Field Depth Matters: Comparing the Valuation of Passes in Football

Leo Martins de Sá-Freire(✉) and Pedro O. S. Vaz-de-Melo

Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
leomartins@dcc.ufmg.br, olmo@dcc.ufmg.br

Abstract. This study delves into the influence of missing spatial context information on the valuation of football actions through event data based metrics. Using actions from an entire Premier League season, we analyze successful passes originating from different field depths, considering the subsequent occurrence of goals. By comparing the value assignments by Valuing Actions by Estimating Probabilities (VAEP), we provide insights into the metric's ability to recognize the quality of passes in the early stages of attacks.

Keywords: Sports analytics · Event data · Machine Learning

1 Introduction

Valuation metrics have emerged in the field of football analytics as a way to quantify the impact of each action in a match [1]. With high performance and a wide range of useful employments, they have generated academic developments and industrial applications over the past years [3,7–12]. Among the state-of-the-art metrics, VAEP (Valuing Actions by Estimating Probabilities) [1] and xT (Expected Threat) [4]) stand out, and they have already been compared in previous research [2].

VAEP is particularly notable for its sensitivity to the specificities of each action [2]. By analyzing event data, it assigns values to actions by estimating goal probabilities at each moment of the game. This approach allows for valuing every move in the game based on the probability variation it generates.

However, event data does not contain the location of all players on the field for every stored event. Only the protagonist player of each action has their coordinates recorded. Consequently, this representation suffers from a significant loss of spatial context information, which inevitably limits the effectiveness of metrics utilizing it. One notable limitation is the neglect of positional advantage, which refers to the advantage resulting from the relation between players' positions from both teams.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

U. Brefeld et al. (Eds.): MLSA 2023, CCIS 2035, pp. 64–73, 2024.
https://doi.org/10.1007/978-3-031-53833-9_6

Positional advantage is among the key aspects of modern football [5,6]. In a game where the offensive process is increasingly integrated, it is possible to observe that many goals and dangerous plays are a consequence of coordinated movements that put the attacking team in equal or numerical superiority situations on the field. Many of these movements start far from the opponent's penalty area, with passes resulting from good readings by defenders, full-backs, and midfielders initiating an attack.

The advantage generated by this kind of pass seems to be mainly positional. Therefore, we raise the following question: Considering that VAEP is based on event data, can it recognize contributive passes in the early stages of an attack? Throughout this work, we will gather evidence to answer this question.

The proposed analysis consists of implementing a framework of VAEP applied to an entire season of the Premier League. From this, we will group successful passes that occurred throughout the championship according to the depth of the field from which they originated and the occurrence or non-occurrence of goals in the near future. Subsequently, we will compare the value assignment by VAEP in the different groups.

2 Methodology

This section presents an overview of the methodology utilized in the study, highlighting the fundamental approaches. Initially, the description of football event data will be provided. Following that, we will introduce the VAEP framework, which utilizes a machine learning algorithm to establish the probabilities of scoring or conceding a goal in each gamestate. Next, the process of assigning a value to each action will be explained. Finally, we will elucidate the grouping of examined passes and list the conducted analyses and tests.

2.1 Event Data

In football, event data refers to a data representation that captures the sequence of on-ball actions throughout a match. In this work we used the SPADL [1] format and the stored attributes of each event are described in Table 1. Although this data type includes a complete overview on each action on the ball, event stream data lacks spatial context information once it ignores the other players' positions during each action.[1]

2.2 VAEP

Event data allows us to represent a football match as a finite sequence of actions $(a_1, a_2, ..., a_n)$, where $n \in \mathbb{N}$. The VAEP framework is based on gamestates consisted of three consecutive actions. A gamestate S_i is defined as (a_{i-2}, a_{i-1}, a_i).

[1] An alternative data representation is tracking data, which constantly captures all players and ball coordinates. Although it details players positioning, it is more expansive, more challenging to acquire, and requires a significantly larger volume of instances to represent full matches.

