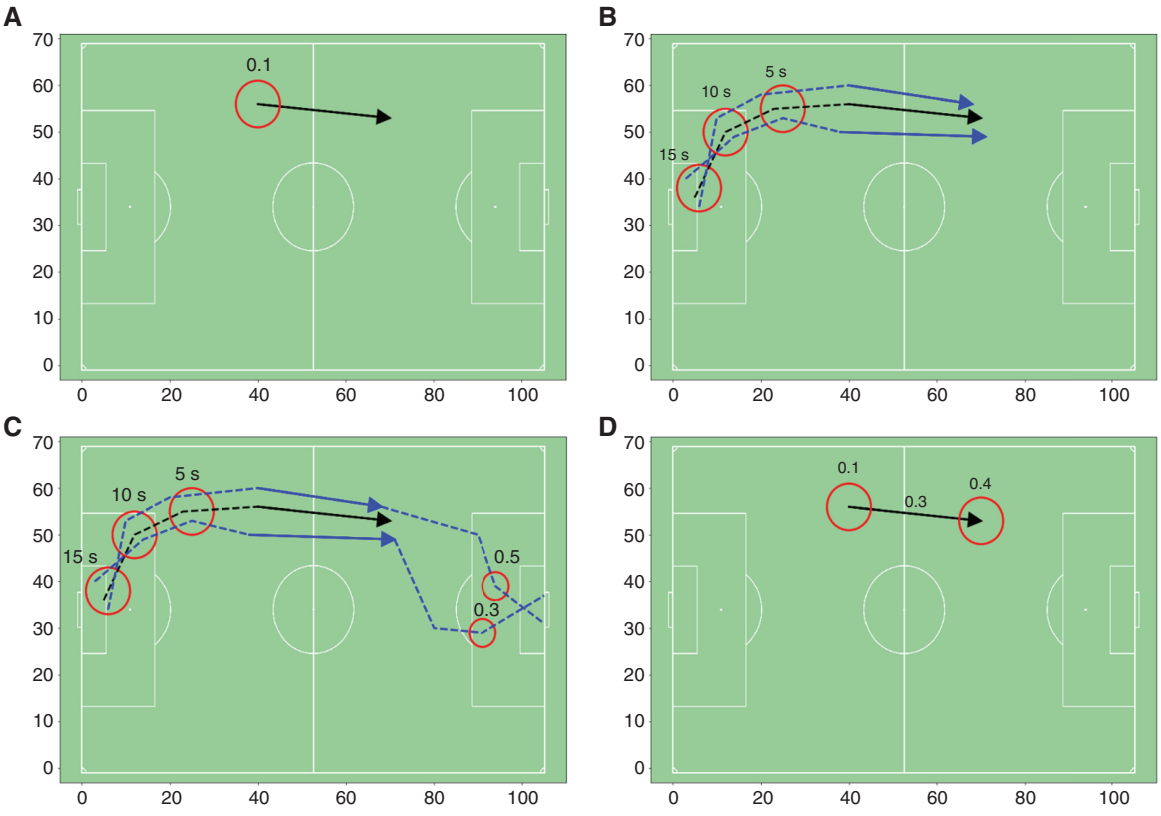

Figure 1: Visualization of our distance-weighted k-nearest-neighbors approach for computing the expected added reward of a successful pass. (A) Compute the expected start reward of the pass by averaging the labels of the passes ending in the start location. (B) Perform a distance-weighted k-nearest-neighbors search to discover the most similar passes. (C) Compute the expected end reward of the pass by averaging the labels of the possession sequences encompassing the most similar passes. (D) Value the pass by subtracting the expected start reward of the pass from the expected end reward of the pass.

unsuccessful passes. For successful passes, we compute a weighted average of the labels for the passes in the corresponding cell, where the weights are given by the similarity function $s(p_i, p_j)$ from Section 4.3. For unsuccessful passes, we set the expected end reward to zero. We exploit the observation that passes resulting in a loss of possession cannot lead to a goal-scoring attempt. In summary, we compute the expected end reward for a given pass p_i as follows:

$$V_e(p_i) = \begin{cases} \dfrac{\sum_{j=1}^{k} s(p_i, p_j) \cdot L(p_j)}{\sum_{j=1}^{k} s(p_i, p_j)} & \text{if } p_i \text{ is successful} \\ 0 & \text{if } p_i \text{ is unsuccessful} \end{cases}$$

Figure 1 visualizes our distance-weighted k-nearest-neighbors approach for computing the expected added reward for a successful pass.

4.5 Rating players

We obtain the ECOM rating for each player by computing their expected added reward from passes per 90 min of play. Intuitively, a player's ECOM rating reflects the number of goals that is expected to arise from the passes the player performs during 90 min of play.

Given a set of passes $\{p_1^r, \ldots, p_{N_r}^r\}$ for a player r during a given time period, where N_r is the number of passes performed by that player, we compute the ECOM rating for a given player r as follows:

$$ECOM(r) = \frac{\sum_{i=1}^{N_r} V(p_i^r)}{T_r} \cdot 90,$$

where T_r is the total number of minutes played by player r during the time period under consideration.

5 Experimental evaluation

We now motivate the design decisions for our approach and evaluate our proposed ECOM metric by comparing its ability to predict match outcomes to the predictive performance of three baseline metrics. We first introduce our methodology and then present our experimental results.

Table 2: The number of seasons, matches, passes, goal attempts and goals in our training set, validation set and test set.

Type	Training set	Validation set	Test set	Total
Seasons	2	1	1	4
Matches	4253	2404	2404	9061
Passes	3,4252,285	1,998,533	2,023,730	7,447,548
Goal attempts	95,381	53,617	54,311	203,309
Goals	9853	5868	5762	21,483

The training set covers two full seasons, whereas the validation set and the test set cover one full season each.

5.1 Methodology

In this section, we explain how we construct the datasets, train our expected-goals model, cluster the passes, implement the baseline metrics and predict the outcomes of future matches.

5.1.1 Constructing the datasets

We divide the available data into three datasets: a training set, a validation set and a test set. In order to evaluate the predictive performance of our ECOM metric, we respect the chronological order of the matches. As a result, the training set covers the 2014/2015 and 2015/2016 seasons, the validation set covers the 2016/2017 season and the test set covers the 2017/2018 season. We omit the matches for which either no play-by-play match data are available or the available data are incomplete (e.g. missing timestamps for the actions). More specifically, we omit 555 matches from the training set, which are mostly matches in the 2014/2015 season in the Belgian Pro League and Dutch Eredivisie.

Table 2 provides the number of seasons, matches, passes, goal attempts and goals in each of the three sets. For the evaluation, we train our models on the training set and optimize the parameter values on the validation set. In Section 6, where we present the results and two concrete use cases for our metric, we train the models on the training set and validation set combined for the optimal parameter values and report results on the test set.

Adopting the SPADL representation for describing player actions from Decroos et al. (2018), we extract all actions that describe an interaction between a player and the ball from the play-by-play data. Thus, our datasets contain all dribbles, passes, crosses, shots, freekicks, penalties, throw-ins, goalkeeper saves, interceptions, clearances and touches that occurred during each match.

In order to measure the similarity between passes using our domain-specific distance function, we also assign to each pass in our datasets the locations of the ball 5, 10 and 15 s before that pass. We obtain these ball locations by performing linear interpolation between the locations of the actions. We omit passes that occur during the first 15 s of each half as the historical ball locations are not available for these passes.

5.1.2 Training the expected-goals model

To label possession sequences resulting in a shot as explained in Section 4.2, we train an expected-goals model that estimates the likelihood of a shot yielding an actual goal. We pose this problem as a binary probabilistic classification task.

We construct a dataset based on the 95,381 shots in the training set to train the model. To increase the number of training examples, we duplicate the shots and mirror their locations along the length of the pitch to obtain a dataset containing 190,762 shots. For each shot, our dataset contains the x-coordinate and y-coordinate of the location, the distance to the center of the goal and the angle between the location and the two goal posts. We label the shots yielding a goal as positive examples and all other shots as negative examples.

We use the `XGBoost` algorithm to train a probabilistic classifier.[2] We optimize the algorithm's hyperparameters using `GridSearchCV` in `scikit-learn`. We try setting the number of estimators to 100, 500 and 1000, restricting the tree depth to 3, 4, 5 and 6, and using learning rates of 0.001, 0.01 and 0.1. We obtain the highest AUC-ROC for 100 estimators, a maximum tree depth of 4, and a learning rate of 0.1. For the 53,617 shots in the validation set, we obtain an AUC-ROC of 0.763, which is in line with the results reported in the literature for slightly more sophisticated expected-goals models (Decroos et al. 2017).

5.1.3 Clustering the passes

To apply the distance-weighted k-nearest-neighbors search as explained in Section 4.4, we need to compute

2 https://xgboost.readthedocs.io/en/latest/

the distance between each pass in the test set and each pass in the training set. This task quickly becomes computationally expensive for datasets containing millions of passes. To reduce the number of distance computations, we exploit the observation that passes starting or ending in entirely different locations on the pitch are unlikely to be similar. Hence, we first cluster the passes based on their spatial locations and then perform the distance-weighted k-nearest-neighbors search within each cluster separately.

