Valuing the Art of Pressing

Pieter Robberechts

KU Leuven
Leuven, Belgium
pieter.robberechts@cs.kuleuven.com

Abstract. Pressing is an essential part of defense in football. Broadly speaking, the goal is to quickly win the ball back by putting pressure on the player in possession. The successes of coaches like Guardiola, Klopp, Sarri and Pochetino in deploying a high press has increased the profile and interest in this strategy. Yet, pressing is a phenomenon that has not yet received much attention from researchers in football analytics. Previous research has focused on the spatial aspect and the intensity of pressing, but we currently lack metrics to quantify its effectiveness in different contexts. This paper introduces a novel metric that quantifies the effectiveness of pressing in different game scenarios as a trade-off between the benefits of recovering the ball versus the cost leaving the defensive structure, which makes passing through the lines easier for the opposition. We show how our metric can be used in practice through a number of use cases in the 2018/19 season of Europe's top leagues.

1 Introduction

Pressing is an essential part of defense in football. The goal is to win the ball or at least deprive the opponents of the opportunity to develop an attack by putting pressure on the player possessing the ball. Effective pressing has played a crucial role in the success of several high-level football clubs, such as Liverpool, Manchester City, and Atletico Madrid. Club and media analysts are therefore very interested in gaining insights into how teams organize and conduct pressing activity, and the circumstances in which this tactic may be effective.

If executed effectively, the pressing team can strangle the opponent and disrupt their rhythm of play. On the other hand, when executed incorrectly, a team's press can be bypassed by a player that keeps his cool and manages to pass through the zone of press or by teams using a more agricultural, direct style of play. If the opposition can switch the direction of attack using a diagonal pass or pass through the lines, they would have lots of space to launch an attack. Therefore, the decision of whether or not to press can be regarded as a risk-reward trade-off. This risk-reward framework raises several interesting questions about the effectiveness of pressing, such as "In which game situations does the risk of pressing outweigh the reward?", "In which game situations does the opponent manage to repeatedly break through a team's pressing?", "Can a team maintain an effective low-risk pressing strategy throughout the entire game or do they get fatigued at some point?", "Which players are making good decisions under pressure?", etc.

Nevertheless, the effectiveness of pressing is a phenomenon that has not yet received much attention from researchers in football data analysis. Previous research focused on the spatial aspect and the intensity of pressing [2, 1], but metrics to quantify its effectiveness are currently lacking. Therefore, this paper introduces the VPEP (Valuing Pressure decisions by Estimating Probabilities) metric, a novel performance metric for valuing pressure actions performed by football players. The VPEP metric is heavily inspired by Decroos et al.'s [3] VAEP metric.

Using this metric, we analyzed event data provided by StatsBomb for 1,826 matches from the 2018/19 seasons of the five European top leagues. At the moment of writing, StatsBomb is the only provider that provides information about pressing in its event data [4]. They describe pressing as a separate event that is triggered when a defending player is within a five-yard radius of an opponent in possession. The radius varies as errors by the opponent would prove more costly, with a maximum range of ten-yards that is usually associated with goalkeepers under pressure. As well as logging the players involved in the pressure event and its location, the duration of the event is also collected. Although these events can't provide any information about the organization and cohesion of the press behind the front players, it is sufficient to uncover interesting insights about the locations and game contexts where pressing can be advantageous.

In the remainder of this paper, we first discuss how the decision about whether or not to press can be rated objectively by formulating pressing as a risk-reward trade-off. Second, we discuss how the risk and reward can be expressed as probabilities by using machine learning. Finally, we evaluate our metric and demonstrate how it can help clubs discover weak links in the pressing of their team and opponents.

2 VPEP: A framework to assess the risk-reward trade-off of pressing

Broadly speaking, the goal of pressing is to quickly win the ball back. If executed efficiently, the team pressing can prevent the opposition from settling into a passing rhythm and deprive them of the opportunity to develop an attack. Playing with a high-pressing style does have a downside, however. Leaving the defensive structure to press can make passing through the lines easy for the opposition. As such, an aggressive high press can be a chance-conceding machine when it fails. When opposing teams have found a way through, they are often able to get to goal at great speed.

