xGad: A Dynamic Rating System for Probabilistic Forecasting in the English Premier League (EPL).

*

*

*

Abstract

This paper proposes a new rating system named xGad, inspired by the Elo rating system. xGad aims to quantify EPL team's relative attacking and defensive strengths, in independent home and away fixtures. Utilising eight seasons of EPL data, from 2014 to 2022, a historic dataset was created following the xGad methodology. This resulted in 3040 matches, with two pre-game xGad ratings for each home and away team. The ratings' relationship with home and away match goals, in addition to total match goals was examined using binary logistic regression. This analysis uncovered an ability for the ratings to explain separate home and away match goals but limited in total match goals. Given this, analysis continued with home and away goals to evaluate the rating's predictive ability. Model 3 was found to be 66.71% accurate in predicting whether a home team would have two or more goals. Secondly, model 4 was 70.82% accurate in predicting the outcome of two or more away team goals. This analysis indicates that given improvements to the accuracy of the model, the xGad system could be used to form a profitable betting strategy within separate home and away team goal markets.

Table of Contents

1. Introduction

Football is the most popular sport in the world, boasting approximately 3.5 billion fans and 250 million players worldwide (WorldPopulationReview, n.d; Fifa, 2007; da Costa et al., 2022). The English Premier League (EPL) is widely considered the highest performing and most competitive professional football league (Ackerson, 2022). The EPL is comprised of 20 teams, each playing 38 matches in a season, evenly spread between home and away venues. There are three outcomes of a match in the EPL, these include a win (3 points), draw (1 point) or loss (zero points). Following the final matchday of the season, the bottom three teams in the league table are relegated to the English Football League (EFL) Championship, the second division in English football. The relegated clubs are replaced by the top two teams from the EFL Championship, with a final team gaining promotion through a playoff system conducted at the conclusion of the EFL Championship regular season.

The EPL draws the highest number of viewers amongst the top European leagues (Walker, 2022). The 2019/20 EPL season drew 3.2 billion viewers worldwide (Walker, 2022). The international audience of the EPL has been a catalyst in making the league the richest domestic competition in the world. By way of comparison, a report produced by UEFA in 2019 revealed that the EPL had over double the revenue per club compared to the top division of Spanish football, the La Liga (UEFA, 2019). One of the business sectors that has capitalised off this large audience base is the sports betting industry. In the 2021/22 EPL season, 19 out the of 20 clubs were sponsored by a betting company (Punters Page, 2021). With an estimated 2.7 million pounds wagered on each match, it is evident that the betting industry has a strong foothold in the EPL (Deutscher et al., 2018).

The prevalence of betting on the EPL, and other top European football leagues has driven an eagerness for statistical models that can accurately forecast outcomes in matches. Interest in this field has grown in recent years, stemming from the greater access to data, the development of computational methods and the increase in the variety of betting markets (The Economist, 2018: quoted in Wheatcroft, 2020). Despite the seemingly modern nature of these methods, the power of statistics in forecasting football outcomes has been examined since at least the 1950's through the work of Moroney (1951). The early papers of this time have provided crucial criterion which are still used in modern statistical sport models (Maher, 1982; Dixon & Coles, 1997).

In this early literature, the use of rating systems to discern the strength of a team in football was introduced (Maher, 1982; Dixon & Coles, 1997). However, these early systems produced by Maher (1982) and Dixon and Coles (1997) were limited by their static nature, being unable to account for fluctuations in form throughout a season. In recent years, dynamic rating systems in football have been developed that seek to address this inflexibility in previous systems, including the Generalised Attacking Performance ratings (Wheatcroft, 2020) and the Elo rating system (Lacy, 2019). These dynamic systems are proposed to provide more robust measurements of team strength, which subsequently creates more accurate predictions (Wheatcroft, 2020; Lacy, 2019).

The use of a rating system in a sporting context requires the researcher to consider the most accurate measure of a team's performance. A commonly occurring conclusion is that goals scored is the most precise measure of performance (Hughes & Barlett, 2002; Szwarc, 2007). However, this assumption has come under scrutiny on the basis that there is a significant element of chance in the scoring of each goal in any football match (Reep et al., 1971). There are many cases in which teams who create more chances and dominate their opposition do not win the game in question. Due to the impact of chance, this study does not consider team goals to be a definitive marker of performance.

In 2012, Opta introduced a statistic called expected goals (xG), which aimed to provide a more accurate measurement of performance rather than relying on raw goals. Opta states that xG:

"measures the quality of a shot based on several variables such as assist type, shot angle and distance from goal, whether it was a headed shot and whether it was defined as a big chance… adding up a player or team's expected goals can give us an indication of how many goals a player or team should have scored on average, given the shots they have taken"

(Opta, n.d.).

Since the introduction of this statistic, it has been received as a viable method to measure performance (Rathke, 2016). Moreover, many companies and researchers have constructed their own xG models that contain different factors and variables which may improve the accuracy of xG (Understat, 2022; StatsBomb, 2022).

The purpose of this paper is to propose a new dynamic rating system, xG attack and defence ratings (xGad). This rating system is explained through four separate ratings that quantify a team's relative attacking and defensive strength: $xGad_{HomeAtt}$, $xGad_{AwayDef}$, $xGad_{AwayAtt}$, and $xGad_{HomeDef}$. Each team has a home and away rating, which are calculated independently from one another. In this paper, a team's attacking, and defensive strength is described by their ability to score and not concede goals in a match, while controlling for their performance through xG.

The xGad system was developed using data from eight seasons of EPL football, between 2014 and 2022. The system uses iterative methods to assess each EPL game in chronological order so that following each EPL fixture, each team's xGad rating was updated based on the number of goals scored or conceded, expected goals for or against, and the strength of the opponent. This process resulted in a database of 3040 fixtures, which are individually supplemented by two pre-game xGad ratings for each team.

The xGad system was inspired by Lacy's (2019) 'Predictaball' model, which is an Elo rating system. xGad shares many similarities with the Elo rating system framework. However, the Elo system infers its ratings from a match outcome only (i.e., team wins, draws and losses). xGad deviates from this as the ratings are inferred from the number of goals scored or conceded while controlling for performance, within each respective match. Moreover, xGad splits its ratings into attacking and defensive disciplines, unlike traditional Elo systems which study teams or players in general. Further discussion regarding the Elo rating system is continued during the *Literature Review* section below.

Following the *Literature Review* section, this paper is split into three further sections:

1. *Methodology* – in which the composition of the xGad ratings is considered.
2. *Results* – in which a review is provided showing the predictive accuracy of the xGad rating system,
3. *Diagnostics* – in which the author considered how the xGad system may be improved or utilised in future.

As explored in this paper, the xGad ratings aim to provide robust statistical information that can be used to produce probabilistic forecasts for separate home and away match goals, in addition to total match goals. Once produced, the ratings were inputted into three separate binary logistic regression models: the first two measured xGad's relationship with home and away team goals, and the final model examined xGad's association with total match goals. These models were then used to analyse xGad's predictive accuracy in the outcome variables. Following the analysis, this paper conducted a diagnostic review of the xGad system, to identify flaws and propose improvements, aiming to increase its accuracy and real-world application.

2. Literature Review

Since the 1950s, research has been conducted to examine the link between statistics and football. The dominating theme in early research focused on the distribution of total match goals and the resulting score-lines. Moroney (1951) concluded that a Poisson distribution may be acceptable to model scores, however, a negative binomial distribution could be a more efficient alternative. Reep et al. (1971) used negative binomial distribution to fit football score-lines against other score-based games. It was determined that the game of football was dominated by the element of chance, making statistical predictions unreliable due to the inherent noise within the data (Reep et al., 1971). The difficulties in creating statistical models in football was summarised by Dixon and Coles (1997). It was stated that, "in the long run, it is not difficult to predict fairly accurately which teams are likely to be successful, but the development of models that have a sufficiently high resolution to exploit this long run predictive capability for individual matches is substantially more difficult" (Dixon and Coles, 1997).

Since the aforementioned conclusion of Reep et al. (1971), there have been considerable advancements in this area. Maher (1982) altered the narrative that football was plagued by chance, stating "that whilst in a single match, chance plays a considerable role (missed scoring opportunities, dubious offside decisions and shots hitting the crossbar can obviously drastically affect the result), over several matches luck plays much less of a part" (Maher, 1982: 109). Furthermore, Maher's (1982) ideology was centred around the assumption that teams are not identical, that they vary in their attacking and defensive capability. Therefore, stronger teams are more likely to beat their weaker counterparts (Maher, 1982).

Maher (1982) continued to explore the varying ability of teams, in both attacking and defensive settings. His beliefs were justified as there is clear evidence that within a division such as the

EPL, there are teams that are stronger than others. This is shown by teams who perform to a high level over multiple seasons. A key example of this is Manchester United, who between the seasons 1992/93 to 2000/01, won the Premier League title seven times (Premier League, n.d.). Maher (1982) aimed to quantify team strengths to highlight disparities between clubs in the top English division. He calculated maximum likelihood estimates for his model, in which "the scores of the home and away teams in any game are independent Poisson distributions, with means modelled as functions of the respective teams' previous performances" (Maher, 1982; quoted in Dixon & Coles, 1997). The work of Maher (1982) inspired further research into this topic thus prompting a wave of new methods which can be used to quantify team strength.

Following Maher's (1982) research, one of the most significant papers released was by Dixon and Coles (1997). They constructed a parametric model utilising Poisson regression to fit English league and cup scores between 1992 and 1995 (Dixon & Coles, 1997). Moreover, they used this model to construct a betting strategy with an aim to exploit the inefficiencies in the betting markets.

Dixon and Coles (1997) highlighted the difficulty in attempting to include all relevant information that may affect a football match. For example, they note that team performances can be affected by external factors such as newly signed players or managerial changes. The data on these factors are often available, but they are "less easily formalized and its qualitative value subjective" (Dixon & Coles, 1997: 267). To combat this, they highlight five impactful factors which must be accounted for within statistical football models.

a) "The model should take into account the different abilities of both teams in a match";
b) "There should be allowance for the fact that teams playing at home generally have some advantage - the so-called `home effect'";

c) “The most reasonable measure of a team's ability is likely to be based on a summary measure of their recent performance”;
d) “The nature of football is such that a team's ability is likely to be best summarized in separate measures of their ability to attack (to score goals) and their ability to defend (not to concede goals)”;
e) “In summarizing a team's performance by recent results, account should be taken of the ability of the teams that they have played against.”

(Dixon & Coles, 1997: 269)

The five factors (Dixon & Coles, 1997:269) mentioned previously are echoed in the work of Stuart Lacy (2019), who constructed a model which produces probabilities for matchups in Europe’s top leagues. His work is centred around the use of the Elo Rating System, “a method of quantifying the skill of competitors in a head-to-head competition” (Lacy, 2019). The Elo rating system was developed by Arpad Elo in the 1950s, a chess master who sought out a more efficient method than the Harkness rating system which was used in chess at the time (Hoekstra, 2021). His system was eventually adopted by the International Chess Federation (FIDE) in 1970 (Hoekstra, 2021). The Elo system is prevalent in modern sport, it is used within FIFA World rankings, collegiate American Football, and Major League Baseball.

Within a football Elo system, the winning team takes points from their opponent. The total points that are at stake in any given match is equal to the difference between the two teams current Elo rating. It is important to highlight that a higher-ranked team risks more points if they lose to a lower ranked opponent, and vice versa (Hoekstra, 2021). This rating system is inherently limited to the group that are within the same rating pool, as the Elo ratings are relative to the other team’s skill (Hoekstra, 2021). Therefore, the comparison of team Elo ratings between leagues are problematic. For example, it would not be recommended to use such a method to compare teams in European competitions such as the Champions League.

This is because each team’s ratings in this competition would be relative to their own domestic league.

Lacy (2019) adopted the Elo system to create a dynamic rating for teams within the EPL, La Liga, Bundesliga, and Serie A to predict match outcomes, he called his model ‘Predictaball’. The three outcomes he focused his analysis around were a win, draw or loss. It should be noted that the leagues were treated independently due to the aforementioned incomparability between leagues. In Lacy’s (2019) system, each team is assigned a home and away Elo rating, which are representative of their respective home and away strengths prior to a game. Each updated rating is assigned to the respective team’s next home or away fixture. Unfortunately, Lacy (2019) does not state the data quantity or source used within the study. However, it can be assumed that it is approximately 15 years of historical data, which is stated by Goddard (2005) as an appropriate number of seasons for improved forecasting in football.