Table 1. SPADL Attribute Descriptions

Attribute	Description
StartTime	The action's start time
EndTime	The action's end time
StartLoc	The (x, y) location where the action started
EndLoc	The (x, y) location where the action ended
Player	The player who performed the action
Team	The player's team
ActionType	The type of the action
BodyPart	The player's body part used for the action
Result	The result of the action

Then, each gamestate S_i is associated with the probabilities of the team t in possession of the ball scoring ($P_{scores}(S_i, t)$) or conceding ($P_{concedes}(S_i, t)$) a goal within the subsequent ten actions. These probabilities are separately estimated through a learning algorithm that utilizes features derived from the SPADL attributes of the 3 gamestate actions, as detailed in [1]. A binary label is used to indicate whether a goal occurs or not shortly after the action sequence. Finally, a value is assigned to each action a_i, denoted as $V(a_i)$. The calculation is performed as follows:

$$\Delta P_{scores}(S_i, t) = P_{scores}(S_i, t) - P_{scores}(S_{i-1}, t)$$

$$\Delta P_{concedes}(S_i, t) = P_{concedes}(S_i, t) - P_{concedes}(S_{i-1}, t)$$

$$V(a_i) = \Delta P_{scores}(S_i, t) - \Delta P_{concedes}(S_i, t)$$

From now on, we will refer to $\Delta P_{scores}(S_i, t)$ as the offensive value of action a_i.

For our implementation, we utilized a free event data sample provided by Wyscout, containing the five main European leagues in the 2017/18 season. We used the matches from La Liga, Ligue 1, Serie A, and Bundesliga for training the models and defined the matches from Premier League as the test dataset. Considering the results reported in [13], XGBoost was the chosen learning algorithm, with default hyperparameters except for those indicated in Table 2. The models performances were measured using the Normalized Brier Score (NBS) and the obtained values are displayed in Table 3. The Brier Score is defined as the mean squared error between the predicted probabilities and the corresponding binary outcomes. The normalization step involves dividing the Brier Score by the baseline Brier Score, where the predicted probabilities are equal to the observed frequency of the event in the dataset. A lower NBS indicates better calibration and higher reliability, signifying that the probabilities align well with the actual outcomes.

Table 2. Models Hyperparameters

Model	Algorithm	Objective	Learning Rate	Max Depth	Number of Estimators
Scores	XGBoost	Binary: Logistic	0.1	9	100
Concedes	XGBoost	Binary: Logistic	0.1	9	100

Table 3. Models Performance

Dataset	Size	Model	NBS
Train	1454120	Scores	0.814
Train	1454120	Concedes	0.876
Test	482901	Scores	0.847
Test	482901	Concedes	0.969

2.3 Groups and Tests

With the set of actions from the entire Premier League 17/18 season properly valuated by the metric, it is possible to conduct the desired investigation. Initially, we separated a dataset including all successful passes throughout the championship. From this dataset, we extracted four groups, 1-G, 1-NG, 3-G, and 3-NG, according to pass origin and the occurrence of goals by the team in possession in the subsequent ten actions. The criteria and the size of each group are presented in Table 4. We used the traditional division of the pitch into thirds, consisting of the defensive third (1st), the midfield third, and the attacking third (3rd), selecting only passes from the initial and the final thirds. The occurrence of goals is represented by the binary label of the scores model.

Table 4. Groups of successful passes.

Group	Pass origin	Label	Size
1-G	1st third	True	572 passes
1-NG	1st third	False	76036 passes
3-G	3rd third	True	1606 passes
3-NG	3rd third	False	56395 passes

We proceeded by first examining the Cumulative Distribution Functions (CDFs) of the offensive values in the groups of passes, comparing groups 1-G $\times$ 1-NG and 3-G $\times$ 3-NG. Additionally, we conducted Kolmogorov-Smirnov tests to analyze the differences between the groups. Next, we compared the CDFs of groups 1-G and 3-NG to gain insights into their offensive value distributions. Lastly, we closely examined a selection of high-valued passes through plots.

3 Results

With only event data attributes, it is not easy to accurately determine which passes in the first third of the field actually significantly increased the real probability of a subsequent goal. Still, it is reasonable to assume that there is a higher concentration of such passes among those that actually resulted in a goal.