Therefore, as any other action in football, the decision whether or not to press in a game situation requires that players make a quick assessment of the rewards (how likely am I to regain possession in this game situation) and the risks (how likely am I to fail to recover the ball and open up space for the opponent to launch an attack). Ultimately, a player should only decide to press if his pressing increases the probability of recovering the ball without significantly increasing the probability of yielding a goal-scoring opportunity.

Correspondingly, we can assess a pressure event by calculating how much it alters the chances of the defending team for both recovering the ball and conceding a goal-scoring opportunity in the near future. We describe how these probabilities can be estimated in the next section. For now, assume that for each game state S_i we have access to the probabilities

- $P_{recovery}(S_i, p_i)$: that the defending team will recover the ball in the near future after game state S_i in which it presses on the opponent as described by the pressure events p_i
- $P_{attack}(S_i, p_i)$: that the defending team will concede a goal-scoring opportunity in the near future after game state S_i in which it presses on the opponent as described by the pressure events p_i

Here, a game state $S_i = [a_1, \ldots, a_i]$ corresponds to a sequence of all on-the-ball actions a_i up to a specific point in the game. We use the SPADL language [3] to describe these actions. SPADL describes each action in terms of nine attributes: the action's start and end time, the (x,y) location where the action started and ended, the player who performed the action and his team, the type of the action (e.g., pass, shot, dribble), the body part used to perform the action, and the result of the action (e.g., success or fail). Additionally, we extend SPADL with a tenth attribute containing a set p_i that describes all pressure events of the opposing team during action a_i. Each of these pressure events is described by the same nine attributes used to describe any other action. We use the empty set symbol $\varnothing$ to denote a lack of pressing.

To value these decisions, our metric will assess how pressing (p_i) would change the probability for both recovering the ball ($\Delta P_{recovery}$) and incurring a goal-scoring opportunity (ΔP_{attack}) in the near future in a game state S_i.

$$\Delta P_{recovery}(S_i) = P_{recovery}(S_i, p_i) - P_{recovery}(S_i, \varnothing) \tag{1}$$

$$\Delta P_{attack}(S_i) = P_{attack}(S_i, \varnothing) - P_{attack}(S_i, p_i) \tag{2}$$

Intuitively, $\Delta P_{recovery}$ will be larger if pressing is more likely to succeed, while ΔP_{attack} will be larger if pressing is more likely to fail. Therefore, we combine both equations to derive a pressure event's total value as

$$V(S_i) = \Delta P_{recovery}(S_i) - C * \Delta P_{attack}(S_i) \tag{3}$$

Since the cost of failure is much higher than the reward of success in this framework, we add a factor C to the risk factor. We chose to set C to 5.

3 Estimating the success and failure probabilities of pressing

To estimate the probabilities $P_{recovery}(S_i, p_i)$ and $P_{attack}(S_i, p_i)$, we train two binary probabilistic classifiers on historical match data.

Given: Game state S_i, where p_i is a set of pressure events by the defending team during S_i
Estimate: (1) $P_{recovery}(S_i, p_i)$, and (2) $P_{attack}(S_i, p_i)$.

In principle, any probabilistic (e.g., Logistic Regression, Random Forest, or Neural Network) classifier could be used to address these tasks. However, it is essential that the probability estimates are well-calibrated [6]. We use XGBoost stacked with a logistic regression model to estimate these probabilities [5]. The logistic regression model is applied to the leaves of the forests in order to improve probability calibration.

To be able to train such a classifier, we first have to convert each game state to the feature-vector format. We now describe how we compute the labels and features for each game state.