Within the ‘Predictaball’ model, each team starts with a home and away Elo rating of 1500 (Lacy, 2019). Within this system, following a game, each of the components listed in table 1 are calculated to update a team’s rating. The following discussion will explore these components and justify their implementation.

Formula Components	Description
Elo_{Home}	Home Team Elo Rating – Strength of team at home
Elo_{Away}	Away Team Elo Rating – Strength of team away
O	Observed outcome – 0 (Loss), 0.5 (Draw), 1 (Win)
E	Expected Outcome - ~0 (Loss), ~0.5 (Draw), ~1 (Win)

G	Margin of Victory Multiplier – The margin of victory within the match (Measure of team performance)
K	Scaling Parameter (20)

Table 1 – System Components

The primary step to update the ratings is to consider the relationship between the observed outcome (*O*) and the expected outcome (*E*). *O* is a component which holds a value which corresponds with the match result. An *O* value of zero represents a loss, 0.5 is a draw, and a value of 1 is equal to a win. This a vital piece of information for Lacy's (2019) model as its aim is to analyse team's likelihood to win a game.

The use of an *O* value alone in this system would be problematic, as such a model must account for the ability of the opposition, as noted by Dixon & Coles (1997). Without this consideration, teams who perform consistently at a high standard would experience over inflated Elo ratings. To counter-act this process, Lacy (2019) included the *Expected Outcome (E)* term, which was a principal component within the original Elo framework. The primary input to calculate *E* is the difference in ratings (*dr*), which serves as a predictor for the outcome of the match (Mittal, 2020).

This can be seen in the formula below.

$$dr = Elo_{Home} - Elo_{Away} + HA.$$

Evaluating *dr* can also be considered as measuring the difference in strength and can therefore be used interchangeably. Through the above calculation, Lacy assumes that "a team's performance is distributed according to a logistic distribution, so the difference in performance (*dr*) also shares the same distribution with μ = 0" (Lacy, 2019).

As can be seen above, the team's Elo ratings are not the only component within the calculation of *dr*. Researchers use the *E* term to integrate the phenomenon of home advantage (*HA)* (Lacy, 2019; World Football Elo, 2022). This is an agreed factor within literature, which is known to affect match outcomes within football (Dixon & Coles, 1997; Pollard, 2008). This was not required in the original Elo framework, as the phenomenon of a home advantage is not noted as a factor in chess.

A recognised definition for home advantage is "a phenomenon in which the home playing team wins over 50% of its matches within a balanced home and away schedule" (Courneya & Carron, 1992). However, this definition is limited to sports which do not have the option for the respective teams to draw. Thus, an alternative definition is required for the outcome of a draw. Pollard (1986) provided this by defining *HA* as "the number of points won at home… expressed as a percentage of all points gained" (Pollard, 1986: 239). Pollard (2008) notes that the home support from the crowd, the length of the journey, and familiarity of surroundings can play a role in the strength of *HA*.

Pollard (1986) utilised a chi-squared test to find the home advantage in the First Division (Premier League) between 1888 and 1984. He found that home advantage varied between seasons and did not remain constant. Pollard (1986) found a maximum home advantage of 67.9% (1888-1900) and a minimum of 62.5% (1946-1960). These figures allowed to conclude that home advantage has remained relatively stable throughout the years (Pollard, 1986). However, in recent years, Peeters and van Ours (2020) observed a "secular decline in home advantage over the past 45 years" (Peeters & van Ours, 2020:123). Many may attribute this decline to the reduction in the power of the referee through the improvement of technologies (i.e the implementation of the VAR decision review system). However, Peeters and van Ours (2020) note that this decline is secular and is not related to these technological advancements.

Many Elo models account for *HA* by adding Elo points when calculating the difference in the ratings. However, disagreements arise when deciding how much of an advantage should be given. World Football Elo (2022) uses a value of 100, whereas FiveThirtyEight's NFL model uses a *HA* value of 65 (Lacy, 2019). Alternatively, Club Elo proposed the use of a dynamic *HA* term which is updated weekly. Lacy (2019) concluded that this degree of detail to the home advantage does not result in a significant improvement to the model. Therefore, Lacy (2019) stratified the home advantage by league and updated it each season.

Lacy (2019) continued to utilise the standard equations in the Elo framework. Following the calculation of *dr*, the value is inputted into the formula shown below to determine a value of *E*. As noted, within Lacy's (2019) model the *E* value corresponds with the observed outcome (*O*), in that a value of 0 is a loss, 0.5 is a draw, and 1 represents a win. This is used to account for upsets, i.e., a worse team beating a stronger team. This ensures that teams are rewarded heavily for beating opponents of a higher skill level opposed to the same or worse. Lacy (2019) states that the use of the constant 400 in the calculation E is suitable to create a desired effect.

$$E = \frac{1}{1 + 10^{\frac{-dr}{400}}}$$

$$O \in \{0, 0.5, 1\}$$

(Lacy, 2019)

The final two parameters to consider in Lacy's (2019) methodology are *K* and *G*. K is a scaling parameter used to dictate the level of impact recent results have on the Elo update. For sports with a small number of games, such as tournaments, it is appropriate to use a large scaling parameter (Lacy, 2019). Long competitions such as the EPL require a smaller parameter as it is not desired for the Elo to fluctuate too much throughout 38 game weeks.

Therefore, Lacy (2019) states 20 as an appropriate scaling parameter for a European league season.

G is an extension to the traditional Elo framework within Lacy's (2019) model. *G* is used to describe a game's margin of victory (MOV), which Lacy (2019) uses to assign a greater number of Elo points to a team who wins convincingly as opposed to a team who wins a close game. For example, a team who wins by a margin of three goals (e.g., 3-0) will be awarded more points than a team who won by one goal (e.g., 2-1). Although not noted by Lacy (2019), *G* can be seen to be a proxy variable to measure the level of a team's performance, a statistic which would otherwise be extremely difficult to quantify. An essential characteristic of *G* is its non-linear nature (Lacy, 2019). In football, the performance of a team cannot be described as linear, as fundamentally, goals are hold a greater significance when there are a small number of goals in the game. For example, the difference between 1-0 to 2-0 is greater than 5-0 to 6-0. This disparity can be explained by the losing team becoming increasingly demotivated, thus reducing their efforts. Therefore, a greater significance is required in the parameter G for smaller values compared to the larger values. Additionally, Lacy (2019) notes that it is important to recognise G as a discrete parameter, not continuous as a large MOV will almost never be met under normal conditions within a football match. For example, the greatest margin of victory recorded in the EPL occurred in 1995, when Manchester United beat Ipswich Town 9-0 (The Analyst, 2022).

The conditions and assumptions associated with the G parameter in Lacy's (2019) model have been noted in other systems, such as the World Football Elo system (2022). Their model calculates the G parameter through the conditions shown below. It should be noted to avoid confusion, the losing team would use the margin of victory of the winning team in their rating update.

$$G = 1 \; if \; MOV \in 0, 1$$

$$G = 1.5 \; if \; MOV = 2$$

$$G = 1.75 \; if \; MOV = 3$$

$$G = 1.75 + \frac{MOV - 3}{8} \; otherwise$$

(World Football Elo (2022), as cited in Lacy, 2019)

Unfortunately, World Football Elo's method has its shortcomings as FiveThirtyEight identified an issue with auto-correlation (FiveThirtyEight, as cited in Lacy, 2019). They found that strong teams have an inflated rating as they often win by a large margin. A method to combat this is to account for the difference in the team's ratings in the G calculations. Inspired by the methods of FiveThirtyEight in their NBA and NFL models which included a MOV parameter, Lacy (2019) produced an equation (below) which satisfied his requirements. A central condition for Lacy (2019) was the need to reward smaller MOVs without overinflating large MOVs while controlling for the difference in the team's qualities.

$$G = log_2(1.7MOV)\frac{2}{2 + 0.001dr}$$

(Lacy, 2019)

Evaluating a team's performance is of the utmost importance to a club, as it can give an indication of future successes. However, in this assessment, many professional leagues, such as the EPL, are plagued by a short term vision. The significant financial implications of success within a single season causes clubs to become obsessed with short-term outcomes (Brechot & Flepp, 2020). This has been evident in the EPL through the unjust sackings of managers such as Claudio Ranieri (Leicester, 2017), Carlo Ancelotti (Chelsea, 2011), or Frank de Boer (Crystal Palace, 2017), to name a few (Jones, 2019). The last manager mentioned, Frank de Boer, oversaw Crystal Palace for just 77 days (Jones, 2019). However, as indicated by Reep et al. (1971), football is often dominated by the element of chance, “in which winning and losing is often determined by a single goal” (Brechot and Flepp, 2020: 336) and therefore only considering results can give a misleading picture to overall performance or likelihood of future success.

To combat this issue of inflating the relationship between match outcome and performance, researchers aimed to create a statistic which could better represent performance and capture the events within a match. In 2012, Opta introduced the first expected goals (xG) model. The role of the xG statistic is to calculate the probability of scoring during a shot at goal, factoring in the minutiae of the situation evidenced below. The sum of every shot’s probability is the number used to quantify xG for that given team. In essence, it represents the expected number of goals from a team in a game given the shots that were attempted. Despite the apparent simplicity of this metric, various factors are considered to calculate an accurate probability for a shot resulting in a goal.

Firstly, historical data is accumulated on shots taken within top European leagues. Most robust xG models contain over 100,000 shots for the shot to be compared against (Understat, 2022). Furthermore, most accurate models consider a wide variety of parameters which may affect the scoring probability. Brechot and Flepp (2020) outline 8 different factors which should be considered. ‘Location on the pitch’ includes the distance and the angle of the shot in relation to the goal. ‘Rule setting’ considers whether the shot occurs in open play, a free kick or penalty kick setting. ‘Body part’ assesses which part of the body the shot comes off. ‘Defensive pressure’ locates the position of the defenders and goalkeeper. ‘Motion sequence’ accounts for the sequence before the shot was taken. ‘Player finishing ability’ assesses how accurate the player is at shooting along with their mental capability. Similarly, the ‘Goalkeeper skills’ are accounted for. Finally, for advanced models, they can consider such factors as pitch conditions, wind, or spin of the ball (Brechot & Flepp, 2020: 340).

The power of the xG metric is that there are various factors included within one variable, making it extremely appealing for statistical models. This appeal is heightened due to the lack of free, readily available historic data on factors such as weather, team line-ups and match situations. Although the xG statistic may not account for the full impact of these factors, it is viewed more favourably than using just goals or shots on target statistics. Moreover, the xG metric is “less prone to the randomness associated with match outcomes because they are based on a much larger number of actions that represent a team’s true performance on the pitch” (Brechot & Flepp, 2020: 345). As noted, models include factors that account for the actions of the defensive team. Therefore, it is suggested that xG can be used to evaluate attacking and defensive performance.

All the discussed components are assigned following the conclusion of a game. These are inputted into the formulae shown below to update the ratings.

$$Elo_{Home(updated)} = Elo_{Home(pre-game)} + KG(O - E)$$

$$Elo_{Away(updated)} = Elo_{Away(pre-game)} + KG(O - E)$$

(Lacy, 2019)

A final consideration that requires attention in all football forecasting models that span over multiple seasons is the phenomenon of relegation and promotion. As mentioned earlier, following a season in the EPL three teams are relegated to the EFL Championship. To account for this, Lacy (2019) assigns the average of the relegated teams Elo to the three promoted teams. Moreover, to account for off-season changes such as player transfers, managerial and staff changes, a decay method was used to bring the team’s ratings back towards the general league mean (Lacy, 2019). This is an important step as the changes during the off-season in football can greatly impact the quality of a team. For example, Bahtia (2020) found that paying a transfer fee within the top 20% for a given players position results in a significant improvement to team performance. The equation for this process can be seen below.

$$Elo_{Home} = 0.8Elo_{Home} + 0.2 \times 1500$$

$$Elo_{Away} = 0.8Elo_{Away} + 0.2 \times 1500$$

3. Methodology

3.1 Data

Within football, like many modern professional sports, a wide variety of data is collected from each game. This is in the aim to quantify and record every action which holds a degree of value throughout an array of industries. Basic information such as team names, score-lines, shots, shots on target are available through many platforms. However, more detailed factors such as team line-ups, weather, or formations are rarely freely available in favourable formats (Dixon & Coles, 1997). Due to the lack of freely available detailed data, this study is limited to the use of team goals and expected goals as observable data from a match.