We assign each pass to a cluster based on its origin and destination. Since passes are unlikely to have the exact same origin and destination locations, we divide the pitch into zones and allocate the origins and destinations of the passes to their corresponding zones. We represent each cluster as an origin-destination pair, which means that two passes belong to the same cluster if their origin and destination locations are both in the same zone. This representation also enforces that passes within the same cluster have similar lengths.

5.1.4 Implementing the baseline metrics

We compare the predictive performance of our ECOM metric for rating players to the predictive performance of three baseline metrics. We implement two variants of the QPass metric introduced by Gyarmati and Stanojevic (2016) and define a metric based on the traditional pass accuracy statistic.

We implement two best-effort approximations of the QPass metric based on the details in the paper. In both variants, we value successful passes by subtracting the value of the origin location for the team in possession from the value of the destination location for the team in possession. Similarly, we value unsuccessful passes by subtracting the value of the origin location for the team in possession from the value of the same location for their opponent, where we multiply the latter value by -1 to reflect the change in possession.

We compute the value of each location for each team in two steps. First, we divide the pitch into a 10-by-10 grid of equal-sized cells. Second, we value each cell by computing the average value for the possession sequences originating from that cell. The first variant named "QPass approximation" follows the paper and assigns a value of 0.7 to possession sequences leading to a shot and a value of 0 otherwise. The second variant named "QPass approximation xG" uses the expected-goals value for the shot instead of a fixed value of 0.7 for possession sequences leading to a shot. In both variants, we rate the players by computing the average pass value per 90 min as explained in Section 4.5.

Furthermore, we implement a metric based on the traditional pass accuracy statistic. We rate the players by computing the ratio between their number of successful passes and their total number of passes.

5.1.5 Predicting the outcomes of matches

Due to the unavailability of a ground truth, we evaluate our metric by predicting future performances from past performances as is commonly done in the sports analytics literature (Schulte, Zhao, and Routley 2015; Liu and Schulte 2018). We expect our metric to be a predictor of future performances, which is vital in the player recruitment process. Therefore, we compare our metric to the baselines introduced in Section 5.1.4 in terms of their ability to predict the outcomes of matches.

Assuming that the number of goals scored by each team in each match follows a Poisson distribution, we represent the number of goals that a team is expected to score in a match by a Poisson random variable (Maher 1982). We use the Skellam distribution to determine the probability that one Poisson random variable is higher than the other Poisson random variable and thus obtain the probabilities of a home win, draw and away win (Karlis and Ntzoufras 2008). We evaluate these probability estimates by computing their logarithmic loss (Langseth 2013; Ley, Van de Wiele, and Van Eetvelde 2017). Logarithmic loss measures how good the probability estimates are and is thus an appropriate evaluation metric for this task (Ferri, Hernández-Orallo, and Modroiu 2009).

We compute the Poisson mean for each team by summing the ECOM ratings for the players in the starting lineup. We only use information that is available prior to kick-off and thus do not consider substitutions. For players who played at least 900 min in the training set, we consider the actual ratings. For the remaining players, we use the average rating of the team's players in the same line. Since the average reward gained from passes (i.e. 0.72 goals per team per match) reflects around 50% of the average reward gained during matches (i.e. 1.42 goals per team per match), we transform the distribution over the player ratings per team per match to follow a similar distribution as the average number of goals scored by each team in each match in the validation set.

5.2 Evaluation

We now present experimental results to motivate the design of our domain-specific distance function for passes, to investigate the optimal grid cell dimensions for the clustering step, to optimize the parameters for the

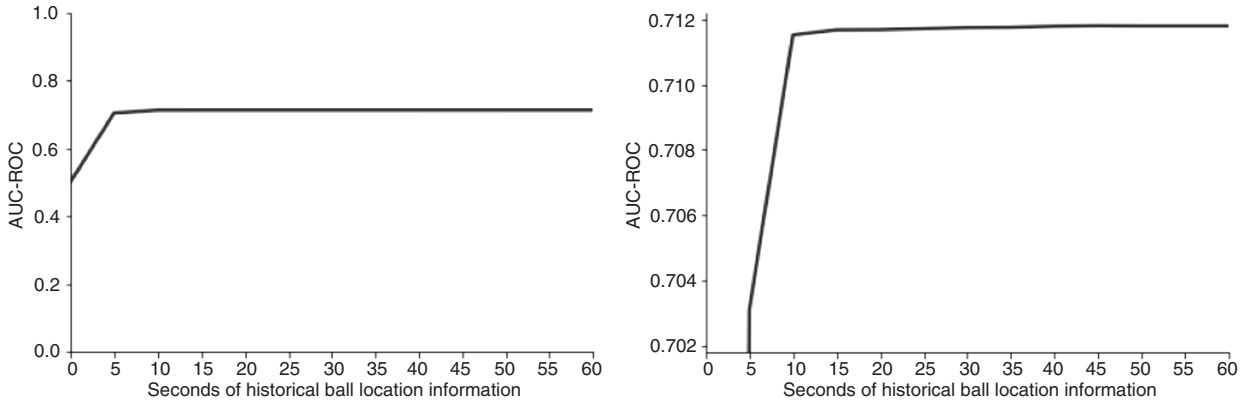

Figure 2: The AUC-ROC scores for an increasing amount of historical ball location information. The graph on the left shows the entire AUC-ROC range, whereas the graph on the right shows a zoomed-in view of the area of interest. The AUC-ROC improvement drops off after 15 s of historical ball location information.

distance function and to investigate the impact of the clustering step. We also compare our ECOM metric to four baselines in terms of predictive performance. For each of our experiments, we restrict our datasets to the matches in the English Premier League. We compute the ECOM player ratings on the validation set (i.e. 2016/2017 season) and report results on the test set (i.e. the 2017/2018 season).

5.2.1 Designing the distance function

In this experiment, we motivate our decision to include the location of the ball 5, 10 and 15 s prior to the pass in our distance function to capture the circumstances under which the pass was performed. More specifically, we investigate the impact of historical ball locations on the current location of the ball. To this end, we design an experiment where the goal is to predict whether the ball is on one half or the other half of the pitch based on an increasing amount of information on the historical location of the ball.

We address this prediction task in an iterative fashion. Starting from a single feature that represents the ball location 5 s ago, we add one additional feature that represents the ball location 5 s earlier in each iteration. Thus, in the first experiment we only consider the ball location 5 s ago, in the second experiment we consider the ball location 5 and 10 s ago, in the third experiment we consider the ball location 5, 10 and 15 s ago, and so on. We consider the ball location up until 60 s before the current location and thus perform twelve experiments.

We use the `XGBoost` algorithm to train the models.[3] We optimize the algorithm's hyperparameters using `GridSearchCV` in `scikit-learn`. Having optimized the parameters for the first experiment, we set the number of estimators to 100, restrict the tree depth to 6 and use a learning rate of 0.1 for all experiments. We train the models on the training set and make predictions for the validation set. We omit the first 60 s of each half to allow for a fair comparison with the model where we include the ball location 60 s ago.

Figure 2 shows the AUC-ROC scores for each of the twelve models and for the case where no historical ball location information is available, which corresponds to random guessing and thus an AUC-ROC of 0.500. The inclusion of the ball location 5 s ago increases the AUC-ROC from 0.500 to 0.703, while the inclusion of the ball location 10 s ago increases the AUC-ROC further to 0.712. Although the inclusion of the ball location 15 s ago yields another subtle AUC-ROC increase, the inclusion of the ball location beyond 15 s ago does not lead to further improvements.

Based on these insights, we only include the ball location 5, 10 and 15 s prior to a pass in our distance function.

5.2.2 Investigating the optimal grid cell dimensions for the clustering step

In this experiment, we investigate the optimal dimension for the grid cells used in the clustering step. Within each cluster, the distance-weighted k-nearest-neighbors search needs to compute the distance between each pass in the training set and each pass in the test set. Thus, we aim to optimize the balance between the number of clusters and the maximum number of passes in each cluster. If the number of clusters increases, the risk of missing a similar pass in the k-nearest-neighbors search also

3 https://xgboost.readthedocs.io/en/latest/

Table 3: Characteristics for five different dimensions of the cells in the clustering step when training on the 759 English Premier League matches in the training set and evaluating on the 380 English Premier League matches in the validation set.

Characteristic	105 × 68	52.5 × 34	15 × 17	7 × 8.5	5 × 4
Total number of grid cells	1	4	28	120	357
Total number of clusters	1	16	784	14,400	127,449
Max. number of labeled passes in cluster	621,547	115,420	10,792	892	635
Max. number of valued passes in cluster	315,914	56,023	5421	501	89
Required amount of memory in gigabyte	1570.8	51.7	0.5	0.0	0.0

As expected, the required amount of memory increases as the size of the grid cells increases.

increases. If the number of clusters decreases, the number of passes within each cluster and thus also the number of distance computations increases. To reduce the total runtime for our approach, we aim to minimize the number of clusters and still be able to compute the distances using vectorization.