This justifies why Group 1-G, which contains initial passes that resulted in a goal in the near future, will be the focus of our analysis. It is not possible to affirm that all the passes in this group actually made a significant contribution to the subsequent goal, but certainly, some of them were decisive.

To understand if the VAEP valuation based on event data can identify them, we will divide this section into three subsections. The first subsection will compare the offensive valuation differences of passes within the same third of the field. In the subsequent subsection, we will compare groups 1-G and 3-NG. Lastly, the final subsection will showcase specific pass examples and highlight notable cases.

3.1 Comparing Same Third Groups

We conducted the mentioned comparisons by analyzing the groups' Cumulative Distribution Functions (CDFs) of the offensive value distribution, as shown in Fig. 1. Additionally, we performed two Kolmogorov-Smirnov tests. The KS Statistic for groups 1-G and 1-NG is 0.10582 while the same calculation between 3-G and 3-NG is 0.23937.

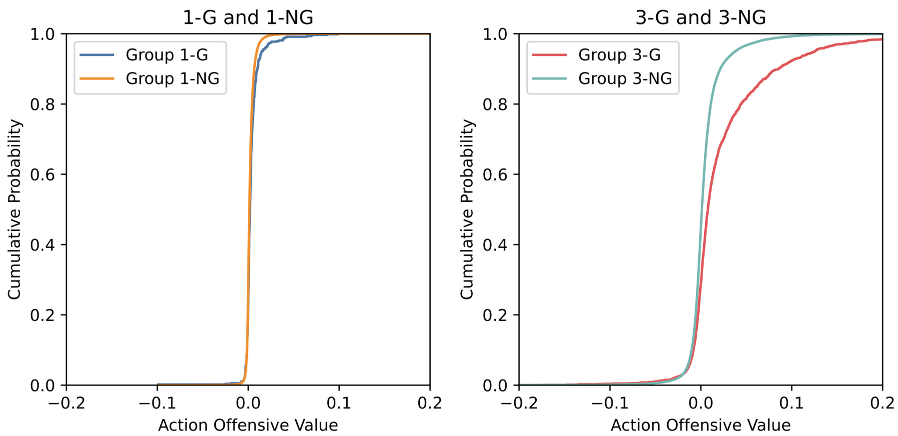

Fig. 1. Comparison of Cumulative Distribution Functions (CDFs) for groups 1-G and 1-NG (left), 3-G and 3-NG (right). The larger discrepancy between 3-G and 3-NG CDFs compared to 1-G and 1-NG CDFs indicates that the metric better differentiates passes that result in goals in the final third than in the first third.

These results show that there is a significant difference between the two comparisons. The performed test and the visualizations indicate that, although both comparisons point out differences between the groups, the discrepancy of values between 3-G and 3-NG is significantly larger than between 1-G and 1-NG. This indicates that the metric can better differentiate passes that result in goals in the final third than in the first third.

This result can be interpreted in different ways. One possible interpretation is that the factors that led to the subsequent goals in group 1-G occurred after the passes themselves. Thus, the variation in the label between groups 1-G and 1-NG is more associated with chance than with the differences between the passes in each group, which results in a similar valuation by VAEP in both groups. On the other hand, it is in the final third where the goal condition is really generated. Therefore, the metric appropriately differentiates good passes in the final third, which are abundant in group 3-G and receive better overall ratings than group 3-NG.

However, this interpretation does not align with what we can observe in contemporary football. In recent years, defenders have been increasingly involved in the offensive phase of the game, starting from their own field, breaking lines, finding long balls, and structuring counter-attacks. This is a cornerstone of various prominent tactical approaches today. Therefore, we believe in an alternative interpretation of the results.

The advantage generated by good passes in the early stages of the field is predominantly positional. Whether it is a pressure release, a first-line break, or a good long pass, the gain for the executing team lies in the relation between players' positions and the ball rather than in the attributes stored in event data. Conversely, in the final third of the field, the features involved in the model can more easily capture dangerous passes. Although positional relations are still relevant, there is also substantial importance on elements such as proximity to the goal and the speed of the sequence of plays. Because of this, VAEP can make a better distinction in the final third.