3.1 Labels

For the first classification problem of estimating $P_{recovery}(S_i, p_i)$, we assign a game state S_i a positive label (= 1) if the defending team in state S_i recovers the ball in the subsequent k actions, and a negative label (= 0) in all other cases. Similarly, for the second classification problem of estimating $P_{attack}(S_i, p_i)$, we assign a game state S_i a positive label (= 1) if the defending team concedes a shot in the subsequent k actions, and a negative label (= 0) in all other cases. In both binary classification problems, k is a user-defined parameter that represents how far ahead in the future we look to determine the effect of an action. In this paper, we chose $k = 4$ for $P_{recovery}$ and $k = 7$ for P_{attack}. For these values of k, we observe a large effect of pressing on $P_{recovery}$ and P_{attack} (Table 1). Also, we prefer $k = 4$ for $P_{recovery}$ instead of $k = 6$ to better capture the short term objective of regaining possession as quickly as possible (usually within five seconds).

3.2 Features

A complete description of game state S_i would encompass all actions $[a_1, \ldots, a_i]$ up to that specific point in the game. Since it is infeasible to define features based on this entire game state, we only consider the previous three actions $[a_{i-2}, a_{i-1}, a_i]$. This small window contains the most relevant aspects of the current context and was empirically found to work well. From these three actions, we define features that impact the probability of recovering the ball and conceding a goal-scoring opportunity in the near future. Based on the SPADL representation, we consider three categories of features.

1. **SPADL features.** For each of the three actions, we define a set of categorical and real-valued features based on information explicitly included in the SPADL representation. We consider categorical features for an action's type and result, and the body part used by the player performing the action. Similarly, we consider real-valued features for the (x,y)-coordinates of

Table 1: The difference in probability (%) to recover the ball ($P_{recovery}$) and to incur a goal-scoring opportunity (P_{attack}) in the next k actions after an action that happened under pressure vs after an action that happened free of any pressure.

k	$P_{recovery}$			P_{attack}		
	With pressing	Without pressing	Difference	With pressing	Without pressing	Difference
1	16.89	13.09	3.80	1.22	1.20	0.03
2	30.93	22.38	8.54	2.39	2.29	0.10
3	39.87	31.23	8.64	3.74	3.28	0.47
4	48.83	37.98	10.85	4.85	4.27	0.58
5	54.95	44.65	10.29	5.98	5.18	0.80
6	61.01	50.05	10.96	6.95	6.11	0.84
7	65.44	55.25	10.19	7.89	6.97	0.92
8	69.75	59.62	10.13	8.72	7.82	0.90
9	73.03	63.76	9.26	9.53	8.64	0.89
10	76.26	67.27	8.99	10.27	9.43	0.84

the action's start and end locations, and the time elapsed since the start of the game.

2. **Complex features.** The complex features combine information within an action and across consecutive actions. Within each action, these features include (1) the distance and angle to the goal for both the action's start and end locations, and (2) the distance covered during the action in both the x and y directions. Between two consecutive actions, we compute the distance and elapsed time between them and whether the ball changed possession. These features provide some intuition about the current speed of play.
3. **Game context features.** The game context features are (1) the number of goals scored in the game by the team possessing the ball after action a_i, (2) the number of goals scored in the game by the defending team after action a_i, and (3) the goal difference after actiona_i. We include these features because teams often adapt their playing style to the current scoreline (e.g., a team that is 1-0 ahead will play more defensively than a team that is 0-1 behind).
4. **Pressure features.** For each pressure event associated with an action, we include (1) the distance between the defender and the ball, (2) the angle between the defender's goal, the ball and the defender and (3) the time delay between the start of the action and the start of the pressure.

4 Experiments

Since no ground truth ratings exist for pressing, it is not straightforward to evaluate our metric. Therefore, our experiments first evaluate the underlying machine learning models of our metric. These are the binary probabilistic classifiers that estimate $P_{recovery}$ and P_{attack}. Second, we provide intuitions into how our framework behaves and compares to other metrics. Third, we present use cases revolving around how teams can use our metric to identify the weaknesses of their pressing strategy.

