Effective models for winning football games: a comparison between xG, xT and VAEP

Milan Klaasman

milanklaasman@gmail.com
Amsterdam, Netherlands
https://mklaasman.github.io/

Abstract. Football analytics has been developing at a rapid pace over the previous years. There has been an increase in the development of different models that aim to model parts of the game. In this paper, we compare three of those models: Expected Goals (xG), Expected Threat (xT) and Valuing Actions by Estimating Probabilities (VAEP) as effective models for winning football games. The models are compared in terms of their interaction with three variables, namely: the difference in the number of goals per match, the number of goals scored per match and the winning team. To compare the models we used Linear Regression and Pearson correlation as evaluation metrics. This study shows that VAEP is the most effective model for winning football games among the three models. Scoring the highest in all approaches and on all metrics.

Keywords: Football Analytics · Data Science · Machine Learninig · Linear Regression.

1 Introduction

Football analytics has been developing at a rapid pace over the previous years. There are multiple factors enabling this improvement, such as an increase in technological development in data storage and data science, increase in freely available data and a steep increase in enthusiasts who contribute to the football analytics space. This paper is one of those. The goal of football analytics should comprise the investigation of winning matches or at least improve the probability of doing so. The means of winning a match is scoring more goals than the opponent team. Effective and useful models in football are supposed to explain an element of the game which makes one team to win more likely (i.e. score more goals) than the other team. Otherwise, it can be coined that the model "only" e.g. describes the style of play of a team or player. In recent years, multiple different models have aimed to succeed. For example, expected Goals and expected Assists (xA) which both are predictive models. The former used to assess every goal-scoring chance, with respect to its likelihood of scoring and the latter used to assess assists. Moreover, there are also possession-based models, such as Valuing Actions by Estimating Probabilities (VAEP) [1], expected Threat (xT) [3], OptaPro's Possession Value (PV) [2] and Goals Added (g+) [4]. These models aim to assign values to a given state of possession. Additionally, the models use the differences between consecutive states to award a corresponding value to the actions that took place between the states.

In a publication, Van Roy et al. [5] stated that "both the traditional metrics (i.e., shots and assists) and their context-dependent successors (i.e., xG and xA) fall short in addressing the impact of the individual actions as they focus on rare actions like shots and goals alone."

The question arises whether possession-based values still manage to explain the notion of winning matches. Van Roy et al. claim that xG and xA fall short in addressing the impact of individual actions, however, we do not know whether they are they still good predictors of winning matches.

Therefore, we will investigate three different models: xG, xT and VAEP. We chose these models since they are publicly available and can be reproduced by others. For sake of reproducibility, we will be using the 2018/2019 and 2019/2020 FA Women's Soccer League (FAWSL) freely available Statsbomb events dataset in this approach [6]. The xG metric is provided within the Statsbomb data, and xT and VAEP are produced using KU Leuven's socceraction implementations [7]. The main research question reads as follows:

RQ: Which model is most effective in predicting winning football games?

In the following section, we will explain the approach taken to compare xG, xT and VAEP.

2 Approach

2.1 Models

xThreat

Firstly, we derived the xThreat (xT) model for each team separately, this to ensure that actions that tend to be effective in a team's style of play will be awarded accordingly. One noteworthy setback of this method is that the amount of data used is only 1.5 seasons worth and thus may consist of significant deviations. For the calculation, we used KULeuven's implementation of xT [7]. This ended up with 13 xT models, one for each team. Two heatmap representations of these models are shown below in Fig. 1.

Using the xT map of each team, it is possible to calculate the xT of actions using the difference in xT value between the starting location (x_{start}, y_{start}) and the end location (x_{end}, y_{end}). Which can be calculated as follows:

$$xT_{action} = xT(x_{start}, y_{end}) - xT(x_{end}, y_{end}) \tag{1}$$

Using the xT implementation only ball-moving actions, such as passes, crosses and dribbles can be evaluated.

