Freie Universität Berlin

Bachelorarbeit am Institut für Informatik der Freien Universität Berlin,

Arbeitsgruppe Künstliche Intelligenz

Development of a Simulator for the FIFA World Cup 2014

David Dormagen
david [dot] dormagen [at] fu-berlin.de

Supervisor: Raúl Rojas

Berlin, 13.05.2014
Revision 1, 29.05.2014

Abstract

In this paper, a probabilistic simulator for the upcoming FIFA World Cup is developed. For that, we develop a simulation model that can predict the outcome of a match by calculating a win expectancy for both participating teams based on several rules, that take into account different rating methods and can be arbitrarily weighted and combined by the user. The process of deriving the rules' formulae, if not available, from historical FIFA World Cup matches is described as well as the methods for calculating the probability for a draw and the expected number of goals in a match. Further, the resulting application, which does not only include the program to run the simulation, but also a web front end and a back end, which allow straight-forward user interaction with the simulator, is described.

Contents

1 Related Work

There have been different approaches to model the outcome of football games, of a complete tournament or league by using statistical models with different rating methods of teams as their basis. *Liu and Zhang* use multinomial logistic models to predict the results of matches and apply that to the English Premier League, being able to give correct predictions around 60% of times, while a random guess would only yield around 33%[1]. *Karlis and Ntzoufras* use Poisson models, incorporating different parameters of both teams, to make predictions about games in the Greek League[2]. *Dyte and Clarke* use a Poisson model based on the official FIFA rating to model the outcome of world cup games[3]. *Reddy and Movva* suggest that machine learning techniques, especially a gradient boosting model, may be suited to predict football game outcomes[4]. *Dixon and Coles* use a Poisson regression model to predict games in the English league. They review bookmakers' odds and show that their model would yield a positive return if used as the base for a betting strategy[5]. *Hvattum and Arntzen* use logit regression models based on Elo rating and verify their model on games of the English league[6].

The main difference from other simulation models is that the simulation model developed and described in this paper is constructed in a very modular way. Instead of relying on a specific rating method, the model itself works with different ratings. The calculations based on the different rating systems are independent from other calculations such as the calculation of the amount of goals, the probability for a draw, and the home advantage for one team. This is achieved by inferring win expectancy calculation formulae for the different rating methods and then calculating the other results and parameters based on that win expectancy. This approach allows for integration of rating systems and rules where either no clear formula for a probability other than a win or loss exists or where the historical data is not enough to derive such a formula. We are also able to combine the results from different rating methods with user-given weights without influencing other calculations, such as the calculation of the draw-probability, the adjustment of the win expectancy for home teams, or the calculation of the expected goals. In contrast, other approaches have, for example, integrated the home advantage directly into their ratings-based probability model by adjusting the home team's rating points, which is not necessary in this model.

2 Background

This section gives a brief introduction to different rating systems and the tournament format that the FIFA World Cup 2014 uses.

2.1 FIFA World Cup

The FIFA World Cup™[1] is the biggest international football competition, which is held every four years among allotted teams.
The tournament itself consists of two stages: the group phase and the knockout stages leading to the finals. In 2014, 32 teams participate in the world cup. In the group phase, the teams are split into 8 groups and each group of four teams plays a round-robin-tournament[2]. Points are given to the teams based on the number of wins, draws and losses. The top two teams of each group advance into the knockout stage. When two teams are tied in points after the group phase, other factors such as overall goal difference are used as a tie-breaker.
In the following knockout stage, the winning team advances into the next stage while the losing team is out of the tournament. Draws are not permitted and will lead to overtime or a penalty shoot-out.
The group compositions are known beforehand, as well as the knockout stage course of both the winner and the second place of each group.

2.2 FIFA Rating System

The official FIFA rating system tries to rate national teams' strength by giving points for matches relative to the performance of each team[7]. The rating of a team is the weighted average over the team's last years' average match rating[7]. This is a contrast to methods which try to find a stable strength estimation of a participant by adjusting the current estimation after a match, such as the Elo rating system.
The FIFA rating procedure has received a lot of criticism in the past. As described by *Kaminski*[8] the team of the world cup's host nation generally drops in rating during the two years of qualification before the event due to the host nation being disallowed to play qualification games, which would award more points than a normal friendly (i.e. a friendly game between two national teams). *Kaminski* also explains counter-intuitive properties of the FIFA rating system, such as that conceding a game instead of continuing to score goals might result in a higher point reward in some situations. *Dyte and Clarke*[3] showed that manually lowered FIFA ratings for teams that regularly played in lesser known federations resulted in a more accurate prediction compared to the actual ratings. This further suggests that there

[1]The FIFA World Cup is a trademark of FIFA
[2]In a round-robin-tournament, each team plays every other team exactly once.

may be inconsistencies in the official FIFA rating when used as an estimation of a team's strength.

2.3 Elo Rating System

Originally developed by the Hungarian physicist Arpad Elo for assessing the strength of chess players[9], the Elo rating system has found widespread use in different types of sport, including association football[6][10]. In the Elo rating system, each team is assigned a rating, which is then updated whenever the team partakes in a match based on the outcome. A win will generally increase a teams rating while a loss generally lowers it, depending on the Elo rating of the opponent. Defeating a higher rated opponent will result in a higher gain in Elo rating than defeating a weaker opponent. In an analogous manner, a loss against a weaker opponent will cost more Elo rating.
While there have been several proposals to change the Elo rating system to reflect a team's skill more accurately, the original method is still used a lot. *Lasek and Szlávik*[11] found that calculating a win probability using different Elo-based methods for a match in association football had a higher accuracy than other rating systems such as the official FIFA rating.

2.4 Soccer Power Index

The Soccer Power Index, abbreviated SPI, is a rating system that was developed by the statistician Nate Silver to rate both a team's offensive and defensive strength[12]. Similar to the Elo rating, the SPI continually adjusts a team's ratings based on its performance in a match. It also allows rating the individual players based on their performance and the overall outcome of the match. The SPI takes more factors into account than f.e. the FIFA rating when calculating the new score after a match, such as the number of goals and the home advantage.

2.5 Market Values

Whenever a player transfers to a different club, large amounts of money change the owner as well. How much money clubs are willing to pay for a player strongly depends on their expectation towards the player's skill and their prediction of how much the single player will benefit the team. It thus seems plausible that, through the sum of the most recent transfer-values of all players from a team, the team's strength can be estimated as well. Unless otherwise stated, the *market value* of a team describes the sum of all the single market values of the team's players or, for better comparison, an average of this sum. Other properties of the team are not taken into account (f.e. who coaches the team). The main source for all market values used is *transfermarkt.de*[13].

3 The Simulation Model

This paragraph describes the simulation model that is used to simulate the tournament. The main focus will be on the execution of a single match, since the general structure of the tournament and its rules are well-known. The simulation of one match is executed as follows: as the first step, all N active rules (except for the home advantage rule, see below) are applied to the teams' respective scores yielding a win expectancy $p \in [0, 1]$ each for both teams. The probability $P'_{VICTORY_i}$ is the weighted sum over the rules' probabilities with the rules' weights w, which had been selected by the user:

$$P'_{VICTORY_i} = \frac{\sum_{i=1}^{N} p_i * w_i}{\sum_{i=1}^{N} w_i}$$

Then, if active with a weight $w > 0$, the home advantage rule $ha(p)$ is applied to $P'_{VICTORY_{1/2}}$.

$$P_{VICTORY_i} = (\frac{w}{w_{max}}) * ha(P'_{VICTORY_i}) + (1 - \frac{w}{w_{max}}) * P'_{VICTORY_i}$$

, where w is the user-selected weight of the home advantage rule, and w_{max} is the maximum weight of all active rules (including the home advantage rule).
Note that if $w = w_{max}$, only the adjusted win expectancy of the home advantage rule is used.

A description of the formulae used for the different rules can be found in the section 3.2.

After calculating the win expectancy, the chance for a draw P_{DRAW} is calculated based on the win expectancy $P_{VICTORY_1}$ (see 3.3). A uniformly random number $roll \in [0, 1]$ is rolled and the game is said to be a draw if $roll < P_{DRAW}$.

If the game is not a draw, a winner is chosen by rolling a uniformly random number $roll \in [0, 1]$. If $roll < P_{VICTORY_1}$, the first team is selected to be the winner. If $roll \geq P_{VICTORY_1}$, the second team is the winner.

Then the number of goals for each side are rolled using a Poisson-distributed random number generator with $\lambda_{1/2}$ calculated from $P_{VICTORY_{1/2}}$ respectively, see 3.4. The rolls for both sides are repeated until the result is either a draw or the previously selected winner has more goals.

3.1 Description of the Data Set

The match data that has been used to derive the formulae and probabilities used in the simulation comes from the match archive of the official FIFA website [14]. For matches that had a penalty shootout, the number of goals prior to the shootout has been used for analysis.
The data set thus consists of every match for all world cups starting from 1930. For every match, the number of goals for both teams at the official end of the match is included. The number of goals are used to decide which team won the match. The data set contains over 700 official FIFA world cup matches. However, it must be noted that for certain rating methods information about the rating points for each team was not available and thus only a subset of those matches could be used for analysis.

3.2 Description of the Rules

This section describes the different rules that can be used to calculate the winning probability for a match between two teams.