We apply our analysis on StatsBomb event data for the 2018/19 season of the English, Spanish, German, Italian and French top divisions. Due to the limited dataset, we did not work with a separate training set. Instead, we use the first 80% of the games in each league to train the two classification models using the XGBoost algorithm and the feature set described previously. The remaining 20% is used as a validation set to avoid overfitting. The learned models are then applied on all 1826 games to produce VPEP ratings for the entire 2018/19 season.

4.1 Evaluation of the probabilistic classifiers

There exist no objective guidelines about the game contexts in which pressing is either a good or bad idea. Each coach and football enthusiast will have different opinions on this. As a result, the only way to evaluate our metric is to evaluate the components it consists of. In our case, we evaluate the VPEP values by evaluating the underlying probabilities for recovering the ball and conceding a goal-scoring opportunity, for which ground truth labels are available.

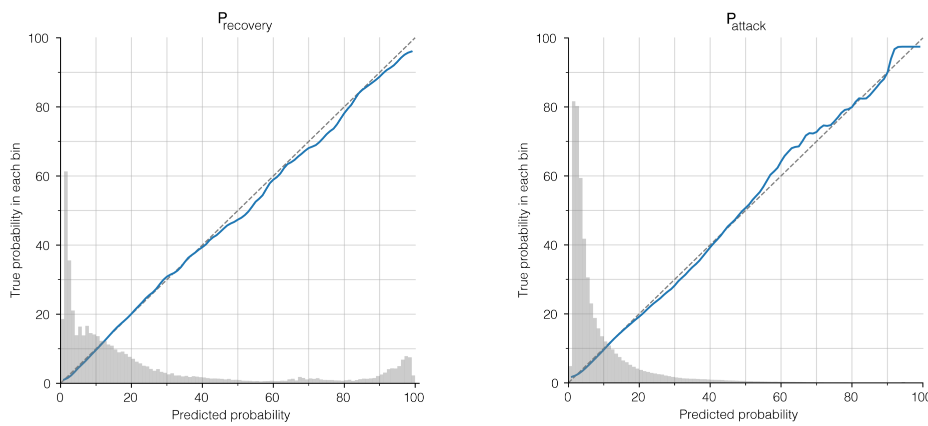


Fig. 1: Probability calibration curves and histograms of the predicted probabilities. Both the probabilities for $P_{recovery}$ (left) and P_{attack} (right) are well-calibrated.

We evaluate two aspects of each classifier: the accuracy and the calibration. First, our classifiers should be able to predict accurately whether the defending team will either recover the ball or concede a goal-scoring opportunity in the near future. We use the area under the receiver operator curve (ROC AUC) to evaluate our classifiers with respect to this goal. An important advantage of

ROC AUC is that the metric is unaffected by unbalanced data sets, as is the case in our data. With a ROC AUC of 0.902 for $P_{recovery}$ and 0.806 for P_{attack}, we can conclude that our classifiers can distinguish between positive and negative examples with reasonably high accuracy.

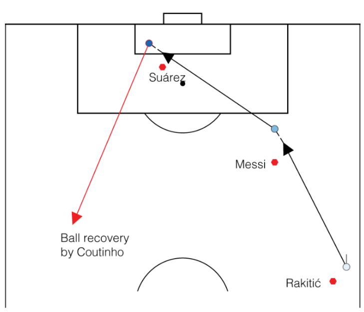
Second, the predicted probabilities should correspond with the true underlying probability distribution of the data. For example, when the defending team is given an 8% probability of recovering the ball at a given state of the game, this essentially means that if the game was played from that state onwards a hundred times, the team is expected to actually recover the ball in eight of these hundred games. However, since each game is played only once, this cannot be assessed for a single game. Therefore, we collect all game states for which our model predicts an 8% recovery probability and then look at whether about 8% of those game states actually resulted in a ball recovery. This is reflected in the probability calibration curves in Figure 1. As can be seen, we achieve a good calibration for both classifiers.