VAEP

Valuing Actions by Estimating Probabilities (VAEP) is a framework for valuing player actions introduced by Decroos et al. [1]. This approach firstly calculates the scoring probability (P_{scores}) and conceding probability ($P_{concedes}$). Given multiple game states $Si = [a1, ..., ai]$, the change in probability for team x scoring can be calculated as follows:

$$\Delta P_{scores}(a_i, x) = P_{scores}(S_i, x) - P_{scores}(S_{i-1}, x) \tag{2}$$

The same holds for the change in probability of conceding a goal. Which in combination leads to the calculation of VAEP, where the total VAEP value of an action is the sum of that action's offensive value and defensive value[1]:

$$V(a_i, x) = \Delta P_{scores}(a_i, x) + (-\Delta P_{concedes}(a_i, x)) \tag{3}$$

2.2 Implementation

For both possession value metrics, xT and VAEP, we calculate the valuation of each action in each match (n=193) in the dataset. In terms of xT, this will be only ball-moving actions. Subsequent, we can derive the cumulative possession value changes by summing all positive and negative actions values per match. This will result in xG values for all shots; xT values for all passes, crosses and dribbles; and VAEP values for player actions.

To compare the three models, we introduce two approaches. Firstly, we compare the models on the predictiveness on *goal_difference* for each match to create a continuous valuation of winning. Secondly, we compare the models on their ability to predict the number of *goal_scored* in a match. For both approaches, a linear regression model is used to evaluate the relationship between higher values of the model and the two output values. Linear regression is a linear approach to modelling the relationship between a dependent variable and one or more independent variables. R-squared (R^2) will be used to validate the linear

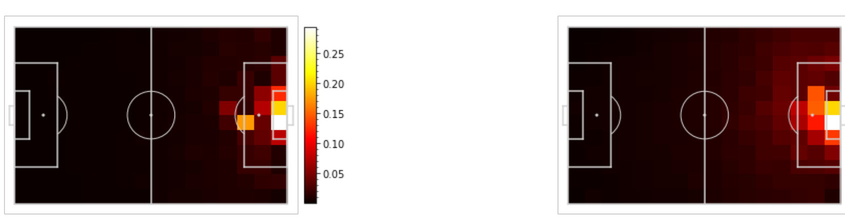

Fig. 1. This figure shows two Expected Threat heatmaps of Liverpool WFC (left) and Chelsea WFC (right). The xThreat valuation is represented by the difference in colour as can be seen by the colour bar accompanied on the right side.

regression model. R^2 is the proportion of the variance in the dependent variable that is predictable from the independent variable(s):

$$R^2 = \frac{\textit{Variance explained by the model}}{\textit{Total variance}} \tag{4}$$

Usually, the larger the R^2, the better the regression model fits your observations. Meaning, that are is a higher likelihood of an existence of a linear relation between the two variables. So our evaluation will be based on the value of R^2 for both scenarios explained above. Additionally, we investigate the correlation between the models in their valuation and their correlation with both *goal_difference* and *goal_scored*. For this Pearson correlation is used. In the following section, we will elaborate on the results found.

3 Results

In this section, we will elaborate on the results of the two approaches taken, namely comparing the models on *goal_difference* and *goal_scored*, by calculating the linear regression's R^2 and the correlations.

3.1 Goal difference

Firstly the *goal_difference* is shown in Fig. 2. The models are plotted against the difference in goals scored per match between the home team and the away team.

From this figure we can draw a few conclusions:

- Firstly, these graphs show a remarkably similar result, which leads to suggest that both xG, xT and VAEP are at least somewhat comparable in their correlation with *goal_difference*.
- Secondly, for all models, a higher value for the model leads to a higher *goal_difference* in favour of that team.
- Thirdly, VAEP clearly shows to be better fitted by the linear regression line.
- Fourthly, R^2 values differ between the models. Where VAEP scores highest, followed by xG and xT, with R^2 scores of 0.96, 0.58 and 0.37 respectively.

Fig. 3 shows a correlation heatmap. This heatmap shows us that VAEP has the highest correlation with *goal_difference* between teams (0.99) and the winning team (0.78), followed by xG having a correlation of 0.76 with *goal_difference* and 0.66 with *winning_team*. Lastly, xT scores near xG on *goal_difference* having a correlation of 0.75 and 0.6 with *winning_team*.

As can be seen, the winning team value is not one to one replaceable by *goal_difference*s between teams since the correlation between these two values is 0.8. This is noteworthy and should be taken into account when looking at the graphs. Finally, there is a high correlation found between the different models as well.