3.2.1 Elo Rating

The Elo rule uses the football Elo rating points from the World Elo Ratings-website [15]. The Elo system was invented for chess and there is a widely accepted formula for calculating the winning probability of two contestants via their respective Elo rating points in chess, which is also used as the basis for the winning probability in the simulation.
Unfortunately the historical Elo rating points of the football teams participating in the current world cup are not easily found, leading to the data for evaluation of the validity of the Elo rating for predicting the outcome of football matches in a world cup to be sparse. The data mainly contains the FIFA matches from only 2010.
For the significance check, all FIFA matches with teams, that had Elo rating support for the time of the match, were grouped into categories C_i of Elo rating difference for every 50 points of difference. C_1 contains all matches where the difference in Elo rating of the two teams $r_1 - r_2 \in (-600, -550]$, C_2 contains the range $(-550, 500]$ and so forth. For those categories the percentage of victories could be calculated.
For plausibility, a regression model $P_{VICTORY} = \frac{1}{1+10^{-RatingDifference/c}}$ has been fitted to the match data via least squares fitting, yielding $c = 444$, see Figure 1. However, for the simulation the original Elo formula

$$P_{VICTORY_1} = \frac{1}{1 + 10^{-\frac{r_1 - r_2}{400}}}$$

, where r_1 and r_2 are the Elo ranking points of the respective teams, has been used for the sake of consistency.

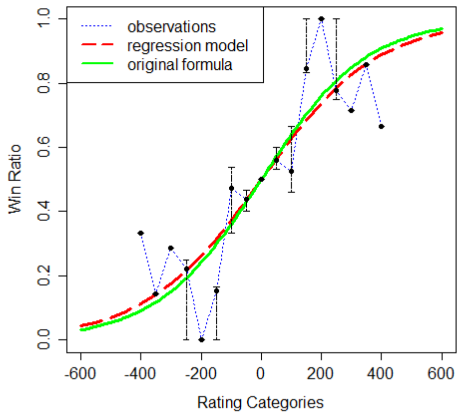

Figure 1: The graph shows the relation between difference in Elo rating and outcome of a match. The original Elo formula $P_{VICTORY} = \frac{1}{1+10^{(-RatingDifference/400)}}$ to calculate a winning probability has been depicted in the graph as well as the results of a regression model, suggesting the use of 444 instead of 400 as the constant. The error bars indicate the minimum and maximum values over the different years.

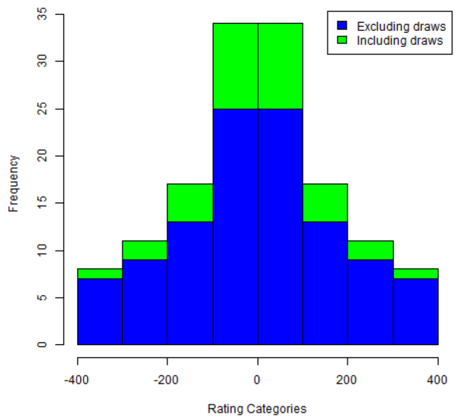

Figure 2: The histogram shows the distribution of the difference in Elo rating in all analysed matches.

3.2.2 FIFA Rating

The FIFA rule uses the FIFA ranking points that are given to football teams by the FIFA, taking into account parameters like the match outcome, the opposing team's strength, the importance of the match and strength of the confederation (in case teams from different confederations meet)[7]. While there are ranking points available for every world cup starting from the one in 1994, only the games from 2006 and 2010 could be used for analysis, since the official ranking method changed in 2006, leading to a very different point distribution compared to the years before[16].
There is no standard formula to calculate the win expectancy of two opposing teams based on the FIFA ranking points.
The approach to finding a formula that could be used in the simulation used the difference of the two teams' ranking points $r_1 - r_2$, similar to the calculation in the Elo system.
Expecting a sigmoid function to be most suited to estimate the win expectancy from the difference of two teams' ranking points, the model $P_{VICTORY} = \frac{1}{1+\exp(-RatingDifference/c)}$ has been fitted to the match data via least squares fitting, see Figure 3. The regression yielded the value $c = 291.5$. Thus, the final formula to calculate the win expectancy of two teams based on the FIFA ranking points, which is now used in the simulation, is

$$P_{VICTORY_1} = \frac{1}{1 + \exp\left(-\frac{r_1 - r_2}{291.5}\right)}$$

, where r_1 and r_2 are the FIFA ranking points of the respective teams.

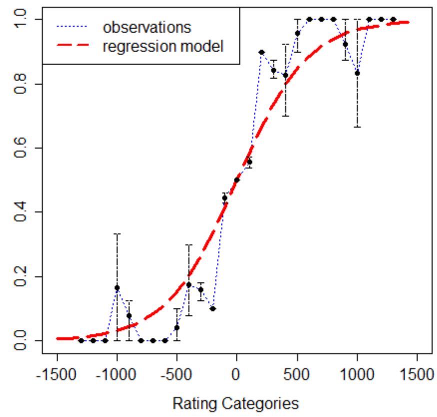

Figure 3: The graph shows the relation between difference of FIFA rating and outcome of a match. The results of a regression model, suggesting the formula $P_{VICTORY} = \frac{1}{1+\exp(-RatingDifference/291.5)}$ as the best approximation, has been depicted. The error bars indicate the minimum and maximum values over the different years.

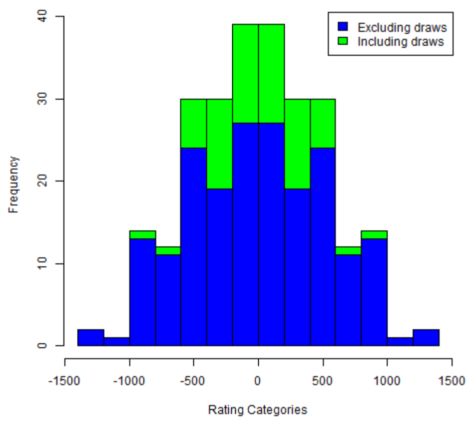

Figure 4: The histogram shows the distribution of the difference in FIFA rating in all analysed matches.

3.2.3 Soccer Power Index

The SPI rule uses both the offensive and defensive rating from the SPI to calculate a win expectancy for a team.
While an intention of the SPI is to be able to calculate the win expectancy for both teams in a match, there is sadly no public official formula and instead the official FAQ reads[17]:

> ”The OFF and DEF ratings for any two teams can be combined to create a prediction about the teams’ chances to win or draw the game. This uses a statistical technique called multiple logit regression; basically, we examine a database of thousands of past games to see what happened when teams with similar OFF and DEF ratings faced one another.”

Due to a lack of this database, an own method to calculate a team’s win expectancy based on the SPI’s offensive and defensive ratings for both teams had to be established. Since there was also not enough historical data available to deduce a formula similar to other rating systems with parameters estimated by non-linear regression, it was decided to calculate the teams’ win expectancy by modelling each team’s goals using a Poisson-distributed random variable with λ calculated from both teams offensive and defensive rating.
The offensive and defensive ratings in the SPI indicate how many goals a team would score and receive on average when a round-robin-tournament between all nations was held. Thus, it does not seem implausible to calculate the win expectancy indirectly through the average number of goals between two teams - especially since we will see later in section 3.3 that a Poisson distribution is indeed a plausible estimation of a team’s goals in a match.
The procedure is as follows:
Firstly, for both teams $Team_{1/2}$ with respective offensive and defensive ratings $r_{OFF1/2}$ and $r_{DEF1/2}$ the combined offensive r_{OFF} and defensive r_{DEF} ratings are calculated

$$r_{OFF} = \frac{r_{OFF1} + r_{DEF2}}{2}$$

$$r_{DEF} = \frac{r_{DEF1} + r_{OFF2}}{2}$$

Then the probabilities $P_{1/2}$ for both teams to win the game are calculated, disregarding draws:

$$P_1 = \sum_{i=0}^{\infty} \left[\left(\frac{e^{-r_{OFF}} * (r_{OFF})^i}{!i}\right) * \sum_{j=0}^{i-1} \left[\left(\frac{e^{-r_{DEF}} * (r_{DEF})^j}{!j}\right) \right] \right]$$

$$P_2 = \sum_{i=0}^{\infty} \left[(\frac{e^{-r_{DEF}} * (r_{DEF})^i}{!i}) * \sum_{j=0}^{i-1} \left[(\frac{e^{-r_{OFF}} * (r_{OFF})^j}{!j}) \right] \right]$$

Since draws have been disregarded before, the sum of P_1 and P_2 will now be smaller than 1. The final step is to normalize the probabilities which results in the following probability $P_{VICTORY_1}$ that $Team_1$ will win:

$$P_{VICTORY_1} = \frac{p_1}{p_1 + p_2}$$

3.2.4 Monetary Value

A team's monetary value is the sum of the team's players' current market values. For the simulation, the average market value of the players has been chosen over the sum, because it made analysing historical data easier, due to market values for some single players not being available. Every team has been taken into account for the analysis where at least ten players had market values available. The values come from the individual player profiles on transfermarkt.de[13].

As the basis for the comparison of two teams, the logarithmic quotient $\ln(\frac{x}{y}) = \ln(x) - \ln(y)$ of the teams' market values $r_{1/2}$ has been used. This provides some stability over multiple years compared to using the difference, because the actual buying power of money changes over time, changing the meaning of a fixed monetary margin between two teams. The ratio of two teams' market values in the same year should, however, show a more stable relationship.