This study utilises data from Understat, a website which provides in-depth match data from the top European leagues. Eight seasons of EPL data was collected for this study, between 2014 and 2022. The study started at 2014 as Understat's xG model was released in 2014, which was a fundamental part of the analysis. Furthermore, the research concluded following the 2021/22 season as this was the last full season of data available. The data collected resulted in 3040 matches, supplemented with the team names, match goals, and xG.

The data was required to construct a new dynamic rating system, xGad, named after the use of the xG statistic to form attacking and defensive team ratings. These ratings aimed to quantify a team's ability to score and to not concede, which was noted by Dixon and Coles (1997) as the best way to summarise a team's ability. These ratings were intended to explain the number of goals each team may score within a game, by producing probabilistic forecasts for goals outcomes. Before the method in which the ratings are formed can be considered, it is important to evaluate the raw statistics that are included within the model.

Statistic	Description
Home Team	Name of Home Team
Away Team	Name of Away Team
Home Goals	Number of Goals for Home Team in Match
Away Goals	Number of Goals for Away Team in Match
Home Expected Goals	Expected Goals for Home Team in Match
Away Expected Goals	Expected Goals for Away Team in Match

Table 2 – Raw Statistics Collected from Understat

3.2 Visualising Data

3.2.1 Home, Away, and Total Team Goals

To score goals against the opposition is the aim of the game, a team cannot win without scoring a goal. Figure 1 highlights the number of goals scored by the respective home and away teams, which is, in nature, count data, in that it takes only non-negative integer values. Figure 1 highlights the existence of home advantage through the higher prevalence of home goals over 1 compared to away goals. Moreover, away team goals were more prevalent in the 0 and 1 categories. This is important information as the strength of the home advantage in the number of goals scored will be considered within the *Methodology* section further below. On average each home team scores 1.5 goals and each away team scores 1.2 goals per game. In addition to highlighting the home advantage, this is important to note approximately how many goals each team scores for distribution purposes within the xGad system.

Figure 1: Home and Away Team Goals in EPL (2014-2022)

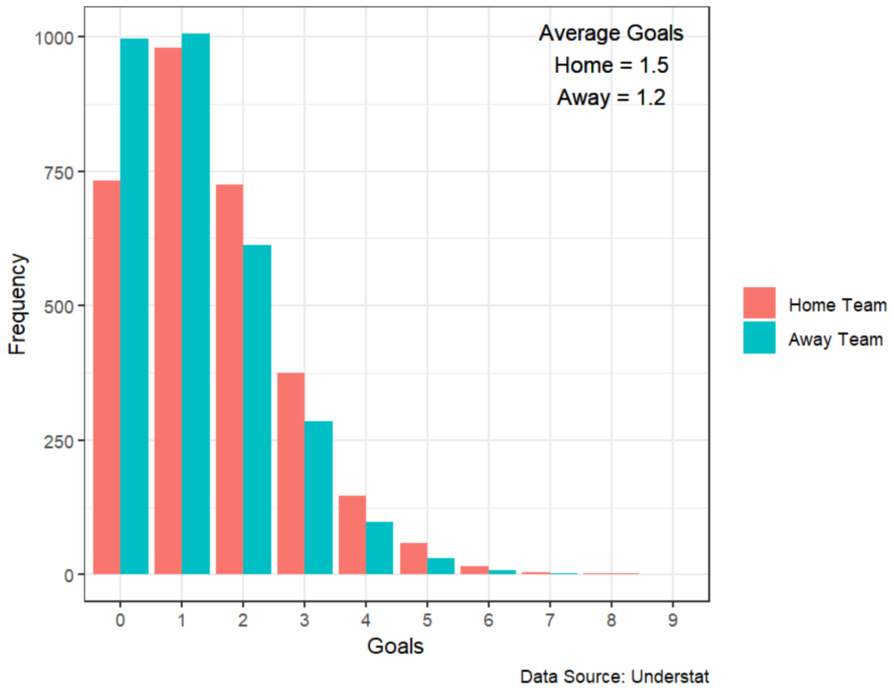


Figure 2: EPL Total Goal Distribution (2014-2022)

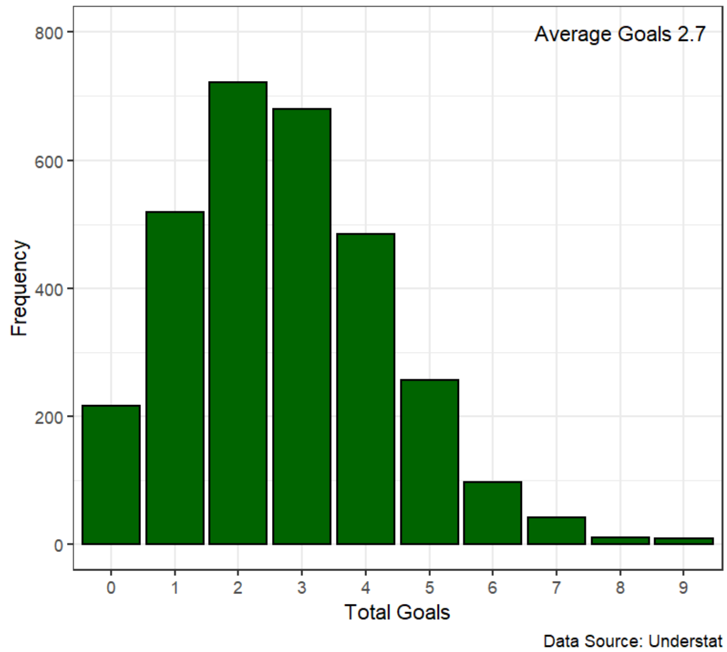
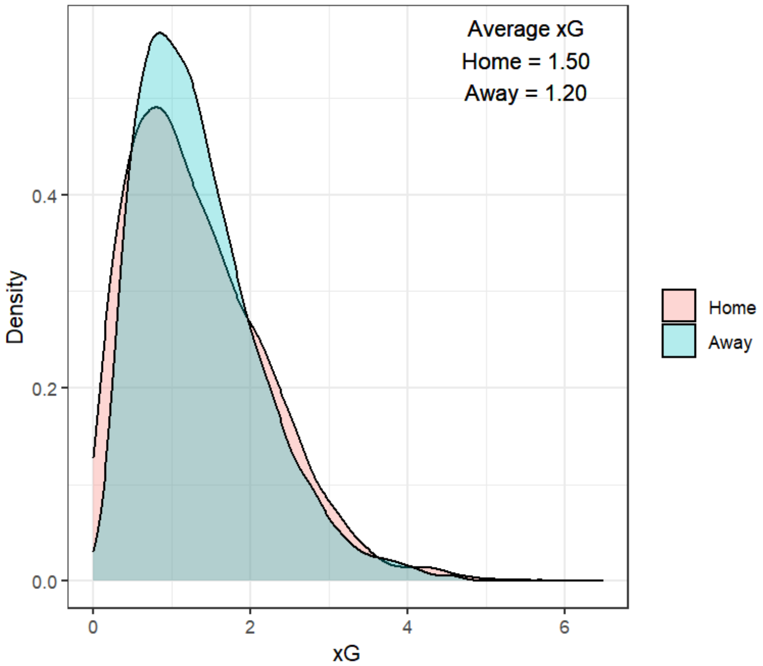


Figure 2 considers the distribution of total goals in the EPL (2014-2022). Total goals presented a similar left-skewed distribution; however, it was slightly more centred than the individual team goals. This can be explained by the nature of total goals, which is the sum of the home and away goals within a fixture. Furthermore, the average total goals in the EPL during the study period was 2.7, which is again an important consideration for the methodology.

3.2.2 Home and Away Expected Goals (xG)

The xG model used for this study was produced by Understat, who define xG as “a statistical measure of the quality of chances created and conceded” (Understat, n.d.). Unfortunately, the factors included in Understat’s model are not publicly available. However, they state that their model utilises neural network algorithms with a dataset which contains over 100,000 shots from the top European leagues (Understat, n.d.). Figure 3 highlights that Understat’s xG values shares a similar distribution and average as team goals. This is promising as an xG model should be representative of team goals over a prolonged period of observation (TacticsNotAntics, 2021).

Figure 3: EPL Expected Goals Distribution (2014-2022)



Data Source: Understat

3.3 xGad Rating Construction

This study used related methods to those proposed by Lacy (2019) and World Football Elo (2022). Utilising historic EPL data, a dynamic rating system was constructed. This was created through iterative methods that evaluated every game in the EPL from 2014 to 2022 in chronological order. Such methods resulted in historic ratings throughout the eight EPL seasons.

3.3.1 Ratings

Using calculations below, a pre-game metric was created to represent each team's relative attacking and defensive strength, in independent home and away fixtures. It was decided to measure attacking and defensive strengths separately, as recommended by Dixon and Coles (1997). They noted "that a team's ability is likely to be best summarized in separate measures of their ability to attack (to score goals) and their ability to defend (not to concede goals)" (Dixon & Coles, 1997: 269). Unlike Lacy (2019), who constructed a total rating for a team, Dixon and Coles (1997) hypothesised that separate ratings could better identify differences between the teams defensive and attacking capabilities. It is possible that team may excel in one of the disciplines but struggle in the other. The use of a total rating can conceal the unique nature of attack and defensive strength.

If two teams (Team A and Team B) of even total rating were matched against each other it is important to appreciate that there may be variation in their defensive and attacking strength. For example, Team A could have a high attacking ability with an average defensive strength, whereas Team B may be proficient in defensive situations but average in attacking ability. As a result, these two teams would have similar total ratings. Through the use of a total rating where attack and defence are combined, these intricacies are lost. The omission of this information may be detrimental to modelling the number of goals a team may score.

Moreover, the exploration into relationships between these intricacies may reveal new patterns.

At the start of the first season, each team was assigned a value of 1500 for their first home and away fixture. After a match was completed, their rating was updated, either rising or falling based on the quality of their performance. As noted previously, Lacy (2019) used margin of victory to scale a team's overall performance. This information was deemed too similar to team goals, given the measurement of attack and defensive ratings separately. Therefore, xG was used to scale the number of points awarded following a match.

The equations below are used in the xGad system to update the ratings following a match. The system is explained through two relationships: the home attacking and the away defensive strength, in addition to the away attacking and home defensive rating. Throughout the methodology, these ratings interact within these pairs, as it is assumed that these relationships have the largest impact on the number of goals a team may score. Within the xGad system, if a home attacking team earned 9 xGad points, the away defending team will have 9 points deducted. This interaction is also true between the away attacking team and the home defending team. This was informed by the assumption that if an attacking team has performed well, the defending team will have performed poorly, and vice versa. However, as will be discussed below, there are many interactions involved within the xGad system which controls this relationship.