4.2 Intuition behind the values

Figure 2 illustrates how our framework works for a successful pressing sequence of Barcelona in the fourth minute of their game against Huesca on September 2, 2018. The sequence starts with Huesca's left-back Luisinho. Under pressure from Barcelona's midfielder Rakitić, Luisinho is forced to pass backwards to his teammate Xabier Etxeita. Immediately, Messi puts pressure on him, forcing Etxeita to pass back to his goalkeeper Werner. Finally, due to the onrushing Luis Suárez, Werner kicks an inaccurate long pass which is recovered by Barcelona's Coutinho.

Barcelona - Huesca

LaLiga | 02/09/2018



Time	Action	Player	Team	$P_{recovers}$	P_{attack}	VPEP
4'32	Pass	Luisinho	Huesca	0.080	0.022	
	Pressing	Rakitić	Barcelona	0.137	0.017	
				Δ 0.056	0.005	0.040
4'36	Pass	Xabier Etxeita	Huesca	0.052	0.010	
	Pressing	Messi	Barcelona	0.095	0.008	
				Δ 0.043	0.002	0.033
4'40	High pass	Axel Werner	Huesca	0.224	0.015	
	Pressing	Suárez	Barcelona	0.409	0.013	
				Δ 0.186	0.001	0.181

Fig. 2: A successful pressing sequence of Barcelona in their game against Huesca on September 2, 2018.

The sequence contains three pressure events. The first two events by Rakitić and by Messi get a rating of respectively 0.040 and 0.033. They are low risk, since both Rakitić and Messi push the opponent backward and to the sideline. Also, the probability of recovering the ball is relatively low, since both Luisinho and Etxeita can easily pass the ball backward where Huesca can switch sides. On the other hand, Suárez pressing the goalkeeper Werner has a much higher rating of 0.181. Suárez can really corner Werner, leaving him no other option but to kick the ball away. Without Suárez pressing, Werner could have kicked a more accurate long ball or passed simply to the right back. Therefore, Suárez's decision to press adds substantial value and gets a high rating.

4.3 Identifying the best pressing teams

Table 2 shows the top-ranked teams according to our VPEP metric. This ranking contains the usual suspects: Bayern Munich, Napoli, Manchester City, Barcelona and Borussia Dortmund. These are all teams that received acclaim for their pressing style. Perhaps more surprisingly is that SD Eibar leads the ranking. Their tactical system is unique. They indulge in a high and very intense wing-oriented pressing. Typically, Eibar crowds one wing, such that the centre-back and full-back at the other end is inaccessible.[1]

Table 2: Teams with the highest average VPEP rating per pressure event. For comparison, the number of pressure events per game and the Passes Per Defensive Action (PPDA) are reported too. The rank of each team according to these metrics is written between parentheses.

	Team	**# pressure events per game**	**PPDA**	$\mathbf{\Delta P_{recovery}}$	$\mathbf{\Delta P_{attack}}$	**VPEP**
1	Eibar	219.895 (77)	7.115 (1)	0.0306	0.00223	0.0195
2	Bayern Munich	183.382 (98)	9.395 (4)	0.0312	0.00252	0.0186
3	Sampdoria	233.868 (52)	10.907 (16)	0.0301	0.00241	0.0181
4	Napoli	231.500 (57)	10.771 (15)	0.0311	0.00279	0.0172
5	Inter Milan	224.105 (67)	9.831 (7)	0.0304	0.00279	0.0165
6	Manchester City	206.763 (91)	9.510 (5)	0.0288	0.00251	0.0163
7	Barcelona	203.763 (93)	10.191 (9)	0.0280	0.00240	0.0161
8	Real Madrid	201.658 (94)	10.245 (11)	0.0278	0.00262	0.0149
9	Juventus	212.447 (86)	11.933 (29)	0.0284	0.00280	0.0144
10	Borussia Dortmund	209.529 (89)	14.781 (73)	0.0288	0.00289	0.0144