The sigmoid regression model $P_{VICTORY} = \frac{1}{1+\exp(-RatingQuotient/c)}$ has been fitted to the match data via least squares fitting, see Figure 5. The regression yielded the value $c = 1.3026$. The final formula to calculate the win expectancy of two teams based on their average monetary value, that is used in the simulation, is

$$P_{VICTORY_1} = \frac{1}{1 + \exp\left(-\frac{\ln(\frac{r_1}{r_2})}{1.3026}\right)}$$

, where r_1 and r_2 are the average value of the respective teams' players in Euro.

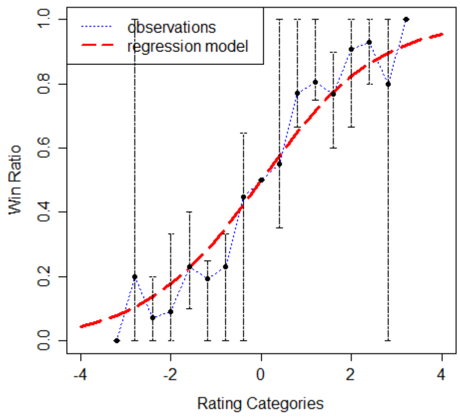

Figure 5: The graph shows the relation between logarithmic quotient of the average monetary value of two teams' players. The results of a regression model, suggesting the formula $P_{VICTORY} = \frac{1}{1+\exp(-RatingQuotient/1.3026)}$ as the best approximation, has been depicted. The error bars indicate the minimum and maximum values over the different years.

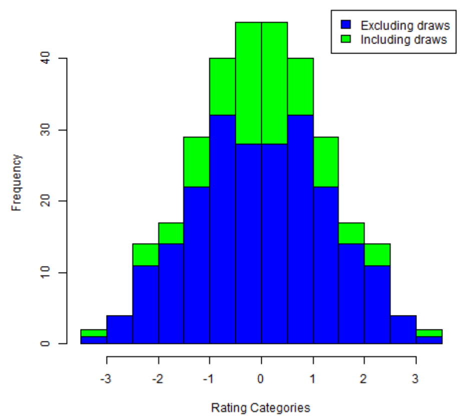

Figure 6: The histogram shows the distribution of the average monetary value of two teams' players in all analysed matches.

3.2.5 Average Age

To analyse the influence of a teams' average age, the average age for each team and year has been calculated using the individual player profiles on transfermarkt.de[18].

As the basis for the comparison of two teams, their difference in average age is used. Since a simple sigmoid function, similar to the ones used in the other rating methods, did not seem to model the data well enough, a more complex regression model $P_{VICTORY} = \frac{1}{1+\exp(-(a*age\_difference - age\_difference^3)/c)}$ has been used and was fitted to the data via least squares fitting, yielding $c \sim 131$ and $a \sim 32$, see Figure 7.

The model suggests that a team's optimal average age is three years older than their opponent's. The final formula to calculate the win expectancy of two teams based on their average age, that is used in the simulation, is

$$P_{VICTORY} = \frac{1}{1 + \exp\left(-\frac{32*(r_1-r_2)-(r_1-r_2)^3}{131}\right)}$$

, where r_1 and r_2 are the average age of the respective teams' players.

Even though, the number of matches in the age categories outside of the range $(-4, +4)$ was only relatively low (see Figure 8) and thus the statistical reliability is drastically reduced, it was decided to keep the formula as described above. This was done not only due to the lack of an alternative, but also because an explanation for the current curve can be given: while an age advantage of around three years with the higher experience of the players increases the team's winning expectancy, a much higher age will reduce the team's win expectancy due to a worse personal fitness in comparison to the younger opponents.

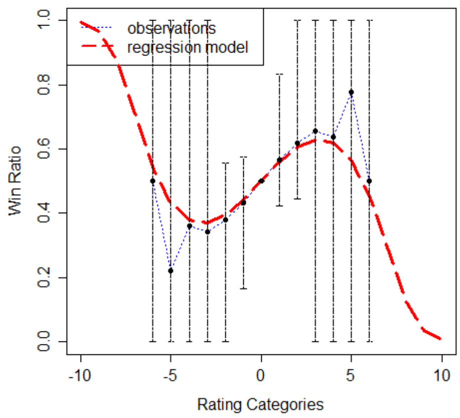

Figure 7: This graph shows the relation between difference in average age of two teams and outcome of a match. The results of a regression model, suggesting the formula $P_{VICTORY} = \frac{1}{1+\exp(-(32*age\_difference+age\_difference^3)/131)}$ as the best approximation, has been depicted. The error bars indicate the minimum and maximum values over the different years. The high error range displays the low significance of the age as the deciding factor for matches.

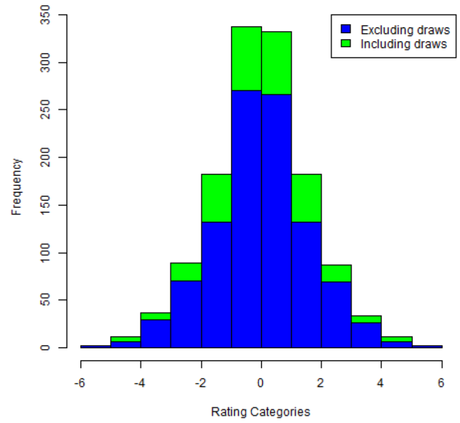

Figure 8: The histogram shows the distribution of the difference in average age of two teams in all analysed matches.

3.2.6 Home Advantage

The home advantage rule uses the property of the teams of being a host nation or not to adjust the win expectancy that was calculated by other rules. When having a look at all of the FIFA world cup matches from 1930 to 2010, teams with a home advantage have a win ratio of 63% when including draws and 75% when omitting draws, see Table 1.

However, this perspective alone does not allow us to deduce a statistically significant formula for the win expectancy yet. One reason for that is, that the number of games with participating host nations per year is only relatively small, making significant conclusions hard. Additionally, the data might be biased if the host nations are generally stronger teams and thus naturally have a higher win expectancy. To reduce that bias, the win rations of host-nation teams have been compared to their win expectancy as calculated via their rating (f.e. FIFA ranking points or Elo rating). Even when accounting for the expected win probability for the different matches, host nations appear to have a higher win ratio than their rating alone would suggest, see Figure 9.
Analysing games in Dutch football from 1990 to 2005, P.C. van der Kruit [19] found that home-teams won around 63% as opposed to the expected 50%, when not counting draws. For the simulation, the average win expectancy for host-nations is increased to $\frac{2}{3}$ from the standard $\frac{1}{2}$, not counting draws. Expecting a sigmoid curve to be the best approximation for the necessary adjustment of the win expectancy p in regards to the home advantage of the teams, the following formula is used in the simulation:

$$p' = \frac{2}{1 + e^{-4*p}} - 1$$

, which satisfies:

$$\int_0^1 \left(\frac{2}{1 + e^{-4*p}} - 1 \right) \mathrm{d}p = \frac{1}{2} \ln \left(\frac{1}{2(1 + e^4)} - 2 \right) \approx 0.66$$

In case both teams have an assigned home advantage $ha_{1/2}$ or the sum of the home advantage of both teams is below 1, the simulation adjusts the weight of the result of the home advantage rule as follows for every team $t \in \{1, 2\}$:

$$w'_t = w * \frac{ha_t}{max(1, ha_1 + ha_2)}$$

, where w is the user-selected weight for the rule.

Year	H.-A.	H.-A. no draws	Other
1930	1.00	1.00	0.50
1934	0.80	1.00	0.50
1938	0.50	0.50	0.41
1950	0.67	0.80	0.44
1954	0.50	0.50	0.45
1958	0.67	0.80	0.34
1962	0.67	0.67	0.40
1966	0.83	1.00	0.42
1970	0.50	0.67	0.43
1974	0.86	0.86	0.34
1978	0.71	0.83	0.37
1982	0.20	0.33	0.34
1986	0.60	1.00	0.37
1990	0.86	1.00	0.38
1994	0.25	0.33	0.40
1998	0.86	1.00	0.34
2002	0.50	0.67	0.38
2006	0.71	0.83	0.38
2010	0.33	0.50	0.38
mean	0.63	0.75	0.40

Table 1: The table shows the win ratios of teams over the years. The second column shows the ratios for teams with a home advantage (being host nations), the third row shows the same data but without counting draws, and the fourth row shows all other games, including draws, without a participating host nation.
An important remark: the data may not be statistically reliable, due to the low number of games with host nations per year.