We investigate the characteristics of five different dimensions for the grid cells in the clustering step: 105 by 68 m, 52.5 by 34 m, 15 by 17 m, 7 by 8.5 m and 5 by 4 m. Since the pitch measures 105 by 68 m, the first dimension corresponds to performing the *k*-nearest-neighbors search without clustering. For each setting, we compute the total number of grid cells, total number of clusters, maximum number of labeled passes per cluster, maximum number of passes to be valued per cluster, and total amount of memory required to compute the distances between the passes using vectorization.

The memory requirement comprises the amount of memory required to represent the labeled passes in the training set and the passes to be valued in the test set as well as the distances between each labeled pass and each pass to be valued. We need six values to represent one pass: one value for each of the six components in the distance function. In addition, we need one value to represent the distance between two passes. For example, in the 105 × 68 setting, the maximum number of labeled passes per cluster is 621,547 and the maximum number of passes to be valued per cluster is 315,914 for the setup where we train on the 759 English Premier League matches in the training set and evaluate on the 380 English Premier League matches in the validation set. As a result, the number of values required to represent the passes is 5,624,766 [$=6 \times (621{,}547 + 315{,}914)$], and the number of values required to represent the distances is 196,355,398,958 ($=1 \times 621{,}547 \times 315{,}914$). Hence, representing each value as a 64-bit float, we would need over 1570 gigabyte of memory to simultaneously store these values.

Table 3 shows the characteristics for each of the five different dimensions of the grid cells when valuing the passes in the 380 English Premier League matches in the validation set using the labeled passes in the 759 English Premier League matches in the training set. The unexpectedly large difference between the maximum number of labeled and valued passes per cluster in the 5 × 4 setting is due to a rule change introduced at the start of the 2017/2018 season. Since then, the ball can move in any direction from kick-off rather than only forward.

As expected, the required amount of memory increases as the size of the grid cells also increases. Since the machine that we use to run our experiments has only 32 gigabyte of memory available, we exclude the 105 × 68 and 52.5 × 34 settings. As mentioned earlier, we prefer larger grid cells over smaller grid cells to minimize the risk of missing highly similar passes. Also, we wish to minimize the number of clusters and thus loops to minimize the overhead caused by reading data from disk and storing data to disk. As a result, we use the 15 × 17 setting as the default grid cell configuration for our approach in all of the following experiments, unless explicitly specified otherwise.

5.2.3 Optimizing the weights for the distance function

In this experiment, we optimize the weights for the six components in our distance function to compare passes: the start and end locations of the passes, the lengths of the passes and the ball locations 5, 10 and 15 s prior to the passes. The search space is large since the weights corresponding to each of the six components can freely range from zero to one. Hence, we use a Bayesian optimization approach to explore the space of candidate weight sets in an efficient way (Brochu, Cora, and De Freitas 2010; Snoek, Larochelle, and Adams 2012). Using the `Bayesian Optimization` package,[4] we run 250 optimization iterations for ten different initial weight sets

4 https://github.com/fmfn/BayesianOptimization

to increase the probability of finding the global optimum. We use the Upper Confidence Bound (UCB) as the acquisition function.

For the default configuration of our approach where we perform clustering with grid cells of 15 m by 17 m, we find the logarithmic loss to be minimal when the weights corresponding to the lengths of the passes and the ball locations 5 s prior to the passes are set to one and all other weights are set to zero. The most likely explanation for the exclusion of the start and end locations of the passes is that this information is already implicitly taken into account by the clustering step. This information is likely to become more important for an increasing number of passes per cluster.

To further investigate this hypothesis, we extend the above experiment to the other clustering settings explored in Section 5.2.2. For each of the five settings, we compute the logarithmic losses for four different sets of weights for the distance function. For computational tractability, we restrict the experiment to four pre-defined sets of weights and do not perform a weight optimization process for each setting. In the 105×68 and 52.5×34 settings, we compute the logarithmic losses in batches to avoid running out of the available memory.

Table 4 provides a description for each of the four predefined weight sets. Set W_{l+5}, which is the optimal weight set for the default configuration of our approach, considers the lengths of the passes and the location of the ball 5 s before the pass. Set W_{o+d} considers the origin and destination locations of the passes. Set W_{10+15} considers the location of the ball 10 and 15 s before the pass. Set W_{all} considers all six components.

Table 5 shows the logarithmic losses for predicting the outcomes of the 2017/2018 English Premier League matches for five different clustering settings and four different weight sets for the distance function. As expected, the origin and destination locations of the passes (W_{o+d}) are more important in the setting without clustering (105×68) than in the settings with clustering. Furthermore, the historical locations of the ball (W_{10+15}) become more important in clustering settings where each cluster contains a reasonably large number of passes per cluster, which is the case for the 15×17 setting.

Table 4: Overview of the four different weight sets considered in the weight optimization experiment.

Distance function component	W_{l+5}	W_{o+d}	W_{10+15}	W_{all}
Lengths of the passes (Δ^1_{ij})	1	0	0	1
Origins of the passes (Δ^2_{ij})	0	1	0	1
Destinations of the passes (Δ^3_{ij})	0	1	0	1
Ball locations 5 s ago (Δ^4_{ij})	1	0	0	1
Ball locations 10 s ago (Δ^5_{ij})	0	0	1	1
Ball locations 15 s ago (Δ^6_{ij})	0	0	1	1

Each weight set W_i considers a different subset of the components.

Table 5: The logarithmic losses for predicting the outcomes of the 2017/2018 English Premier League matches for five different clustering settings and four different weight sets for the distance function.

Weight set	105×68	52.5×34	15×17	7×8.5	5×4
W_{l+5}	1.0654	**0.9989**	**1.0057**	**1.0123**	**1.0067**
W_{o+d}	**1.0353**	1.0187	1.0086	1.0165	1.0091
W_{10+15}	1.0488	1.0172	1.0069	1.0144	1.0165
W_{all}	1.0398	1.0132	1.0068	1.0139	1.0137

The best result for each clustering setting is in bold.

5.2.4 Investigating the impact of the clustering step

We now investigate the impact of the clustering step in a qualitative fashion. For four arbitrary passes, we obtain the four nearest neighbors in the setting without clustering and the clustering setting with grid cells of 15 by 17 m. We use the optimal weights for the distance function obtained in Section 5.2.3.

Figure 3 shows the four nearest neighbors of the red pass when not clustering the passes before performing the *k*-nearest-neighbors search. Similarly, Figure 4 shows the four nearest neighbors of the same pass when clustering the passes with grid cells of 15 by 17 m. Although the obtained passes are different, the four-nearest-neighbors search obtains highly similar neighbors in both settings.

Appendix A shows the four nearest neighbors for three other passes.

5.2.5 Comparing the ECOM metric to the baselines

In this experiment, we compare the performance of our ECOM metric to the performance of the three baseline metrics introduced in Section 5.1.4. For our ECOM metric, we use the default configuration with optimal weights. We also obtain prior probabilities for a home win, a draw and an away win by computing the historical distribution over the match outcomes in the validation set.

Table 6 shows the logarithmic losses for predicting the outcomes of the 2017/2018 English Premier League matches for our ECOM metric as well as the four baselines. Our ECOM metric clearly outperforms each of the baselines. To put these results into perspective, we also compare our loss values to those reported in the literature. Langseth (2013) reports logarithmic loss values ranging

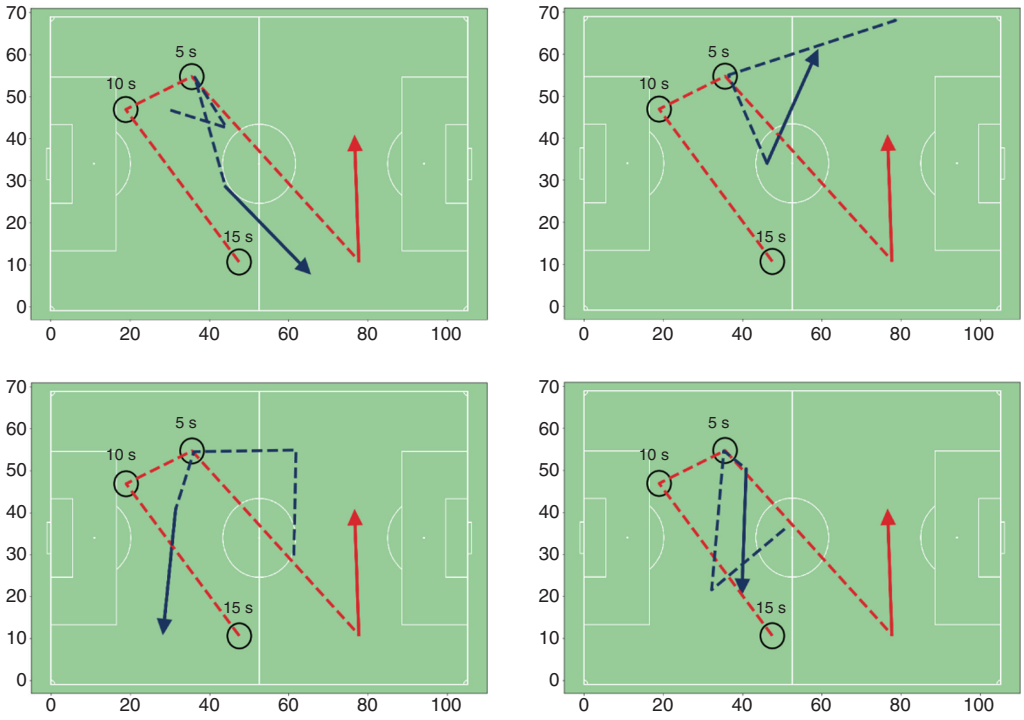

Figure 3: Visualization of the four nearest neighbors of the red pass when not clustering the passes before performing the *k*-nearest-neighbors search.