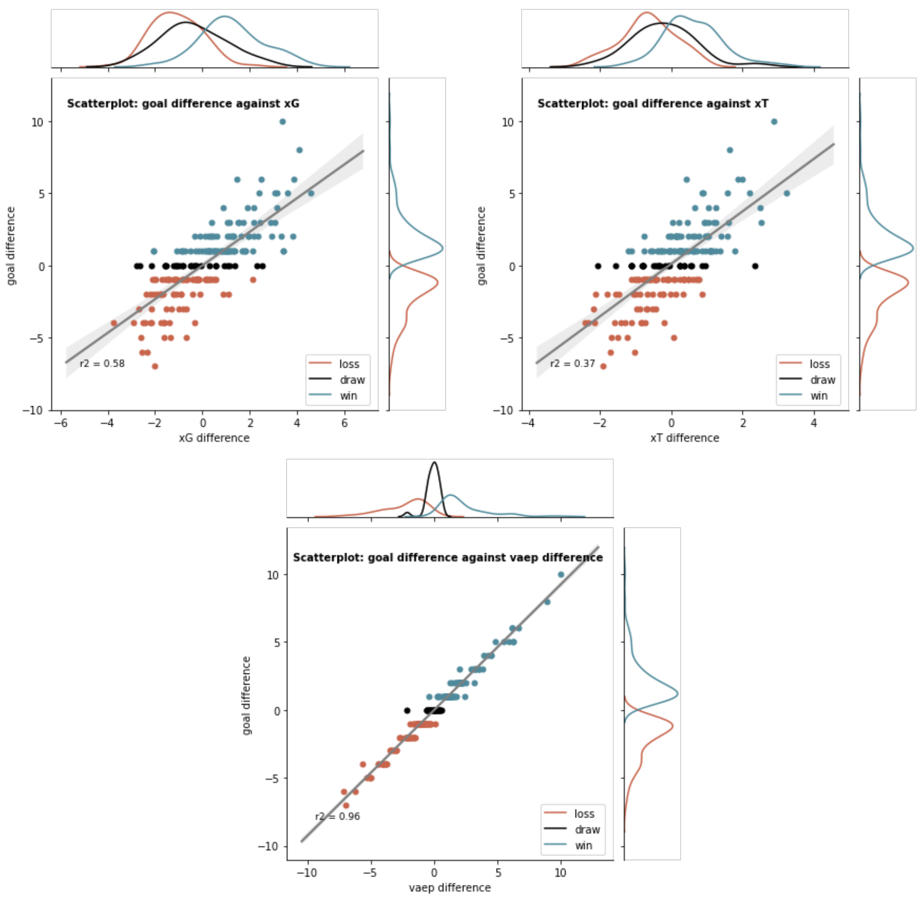

Fig. 2. These scatter plots of the *goal_difference* against the cumulative xG, xT and VAEP, from top left to right to bottom respectively, for all matches in the dataset. Graphs are represented including distribution and regression plots. With the blue dots representing the home team winning the game, black representing the teams drawing and red the away team winning the game.

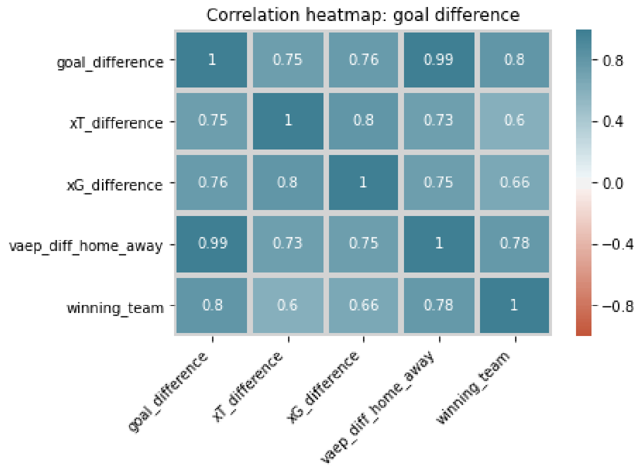

Fig. 3. Correlation heatmap of the variables: *goal_difference*, *xT_difference*, *xG_difference*, *VAEP_difference* and *winning_team*.

3.2 Goals scored

To further investigate the hypothesis, we will plot the models against the goals scored of a team. Thus, instead of taking the difference between the goals scored, we look at each team's performance, while neglecting the opponent's performance and the team's own defensive performance. Therefore this approach gives us two times as much data points (n=386), for each team one per game. Using this approach we can investigate whether higher cumulative scores given by the models for the actions will lead to scoring more goals. Fig. 4 shows the results of this approach.