The four ratings are updated through separate equations. The system is explained by four formulaic components: K, G, O and E. Each of these terms will be thoroughly discussed and justified below. Note that the abbreviations HA, HD, AA, and AD stand for home attacking, home defending, away attacking, and away defending.

$$xGad_{HomeAtt\ (Updated)} = xGad_{HomeAtt\ (Pre\text{-}game)} + (K \times GHA)(OHA - EHA)$$

$$xGad_{AwayDef\ (Updated)} = xGad_{AwayDef\ (Pre\text{-}game)} + (K \times GAD)(OAD - EAD)$$

$$xGad_{AwayAtt\ (Updated)} = xGad_{AwayAtt\ (Pre\text{-}game)} + (K \times GAA)(OAA - EAA)$$

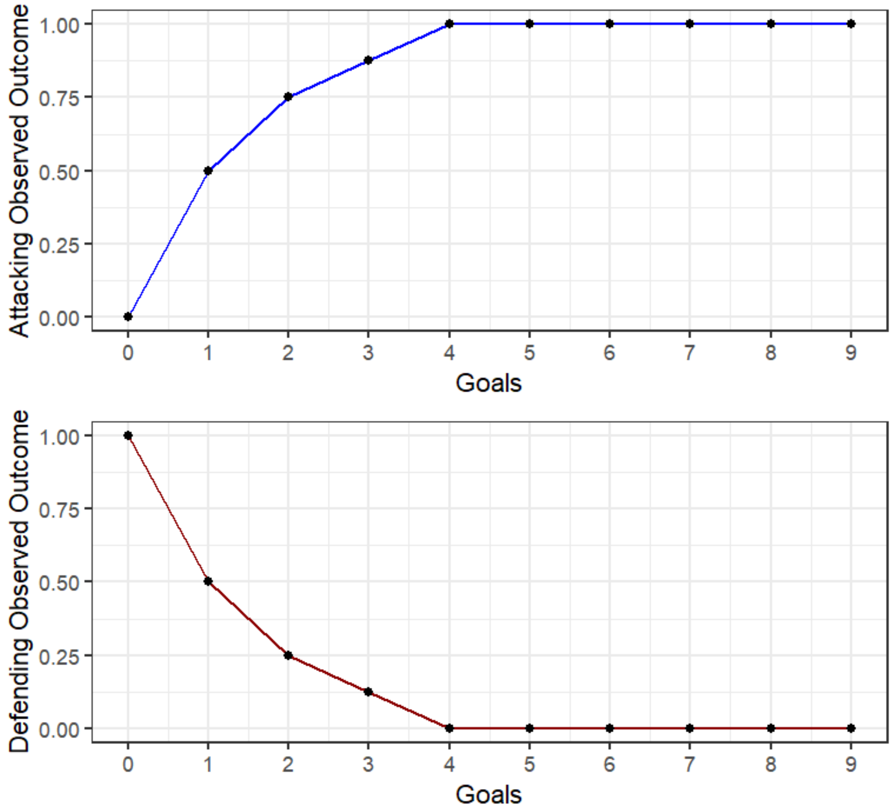
$$xGad_{HomeDef\ (Updated)} = xGad_{HomeDef\ (Pre\text{-}game)} + (K \times GHD)(OHD - EHD)$$

3.3.2 Observed Outcome (OHA/OAD/OAA/OHD)

To score or to stop the opposition from scoring is the central objective for each of the attacking or defensive sides. The *observed outcome (O)* term contains the information regarding the number of goals each team scores but converts it to a desirable form. *O* lies between 0 and 1, which represents how many goals were scored, or conceded, by the respective teams in a game.

In the calculation of the *O* term, it is assumed that the number of goals scored holds a non-linear relationship with the importance of the goal. This is an assumption within the margin of victory term in Lacy’s (2019) model. Football, in general, is a low scoring game, in the data used 70.33% of EPL matches contained 3 goals or less. Therefore, it is important to heavily reward teams for scoring goals of a low value. Furthermore, from 4 goals onwards, a team is awarded the same *O* value of 1. This is to reflect that there is little difference between the scoring 4 goals and 9 goals on the outcome of the game. These goals often do not impact the game and occur when team defences are dejected. Moreover, there are few observations in which 4 or more goals were scored by a single team. In fact, within the data only 6.04% of teams scored four or more goals in a match.

Figure 4: Observed Outcome Distribution



$$OHA\ /\ OAA = 0\ if\ Attacking\ Goals = 0$$

$$OHA\ /\ OAA = 0.5\ if\ Attacking\ Goals = 1$$

$$OHA\ /\ OAA = 0.75\ if\ Attacking\ Goals = 2$$

$$OHA\ /\ OAA = 0.875\ if\ Attacking\ Goals = 3$$

$$OHA\ /\ OAA = 1\ if\ Attacking\ Goals > 3.5$$

$$OHD\ /\ OAD = 1\ if\ Opposition\ Goals = 0$$

$$OHD\ /\ OAD = 0.5\ if\ Opposition\ Goals = 1$$

$$OHD\ /\ OAD = 0.25\ if\ Opposition\ Goals = 2$$

$$OHD\ /\ OAD = 0.125\ if\ Opposition\ Goals = 3$$

$$OHD\ /\ OAD = 0\ if\ Opposition\ Goals > 3.5$$

3.3.3 Expected Outcome (EAH, EDH, EAA, EDA)

The second component in this system is the expected outcome *(E)*. To understand *E*, the difference in ratings (*D*) must first be examined as it is a variable within the calculation of *E*. *D* is defined as the difference in rating of two teams in any given match. *D* is composed of four terms, each paired to their respective opponent xGad rating. As noted previously, $xGad_{HomeAtt}$ is associated with the $xGad_{AwayDef}$, and vice versa. Furthermore, each pair is reflective in nature. For example, if DHA is equal to -10, DAD is equal to 10.

$$DHA = (xGad_{HomeAtt} + HAdv) - xGad_{AwayDef}$$

$$DAD = xGad_{AwayDef} - (xGad_{HomeAtt} + HAdv)$$

$$DAA = xGad_{AwayAtt} - (xGad_{HomeDef} + HAdv)$$

$$DHD = (xGad_{HomeDef} + HAdv) - xGad_{AwayAtt}$$

Within such a system, it is important to consider the home advantage (*HAdv*) effect. As discussed in the *Literature Review*, both Lacy (2019) and World Football Elo (2022) consider that it is appropriate to include this consideration in *E*. The *HAdv* term adds xGad points to the home team, resulting in a *D* term which has bias towards the home team. World Football Elo (2022) used a fixed value of 100 for their model, however separating the attacking and defensive ratings requires a consideration of how the home advantage may affect each of these disciplines. A lower *HAdv* term of 25 was found to be appropriate when splitting attack and defence. The justifications for this are discussed in the *Diagnostics* section.

$$EHA = \frac{1}{1 + 10^{\frac{-DHA}{400}}}$$

$$EAD = \frac{1}{1 + 10^{\frac{-DAD}{400}}}$$

$$EAA = \frac{1}{1 + 10^{\frac{-DAA}{400}}}$$

$$EHD = \frac{1}{1 + 10^{\frac{-DHD}{400}}}$$

E is a necessary term to control for the difference in ratings between teams in a matchup. As *E* is associated with *O*, it was important that it shares a minimum and maximum value of zero to 1. As shown by Figure 5, as the *D* term increases, *E* rises. This relationship was desired as it was assumed that as an attacking *D* term increases, the probability of the attacking team scoring in the game rises. Therefore, *E* increases to control the impact of the *O* term, which is expected to be high due to their rating differential. This assumption holds for the defending *D* term, in which it is assumed a high defensive *D* term increases the probability of the attacking team not scoring. Therefore, an attacking *E* term can be viewed as the probability of a team scoring a goal, given the rating differential. Secondly, the defending *E* term is representative the probability that the defence stops the opposition attack from scoring. This assumption is tested in the *Results* section.

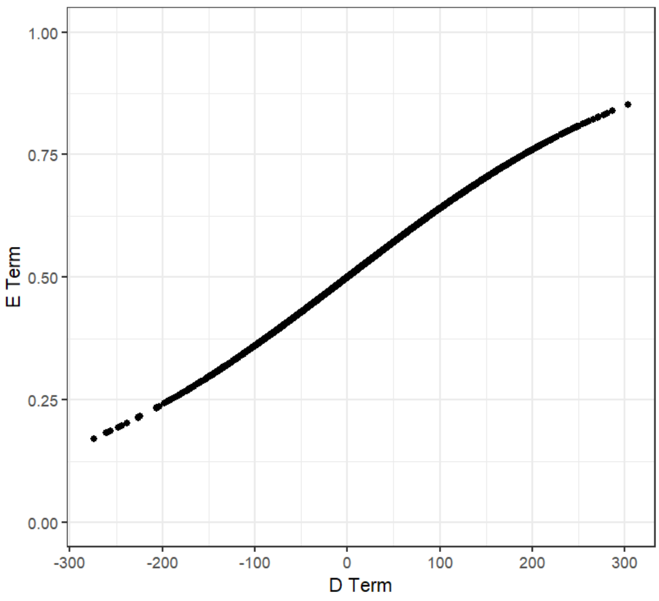
Figure 5 highlights the fact that if there is no difference between the attacking and defensive ratings, *E* will be equal to 0.5. Therefore, as *D* shifts, *E* increases or decreases from 0.5. As noted previously, DHA is reflective of DAD, and DAA is reflective of DHD.

Therefore,

$$If\ EHA = 0.42\ then\ EAD = 0.58$$

$$If\ EAA = 0.61\ \ then\ EHD = 0.39$$

Figure 5: D to E Term Distribution

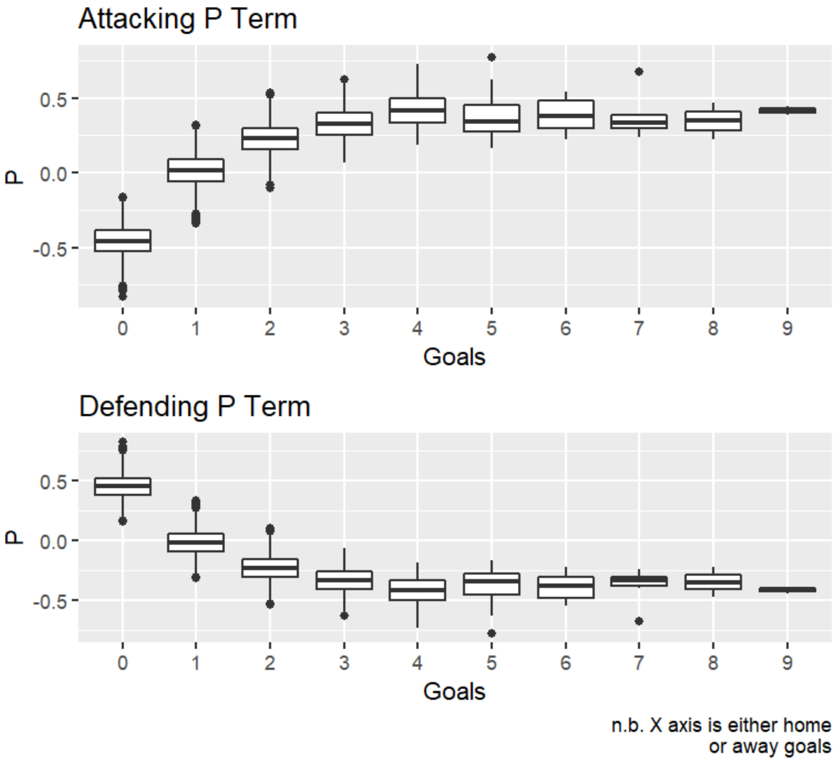


The interaction above results in teams of a higher skill level than their opponent risking losing more xGad points should they perform poorly. Moreover, if they perform to a high level, as expected, they will not gain as many points, and the opposition will not lose such a large degree of xGad points. This follows a key characteristic of an Elo rating system.

3.3.4 Interaction Between Observed Outcome and Expected Outcome

A major component in the updating of the ratings is the interaction between *O* and *E*. The relationship between the two parameters dictates whether the rating will be updated in a positive or negative manner. Figure 6 visualises the relationship between the number of goals scored or conceded and P, which is equal to $O - E$. This reveals the difficultly in rating a team's attacking and defensive strengths separately. In using the team goals as a measure for *O*, the attacking team is favoured. This is highlighted by the many more predominantly positive boxplots in the attacking term than defending. However, this should not be a large issue as most observations lie between 0 and 3 goals. The variation in the boxplots is explained by the *E* term.