This second perspective on the results seems more plausible, but analyzing just one season is not sufficient to claim that the metric indeed underestimates good passes in the first third of the field. There is a significant difference in the number of passes in each group, which makes it challenging to employ certain approaches in this comparison. Nonetheless, even though they are not conclusive, the presented results serve as evidence to believe that underestimation does occur.

3.2 Comparing 1-G and 3-NG

Another comparison we conducted was between the CDFs of the offensive value distribution of passes from groups 1-G and 3-NG. The graph with both functions is displayed in Fig. 2. Clearly, the two groups contain actions that occurred in very different contexts. Still, this comparison provides us with some interesting insights. Examining the curves, we can observe a higher frequency of low values, particularly negative values, in group 3-NG. This aligns with expectations since,

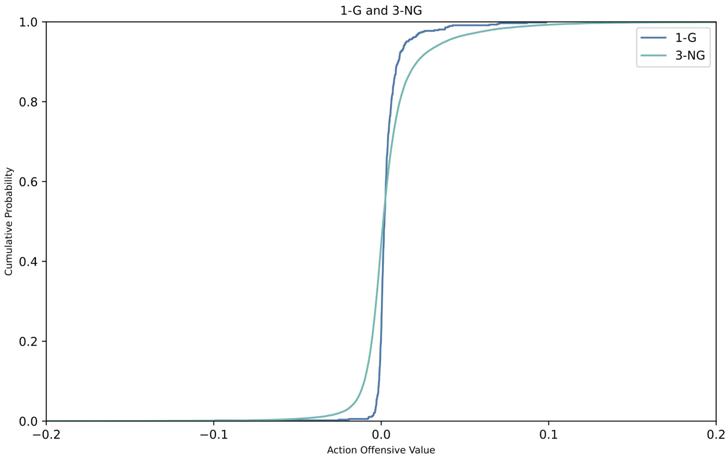

Fig. 2. Comparing CDFs of groups 1-G and 3-NG. Group 3-NG shows more frequent high values than group 1-G. This observation is intriguing since all passes in group 1-G resulted in goals, while none of the passes in group 3-NG did.

in the final third of the field, the probabilities of scoring are higher and, therefore, can be decreased by some passes. Though, when we look at the other end of the graph, we can see that the frequency of higher values is also greater in group 3-NG. This observation is intriguing since all passes in group 1-G resulted in goals, while none of the passes in group 3-NG did.

Once again, this analysis cannot be taken as conclusive. In addition to the inherent contextual difference between these two groups, there is a significant disparity in the number of passes in each. Even so, it is a surprising fact that corroborates the proposed investigation.

3.3 Notable Cases

To analyze some examples more closely, we selected the top 10 passes from group 1-G with the highest offensive values assigned by VAEP. These passes are represented in Fig. 3, where the bottom part of each plot represents the defending goal, and the top part represents the attacking goal.

The most prominent characteristic in this set is the depth of the passes. Except for the first one, all of them are long balls that cover a significant portion of the field. This observation aligns with expectations because ball advance is a feature that event data can represent and is intuitively correlated with the occurrence of a subsequent goal. At first glance, it is plausible that these passes were influential in creating the resulting goal.

Still, when we look at group 3-NG, we observe that 434 passes are better valuated than the one occupying the first position in group 1-G. In Fig. 4, we can see a random sample of 10 of these 434 passes.