[1] For a detailed analysis of Eibar's pressing strategy in the 2018/19 season see https://sport360.com/article/football/la-liga/319752/how-eibar-is-handing-out-pressing-lessons-to-barcelona-and-real-madrid and https://thefutebolist.wordpress.com/2018/01/27/sd-eibar-unique-thrilling-and-fascinating/

Even more interestingly, Eibar have conceded the 4th-lowest shots per game in the big-5 leagues in the 2018/19 season with just 8.4 shots conceded per game. Only Manchester City (6.2), Bayern Munich (7.1) and Liverpool (8.0) conceded fewer shots. This means that their high press is also very effective at limiting shots. But then again, Eibar are easy to score against once the press is broken. Although few, the shots they concede are of high quality (0.16 xG/shot against according to Understat.com; highest in the league).

Additionally, Table 2 compares our VPEP metric with existing metrics to quantify pressing. Perhaps the most widely-used example is “passes per defensive action” or PPDA [7]. The PPDA metric is calculated by dividing the number of passes allowed by the defending team by the total number of defensive actions. As such, this metric expresses how often the defending team decides to press, relative to the number of times that it is possible for the team to press. A smaller PPDA value signifies a greater level of defensive intensity, as in essence, the defence has allowed a smaller ratio of uncontested passes to be made.

There is a strong correlation between the PPDA metric and the VPEP metric. Teams with a low PPDA value often score good VPEP ratings too. Yet they measure different things. While PPDA only measures the intensity of pressing, VPEP measures whether the pressing is done intelligently. Therefore, teams with a high PPDA can still obtain a high VPEP score. For example, Dortmund adopted a more passive defensive strategy during the 2018/19 season, pressing less in the opposition half. Instead they wait for certain pressing triggers and press intelligently. Therefore, they have a relatively high PPDA value of 14.781, but rank 10th in terms of VPEP.

4.4 How fatigue affects the effectiveness of pressing

Pressing requires extremely high fitness levels in order to be executed. Players have to be able to close down options for the opposition by applying pressure to the player in possession. It is often hard to execute this style of play for the full 90 minutes without having trained extensively. This also becomes evident in our results. Figure 3 shows the average VPEP rating for all teams grouped by 15-minute periods, as well as the total number of pressure events in each period. Although the number of pressure events does not change significantly, the average VPEP value decreases near the end of a game.

One example of a team playing implementing this style of play without the required fitness levels to press for the full match is the Liverpool under Jürgen Klopp (Figure 4). Whilst Klopp has received acclaim across his career for his pressing tactics, Liverpool’s pressing seems to taper off after the 30 minute mark. This is in contrast with their 2018/19 competitor for the Premier League title, Manchester City. Pep Guardiola’s side managed to keep pressing effectively during the entire game, only slightly tapering off during the final fifteen minutes of each half.

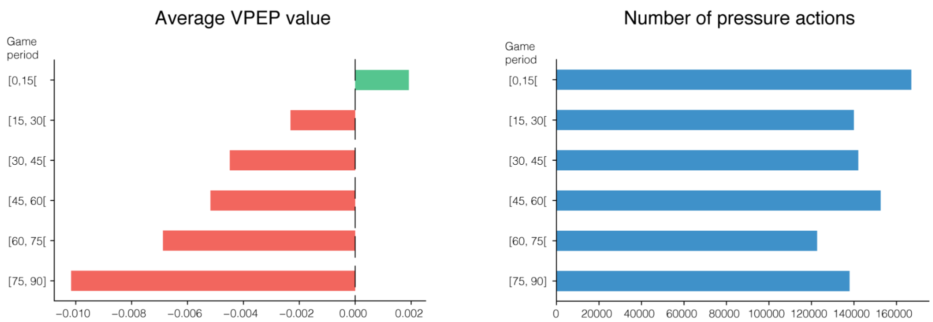


Fig. 3: Average VPEP value during the 2018/19 per fifteen-minute game period for all teams (Left) and the total number of pressing events during each period (right). While the total number of pressing events does not change significantly as the games progress, the VPEP values decrease consistently over the course of a game.