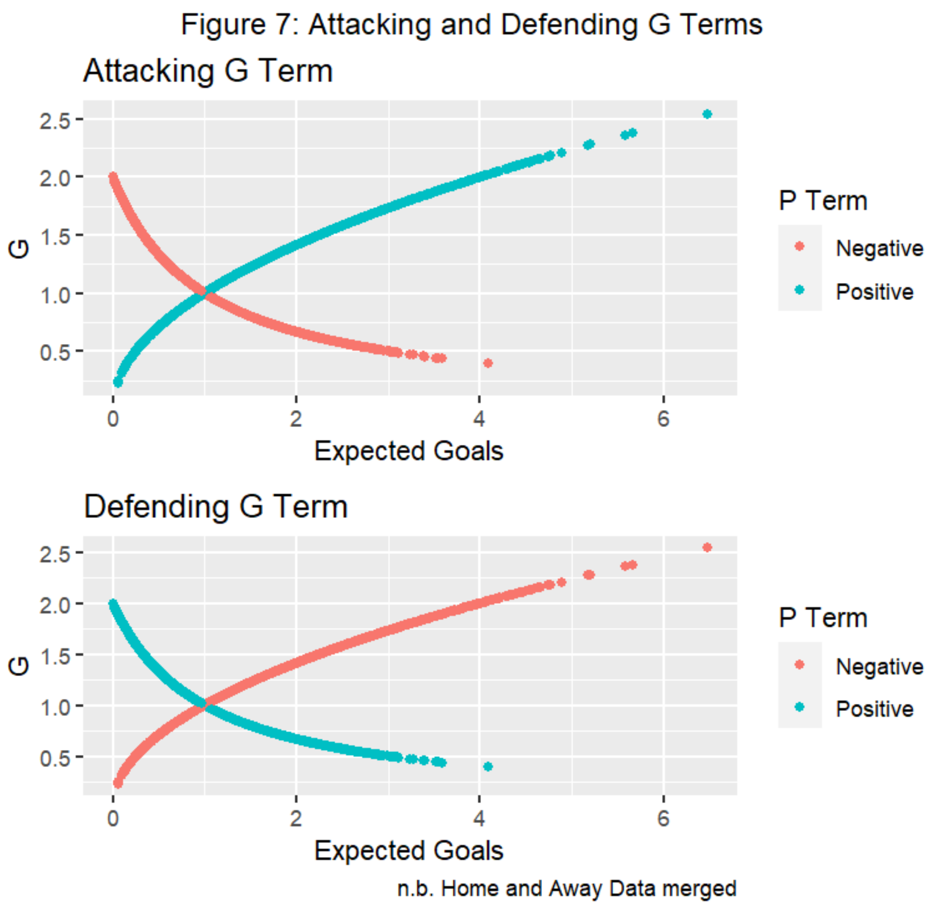
Figure 6: Attacking and Defending P Terms



3.3.5 Expected Goals (GHA, GDA, GAA, GDH)

Within this system, xG is utilised in the form of the G component, a multiplier which scales the points awarded to a team based on their performance. xG is a powerful statistic that can be used to reveal events in the game which are often lost in more traditional match statistics such as goals, shots, or corners.

G must account for positive or negative values that have resulted from $O - E$ interaction. Therefore, the G term is calculated to account for the varying scenarios which would require the xGad points to be scaled up or down. For example, consider a team of a higher level than their opposition, who scored 0 goals in a game, but had an xG of 2.8. This would indicate that their performance did not reflect their zero goals. The interaction between O and E would result in a negative number which is multiplied against the G term. Following this scenario, the team's attacking rating should be reduced. However, their performance warrants a smaller reduction to their xGad rating as they still created numerous goalscoring chances. It should be noted that although xG can uncover a positive performance despite a poor outcome, it is not desired to change a negative P term due to a favourable xG. The aim of G is to scale the observed and expected outcome of the game by the team's performance.



The G term also required scaling for the home effect as xG also contains an inherent advantage to the home team, as shown by Figure 3. The home bias must be accounted for, otherwise the home ratings will be scaled unevenly compared to away teams. The home and away xG was standardised to represent the overall average xG. The standardisation was achieved via the process shown below. The different equations avoided negative xG values and adding too much value to poor performances. This resulted in a home and away xG average of 1.35.

$$if\ Home\ xG > 0.3\ \ then\ HxG - 0.15$$

$$if\ Home\ xG \leq 0.3\ \ then \frac{HxG}{2}$$

$$if Away\ xG > 0.16\ \ then\ AxG + 0.15$$

$$if\ Away\ xG \leq 0.15\ \ then\ AxG \times\ 2$$

Figure 7 visualises the desired relationship between the P term (O – E) and the G multiplier. It should be noted that within the xGad system, O and E were calculated prior to G, which allowed the formula below to be possible.

$$If\ PHA/PAA < 0\ then\ GHA/GAA = \frac{2}{1 + xG}$$

$$If\ PHA/PAA > 0\ then\ GHA/GAA = \sqrt{xG}$$

$$If\ PHD/PAD < 0\ then\ GHD/GAD = \sqrt{xG}$$

$$If\ PHD/PAD\ > 0\ then\ GHD/GAD = \frac{2}{1 + xG}.$$

The calculation of the defending and attacking and G terms were opposite. This was caused by a low xG indicating a good performance by a defending team and a high xG suggesting a good performance for an attacking team.

3.3.6 K

The final parameter within the model was the simplest of all, which was a scaling constant. As seen in Lacy's (2019) model, the standard practice is to use a constant of 20. This constant was also set at 20 within this methodology, which was also deemed appropriate for the length of the EPL season. K ensured that an adequate amount of xGad points were awarded for the array of performances that are possible in football. Moreover, this value was successful in causing an appropriate affect over the eight seasons of the study. The value of 20 was multiplied against G(O-E) to produce the updated xGad rating for the relevant rating.

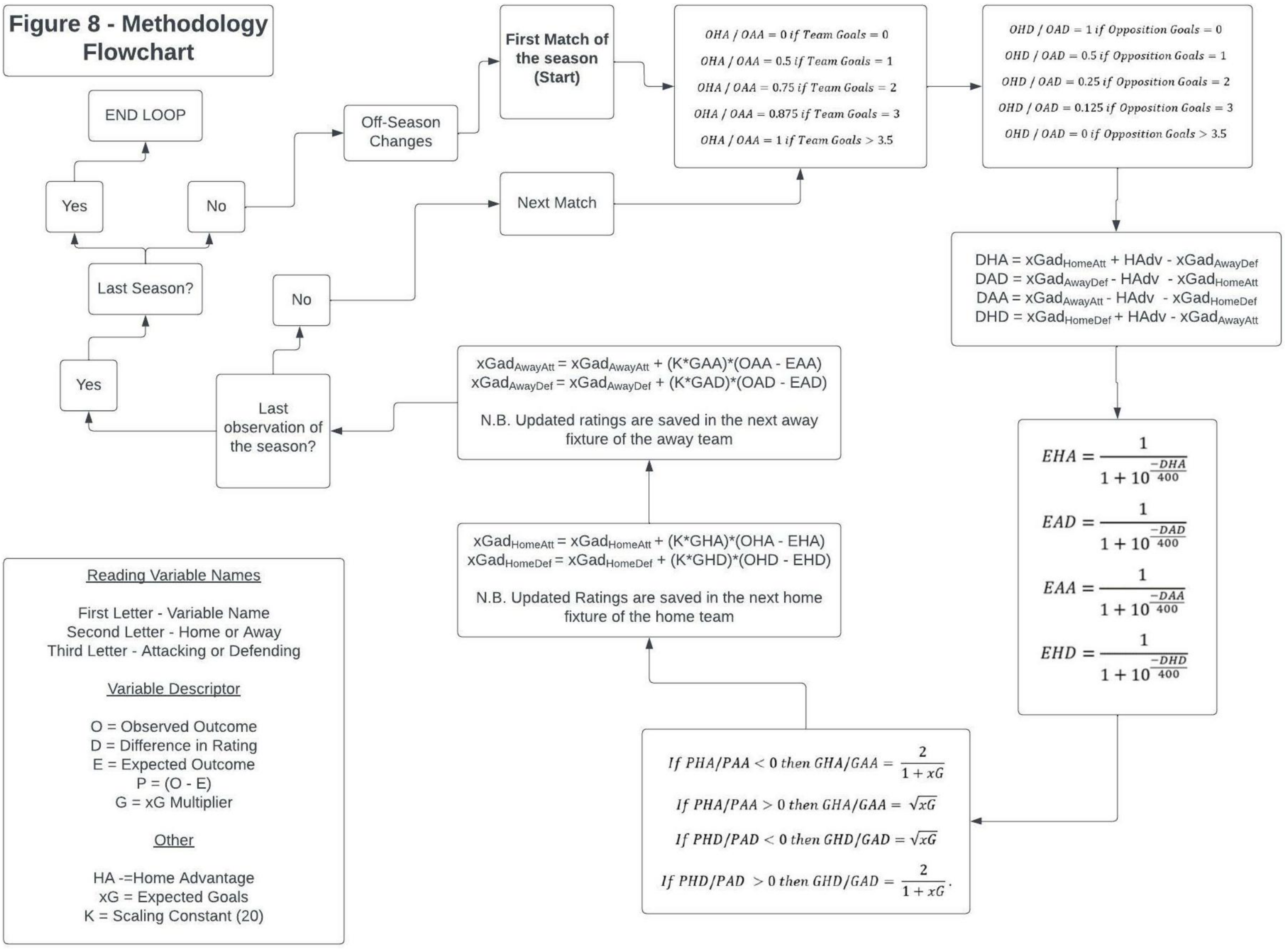
3.3.7 Procedure for Off-Season

At the end of an EPL season, a consideration for the processes which occur in the off-season is required. As noted earlier, following a season the bottom three teams in the EPL table are relegated and replaced by three teams from the EFL Championship. Within this paper, there is not a simultaneous model running for the EFL Championship. Therefore, the promoted teams do not have a rating of their own to start the EPL season with. To account for this, it is assumed that the promoted teams' ratings are equal to the average of the relegated teams (World Football Elo, 2022). Following the calculation of these averages, they are assigned to the promoted teams first games of the new EPL season.

Promoted teams aside, each team is assigned their final ratings from their last home and away games in the previous season. However, during the off-season, a club may undergo a series of changes. Commonly, change can result from the buying and selling of players. It is rare for a team to start a new season in the EPL with the same squad that they finished the prior season with. This change can be positive or negative for the club as it is dependent on if they have a large budget to improve their team. Moreover, there are often managerial and staff changes during this period which can affect club success (Audus, et al., 2002). To account for these processes, a decay method provided by FiveThirtyEight (n.d.) was used. The following formula is used to bring the ratings back towards the starting point of 1500.

$$xGad = 0.8xGad + 0.2 \times 1500$$

The ratings were shifted back towards 1500 as this was also the desired average throughout the xGad ratings. Any team who possessed a rating above 1500 were above average in the respective discipline.



4. Results

The *Methodology* above resulted in four xGad ratings for each EPL team, quantifying their relative attacking and defensive strength in home and away fixtures. This system was run through eight seasons of the EPL, resulting in 3040 matches which contain two pre-game xGad ratings for each team. Figure 9 visualises the final ratings for each of the teams for the 2021/22 season. The final ratings were utilised as xGad is dynamic in nature, the last rating is intended to provide an up-to-date assessment of a team’s strength. Moreover, Figure 8 contains information of where the team placed in the final EPL table. Although xGad measures whether a team is likely to score or concede during a match, when attacking and defensive ratings are paired, they can provide a metric for the overall strength of a team. This generalisation is justified by the fact that results depend on a team's ability to attack and defend. It would be expected that there will be a clear divide between the teams who finished at the top and those who finished at the bottom of the EPL table.