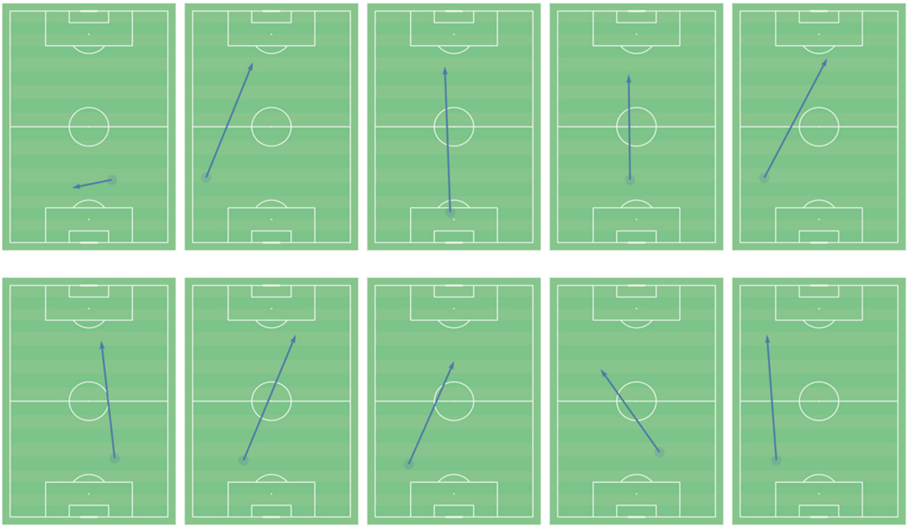

Fig. 3. Top 10 valued passes in group 1-G. There is a prominent characteristic of depth, with all but the first pass being long balls covering a substantial portion of the field.

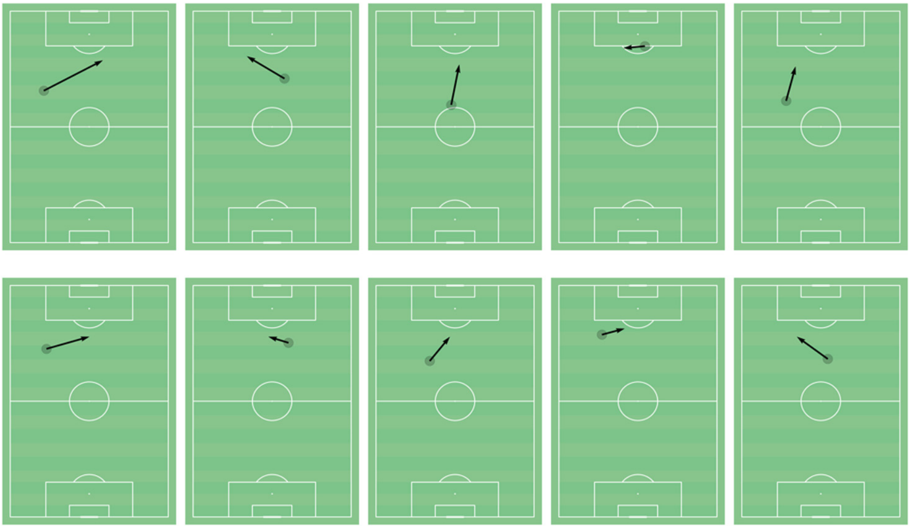

Fig. 4. Random sample of the top 434 valued passes in group 3-NG. In general the passes are short and there is a slight trend of progressiveness.

We can observe a notable trend in these passes: they are generally progressive, aiming for areas of the field closer to the opponent's goal, despite being shorter than the selected passes from group 1-G. While none resulted in a goal, they give the initial impression of increasing the likelihood of scoring. However, it is astonishing to find such a significant absolute number of passes that supposedly enhanced the goal probability more than any pass from the first third that actually resulted in a goal.

4 Conclusion and Future Work

After presenting the results, this section concludes the work by presenting conclusions and directions for future work.

4.1 Conclusion

Event data has considerable advantages, such as more accessible collection, greater availability, and lower memory demand. Despite its inherent limitations, efficient models based on this structure have already been shown to be feasible and are of great scientific and practical importance. This reinforces the need to explore characteristics, identify flaws, and try to improve upon them.

Throughout the scope of this work, we conducted an analysis aiming to understand if VAEP, with its original features, can distinguish good passes in the initial third of the field. Although the scope and depth of the study do not allow for a categorical conclusion, some relevant evidence has emerged.