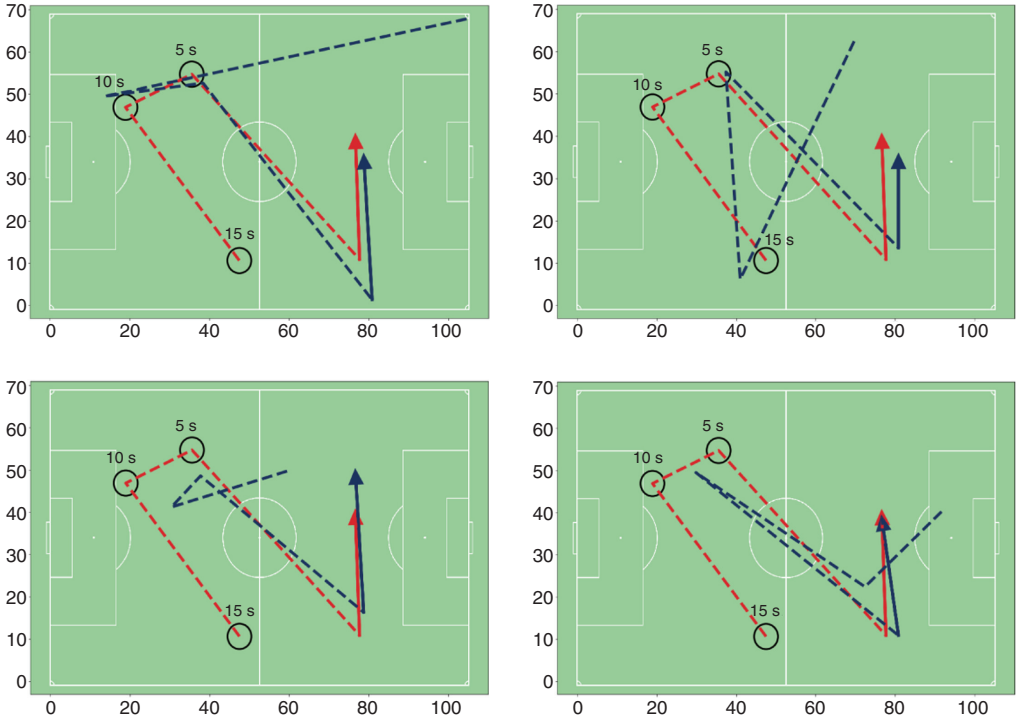

Figure 4: Visualization of the four nearest neighbors of the red pass when clustering the passes with grid cells of 15 by 17 m before performing the *k*-nearest-neighbors search.

from 0.9685 to 1.0041 for predicting the matches in the 2011/2012 and 2012/2013 English Premier League seasons. Ley et al. (2017) report values ranging from 0.9776 to 1.0845 for predicting the matches in the second half of the 2000/2001 through 2016/2017 English Premier League seasons.

Table 6: The logarithmic losses for predicting the outcomes of the 2017/2018 English Premier League matches.

Metric	Logarithmic loss
ECOM default configuration	**1.0057**
Historical prior distribution	1.0738
QPass approximation xG	1.0758
Pass accuracy	1.0765
QPass approximation	1.1263

Our ECOM metric clearly outperforms the four baselines. The best result is in bold.

6 Results for the 2017/2018 season

We present the top-ranked players in terms of ECOM, investigate the characteristics of the ratings and present two concrete use cases for our metric. The analyses in this section include ECOM ratings for 2129 players who played at least 900 min in the 2017/2018 season.

6.1 Identification of top-ranked players

To provide more insight into the ratings for top-rated players, we present the top-fifteen-ranked players and the top-five-ranked players under the age of 23.

Table 7 shows the top-fifteen-ranked players in terms of ECOM rating across all 2129 players. Arsenal playmaker Mesut Özil tops the list with an ECOM rating of 0.3440 per 90 min. Manchester City playmaker David Silva ranks second and FC Barcelona forward Lionel Messi ranks third.

For scouting purposes, we are also particularly interested in identifying young players who have the potential to become the stars of the future. Table 8 presents the top-five-ranked players in terms of ECOM rating across all 352 players under the age of 23. FC Bayern München winger Kingsley Coman tops the list with a rating of 0.2423 per 90 min. Real Madrid midfielder Marco Asensio ranks second, while Ajax midfielder Frenkie de Jong ranks third.

Table 7: The top-fifteen-ranked players in terms of ECOM during the 2017/2018 season.

Rank	Player	Team	ECOM per 90 min
1	M. Özil	Arsenal	0.3440
2	D. Silva	Manchester City	0.3156
3	L. Messi	FC Barcelona	0.3055
4	E. Hazard	Chelsea	0.2951
5	Neymar	Paris Saint-Germain	0.2910
6	A. Sánchez	Arsenal	0.2866
7	O. Kaya	SV Zulte-Waregem	0.2789
8	H. Ziyech	AFC Ajax	0.2716
9	Isco	Real Madrid	0.2706
10	L. Vázquez	Real Madrid	0.2704
11	Marcelo	Real Madrid	0.2592
12	A. Robben	FC Bayern München	0.2576
13	K. De Bruyne	Manchester City	0.2543
14	C. Fàbregas	Chelsea	0.2536
15	A. Iniesta	FC Barcelona	0.2511

Table 8: The top-five-ranked players under the age of 23 in terms of ECOM rating during the 2017/2018 season.

Rank	Player	Team	ECOM per 90 min
1	K. Coman	FC Bayern München	0.2423
2	M. Asensio	Real Madrid	0.2269
3	F. de Jong	AFC Ajax	0.2122
4	M. Lopez	Olympique de Marseille	0.1998
5	D. Neres	AFC Ajax	0.1971

6.2 Characteristics of the ECOM player ratings

In this section, we investigate the distribution of the ECOM ratings, the relationship between ECOM ratings and pass accuracies, the relationship between the number of passes and average value per pass as well as the relationship between the value obtained from successful and unsuccessful passes.

6.2.1 Distribution of the ECOM player ratings

We investigate the distribution of the player ECOM ratings per position. Figure 5 shows the distribution of the player ECOM ratings for goalkeepers, defenders, midfielders and forwards.

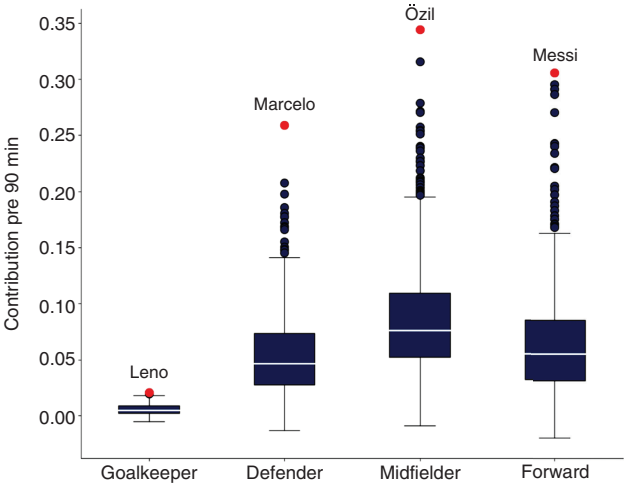

Figure 5: Box plot showing the distribution of the ECOM player ratings per position. On average, midfielders obtain higher ratings than forwards, defenders and goalkeepers.