Figure 9: 2021/22 EPL Final xGad Ratings

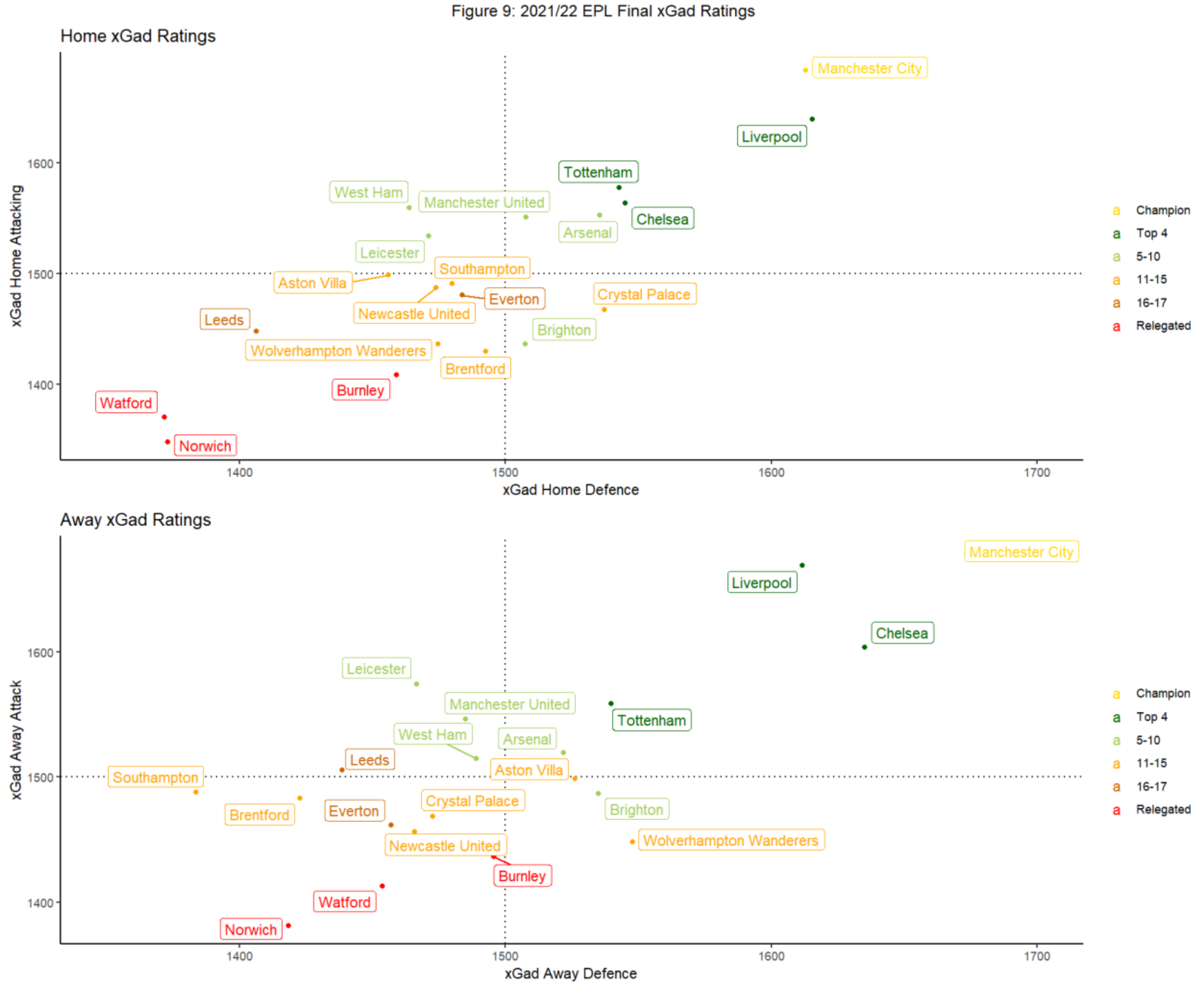


Figure 9 highlights that the xGad ratings were in fact representative of the league table at the end of the season. This is promising as no information was inputted into the methodology regarding the match outcome, i.e., who won the game. The ratings particularly identified the teams who excelled in the league and those who struggled. The teams in the middle of the table are slightly unordered. However, this can be explained by the difference in team's form in home and away fixtures. Moreover, the variation in the centre of the plot can be explained by how tight the EPL table is at the end of the season. For example, in 2021/22 there were only 7 points separating 8th place from 14th.

Finally, the plot identifies how certain teams excel in a discipline but struggle in another. For example, Leicester evidently was successful at attacking in away fixtures but struggled in defending within these matches. This pattern was justified as in the 2021/22 EPL, Leicester scored 1.47 goals, but conceded 1.89 goals per away match. An average of 1.47 away goals scored per game is above average as the league average for that year was 1.31. Similarly, Leicester's 1.89 goals conceded per away game was above the league average of 1.51.

4.1 Assumption Testing

Figure 9 provided a promising first glance at the xGad ratings as they reflected the observed ability of the teams in the EPL. With this satisfied, a central assumption used within the formulation of the xGad ratings requires attention. This is that the difference in Team A's attacking and Team B's defensive rating, has an impact on Team A's goals. This relationship was assumed to be positive, in that an increase in the difference in rating, will increase the number of Team A's goals. Within the *Methodology*, the difference in rating was captured by the *D* term, which impacted *E*.

A secondary assumption within this study is that there is no relationship between the difference in the Team B's attacking and Team A's defensive strength on the number of Team A's goals. In the calculation of *E*, there was no consideration of the opponents attacking rating.

To test these assumptions, two separate regression models were constructed studying home and away team goals, respectively. A generalised linear regression method with a Poisson distribution was chosen, as indicated by early literature as an efficient method to fit football goals (Maher, 1982; Dixon & Coles, 1997). Model 1, seen in table 3, studied DHA and

DAA's relationship with home goals. Interestingly, it found that both variables had a highly significant relationship with home goals. DHA, or the difference in home attacking and away defensive ratings, had a positive intercept. This indicates that a one unit increase in DHA, increases the expected log count of home goals by 0.0028. Alternatively, a one unit increase in DAA, is associated with a 0.0011 decrease in the expected log count of home goals.

Table 3:
Mod1 = Home Goals ~ DHA + DAA, family = Poisson

	Intercept	Standard Deviation	Z value	P value
DHA	0.0028285	0.0002508	11.278	< 2e-16 ***
DAA	-0.0010591	0.0002692	-3.934	8.34e-05 ***
Residual deviance: 3424.1 on 3037 degrees of freedom				
AIC: 9171.2				
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1				

Model 2 below continues the assumption testing for the away goals and found a very similar result. Namely, a one unit increase in DHA is associated with an increase of 0.0027 in the expected log count of away team goals. Secondly, a one unit increase of DHA decreases the expected log count of away team goals by 0.0015. Therefore, the assumption that the difference in a team's attacking rating and their opponents defensive rating has an impact on individual team goals was justified. However, the significant relationship with the oppositions attacking *D* term was unexpected. This finding suggests there may be a more complex relationship between the four ratings than the xGad system has accounted for.

Table 4:
Mod2 = Away Goals ~ DAA + DAH, family = Poisson

	Intercept	Standard Deviation	Z value	P value
DAA	0.0027205	0.0003001	9.065	< 2e-16 ***
DHA	-0.0014745	0.0002899	-5.087	3.64e-07 ***
Residual deviance: 3589.2 on 3037 degrees of freedom				
AIC: 8532				
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1				

4.2 Testing xGad’s Predictive Accuracy

Within football, three types of goal statistics can be considered: home team goals, away team goals, and total match goals. These vary in complexity which can hinder the ability for the ratings to explain the variation within them. Total match goals is the most complex statistic as it is the sum of home and away team goals, resulting in various score lines which could explain the total goals figure. For example, three goals in a match can be explained by four different match results (3-0, 0-3, 2-1 or 1-2).

4.2.1 Team Goals

As noted, the xGad ratings can serve as data to use for predictive purposes, especially within betting markets. Binary logistic regression models were constructed to examine the predictive accuracy of the four xGad ratings on whether there were two or more goals by a team in a game. The average goals scored by home teams in the EPL during the study period was 1.5, and 1.2 for away teams. These averages indicated a relatively even distribution between matches where the team goals were over or under, albeit a higher prevalence for under cases for away goals. Both models included all four xGad ratings as models 1 and 2 suggested that there may be a relationship between the other ratings that were not accounted for in the rating calculations. For example, $xGad_{HomeAtt}$ and $xGad_{AwayDef}$ were associated with home team goals, but models 1 and 2 suggests that $xGad_{AwayAtt}$ and $xGad_{HomeDef}$ may impact home goals as well. The results from the binary logistic regression models can be seen below in tables 5 and 6. To ensure that there was no bias in the models' predictive accuracy, data was partitioned into training and testing sets at a 70% split. This resulted in 2128 matches in the training dataset and 912 in the testing dataset.

Table 5:
Mod3 = Over 1.5 Home Goals ~ $xGad_{HomeAtt} + xGad_{AwayDef} + xGad_{AwayAtt} + xGad_{HomeDef}$, family = binomial

	Intercept	Standard Deviation	Z value	P value
$xGad_{HomeAtt}$	0.008299	0.001123	7.391	1.45e-13 ***
$xGad_{AwayDef}$	-0.007473	0.001357	-5.508	3.63e-08 ***
$xGad_{AwayAtt}$	-0.001402	0.001057	-1.327	0.185
$xGad_{HomeDef}$	0.001505	0.001490	1.010	0.312
Residual deviance: 2683.1 on 2123 degrees of freedom				
AIC: 2693.1				
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1				

The home attacking xGad rating was highly significant, and for every one unit increase in the rating, the log odds of there being more than 2 goals for the home team in a game increases by 0.0083. This is the opposite for the away defending xGad rating, which is associated with a 0.0075 decrease. The other xGad ratings, $xGad_{AwayAtt}$ and $xGad_{HomeDef}$ were not significant, contrary to what was indicated in model 1.

Table 6 continues the analysis to study the away goals in a match. However, different patterns were found within model 4. Firstly, as expected $xGad_{AwayAtt}$ was highly significant, in that a one unit increase in the rating is associated with an increase of 0.0066 in the log odds of there being 2 or more away goals in a match. $xGad_{HomeDef}$ was not found to be significant, however, it was close to the 0.05 P-value required to conclude significance. This was not desirable as it would have been expected that this rating would be highly significant with a negative intercept.

Table 6:
Mod4 = Over 1.5 Away Goals ~ $xGad_{AwayAtt}$ + $xGad_{HomeDef}$ + $xGad_{HomeAtt}$ + $xGad_{AwayDef}$, family = binomial

	Intercept	Standard Deviation	Z value	P value
$xGad_{AwayAtt}$	0.006603	0.001120	5.897	3.70e-09 ***
$xGad_{HomeDef}$	-0.002714	0.001569	-1.730	0.0836 .
$xGad_{HomeAtt}$	-0.004760	0.001136	-4.191	2.78e-05 ***
$xGad_{AwayDef}$	0.002302	0.001330	1.731	0.0835 .
Residual deviance: 2550.5 on 2123 degrees of freedom				
AIC: 2560.5				
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1				

Interestingly, $xGad_{HomeAtt}$ was found to have a highly significant relationship with whether there were two or more away team goals in a match. Given a one unit increase in the rating, it was associated with a 0.0048 decrease in the log odds of the over 1.5 away goals outcome. This finding indicates that a home team who is proficient in attacking, has an impact on the performance of the away attacking team. This could be caused by various reasons, such as the interplay between the midfield and attack in a team. This relationship could be explored in future research to explain this phenomenon.

Given that significant relationships were found between the xGad ratings and team goals, models 3 and 4's predictive accuracy was assessed. Utilising the predict function, two confusion matrices were constructed using the testing sets from each of the models. For this analysis, the matches which the models produced a probability of below 45%, were noted as 'Under', indicating that the model believed that there would be less than 1.5 team goals in the game. Alternatively, the probabilities above 55% were noted as 'Over'. The observations between 45% and 55% were noted as 'Uncertain'. This was conducted to account for the degree of uncertainty within football. The results of this analysis can be seen in tables 7 and 8 below.

Table 7: Model 3 Confusion Matrix (O/U 1.5 Home Goals)

N = 912	Pred. Uncertain	Pred. Over	Pred. Under
Obs. Under	96	94	339
Obs. Over	89	146	148

Table 8: Model 4 Confusion Matrix (O/U 1.5 Away Goals)

N = 912	Pred. Uncertain	Pred. Over	Pred. Under
Obs. Under	53	31	518
Obs. Over	57	50	203

Given the omission of matches labelled as uncertain, the models 3 and 4 were 66.71% and 70.82% accurate, respectively. These results are promising given the uncertainty and the difficulty in predicting outcomes in football. As noted by Dixon and Coles (1997), "in the long run, it is not difficult to predict fairly accurately which teams are likely to be successful, but the development of models that have a sufficiently high resolution to exploit this long run predictive capability for individual matches is substantially more difficult". These results are promising, but there is an indication of improvement that can be made to make the ratings more robust. Namely, improving $xGad_{HomeDef}$ to gain significance with away team goals. Alternatively, this could indicate the need for additional variables to supplement the ratings, such as weather or team line-ups.

It should be noted that the increase in accuracy from model 3 to 4 can be explained by the skewed distribution for away goals. In the model 4 train dataset 66% of games contained less than 2 away goals. Moreover, within model 3's train dataset, 58% of games contained less than 2 home goals. Despite this, both models provided an accurate prediction of these two outcomes, given the degree of uncertainty that is within the game of football.