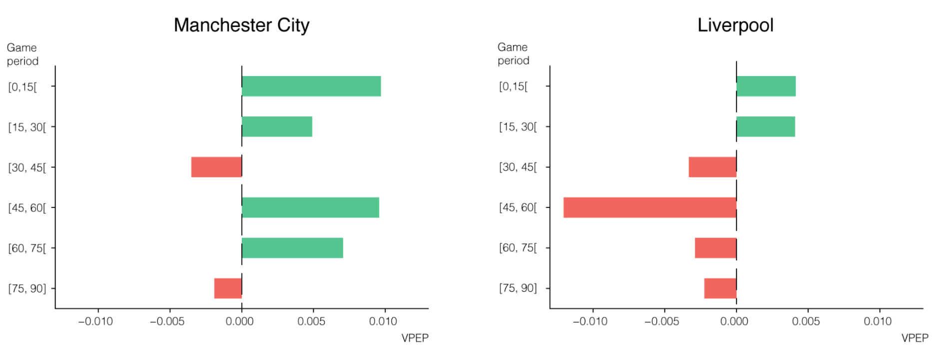


Fig. 4: Average VPEP value during the 2018/19 per fifteen-minute game period for Manchester City (Left) and Liverpool (right). While Manchester City's pressing only tapers off slightly during the final quarter of each half, Liverpool's pressing already tapers off after the 30 minute mark.

5 Conclusion

This paper introduced the VPEP metric as a novel metric to quantify the effectiveness of pressing in different game scenarios as a trade-off between the benefits of recovering the ball versus the risk of leaving the defensive structure. In contrast to existing metrics for assessing pressing, the VPEP metric takes the context of the pressing into account. Although the pressing-related information included in event stream data is still limited, our metric can provide club analysts with useful information for assessing their defensive strategy, as well as provide insights into the weaknesses of their opponents.

Acknowledgements

Pieter Robberechts is supported by the EU Interreg VA project Nano4Sports.

References

1. Andrienko, G., Andrienko, N., Budziak, G., Dykes, J., Fuchs, G., von Landesberger, T., Weber, H.: Visual analysis of pressure in football. Data Mining and Knowledge Discovery **31**(6), 1793–1839 (Nov 2017). https://doi.org/10.1007/s10618-017-0513-2
2. Bojinov, I., Bornn, L.: The Pressing Game: Optimal Defensive Disruption in Soccer. In: MIT Sloan Sports Analytics Conference. p. 8. Boston, USA (2016)
3. Decroos, T., Bransen, L., Van Haaren, J., Davis, J.: Actions Speak Louder than Goals: Valuing Player Actions in Soccer. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19. pp. 1851–1861. ACM Press, Anchorage, AK, USA (2019). https://doi.org/10.1145/3292500.3330758
4. Gürpınar-Morgan, M.: How StatsBomb Data Helps Measure Counter-Pressing (May 2018)
5. He, X., Bowers, S., Candela, J.Q., Pan, J., Jin, O., Xu, T., Liu, B., Xu, T., Shi, Y., Atallah, A., Herbrich, R.: Practical Lessons from Predicting Clicks on Ads at Facebook. In: Proceedings of 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining - ADKDD'14. pp. 1–9. ACM Press, New York, NY, USA (2014). https://doi.org/10.1145/2648584.2648589
6. Niculescu-Mizil, A., Caruana, R.: Predicting Good Probabilities with Supervised Learning. In: Proceedings of the 22Nd International Conference on Machine Learning. pp. 625–632. ICML '05, ACM, New York, NY, USA (2005). https://doi.org/10.1145/1102351.1102430
7. Trainor, C.: Defensive Metrics: Measuring the Intensity of a High Press (Jul 2014)