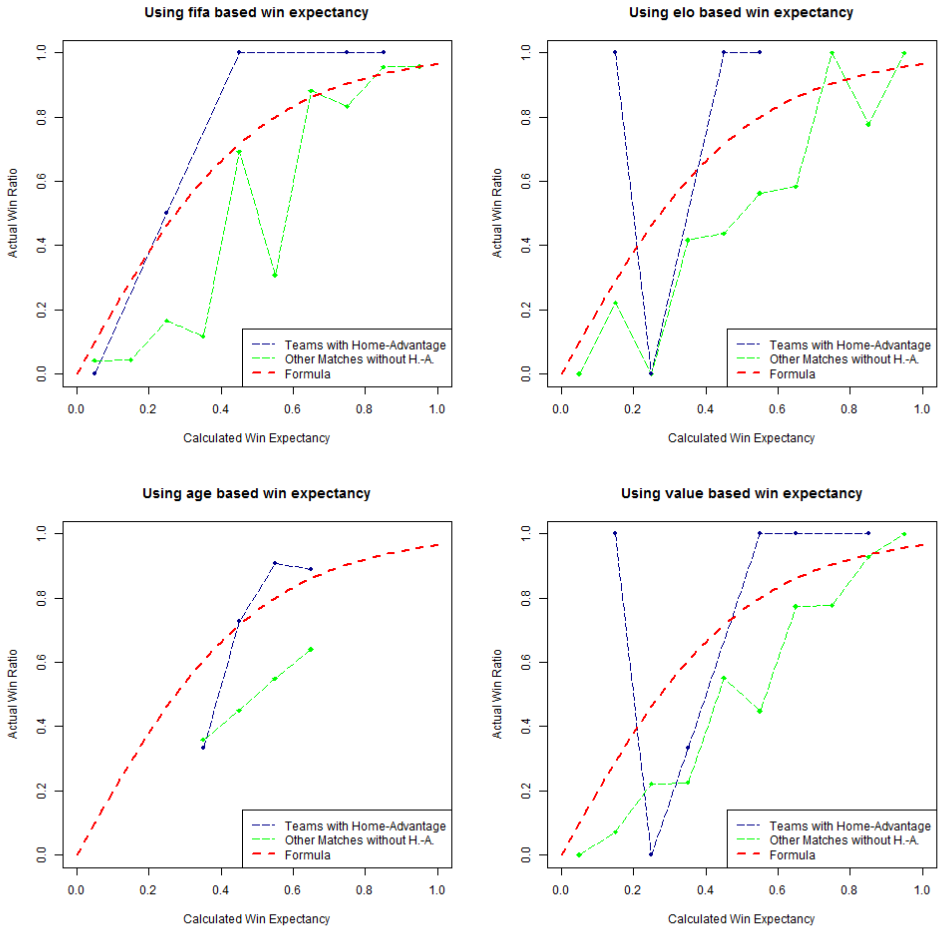

Figure 9: Comparing the calculated win expectancies of the host nations' teams with their actual win ratio shows, that host nations are generally more likely to win than their ranking would suggest. The dashed red line depicts the formula to adjust the win expectancy based on the home advantage of the teams $f(p) = \frac{2}{1+e^{-4*p}} - 1$, which is used in the simulation.

3.3 Draw Probability

For the simulation, a formula that gives the probability for a draw based on the win expectancy of the participating teams is needed. Separating the calculation of the win expectancy from the calculation of draws is necessary not only because some calculation methods might not directly yield the probability for draws, but also to freely allow mixing the results of the rules that calculate the win expectancy without influencing the calculation of the draw probability in an unwanted way. To derive the formula for the calculation of the draw probability, the historical FIFA matches [14] have been analysed *after* the method of calculating the win expectancy from teams' respective ratings was established. The data thus consisted of matches and the assigned win expectancy p, making it straight-forward to find a correlation between the two variables.
Expecting a Gaussian distribution, the regression model $a * \exp(-\frac{(p-0.5)^2}{2*c^2})$ has been fitted to the data using least squares fitting. The results vary slightly depending on the underlying rating system (f.e. Elo rating or FIFA ranking points), see Table 2 and Figure 10.
The value $\frac{1}{3}$ for a has been chosen, because not only is it the mean of the observed values but it also reflects the idea that in an equal game every outcome (win, loss, and draw) should have an equal probability, which also seems to be reflected by the results of the regression model. 2.8 has been chosen for c as the mean of the model results. The final formula to calculate the probability for draws based on the win expectancy thus is:

$$P_{DRAW} = \frac{1}{3} * \exp(-\frac{(P_{VICTORY} - 0.5)^2}{2 * 0.28^2})$$

, where $P_{VICTORY}$ is the win expectancy of one of the teams.
Since the formula is symmetric, the results are the same from each of the teams' perspective: $P_{DRAW}(P_{VICTORY}) = P_{DRAW}(1 - P_{VICTORY})$.

Parameter	Est. (FIFA)	Est. (ELO)	Est. (Value)	Chosen Values
a	0.37994	0.29736	0.31979	$\frac{1}{3}$
c	0.29234	0.26339	0.29036	2.8

Table 2: The table shows the parameter estimations resulting from regression models based on the data for different ranking methods as well as the finally selected parameter values.

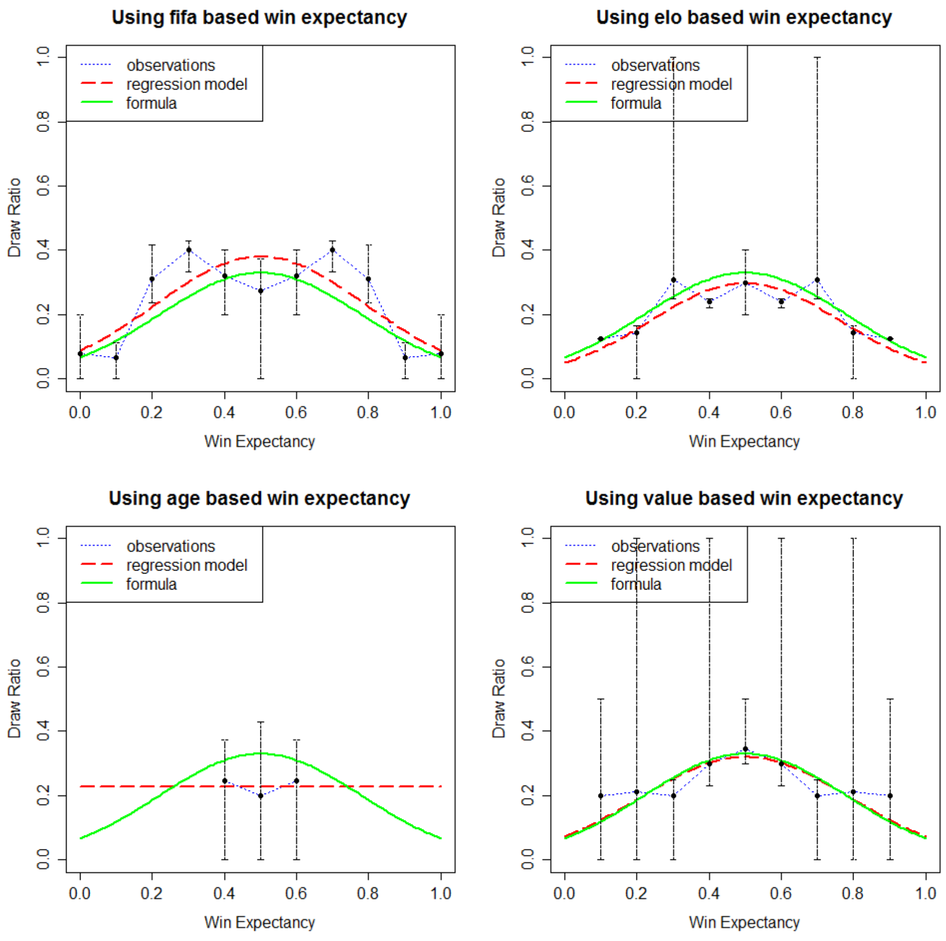

Figure 10: This graph shows the relation between win expectancy based on different rankings and the probability of draws. The result of the regression models for the different scores have been depicted as well as the final distribution formula that is used in the simulation. The error bars indicate the minimum and maximum values over the different years.

3.4 Number of Goals

Since the simulation also simulates the goals for both teams in every match, a model for the goal distribution based on the win expectancy of a team is needed. Again, the historical FIFA matches [14] have been analysed *after* the methods of calculating the win expectancy from teams' respective ratings were established.
The data used for deriving the goal distribution thus consisted of matches with win expectancy and goal amount for both teams. As expected the average amount of goals for a team increases the higher the probability to win the match is, see Figure 11.
To find out the underlying distribution, the data has been divided into win-expectancy-categories C_i, with C_1 containing games with a win expectancy in $(0, 0.1)$, C_2 containing $(0.1, 0.2)$, etc.
For each category C_i, the frequency of every number of goals was calculated. In expectation of a Poisson distribution, the regression model $P_{C_i}(G) = \frac{\lambda^G * \exp(-\lambda)}{!G}$ has been fitted to the matches in the category using least squares fitting. The resulting values for λ can be seen in Table 3.
To find a formula for the λ-parameter and in expectation of a linear correlation, the regression model $\lambda(p) = m * p + b$ has been fitted to the estimated λ-parameters using the centres of the categories as the x-parameters, see Figure 12.
Following the results of the regression models, the formula that is used in the simulation to calculate the λ-parameter for the goal distribution based on the win expectancy p of a team is:

$$\lambda(p) = 1.8 * p + 0.27$$

Figure 13 shows the Poisson-distributed model for the number of goals with the λ chosen as described above in comparison to the observed distribution with the win expectancy based on the FIFA ranking points.

It has to be noted that according to [20] the goals in a football match are not perfectly Poisson-distributed. A team, that already scored goals, seems to be more likely to score further goals. This is attributed to an effect of self-affirmation. However, for simplicity, the pure Poisson model is used for this simulation.