4.2.2 Total Match Goals

Models 3 and 4 produced some promising results, which indicated that the study of a more complex outcome, such as total goals may be possible. Table 9 contains the results from a binary logistic regression model with the over 2.5 total match goals as the outcome variable. Similarly, to models 3 and 4, the value of 2.5 was chosen as the average total match goals within the EPL data period was 2.725. Due to the complexity of the total goals variable, the full dataset was used to form model 5. To evaluate the model's accuracy, a receiver operating characteristic (ROC) plot was constructed, see Figure 10.

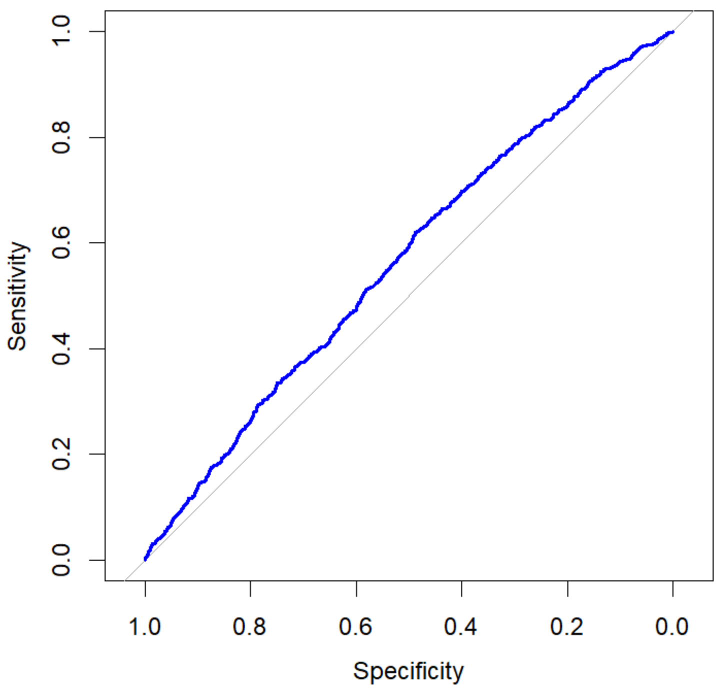
Table 9:
Mod5 = Over 2.5 Total Goals ~ $xGad_{HomeAtt}$ + $xGad_{AwayDef}$ + $xGad_{AwayAtt}$ + $xGad_{HomeDef}$, family = binomial

	Intercept	Standard Deviation	Z value	P value
$xGad_{HomeAtt}$	0.0029679	0.0008566	3.465	0.000531 ***
$xGad_{AwayDef}$	-0.0010540	0.0010448	-1.009	0.313055
$xGad_{AwayAtt}$	0.0031938	0.0008486	3.764	0.000167 ***
$xGad_{HomeDef}$	-0.0008026	0.0011748	-0.683	0.494481
Residual deviance: 4170.7 on 3035 degrees of freedom				
AIC: 4180.7				
Significance codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1				

Table 9 highlights that, as mentioned, the total goals statistic is more complex than team goals. In fact, the defensive xGad ratings were not significant in explaining the variation in whether there would be over 2.5 goals in an EPL match. Despite this, the xGad attacking ratings were highly significant. A one unit increase in $xGad_{HomeAtt}$ increased the log odds of a match being over 2.5 goals by 0.0030. Secondly, a one unit increase in $xGad_{AwayAtt}$ increased the log odds of a fixture being over 2.5 goals by 0.0032. It should be highlighted that these intercepts are notably lower than those found in models 3 and 4. This further indicates issues with the ratings measuring the binary total goals match variable.

Figure 10 visualises the inefficiencies in model 5. A ROC plot analyses the sensitivity against the specificity of the model’s predictions. The area under the curve (AUC) is an indicator of the accuracy of the model, which was 0.5643. This again highlights the inefficiencies within model 5. Despite this finding, the xGad ratings have shown a promising ability to accurately predict the outcomes for individual team goals. For discussion regarding possibilities for future use of these ratings see *Conclusion*.

Figure 10: Over 2.5 Total Goals ROC Curve

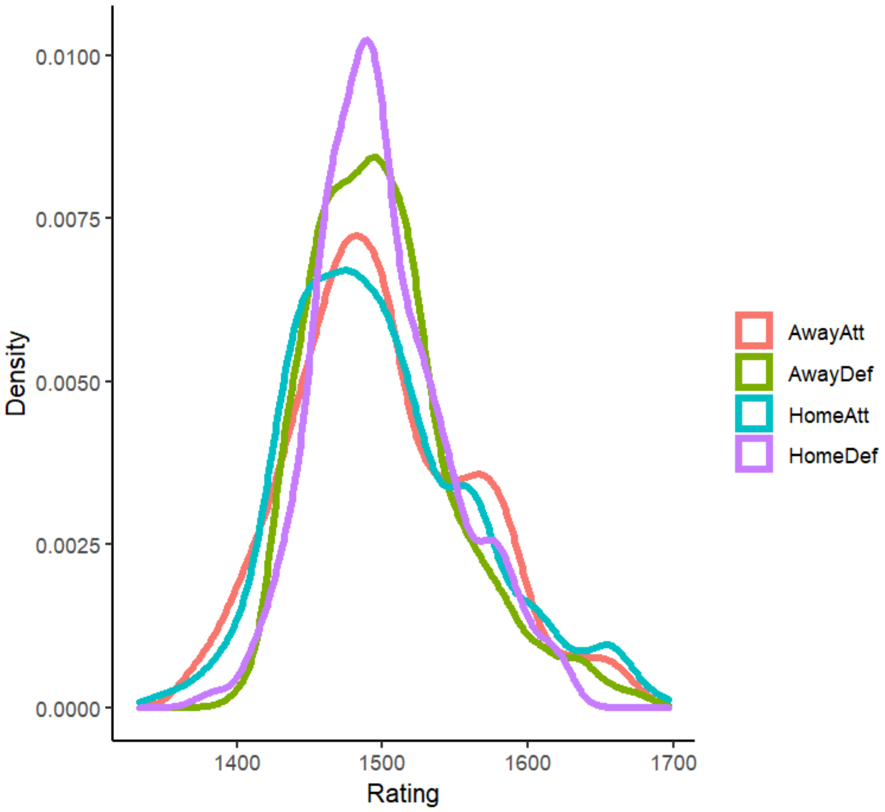


5. Diagnostics

5.1 Post-Season Changes

Following the analysis, it is important to consider the inefficiencies and improvements that can be made to the xGad system. Figure 11 visualises the density of each of the xGad ratings throughout the 8 EPL seasons. There is a difference in the distribution less than and more than 1500 xGad points. Higher ratings have a more prolonged distribution compared to the lower ratings. This can be explained by the post-season procedures. As seen in the *Methodology* section above, the promoted clubs from the EFL Championship are assigned the average of the relegated teams from the EPL. This can be problematic as it assumes that the promoted teams are at an even level to the relegated teams. Moreover, in using the average of the three relegated teams, the value assigned will be more than the lowest value from the relegated teams. This is contrasted by the high ratings, which only experience reduction through the decay method following the season. This leads to consistently strong teams such as Manchester City and Liverpool which have the ability to reach the higher levels on the xGad scale.

Figure 11: Distribution of EPL xGad Ratings (2014-2022)



To combat this phenomenon, it would be proposed that the xGad system is also run for the EFL Championship. The ratings within this system would be in relation to the EPL teams. Through this method, the promoted teams would have accurate ratings when they are placed in the Premier League. Moreover, it would be advisable to revise the decay method, as it is a slightly naïve assumption to assume that following an off-season, the team's become more equal in strength. Clubs have varied off-seasons based on their financial situations; it would be recommended to update clubs' ratings for the start of the season considering their activity in the transfer window. As mentioned, Bahtia (2020) found that paying a transfer fee within the top 20% for a given players position results in a significant improvement to team performance. Taking a club specific approach in could provide a more accurate portrayal of these off-season changes, thus improving the accuracy of the system.

5.2 Positive and Negative Interaction

As noted previously, P, which is equal to *O* minus *E*, determines whether the respective rating will be updated in a positive or negative direction. Table 3 shows the proportion of positive to negative interactions within the 4 xGad ratings. This reveals an issue with a disproportionate number of positive interactions for the attacking ratings compared to defensive. This is the issue with *O*, which representative of team goals, a discrete variable with an inherently low count. The average number of home team goals within the data is 1.5, and 1.2 for away team goals. This raises an issue in finding an even distribution, as either side of 1 or 2 would cause a differential in the *O* term for attacking and defending teams. This could indicate that the use of another variable to measure attacking, and defensive outcomes could be utilised in place of team goals. However, there is the risk of losing the most important outcome of a game, which is the goals. The use of alternative metrics, even xG, in the place of observed goals could lead to issues of over or underestimation. This is due to teams who consistently over or under perform their xG. Such as Tottenham who over the eight-year period, had a home attacking xG average of 1.63 and a home goal average of 1.95.

It should also be noted that the similar distribution between home and away ratings is evidence that the *E* term has appropriately controlled the home advantage. This is despite Lacy (2019) and World Football Elo's (2022) systems which recommended the use of 100. However, the use of separate ratings justified the use of the smaller HA value of 25.

Table 10 – P Interactions

	Negative	Positive
PHA	1312 (43%)	1728 (57%)
PAD	1728 (57%)	1312 (43%)
PAA	1292 (43%)	1748 (57%)
PHD	1748 (57%)	1292 (43%)

5.3 Defensive xGad ratings

Within the xGad system, attacking and defensive ratings are reflective in nature when they are updated. This assumes that if an attacking team performs to a high level, the defending team performed to the same degree in a negative manner, and vice versa. This is caused by the calculation of the defensive ratings being mirrored by those of the attacking ratings. This is an overgeneralisation which may explain the issues in the defensive ratings. This indicates a separate method to assess defensive outcomes and performance may be fruitful in creating a more robust system.

Firstly, an issue with using the goals conceded as an outcome is that football is biased towards the attacking side of the game. This is reflected by only 24% of games in which the home team do not score, and 33% of matches that the away team do not score during the study period in the EPL. This results in an uneven assessment of attacking and defensive outcomes once a first goal is scored as it is more likely for a goal to be scored than not. This indicates that the defensive *O* term may need to be revised to account for these probabilities.

Secondly, it assumed that xG has the same power in evaluating defensive performance as it does for attacking. This is a large assumption can be justified by the fact that most xG models include defensive factors which affect the probability of a shot resulting in a goal. For example, a goalkeeper playing to a high standard will ensure that they are positioned well for a shot, thus reducing the xG. In theory, good defensive teams will cause their opponents to have lower xGs. This is confirmed as Understat data showed that for the 2021/22 season, the top three teams in the EPL had the lowest xG against average (Understat, 2022). However, the accuracy of this in terms of the level of performance is unclear. Unfortunately, there is yet to be a statistic as powerful as xG to evaluate defensive performance aside from standard metrics such as tackles, interceptions, and blocks. Given that the xG statistic was only introduced by Opta in 2012, and has swiftly been adopted by multiple platforms, it can be assumed that within the coming years new models will be developed which are able to measure all areas of the game to a large degree. Using these metrics as an alternative to xG could improve the accuracy of the model.

5.4 Pre-Game Factors

This study attempted to use team's ratings alone to explain the variation within binary goal variables. For future use following improvements to the xGad system, it would also be recommended to include other pre-game factors which affect game outcomes. There are numerous factors such as injuries, weather, playing styles, rest days and pre-game motivations which can play a role in team performance (Hägglund, et al., 2014; Iskandaryan, et al., 2020; Castellano & Pic, 2019; Aughey, et al., 2014). If historical data is available, and the factors listed can be quantified, they could provide vital information to increase the accuracy of the xGad ratings.