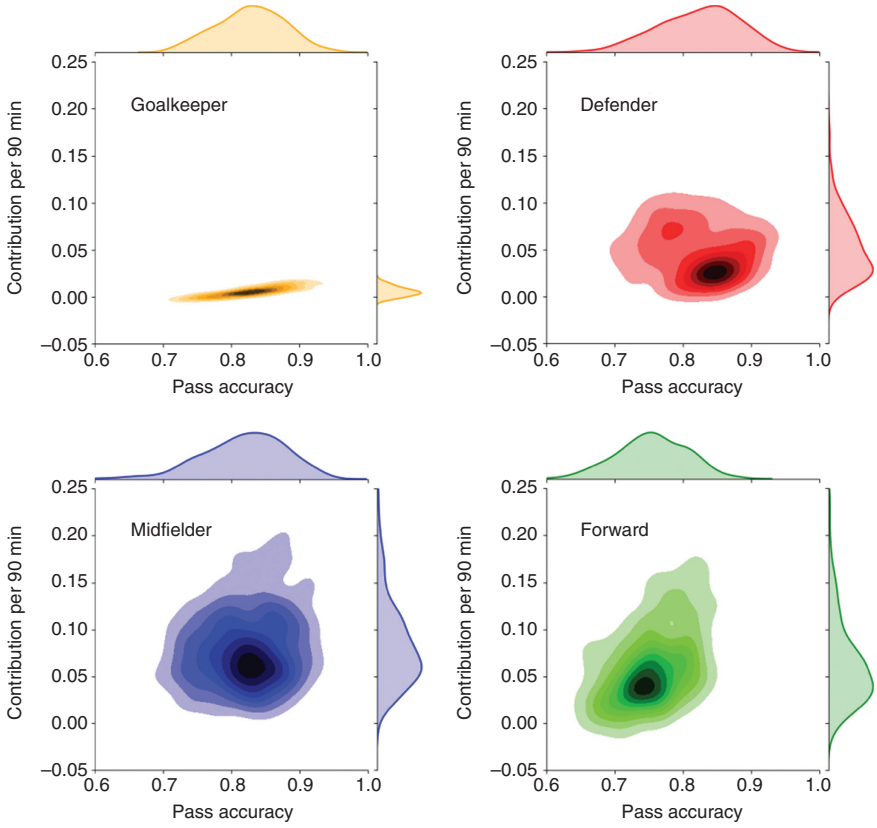

Figure 6: Two-dimensional kernel density plots showing ECOM contributions per 90 min and pass accuracies for goalkeepers, defenders, midfielders and forwards.

The box plot shows that midfielders obtain higher average ECOM ratings than goalkeepers, defenders and forwards. The lower ECOM ratings for goalkeepers and defenders are due to the fact that they contribute less to the offense. Their primary task is to prevent their opponents from scoring goals rather than creating goal-scoring opportunities themselves. The lower ECOM ratings for forwards are due to the fact that their primary task is to score goals themselves instead of providing opportunities to their teammates.

6.2.2 Relationship between ECOM ratings and pass accuracies

We investigate whether players obtaining high pass accuracies also rate high on our ECOM metric and vice versa. In particular, we explore how the distributions of both metrics relate to each other. Figure 6 presents two-dimensional kernel density plots showing ECOM ratings per 90 min and pass accuracies for goalkeepers, defenders, midfielders and forwards. As expected, midfielders rate highest in terms of ECOM rating per 90 min. Goalkeepers exhibit high pass accuracies but obtain low ECOM ratings. Conversely, forwards exhibit lower pass accuracies but obtain higher ECOM ratings. Although defenders exhibit comparable pass accuracies to goalkeepers, they obtain higher ECOM ratings.

6.2.3 Relationship between number of passes and average expected added reward per pass

We investigate whether players who rate high on our ECOM metric obtain their ratings mostly by performing a large number of passes or by performing high-value passes. Figure 7 presents a scatter plot showing the total number of passes per 90 min and the average expected added reward per pass for each player. The multiplication of these two numbers yields the ECOM rating for a player. The dotted line goes through the points that yield a rating of 0.3440, which is the rating for top-ranked player Mesut Özil. The red dots indicate the top-five-ranked players. The orange dots highlight three special cases.

FC Barcelona forward Lionel Messi and Chelsea winger Eden Hazard obtain a high average value per pass but perform fewer passes per match. In contrast, Manchester City midfielder David Silva obtains a lower average value per pass but performs more passes per match. AFC Bournemouth forward Jermain Defoe obtains the highest

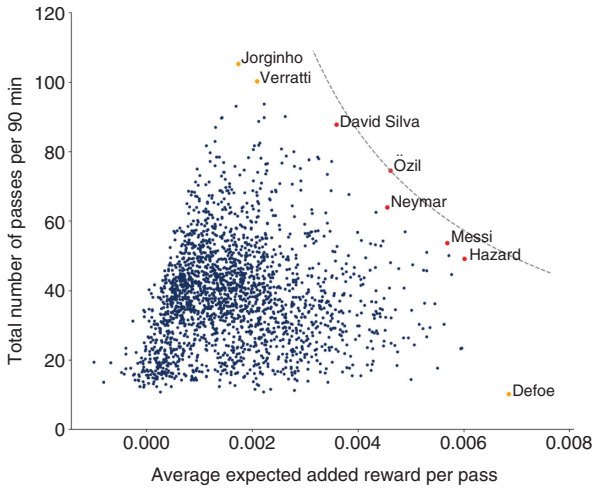

Figure 7: Scatter plot showing the total number of passes per 90 min and the average expected added reward per pass for each player. The red dots indicate the top-ranked players, while the orange dots highlight three special cases.

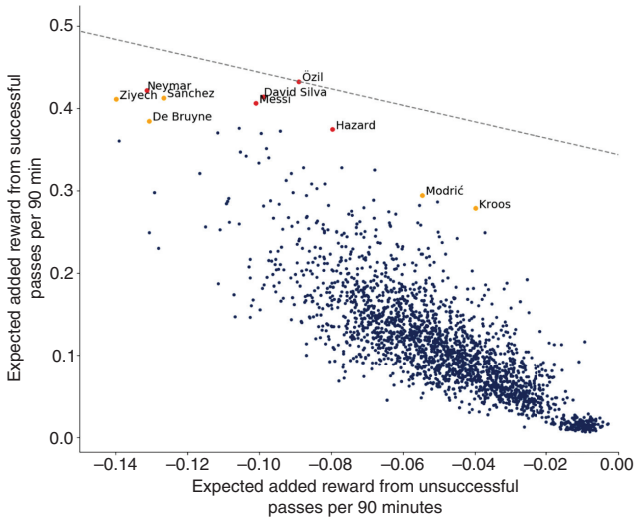

Figure 8: Scatter plot showing the expected added reward from successful and unsuccessful passes per 90 min for each player.

value per pass but performs only 11 passes per 90 min on average. Midfielders Jorginho, who played for Napoli in the 2017/2018 season, and Marco Verratti of Paris Saint-Germain perform most passes but obtain a moderate average value per pass only.

6.2.4 Relationship between expected added reward from successful and unsuccessful passes

We investigate whether players who rate high on our ECOM metric obtain their ratings by receiving positive rewards for performing successful passes or by avoiding negative rewards for performing unsuccessful passes. Figure 8 presents a scatter plot showing the value from successful and unsuccessful passes per 90 min for each player. The sum of these two numbers yields the ECOM rating for a player. The dotted line goes through the points that yield a rating of 0.3440, which is the rating for top-ranked player Mesut Özil. The red dots indicate the top-five-ranked players. The orange dots highlight five special cases.

Clearly, different types of players achieve their ECOM ratings in different ways. For instance, Hakim Ziyech (Ajax), Neymar (Paris Saint-Germain), Alexis Sánchez (Manchester United) and Kevin De Bruyne (Manchester City) compensate their large amount of negative reward from unsuccesful passes by a large amount of positive reward from successful passes. In contrast, Luka Modric (Real Madrid) and Toni Kroos (Real Madrid) achieve comparable ECOM ratings by collecting both a smaller amount of negative reward from unsuccessful passes and a smaller amount of positive reward from successful passes.

6.3 Use cases

We now present two concrete use cases for our proposed ECOM metric. We first use our metric to find a suitable replacement for Andrés Iniesta at FC Barcelona and then use our metric to estimate player market values.

6.3.1 Replacing Andrés Iniesta at FC Barcelona

Prior to the 2018/2019 season, Andrés Iniesta moved from FC Barcelona to Japanese side Vissel Kobe. The midfielder was of vital importance to FC Barcelona in winning the Spanish domestic championship and cup during the 2017/2018 season. Within the FC Barcelona squad, the Spaniard ranks second behind Lionel Messi with an ECOM rating of 0.2511. In this use case, we assume FC Barcelona aims to sign a young player who has the potential to achieve the same passing performance as Iniesta. In particular, we restrict our search to players aged 25 or younger who exhibit a similar pass behavior and impact.

We define a distance function that captures the characteristics of Andrés Iniesta's pass behavior and impact. More specifically, our distance function considers the ECOM rating, the pass accuracy, the number of passes per 90 min, and the ratio between the number of crosses and total number of passes. We normalize the four features to have values between zero and one. We compute the similarity score as one minus the Euclidean distance between these features.

Table 9 presents the top-five-ranked players under the age of 25 who most resemble Andrés Iniesta's pass behavior and impact. Ajax midfielder Frenkie de Jong tops

Table 9: The top-five-ranked players under the age of 25 who most closely resemble Andrés Iniesta's pass behavior and impact.

Rank	Player	Team	Similarity	ECOM	PA	P90	RCP
1	F. de Jong	AFC Ajax	0.9734	0.2122	93.52%	75.84	1.22%
2	C. Tolisso	FC Bayern München	0.9711	0.1946	90.97%	70.72	1.48%
3	J. Kimmich	FC Bayern München	0.9675	0.2089	87.97%	71.55	7.17%
4	M. Lopez	Olympique de Marseille	0.9440	0.1998	91.90%	89.87	2.33%
5	J. Draxler	Paris Saint-Germain	0.9364	0.1680	92.35%	68.95	1.15%
	A. Iniesta	FC Barcelona		0.2511	88.38%	73.50	2.12%

The **ECOM** column shows the players' ECOM ratings per 90 min, the **PA** column shows their pass accuracies, the **P90** column shows their numbers of passes per 90 min, and the **RCP** column shows the ratios between their number of crosses and their total number of passes.

the ranking with a similarity score of 0.9734. FC Bayern München midfielders Corentin Tolisso and Joshua Kimmich rank second and third.