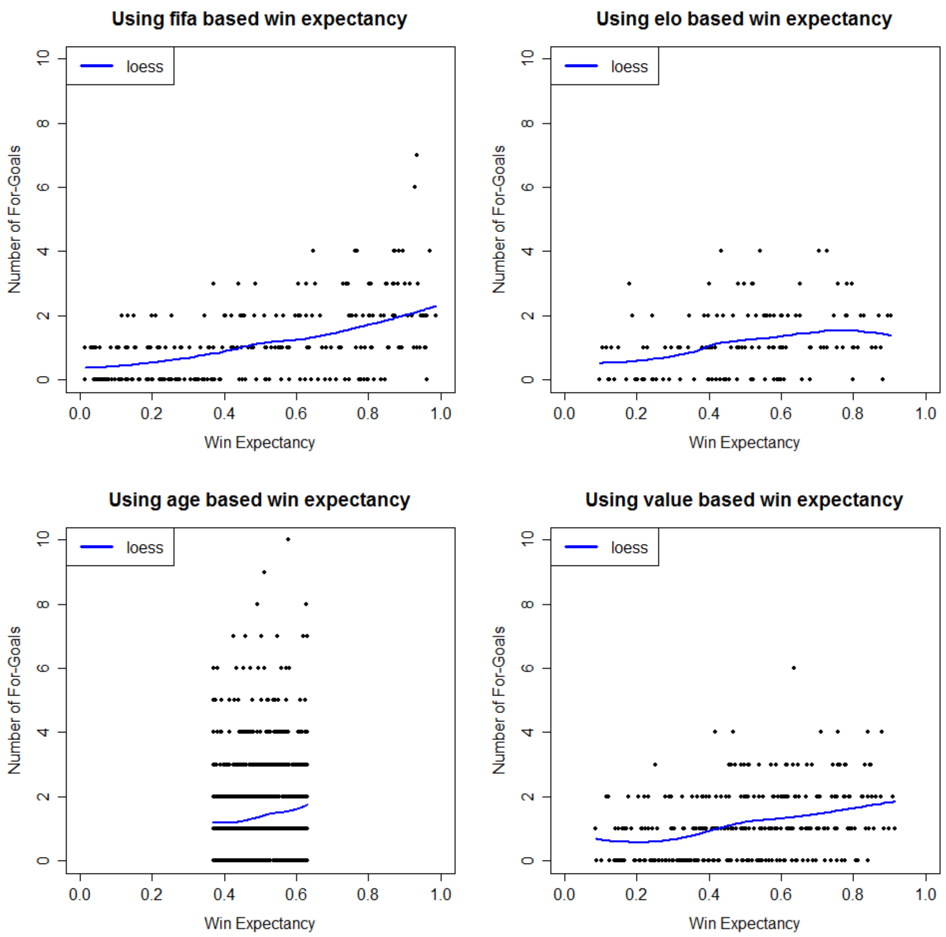

Figure 11: The graphs show the available data that has been used to derive the goal distribution per win expectancy. As expected, a relationship between win expectancy and number of goals can be seen.

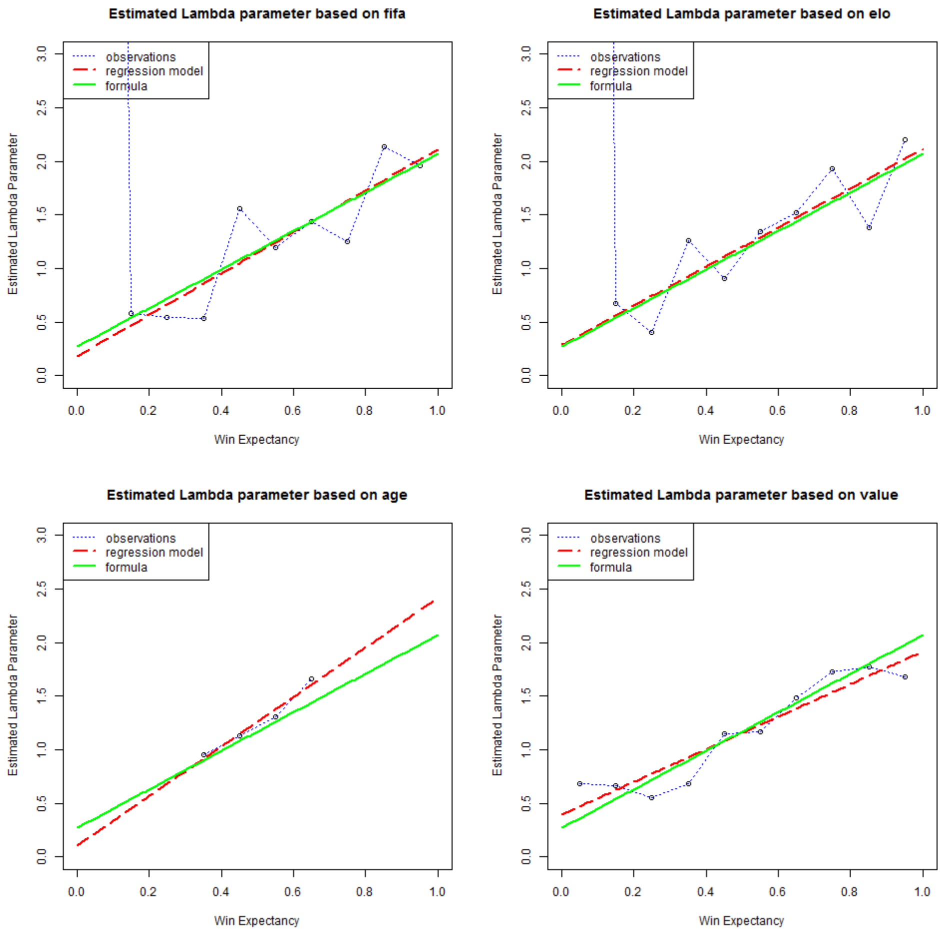

Figure 12: The graphs show the estimated λ-parameters of the Poisson distributed number of goals per win-expectancy-category. A linear regression model has been fitted to the data. The formula depicted in the graph is $\lambda(p) = 1.8 * p + 0.27$. For the expectancies based on Elo and FIFA, the first data point has been disregarded as an outlier.

Category	λ (FIFA)	λ (Elo)	λ (value)	λ (age)
0 - 0.1	29.900	29.600	0.679	
0.1 - 0.2	0.583	0.677	0.659	
0.2 - 0.3	0.540	0.407	0.551	
0.3 - 0.4	0.533	1.260	0.686	0.951
0.4 - 0.5	1.560	0.906	1.150	1.130
0.5 - 0.6	1.200	1.340	1.170	1.310
0.6 - 0.7	1.440	1.520	1.490	1.660
0.7 - 0.8	1.250	1.930	1.730	
0.8 - 0.9	2.130	1.380	1.780	
0.9 - 1	1.960	2.200	1.680	

Table 3: The table shows the estimations for the λ-parameter of a Poisson distribution for the number of goals resulting from regression models based on the data for different ranking methods.

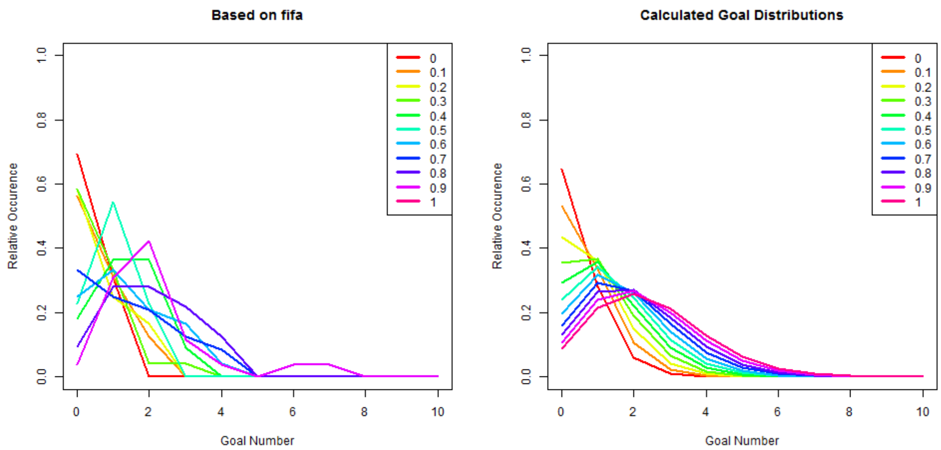

Figure 13: The left graph shows the observed goal distribution in the FIFA matches for different win expectancies based on FIFA ranking points. The right side shows the model with a Poisson distribution as used in the simulation.

Why the number of goals is not the only characteristical number used

When a valid formula for the number of goals based on the win expectancy of two teams is established, the separate random roll for a draw and for the winner of the game could be omitted, in theory. The only random number to be rolled would be the number of goals for each side, leading to draws and a winner as a native result of the rolls.
From the Poisson distributed number of goals, the chance of a draw can be calculated from the winning expectancy p of one of the teams as follows:

$$P'_{DRAW} = \sum_{i=0}^{\infty} \left[(\frac{e^{-\lambda(p)} * \lambda(p)^i}{!i}) * (\frac{e^{-\lambda(1-p)} * \lambda(1-p)^i}{!i}) \right]$$

where $\lambda(p) = 1.8 * p + 0.27$.
The probability for a team with win expectancy p to score more goals than the opposing team and thus, to win the game, can be calculated similarly:

$$P'_{VICTORY} = \sum_{i=0}^{\infty} \left[(\frac{e^{-\lambda(p)} * \lambda(p)^i}{!i}) * \sum_{j=0}^{i-1} \left[(\frac{e^{-\lambda(1-p)} * \lambda(1-p)^j}{!j}) \right] \right]$$

where $\lambda(p) = 1.8 * p + 0.27$.
However, comparing that with the originally derived formula for P_{DRAW} and the original win expectancy p, there is a discrepancy as can be seen in Figure 14. For that reason, the originally derived formulas were kept as the foundation to decide the result of a match. That way, in case the calculation of goals turns out to be inaccurate, the remaining model will not be influenced.