6. Future Use of the xGad System

The *Results* section above suggests a need to improve the xGad system to increase its predictive accuracy. However, given improvements are made to the system, it is important to note how it may be used as a betting model. Prior to use in real-world betting markets, its effectiveness should be evaluated on historical data in conjunction with a betting strategy. A popularised method which could be utilised is Kelly's Criterion (Kelly, 1956). Firstly, a historical xGad dataset should be created, as seen throughout this study. The fixtures in the dataset should also be supplemented by historical odds for each of the matches on a particular outcome that the model is accurate in predicting. Secondly, utilising binary logistic regression, probabilities of the outcome should be calculated, as was conducted in the *Results* section. Next, the odds can be converted to odds-implied probability (OIP) through the equation below.

$$OIP = \frac{1}{Decimal\ Odds}$$

Any cases in which the OIP is higher than the MP should be removed, as this implies that it should not be selected. Finally, the remaining calculated figures are inputted into the formula shown below to determine a maximum stake to place on each remaining selection.

$$Max\ Stake = \frac{(Decimal\ Odds\ \times MP) - 1}{Decimal\ Odds - 1}$$

Kelly (1956)

Given the nature of the historical data, it can be observed whether the event in question occurred or not. To calculate the profit of a winning selection, the formula below can be used.

$$Winning\ Selection\ Profit = (Decimal\ Odds\ \times Stake) - Stake$$

Finally, to calculate the overall profit or loss of the system, the formula below should be used.

$$Overall\ Profit\ /\ Loss = Total\ Winning\ Selection\ Profit - Total\ Losing\ Selection\ Stake$$

Should a profitable method be found, it is important to note how this system can be used in an ongoing season. The EPL, like every domestic European league, runs on a game-week basis. In an EPL season there are 38 game-weeks, in which each team plays one match. Prior to a game-week, the xGad system will have assigned the pre-game ratings to upcoming matches. Utilising a binary logistic regression model containing historic xGad ratings, probabilities can be produced for those upcoming games. These probabilities can then be utilised with the betting strategy used in the previously.

7. Conclusion

This research aimed to construct a dynamic attacking and defensive rating system for EPL teams. This rating system, named xGad, quantified the relative team's ability to score or to not concede, while controlling for their level of performance. Team performance was evaluated using the xG statistic, a powerful metric which can reveal events in a game which are often lost in standard statistics such as shots and goals. The xGad ratings can be utilised as pre-game predictors to produce probabilistic forecasts for goal outcomes in future EPL games.

To evaluate the accuracy of the xGad ratings, binary logistic regression models were constructed, utilising the ratings alone to predict team and total match goals. The dependent variables were converted to a binary form to reflect the betting markets. The primary regression models focused on team goals, which was determined as the simplest form of study for the xGad ratings. $xGad_{HomeAtt}$ and $xGad_{AwayDef}$ were found to be highly significant with whether there were two or more goals in a game. This was a promising finding as these ratings were closely associated with home team goals throughout the xGad system. The away team goals revealed slightly different patterns, in that $xGad_{HomeAtt}$ and $xGad_{AwayAtt}$ were highly significant. These interactions were positive and negatively associated with away goals, respectively. Ideally, $xGad_{HomeDef}$ would have also been significant, as this variable was heavily associated with the away goals. However, it should be noted that the two defensive ratings were close to the level of significance ($P < 0.05$). Both models were used to assess the ratings predictive capabilities and found that models 3 and 4 were 66.71% and 70.82% accurate, respectively. These were impressive figures given that these probabilities were formed only from the xGad ratings. However, for it's use in conjunction with a betting strategy, such as Kelly's Criterion (1956), the predictive accuracy is recommended to be higher.

The ratings were also used to assess their relationship with whether there were three or more total goals in the game. This analysis found that both the attacking ratings were significantly associated, but the defensive ratings lacked significance. It was expected that the ratings would struggle to account for the variation in the total goals variable due to its complexity. This was confirmed by the ROC plot (Figure 10), which highlighted that the model lacked accuracy. The improvements mentioned in the *Diagnostics* section could aid the ratings ability to measure total goals.

The explanations above outline the possible future uses of the xGad ratings. However, it should be emphasized that this is not the aim of this paper. To create a sophisticated statistical model, able to provide accurate probabilistic predictions, more accurate than those produced by bookmakers takes time and extensive research. This paper aims to introduce the xGad system, in the hopes that future research into xGad and other systems will deepen and broaden the academic discussion regarding rating systems in sport. The improvements mentioned in the *Diagnostics* section could be key in improving the accuracy of the xGad system.

8. Bibliography

Ackerson, K., (2022). Football League Rankings. [online] GlobalFootballRankings. Available at: https://www.globalfootballrankings.com [Accessed 2 September 2022].

Audas, R., Dobson, S. and Goddard, J., (2002). The impact of managerial change on team performance in professional sports. Journal of Economics and Business, 54(6), pp.633-650.

Aughey, R., Elias, G., Esmaeili, A., Lazarus, B. and Stewart, A., (2014). Does the recent internal load and strain on players affect match outcome in elite Australian football?. Journal of Science and Medicine in Sport, 19(2), pp.182-186.

Bahtia, A. (2020). Do Football Transfers Actually Improve Performance? Examining the Effect of Football Transfers on Player and Club Performance in Europe, the United States, and China. Bachelor's thesis, Harvard College.

Brechot, M. and Flepp, R., (2020). Dealing With Randomness in Match Outcomes: How to Rethink Performance Evaluation in European Club Football Using Expected Goals. Journal of Sports Economics, 21(4), pp.335-362.

Castellano, J. and Pic, M., (2019). Identification and Preference of Game Styles in LaLiga Associated with Match Outcomes. International Journal of Environmental Research and Public Health, 16(24), p.1-13.

Courneya, K. and Carron, A., (1992). The Home Advantage In Sport Competitions: A Literature Review. Journal of Sport and Exercise Psychology, 14(1), pp.13-27.

da Costa, I., Marinho, L. and Pires, C., (2022). Forecasting football results and exploiting betting markets: The case of "both teams to score". International Journal of Forecasting, 38(3), pp.895-909.

Deutscher, C., Ötting, M., Schneemann, S. and Scholten, H., (2018). The Demand for English Premier League Soccer Betting. Journal of Sports Economics, 20(4), pp.556-579.

Dixon, M. and Coles, S., (1997). Modelling Association Football Scores and Inefficiencies in the Football Betting Market. Journal of the Royal Statistical Society: Series C (Applied Statistics), 46(2), pp.265-280.

FIFA. (2007). FIFA Big Count 2006: 270 Million People Active in Football. [online] Available at: https://digitalhub.fifa.com/m/55621f9fdc8ea7b4/original/mzid0qmguixkcmruvema-pdf.pdf [Accessed 2 September 2022]

FiveThirtyEight. (2022). Club Soccer Predictions. [online] Available at: https://projects.fivethirtyeight.com/soccer-predictions/ [Accessed 2 September 2022].

FiveThirtyEight. (2022). NBA Predictions. [online] Available at: https://projects.fivethirtyeight.com/2022-nba-predictions/ [Accessed 2 September 2022].

FiveThirtyEight. (2022). NFL Predictions. [online] Available at: https://projects.fivethirtyeight.com/2022-nfl-predictions/ [Accessed 2 September 2022].

Goddard, J., (2005). Regression models for forecasting goals and match results in association football. International Journal of Forecasting, 21(2), pp.331-340.

Hägglund, M., Waldén, M., Magnusson, H., Kristenson, K., Bengtsson, H. and Ekstrand, J., (2013). Injuries affect team performance negatively in professional football: an 11-year follow-up of the UEFA Champions League injury study. British Journal of Sports Medicine, 47(12), pp.738-742.

Hoekstra, K., (2021). Who Was Chess Master Arpad Elo, and What is the Elo Rating System?. [online] History Hit. Available at: https://www.historyhit.com/gaming/arpad-elo-rating-system/ [Accessed 2 September 2022].

Hughes, M. and Bartlett, R., (2002). The use of performance indicators in performance analysis. Journal of Sports Sciences, 20(10), pp.739-754.

Iskandaryan, D., Ramos, F., Palinggi, D. and Trilles, S., (2020). The Effect of Weather in Soccer Results: An Approach Using Machine Learning Techniques. Applied Sciences, 10(19).

Jones, R., (2022). 10 most brutal sackings in Premier League history and how they turned out. [online] mirror. Available at: https://www.mirror.co.uk/sport/football/news/10-most-brutal-sackings-premier-20045272 [Accessed 2 September 2022].

Kelly, J., (1956). A new interpretation of information rate. IEEE Transactions on Information Theory, 2(3), pp.185-189.

Lacy, S., (2019). Implementing an Elo rating system for European football | Stuart Lacy. [online] Stuart Lacy. Available at: https://stuartlacy.co.uk/2017/08/31/implementing-an-elo-rating-system-for-european-football/ [Accessed 2 September 2022].

Maher, M., (1982). Modelling association football scores. Statistica Neerlandica, 36(3), pp.109-118.

Mittal, R., (2020). What is an ELO Rating?. [online] Medium. Available at: https://medium.com/purple-theory/what-is-elo-rating-c4eb7a9061e0 [Accessed 2 September 2022

Moroney, M. (1951), Facts from figures, London, Pelican

Opta. (n.d.). Opta Event Definitions - Stats Perform. [online] Available at: https://www.statsperform.com/opta-event-definitions/#:~:text=Expected%20Goals%20(xG)%20measures%20the,defined%20as%20a%20big%20chance. [Accessed 2 September 2022].

Peeters, T. and van Ours, J., (2020). Seasonal Home Advantage in English Professional Football; 1974–2018. De Economist, 169(1), pp.107-126.

Pollard, R., (1986). Home advantage in soccer: A retrospective analysis. Journal of Sports Sciences, 4(3), pp.237-248.

Pollard, R., (2008). Home Advantage in Football: A Current Review of an Unsolved Puzzle. The Open Sports Sciences Journal, 1(1), pp.12-14.

Premier League. (n.d.). Tables. Available at: https://www.premierleague.com/tables. [Accessed 2 September 2022].

PuntersPage. (2022). All Premier League Club Betting Sponsors Listed - 2021/22 Season. [online] Available at: https://www.thepunterspage.com/evolution-premier-league-betting-sponsorship/ [Accessed 2 September 2022].

Rathke, A., (2017). An examination of expected goals and shot efficiency in soccer. Journal of Human Sport and Exercise, 12(Proc2).

Reep, C., Pollard, R. and Benjamin, B. (1971) Skill and chance in ball games. J. R. Statist. Soc. A, 134, 623-629

StatsBomb. (2022). What is xG? How is it calculated? | StatsBomb | Data Champions. [online] Available at: https://statsbomb.com/soccer-metrics/expected-goals-xg-explained/ [Accessed 2 September 2022].

Szwarc, A. (2007). Efficacy of Successful and Unsuccessful Soccer Teams Taking Part in Finals of Champions League, MEDSPORTPRESS, 13(2), 221-225.

TacticsNotAntics. (2022). xG Model Comparison - Understat, StatsBomb & Opta. [online] Available at: https://tacticsnotantics.org/statistical-models-and-analyses/xg-model-comparison/ [Accessed 2 September 2022].

The Analyst. (2022). What Are Expected Goals (xG)?. [online] Available at: https://theanalyst.com/eu/2021/07/what-are-expected-goals-xg/ [Accessed 2 September 2022].

The Economist. (2018). https://www.economist.com/the-economist-explains/2017/07/19/how-data-changed-gambling [online]. [Accessed 2 September 2022].

UEFA. (2019) 2019 UEFA CLUB LICENSING BENCHMARKING REPORT. [online] Available at: https://www.footballbenchmark.com/library/2019_uefa_club_licensing_benchmarking_report . [Accessed 2 September 2022]

Understat. (2022). EPL xG Table and Scorers for the 2021/2022 season | Understat.com. [online] Available at: https://understat.com/league/EPL/2021 [Accessed 2 September 2022].

Understat. (2022). xG stats for teams and players from the TOP European leagues. [online] Available at: https://understat.com/ [Accessed 2 September 2022].

Walker, C., 2022. Prem dominates Bundesliga and La Liga attracting 3.2BN TV viewers. [online] DailyMail. Available at: https://www.dailymail.co.uk/sport/sportsnews/article-10409843/Premier-League-dominates-Bundesliga-Ligue-1-La-Liga-Serie-attracting-3-2-billion-TV-viewers.html [Accessed 2 September 2022].

Wheatcroft, E., (2020). A profitable model for predicting the over/under market in football. International Journal of Forecasting, 36(3), pp.916-932.

World Football Elo (2022). World Football Elo Ratings. [online] Available at: https://www.eloratings.net/ [Accessed 2 September 2022].

WorldPopulationReview. (2022). Most Popular Sport by Country 2022. [online] Available at: https://worldpopulationreview.com/country-rankings/most-popular-sport-by-country [Accessed 2 September 2022].