Since we only take goals scored and cumulative values of xG, xT and VAEP into account, these plots show only positive values. Yet again we see a trend of higher values corresponding with more goals scored. Both xG and xT R^2 scores are relatively low with R^2 scores of 0.56 and 0.45 respectively. Meaning that there is still a lot of unexplained variance in goals scored. Therefore, making the models less robust for predicting a winning team. However, VAEP shows a R^2 score of 0.84, from which we can conclude that also in terms of goals scored, VAEP shows the highest predictive characteristics for winning football games. To get a better understanding of how the values relate to each other, Fig. 5 shows a correlation heatmap of *goals_scored*, cumulative xT, cumulative xG, and cumulative VAEP scores calculated for each match. Once again, VAEP shows the highest correlation (0.89) with the target value, i.e. *goals_scored*. Followed by xG (0.73) and xT (0.68). Noteworthy is that xT does not seem to be significantly higher correlated to VAEP than xG, whereas both xT and VAEP are Possession Value models that aim to value actions in football matches.

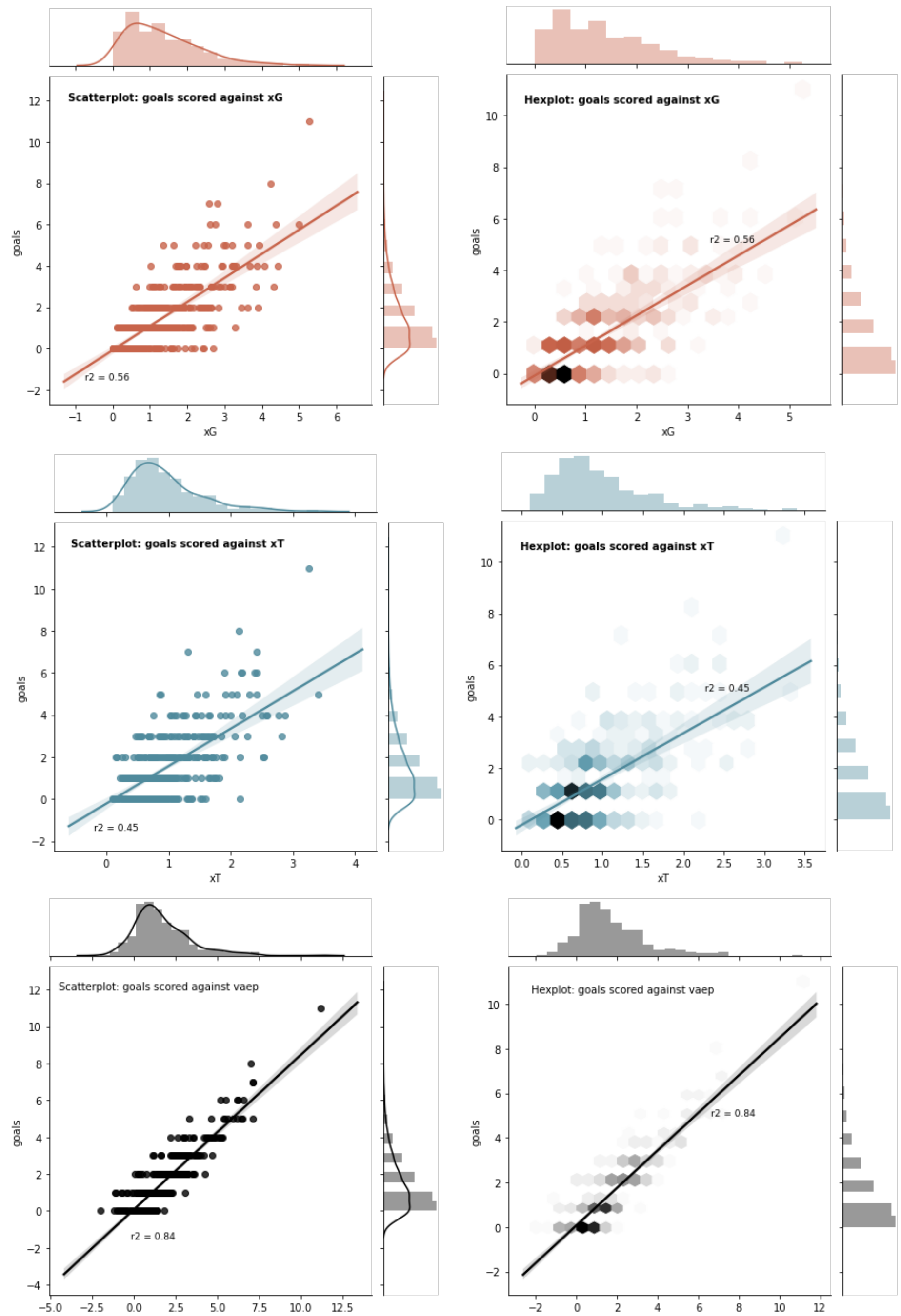

Fig. 4. A collection of six plots representing the *goal_scored* against xG (in red), xT (in blue) and VAEP (in black). The same data is plotted using a scatter plot (left) and a hex plot (right), since scatter plots tend to clutter.

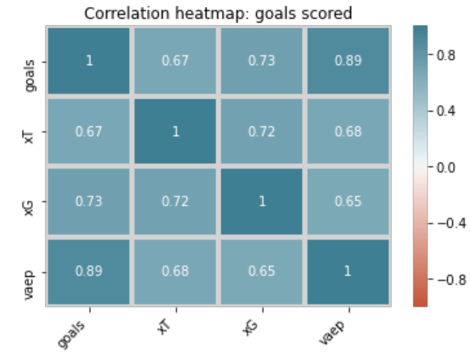

Fig. 5. Correlation heatmap of the variables: *goal_scored*, *xT_score*, *xG_score* *VAEP_score* and *winning_team*

Finally, Table 1 summarizes the R^2 values of the two approaches and their correlations with the models. From this table, we can clearly conclude that VAEP scores highest on both valuation metrics, R^2 and Pearson correlation, for both *goal_difference* and *goal_scored*. Subsequently, xG scores higher than xT on all elements.

	R^2		Correlation	
Model	*goal_difference*	*goal_scored*	*goal_difference*	*goal_scored*
Expected Goals (xG)	0.58	0.56	0.76	0.73
Expected Threat (xT)	0.37	0.45	0.75	0.67
Valuing Actions by Estimating Probabilities (VAEP)	0.96	0.84	0.99	0.89

Table 1. This table shows the R^2 values of the linear regressions of the three models and correlations with *goal_difference* and *goal_scored*