6.3.2 Estimating player market values

We investigate whether our ECOM metric can help to estimate the market values of soccer players more accurately. In particular, we investigate whether the inclusion of our ECOM ratings alongside more traditional performance statistics into a predictive model improves the model's performance.

We collect the market values on the last day of the 2017/2018 season for the players in our dataset from the Transfermarkt website,[5] which we use as the ground truth. We omit five players for whom the market values are missing from the Transfermarkt website and obtain a dataset comprising 2124 players.

We address this problem as a regression task. For each player, our dataset contains the following information: the age in years, the number of minutes played in the 2017/2018 season, the number of assists per 90 min in the 2017/2018 season, the number of goals per 90 min in the 2017/2018 season, an indicator whether the player plays for a club that finished in the top three of their respective league, the position, the ECOM rating for the 2017/2018 season, and the market value on July 1st, 2018.

We use the `XGBoost` algorithm to train the models.[6] We optimize the algorithm's hyperparameters using `GridSearchCV` in `scikit-learn`. We try setting the number of estimators to 100, 500, 1000 and 2000, restricting the tree depth to 1, 2, 3, 4, 5 and 6, enforcing the number of examples per child to 1, 2, 3 and 4, and using learning rates of 0.001, 0.01, 0.1 and 0.5. We randomly split the available data in a training set containing 80% of the examples and a test set containing the remaining 20% of the examples.

Table 10: The mean absolute errors (MAE) for estimating the market values for players in the test set both with and without the ECOM metric.

Players	Examples	*MAE* without ECOM	*MAE* with ECOM
Goalkeepers	160	6.50 million	6.39 million
Defenders	777	6.54 million	6.14 million
Midfielders	760	8.68 million	7.77 million
Forwards	427	13.67 million	12.71 million
All	2124	7.18 million	6.95 million

The inclusion of the ECOM metric consistently leads to better models in terms of MAE.

We train two sets of models. The first set of models considers all available features, whereas the second set of features considers all features but the ECOM rating for the 2017/2018 season. Within each set, we train five different models: one model for each of the four positions and one model considering all players. When training the model considering all players, we include a dummy feature for each of the four positions.

Table 10 shows the mean absolute errors (MAE) for predicting the market values for players in the test set in ten different settings. The inclusion of the ECOM metric consistently leads to more accurate models in terms of MAE. Across all players, the MAE drops with 0.23 million euro from 7.18 million to 6.95 million. Unsurprisingly, we observe the largest effect for midfielders and forwards. Midfielders and forwards are primarily tasked with creating goal-scoring opportunities and scoring goals, whereas goalkeepers and defenders are primarily tasked with preventing goal-scoring opportunities and goals.

7 Conclusion and future work

This paper introduced a player performance metric for soccer named ECOM that measures players' involvement in creating goal-scoring chances by computing the expected added rewards from their passes during matches. To compute the expected added reward for

5 https://www.transfermarkt.com

6 https://xgboost.readthedocs.io/en/latest/

a pass, our approach leverages a distance-weighted *k*-nearest-neighbors search with a domain-specific distance function, whichs considers both the characteristics of the pass and the circumstances under which the pass was performed. Intuitively, passes that increase a team's likelihood of scoring receive positive expected added rewards while those that decrease a team's likelihood of scoring receive negative expected added rewards. A player's ECOM rating reflects his expected added reward per 90 min of play.

We evaluated our ECOM metric on play-by-play match data for the 2014/2015 through 2017/2018 seasons in seven European top-tier leagues. Our experiments demonstrate that our ECOM metric outperforms four baselines for predicting the outcomes of matches and carries valuable information for estimating the market values of players. Furthermore, we identified German midfielder Mesut Özil (Arsenal) as the most impactful passer during the 2017/2018 season and Dutch youngster Frenkie de Jong (Ajax) as a suitable replacement for Spanish midfielder Andrés Iniesta at FC Barcelona.

In the future, we plan to include spatio-temporal player tracking data into our distance function to better capture the circumstances under which each pass is performed. This extension should lead to more accurate expected added rewards for the passes and thus also more accurate ECOM ratings. We will also explore techniques to learn the optimal dimensions for the grid cells from the data and experiment with grid cells that vary in size depending on the distribution of the passes over the pitch.

A Qualitative analysis of the clustering step

A.1 Example 1

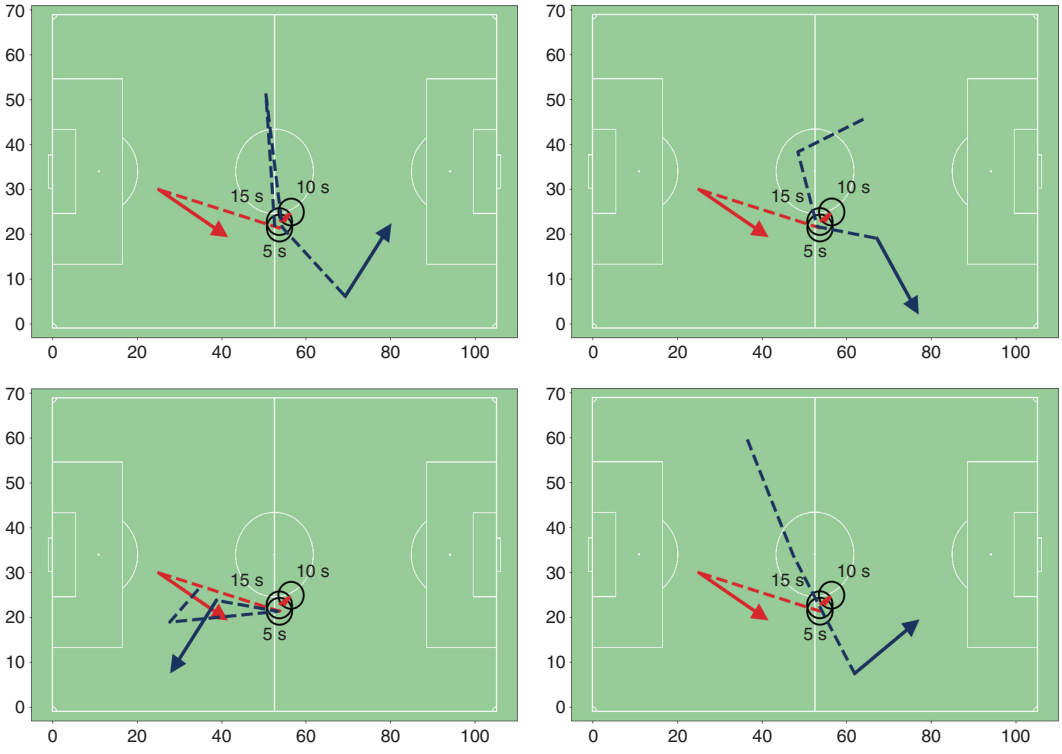

Figure 9: Visualization of the four nearest neighbors of the red pass when not clustering the passes before performing the *k*-nearest-neighbors search.

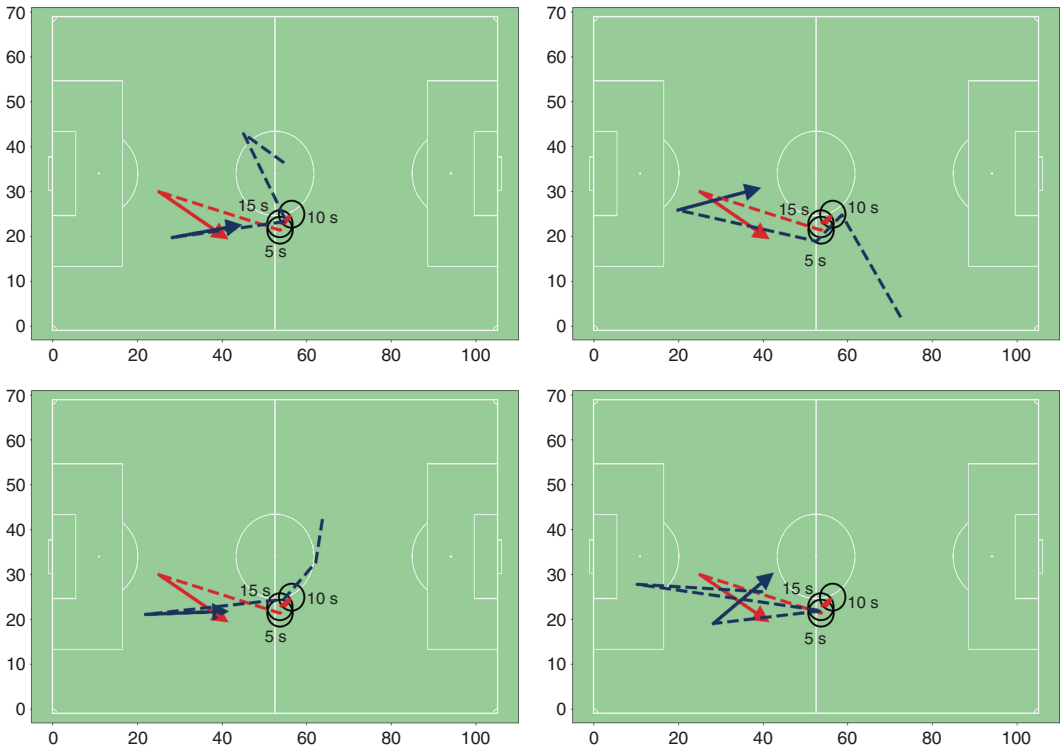

Figure 10: Visualization of the four nearest neighbors of the red pass when clustering the passes with grid cells of 15 by 17 m before performing the *k*-nearest-neighbors search.