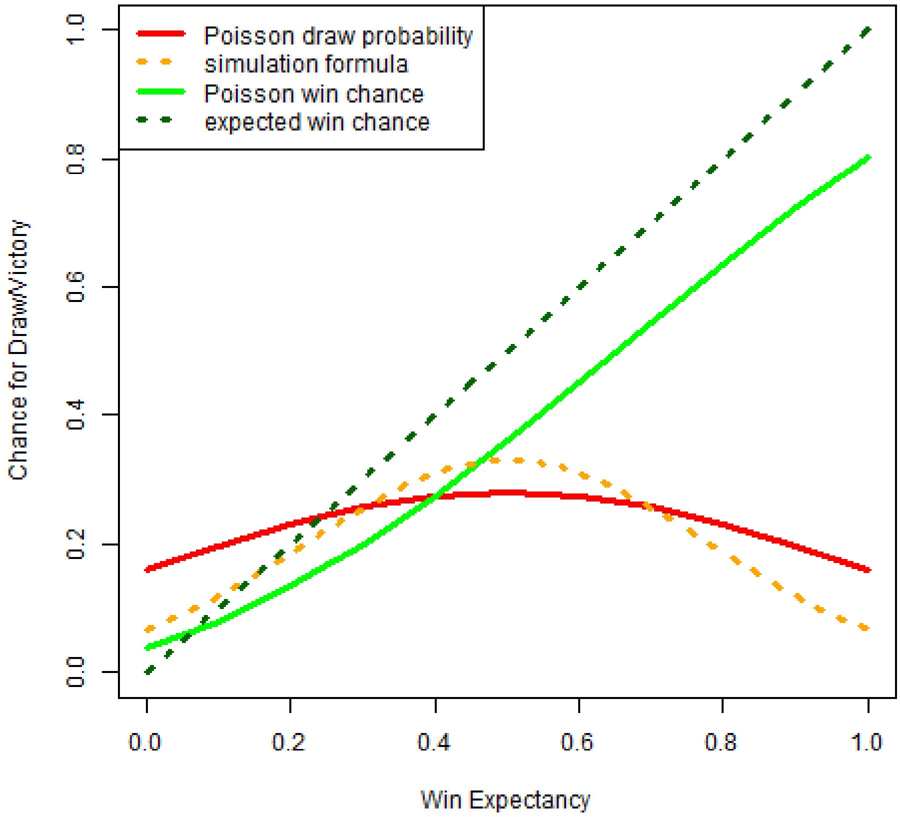

Figure 14: The graph compares the outcomes of a match that were calculated from the Poisson-distributed goal number (solid lines) to the expected results (dashed lines).

3.5 Choosing the default weights

For the simulation, a set of default weights for the rules is needed, because the win expectancy for a match will be calculated from a combination of different rules which have certain weights assigned. The default weights will be presented to the user prior to any customization. Ideally, those defaults should be chosen in a way to best reflect the reliability of the rules and give the user a good basis to start from.

The weight of the age rule has been set to 0, because it appears to have the lowest significance of all rules. Similarly, the weight of the value rule is set to only 0.25, because it is not a ranking method per se and would need further analysis to make reliable statements whether it is suited for prediction.

The weights of the Elo and the SPI rule are both set to 1.0. The weight of the FIFA rule is set to only 0.5, because of known weaknesses as described in section 2.2.

This results in a combined win expectancy, that is hopefully a lot more resistant against outliers in the different rating methods. Figure 15 shows the correlation between the different win expectancies based on the different rating systems. Note that not the rating points from the different rating systems are the basis for this correlation, but the win expectancies based on those rating points. The correlated win expectancies come from a simulated round-robin-tournament between all nations participating in the world cup 2014. Except for the calculation based on the average age of a team, all rating systems show clear correlation. However, especially in the plot that compares the win expectancy based on FIFA ranking points and Elo rating, outliers can be seen. The combined win expectancy has a correlation value of at least 0.9 with the calculations based on Elo, FIFA ranking points and SPI.

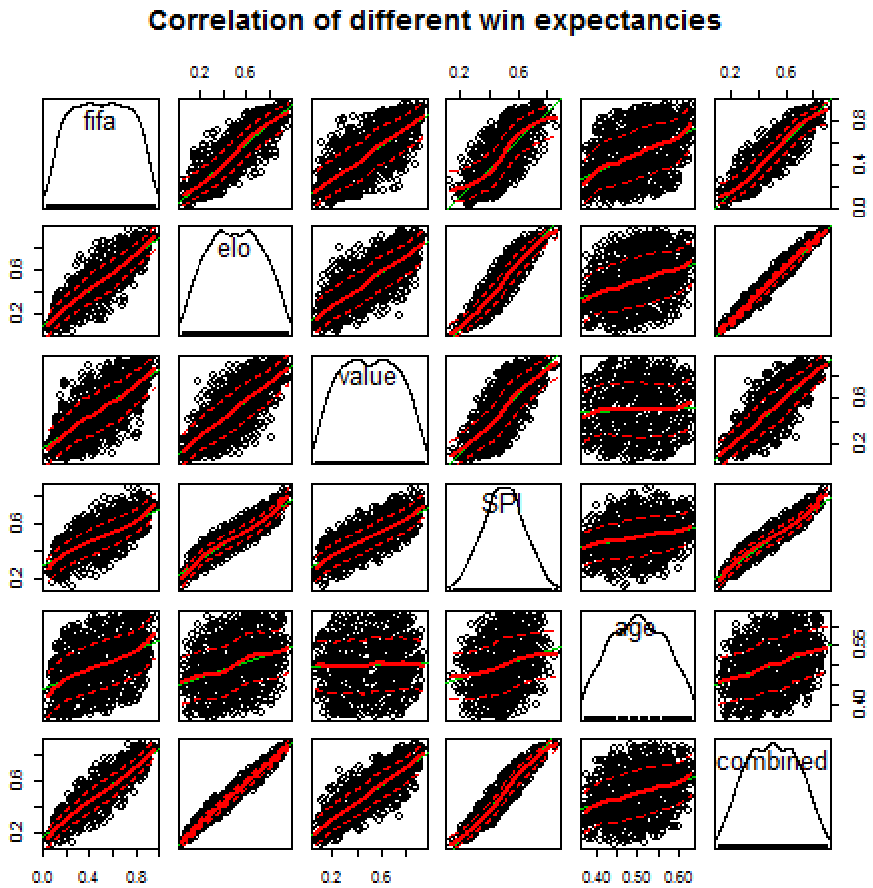

Figure 15: The plot shows the correlation between the win expectancies calculated from different ratings. A round-robin-tournament between all nations that participate in the world cup 2014 has been used as the basis for the calculation.

expectancies	elo	fifa	value	SPI	age	combined
elo	1.00	0.83	0.75	0.91	0.33	0.98
fifa	0.83	1.00	0.70	0.72	0.46	0.90
value	0.75	0.70	1.00	0.76	0.04	0.82
SPI	0.91	0.72	0.76	1.00	0.24	0.93
age	0.33	0.46	0.04	0.24	1.00	0.34
combined	0.98	0.90	0.82	0.93	0.34	1.00

Table 4: The table shows the correlation of different win expectancies based on a round-robin-tournament between all nations participating in the world-cup 2014.

4 The Software

The software, which was created during this thesis, mainly consists of three parts: the web-*frontend* that consists of the user-visible parts of the software, the *backend* that cares for the management and website-generation, and the *simulator* which runs the actual simulation and calculates the statistic results.
The whole application is written with scalability and flexibility in mind, as one requirement was, that future adjustments in any direction should be easy to make. The different parts of the application communicate over clearly-defined and simple interfaces, see Figure 16. The following subsections explain the different parts of the simulator application.