4 Conclusion

In this paper, we compare Expected Goals (xG), Expected Threat (xT) and Valuing Actions by Estimating Probabilities (VAEP) as effective models for winning football games. To compare the models we used linear regression and Pearson correlation as evaluation metrics. We found that all models show correlation with *goal_difference*, *goal_scored* and the winning team. Also, the outcomes of the correlations add strength that all three models can be used as models for winning games. However, VAEP scores highest in all approaches and on all metrics. Making it the most effective model for winning football games among the three models. This outcome points out that different design choices will lead to different usability factors of the model. Where VAEP, based on these findings, shows better capabilities in predicting winning teams. In a blog post comparing VAEP and xT, Van Roy et al. mention that VAEP is better in capturing the risk-reward trade-off of actions and xT being more interpretable [5]. For this reason, it is important to further investigate the possibilities and limitations of models used in the field of football. In a world where all models are wrong, but some models are useful.

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 1851-1861 (2019).
2. Mackay, N.: Introducing a possession value framework (2019) https://www.optasportspro.com/news-analysis/blog-introducing-a-possession-value-framework/. Last accessed 15 June 2019
3. Singh, K. Introducing Expected Threat (xT): Modelling team behaviour in possession to gain a deeper understanding of buildup play. https://karun.in/blog/expected-threat.html. Last accessed 15 June 2019
4. Muller, J. Goals added: Introducing a new way to measure soccer. (2020) https://www.americansocceranalysis.com/home/2020/4/22/37ucr0d5urxxtryn2cfhzormdziphq. Last accessed 15 June 2019
5. Van Roy, M., Robberechts, P., Decroos, T., Davis, J. Valuing On-the-Ball Actions in Soccer: A Critical Comparison of xT and VAEP. (2019)
6. StatsBomb open data Github Repository https://github.com/statsbomb/open-data. Last accessed 15 June 2019
7. KU Leuven socceraction Github Repository https://github.com/ML-KULeuven/socceraction. Last accessed 15 June 2019