A.2 Example 2

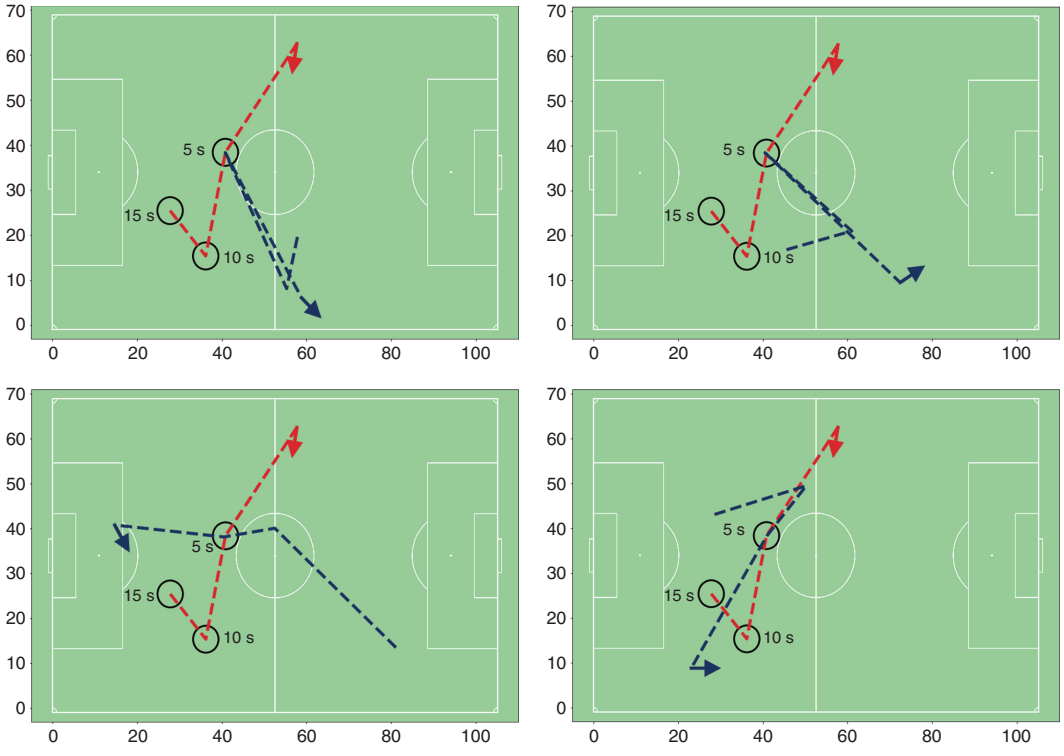

Figure 11: Visualization of the four nearest neighbors of the red pass when not clustering the passes before performing the *k*-nearest-neighbors search.

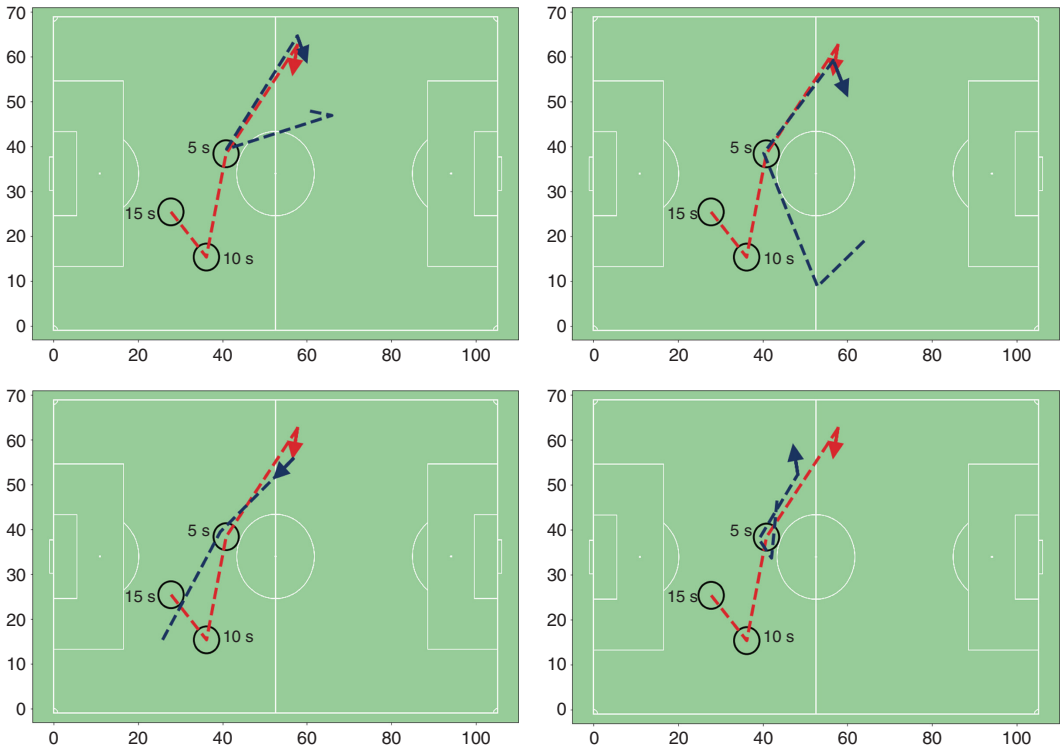

Figure 12: Visualization of the four nearest neighbors of the red pass when clustering the passes with grid cells of 15 by 17 m before performing the *k*-nearest-neighbors search.

A.3 Example 3

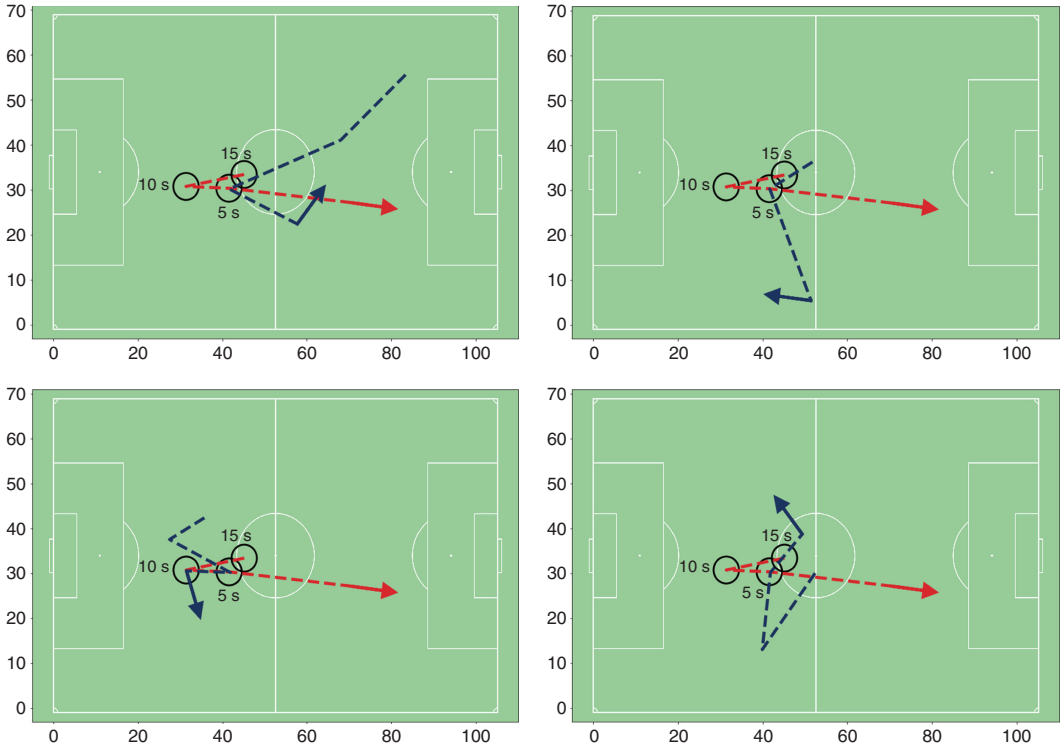

Figure 13: Visualization of the four nearest neighbors of the red pass when not clustering the passes before performing the *k*-nearest-neighbors search.