Initially, it was possible to observe a significant discrepancy in the offensive valuation of passes in the final third of the field when comparing actions with positive and negative labels. The same was not observed in the initial third, where the value distributions are more similar. Additionally, we observed that, in general, there are more well-valuated passes among the negative-labeled passes in the final third (group 3-NG) than among the positive-labeled passes in the initial third (group 1-G). Finally, we presented some passes from these two groups on the field and found that the best-valuated pass by the metric in group 1-G has a lower score than 434 passes from group 3-NG.

These three stages of analysis support the hypothesis that the event-based VAEP is limited in distinguishing good passes in the initial third of the field. As mentioned before, this limitation may be due to the spatial context limitation of the data format, which has a greater impact on initial passes than on final passes. Thus, this study encourages further investigation in this direction.

4.2 Future Work

The evidence raised points to directions for future work. One possibility is to repeat the analysis on different datasets to ensure that what was observed in this Premier League season is not an exception. Additionally, making slight variations in the VAEP model, such as changing the outcome window of an action to determine labels, could be interesting. This would increase the likelihood that the grouped passes are more directly related to goal occurrence.

A longer-term and ongoing approach as a continuation of this work involves developing a second framework for VAEP that includes features extracted from tracking data. By re-valuating the actions using this second form of the metric, it would be possible to identify which passes had a significant change in value. This approach could guide a search for patterns within the original features to enhance and balance the functioning of the metric using event data.

References

1. 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, pp. 1851–1861. Association for Computing Machinery, New York (2019). https://doi.org/10.1145/3292500.3330758
2. Van Roy, M., Robberechts, P., Decroos, T., Davis, J.: Valuing on-the-ball actions in soccer: a critical comparison of XT and VAEP. In: Proceedings of the AAAI-20 Workshop on Artificial Intelligence in Team Sports. AI in Team Sports Organizing Committee (2020)
3. Robberechts, P., Van Roy, M., Davis, J.: un-xPass: measuring soccer player's creativity. In: StatsBomb Conference (2022). https://statsbomb.com/wp-content/uploads/2022/09/Pieter
4. Singh, K.: Introducing Expected Threat (xT) (2023). https://karun.in/blog/expected-threat.html
5. Crow, B.: How the creation of different superiorities generates opportunities to progress the ball (2023). https://soccerdetail.com/2020/10/26/how-the-creation-of-different-superiorities-generates-opportunities-to-progress-the-ball/
6. Hodson, T.: Positional play: football tactics explained (2023). https://www.coachesvoice.com/cv/positional-play-football-tactics-explained-guardiola-cruyff-manchester-city/
7. Fernández, J., Bornn, L., Cervone, D.: Decomposing the immeasurable sport: A deep learning expected possession value framework for soccer. In: 13th MIT Sloan Sports Analytics Conference (2019)
8. Decroos, T., Bransen, L., Van Haaren, J., Davis, J.: VAEP: an objective approach to valuing on-the-ball actions in soccer. In: IJCAI, pp. 4696–4700 (2020)
9. Davis, J., et al.: Evaluating sports analytics models: challenges, approaches, and lessons learned. In: AI Evaluation Beyond Metrics Workshop at IJCAI 2022, vol. 3169, pp. 1–11. CEUR Workshop Proceedings (2022)
10. Pappalardo, L., Cintia, P., Ferragina, P., Massucco, E., Pedreschi, D., Giannotti, F.: PlayeRank: data-driven performance evaluation and player ranking in soccer via a machine learning approach. ACM Trans. Intell. Syst. Technol. (TIST) **10**(5), 1–27 (2019). https://doi.org/10.1007/1234567890
11. Fernández, J., Bornn, L., Cervone, D.: A framework for the fine-grained evaluation of the instantaneous expected value of soccer possessions. Mach. Learn. **110**(6), 1389–1427 (2021). https://doi.org/10.1007/1234567890
12. Nascimento, R.F.M., Rios-Neto, H.: Generalized action-based ball recovery model using 360 data. In: StatsBomb Conference (2022). https://statsbomb.com/wp-content/uploads/2022/09/Ricardo. Accessed 6 June 2023
13. Decroos, T., Davis, J.: Interpretable prediction of goals in soccer. In: Proceedings of the AAAI-20 Workshop on Artificial Intelligence in Team Sports (2019)