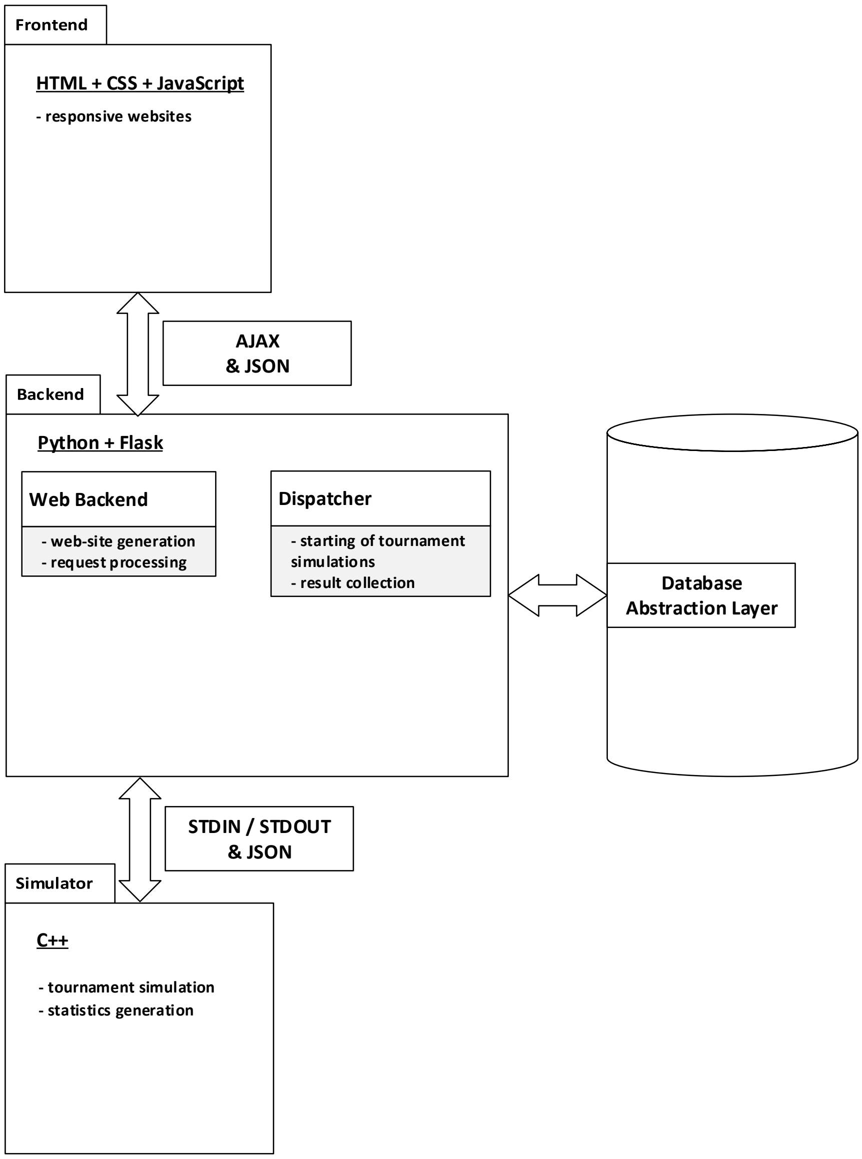

Figure 16: The diagram shows the general setup of the created application. The different main parts of the application are outlined as well as the interfaces between those parts.

4.1 Frontend

The frontend is the part of the application, the user can interact with. It consists of several web pages that allow the user to interact with the backend and, for example, start new simulations.
The web pages are generated dynamically by the *backend* from HTML templates. For user-interaction and general responsiveness, JavaScript is used in combination with dynamic Ajax requests that allow for a fluent user interface.
JavaScript libraries that are used for an improved user experience are:

JQuery The JQuery library allows for easy DOM traversal in HTML documents, event handling, and animations[21].

ChartJS The ChartJS library allows generation of charts and diagrams via JavaScript[22].

As a CSS library, Bootstrap[23] is used, which allows for a responsive and flexible design, that works well with both desktop and mobile clients.

The web pages pages, that are necessary for the simulation process, unlike pages such as the index page or the impressum, are the following:

New World Cup Run Page This is the main entry point for the user, which allows customizing a world cup simulation by setting rule weights or even adding a custom rule with custom ratings for the teams.
After finishing the custom setup, the tournament simulation is sent to the backend as a JSON string via Ajax and a loading dialogue is shown until the tournament simulation is finished. The user will then be redirected to the tournament page.

My Tournaments Page This page shows the user an overview over their past simulations. A simulation is linked to a user via a session cookie that contains solely a user-id.
The tournaments are ordered by the date and time the user requested the simulation with their custom parameters. Note that this date and time is not necessarily the original date of simulation, since no two simulations with the exact same parameters are ever run. In case users try to run an existing simulation another time, they are instead presented the old results with an updated timestamp.

Teams Page This page shows a list of the available teams that participate in the world cup 2014 and their respective rating points in the different available rating systems.

New Custom Tournament This is a page mainly for testing purposes and as a prove-of-concept. The user can iteratively set up and customize a new tournament simulation with full control over the type of simulation, the available rules, and the participating teams.
When one stage of customizing the tournament is complete, additional informations for the next stage are requested from the backend via Ajax. This page is usually hidden in the deployed version of the simulator.

4.2 Backend

The backend is the main part of the application that not only cares about website generation but also handles request processing and tournament simulation dispatching. The backend is written in Python and makes use of the following main libraries:

Flask Flask is a *Python web application microframework* that is used for the communication with the frontend[21].

SQLAlchemy SQLAlchemy is a library for database abstraction[22].

The backend can be divided into two sub-parts: firstly, the part that cares about the frontend interaction, and secondly, the part that starts the simulator when a new user request is available: the *simulation dispatcher*.
The simulation dispatcher will be described in the next section. The remainder of this section will describe the part that is responsible for user-interaction and request processing.
The backend generates the user-visible web pages from a set of Jinja2 templates, that can be cached as necessary. The backend also implements a very simple user management: whenever users simulate their first tournament, a new user id is created and linked to the actual user with a session cookie without an expiration date. All future tournaments simulated by that user can now be linked to that user id, which enables the application to show a user all their past tournament simulations without requiring a sophisticated password-based user management. As long as the users do not delete their session cookies, their tournaments will be linked to them.

4.2.1 The Simulation Dispatcher

The simulation dispatcher is the part of the Python backend that starts the simulator when a user requested a new simulation run and eventually, after the simulation finished, writes the results back into the database.
There are two available dispatchers: the first one is the local dispatcher that starts a simulation on the same computer as the backend. The second one is a QLess-py[24] based simulator which makes use of a redis[25] database to enable running tournament simulations on distributed worker machines for maximum scalability, see Figure 17. In the decentralized case, workers

can be dynamically added and removed at any time without the need to manually register them with the backend.
The communication between the simulation dispatcher and the actual simulator happens entirely over standard input and standard output using JSON-formatted strings. This approach was chosen to keep a maximum flexibility in both the dispatcher and the simulator and prevent them from being too intertwined.
The dispatcher will start the simulator and send all of the user configuration, teams' score, and anything else that is needed for the simulation over the standard input. After the simulation is done, the simulator will format the calculated results in JSON format and send them through the standard output.
This implies that the simulator does not need a database connection. It is entirely possible to use the simulator as a stand-alone program completely independent from the rest of the application and without using any type of dispatcher as an interface.

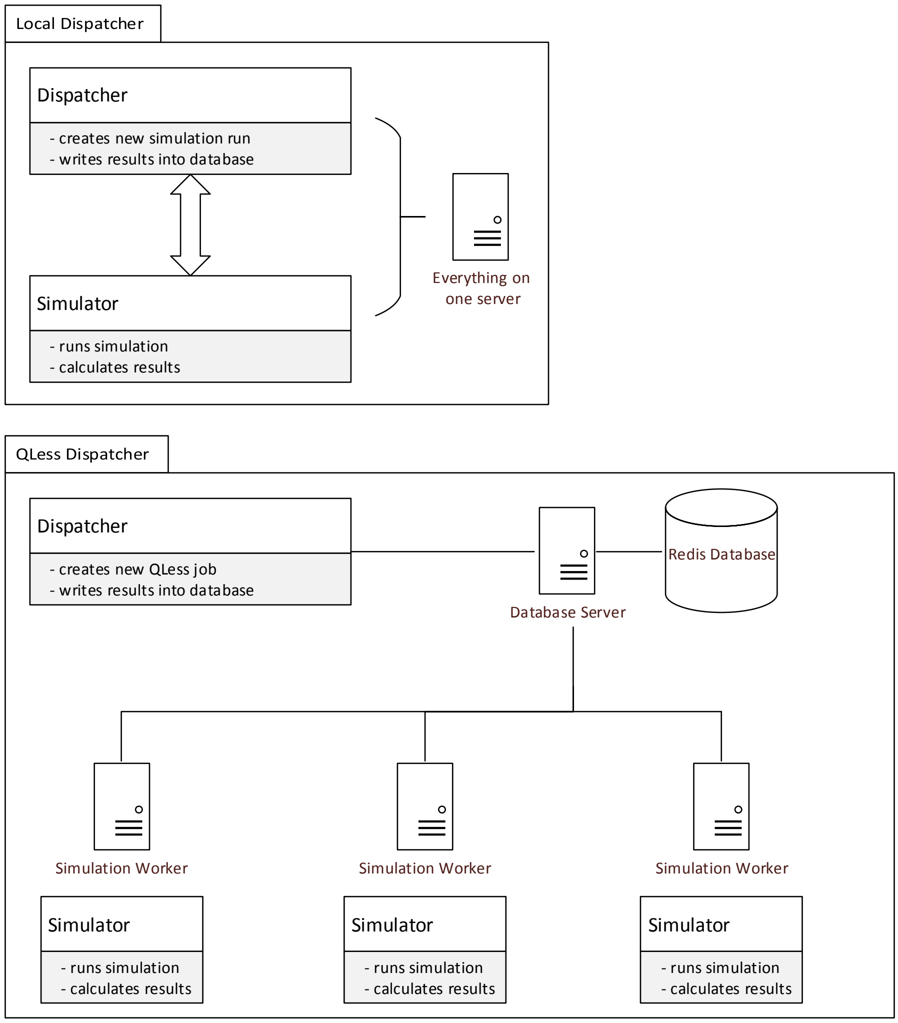

Figure 17: The diagram shows the two different types of available dispatchers. The local dispatcher is ideal for low-traffic applications or testing and the QLess dispatcher is ideal for a decentralized setup providing good scalability with an arbitrary number of possible workers.