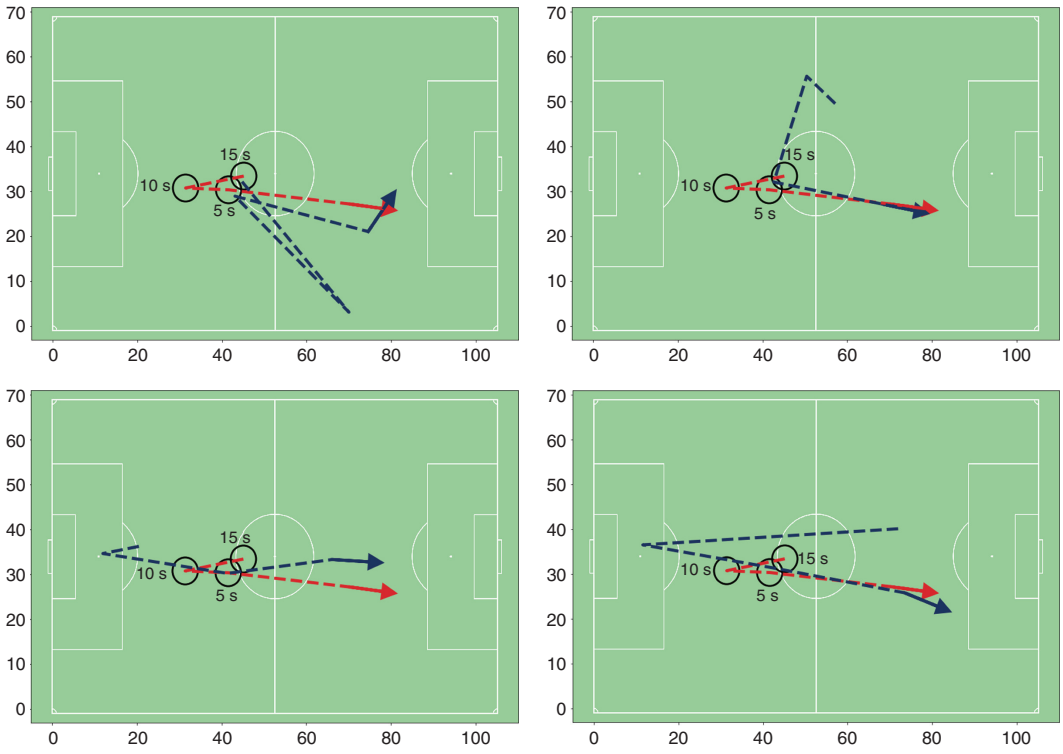

Figure 14: Visualization of the four nearest neighbors of the red pass when clustering the passes with grid cells of 15 by 17 m before performing the *k*-nearest-neighbors search.

References

Barnard, M., M. Dwyer, J. Wilson, and C. Winn. 2018. "Annual Review of Football Finance 2018." https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/sports-business-group/deloitte-uk-sbg-annual-review-of-football-finance-2018.PDF.

Beetz, M., N. von Hoyningen-Huene, B. Kirchlechner, S. Gedikli, F. Siles, M. Durus, and M. Lames. 2009. "ASPOGAMO: Automated Sports Game Analysis Models." *International Journal of Computer Science in Sport* 8(1):1–21.

Brochu, E., V. M. Cora, and N. De Freitas. 2010. "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning." *arXiv preprints 1012.2599*.

Cervone, D., A. D'Amour, L. Bornn, and K. Goldsberry. 2016. "A Multiresolution Stochastic Process Model for Predicting Basketball Possession Outcomes." *Journal of the American Statistical Association* 111(514):585–599.

Chawla, S., J. Estephan, J. Gudmundsson, and M. Horton. 2017. "Classification of Passes in Football Matches Using Spatiotemporal Data." *ACM Transactions on Spatial Algorithms and Systems (TSAS)* 3(2):6.

Decroos, T., V. Dzyuba, J. Van Haaren, and J. Davis. 2017. "Predicting Soccer Highlights from Spatio-Temporal Match Event Streams." in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, California, USA: The AAAI Press, pp. 1302–1308.

Decroos, T., L. Bransen, J. Van Haaren, and J. Davis. 2018. "Actions Speak Louder Than Goals: Valuing Player Actions in Soccer." *arXiv e-prints 1802.07127*.

Deutscher Fussball-Bund. 2017. "Fussball-Regeln 2017/2018". Online, accessed 1 August 2018. https://www.dfb.de/fileadmin/_dfbdam/143897-Fussballregeln_2017_WebPDF.pdf.

Duch, J., J. Waitzman, and L. Nunes Amaral. 2010. "Quantifying the Performance of Individual Players in a Team Activity." *PLoS One* 5(6):1–7.

Eggels, H., R. van Elk, and M. Pechenizkiy. 2016. "Explaining Soccer Match Outcomes with Goal Scoring Opportunities Predictive Analytics." in *Proceedings of the 3rd Workshop on Mining and Learning for Sports Analytics*, Riva del Garda, Italy. http://ceur-ws.org/Vol-1842/.

Fédération Internationale de Football Association. 2018. "2018 FIFA World Cup Regulations". Online, accessed 1 August 2018. https://www.uefa.com/MultimediaFiles/Download/Regulations/uefaorg/Regulations/01/87/54/21/1875421_DOWNLOAD.pdf.

Ferri, C., J. Hernández-Orallo, and R. Modroiu. 2009. "An Experimental Comparison of Performance Measures for Classification." *Pattern Recognition Letters* 30(1):27–38.

Grund, T. 2012. "Network Structure and Team Performance: The Case of English Premier League Soccer Teams." *Social Networks* 34(4):682–690.

Gudmundsson, J. and M. Horton. 2017. "Spatio-Temporal Analysis of Team Sports." *ACM Computing Surveys* 50(2):22.

Gyarmati, L. and R. Stanojevic. 2016. "QPass: A Merit-based Evaluation of Soccer Passes." *arXiv e-prints 1608.03532*.

Karlis, D. and I. Ntzoufras. 2008. "Bayesian Modelling of Football Outcomes: Using the Skellam's Distribution for the

Goal Difference." *IMA Journal of Management Mathematics* 20(2):133–145.

Langseth, H. 2013. "Beating the Bookie: A Look at Statistical Models for Prediction of Football Matches." in *Proceedings of the 12th Scandinavian AI Conference*, volume 257, pp. 165–174. http://ebooks.iospress.nl/volume/twelfth-scandinavian-conference-on-artificial-intelligence-scai-2013 and http://ebooks.iospress.nl/volumearticle/35457.

Ley, C., T. Van de Wiele, and H. Van Eetvelde. 2017. "Ranking Soccer Teams on Basis of Their Current Strength: A Comparison of Maximum Likelihood Approaches." *arXiv e-prints 1705.09575*.

Liu, G. and O. Schulte. 2018. "Deep Reinforcement Learning in Ice Hockey for Context-Aware Player Evaluation." *arXiv e-prints 1805.11088*.

Lucey, P., A. Bialkowski, M. Monfort, P. Carr, and I. Matthews. 2014. "Quality vs. Quantity: Improved Shot Prediction in Soccer Using Strategic Features from Spatiotemporal Data." in *MIT Sloan Sports Analytics Conference*, Boston, MA, USA.

Maher, M. 1982. "Modelling Association Football Scores." *Statistica Neerlandica* 36(3):109–118.

Power, P., H. Ruiz, X. Wei, and P. Lucey. 2017. "Not All Passes Are Created Equal: Objectively Measuring the Risk and Reward of Passes in Soccer from Tracking Data." in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, USA: ACM Press, pp. 1605–1613.

Rein, R., D. Raabe, and D. Memmert. 2017. "'Which Pass Is Better?' Novel Approaches to Assess Passing Effectiveness in Elite Soccer." *Human Movement Science* 55:172–181.

Schulte, O., Z. Zhao, and K. Routley. 2015. "What is the Value of an Action in Ice Hockey? Learning a Q-function for the NHL." in *Proceedings of the 2nd Workshop on Machine Learning and Data Mining for Sports Analytics*, Porto, Portugal. http://ceur-ws.org/Vol-1970/.

Snoek, J., H. Larochelle, and R. Adams. 2012. "Practical Bayesian Optimization of Machine Learning Algorithms." in *Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, pp. 2951–2959. https://papers.nips.cc/book/advances-in-neural-information-processing-systems-25-2012.

Union of European Football Associations. 2018. "UEFA Pitch Quality Guidelines." Online, accessed 1 August 2018. https://www.uefa.com/MultimediaFiles/Download/uefaorg/Stadium&Security/02/54/11/97/2541197_DOWNLOAD.pdf.

Van Haaren, J., V. Dzyuba, S. Hannosset, and J. Davis. 2015. "Automatically Discovering Offensive Patterns in Soccer Match Data." in *Advances in Intelligent Data Analysis XIV*, Vol. 9385, Saint Etienne, France: Springer Verlag, pp. 286–297. https://link.springer.com/chapter/10.1007/978-3-319-24465-5_25#citeas.