4.3 The Simulator

The *simulator* is the part of the application that runs the actual simulation and collects the statistics to generate the output that is sent back to the backend. The simulation is written in C++11 and makes use of the JSON Spirit library[26] to both parse and generate the necessary JSON-encoded strings for communication with the backend. Threads are used to run single tournament executions concurrently for an optimal efficiency. The results are merged to generate the final statistics once the simulation threads have finished.
The simulator is a very light-weight program where the main part consists of the definition of the tournament's rules and the fomulae that are used to calculate the win expectancies as described in section 3, and the collection of the statistical data.

5 Results

In this paper, we have not only developed a simulation model to simulate the FIFA World Cup by combining multiple rating systems into win expectancies that yield the probability to win for any team in any match, but also implemented a very accessible simulator application that implements the official FIFA tournament rules and uses the discussed simulation model to run the complete tournament multiple times, generating a statistical probability distribution for the outcomes of the tournament from the observations made during all matches. Since the tournament is simulated following the official rules, the resulting distribution for every team also takes into account tournament-specific attributes such as the possibility of meeting stronger opponents most of the matches due to a possibly more unfavourable initial group draw. Table 6 shows the resulting distribution collected from 100000 executions of the whole tournament. The rule weights are configured as described in section 3.5 and the different scores for each team were collected on 8th May 2014.

This distribution can be seen as the estimate of the world cup's results under this simulation model. According to this distribution, there are four clear favourites for the first place: Brazil, Spain, Argentina and Germany. There is only a chance of around one third that the first place will be neither of those four teams.

Team	1^{st}	2^{nd}	3^{rd}	4^{th}	Avg. Goals
Brazil	25%	10%	10%	3%	1.8
Spain	15%	10%	8%	4%	1.6
Germany	14%	10%	13%	6%	1.6
Argentina	10%	10%	11%	7%	1.6
Portugal	4%	5%	5%	6%	1.3
Colombia	4%	5%	4%	4%	1.4
France	3%	4%	5%	6%	1.3
Uruguay	3%	4%	4%	4%	1.3
Chile	3%	4%	3%	4%	1.2
England	3%	4%	3%	3%	1.3
Netherlands	2%	4%	3%	3%	1.2
Italy	2%	4%	3%	3%	1.3
Belgium	2%	3%	4%	5%	1.3
Switzerland	2%	3%	3%	4%	1.2
Russia	1%	3%	3%	4%	1.3
Ecuador	1%	2%	2%	4%	1.1
USA	1%	2%	2%	4%	1.1
Bosnia and Herzegovina	1%	2%	2%	3%	1.1
Ivory Coast	1%	2%	1%	2%	1.1
Mexico	1%	1%	1%	2%	1.0
Croatia	1%	1%	1%	2%	1.0
Greece	1%	1%	1%	2%	1.1
Honduras	0%	1%	1%	2%	0.9
Iran	0%	1%	1%	2%	0.9
Korea Republic	0%	1%	1%	2%	1.0
Nigeria	0%	1%	1%	2%	1.0
Ghana	0%	1%	1%	2%	0.9
Japan	0%	1%	1%	1%	0.9
Costa Rica	0%	1%	0%	1%	0.8
Cameroon	0%	0%	0%	1%	0.8
Australia	0%	0%	0%	1%	0.7
Algeria	0%	0%	0%	1%	0.9

Table 6: This table shows the resulting distribution from 100000 tournament runs. The chances shown are direct observations from the simulation and represent how many times a team had a certain rank in the tournament. The average goals in the table are the average goals for every team over all matches.

6 Conclusion and Future Work

We have developed a probabilistic simulation model to simulate the football world cup by means of different rules, which take into account different rating systems and can be combined arbitrarily with user-defined weights.
We also developed a simulation application that employs the described simulation model. The source code is released under the MIT license[27] and can be found on github.com[3], as of May 2014.

A lot of different rating methods have been looked at in this thesis. However, there remain more and it is possible that yet another rating method might turn out to provide not only a better prediction of the results of a match, but also a more straight-forward way to calculate the odds between two teams. In particular, it might be interesting to check how well bookmaker's odds as a method of calculating the win expectancy for a team in a particular match perform. Bookmaker's odds are designed to do exactly what other rating systems might only yield as a secondary by-product: they rate the chance of success in a particular match up between two teams or the probability that a team wins a certain competition. They usually not only take different rating systems into account, but are also adjusted by experts to be as accurate as possible. For this version of the simulation and the thesis, bookmaker's odds have not been looked at specifically. The main reason is that the moral obligations coming from an accurate simulation model must not be forgot: allowing a user to directly compare side-by-side the calculated odds for a team to win and what a bookmaker would predict, might directly increase the willingness of the user to place bets on certain outcomes. At least as long as the simulator and the simulation model have not proven to be extremely accurate, we think nobody should be tempted to place bets based on the results of this simulation. Another issue that would have to be dealt with is, that the bookmaker's odds for a specific tournament usually already take into account things like the group draws and the different matches a team has to play throughout the tournament. For the simulation model this is not ideal, since the winning probabilities for two teams that face each other are needed, regardless of which other games each team might play in the tournament. When using bookmaker's odds to predict the outcome of a tournament with the simulation model discussed in this paper, it would be crucial to have most neutral ratings to, for example, not overrate teams, which are favoured by bookmakers due to an easy group draw, throughout all single matches in the complete tournament. Further, bookmaker's odds are not designed to be fair. They are designed to generate revenue for the bookmakers. Bookmakers might for example overestimate certain teams to keep the return rate lower. However, since the bookmaker's odds might be

[3] https://github.com/walachey/football-worldcup-simulator

an extremely good model of a team's chance of winning a certain event, it might still be interesting to look into it. For example, [28] states that bookmaker's odds might outperform the Elo rating system and the FIFA rating in regards to predicting the outcome of a tournament.

Another open topic is the verification of the simulation model with the rules and default weights as described. The straight-forward way of checking the simulation results is obviously comparing the actual results of the world cup with the prediction from section 5. This is not possible at the moment. Another method would be to simulate a previous world cup with known outcome and then do the comparison. However, the following general issue with comparing a probabilistic prediction with the actual outcome of a tournament must not be overlooked when doing such a comparison: a statistical prediction has only little significance when comparing it to one single event. The model could be extremely accurate and the results of the event could still differ drastically. Only a long-term comparison with multiple tournaments using the same rating method could show whether the model is valid or even refute it.

References

[1] F. Liu and Z. Zhang, "Predicting Soccer League Games using Multinomial Logistic Models," 2008.

[2] D. Karlis and I. Ntzoufras, "Statistical Modelling for Soccer Games: the Greek League," in *Proceedings of Hellenic European Conference on Computer Mathematics and its Application*, 1998.

[3] D. Dyte and S. R. Clarke, "A ratings based poisson model for world cup soccer simulation," *Journal of the Operational Research Society*, vol. 51, no. 8, pp. 993–998, 2000. [Online]. Available: http://dx.doi.org/10.1057/palgrave.jors.2600997

[4] V. Reddy and S. V. K. Movva, "The Soccer Oracle: Predicting Soccer Game Outcomes Using SAS®Enterprise Miner™."

[5] M. J. Dixon and S. G. Coles, "Modelling Association Football Scores and Inefficiencies in the Football Betting Market," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 46, no. 2, pp. 265–280, 1997. [Online]. Available: http://www.jstor.org/stable/2986290

[6] L. M. Hvattum and H. Arntzen, "Using ELO ratings for match result prediction in association football," *International Journal of Forecasting*, vol. 26, no. 3, pp. 460 – 470, 2010, sports Forecasting. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169207009001708

[7] FIFA, "FIFA/Coca-Cola World Ranking Procedure," http://www.fifa.com/worldranking/procedureandschedule/menprocedure/index.html. [Online]. Available: http://www.fifa.com/worldranking/procedureandschedule/menprocedure/index.html

[8] M. M. Kaminski, "How Strong are Soccer Teams? "Host Paradox" in FIFA's ranking," 2012.

[9] A. E. Elo, *The rating of chessplayers, past and present*. New York: Arco Pub., 1978. [Online]. Available: http://www.amazon.com/Rating-Chess-Players-Past-Present/dp/0668047216

[10] D. Ross, "Arpad Elo and the Elo Rating System," http://en.chessbase.com/post/arpad-elo-and-the-elo-rating-system. [Online]. Available: http://en.chessbase.com/post/arpad-elo-and-the-elo-rating-system

[11] S. B. J. Lasek, Z. Szlávik, "The predictive power of ranking systems in association football," *Int. J. of Applied Pattern Recognition*, to appear.

[12] N. Silver, "A guide to ESPN's SPI ratings," http://espnfc.com/world-cup/story/_/id/4447078/ce/us/guide-espn-spi-ratings, 2009. [Online]. Available: http://espnfc.com/world-cup/story/_/id/4447078/ce/us/guide-espn-spi-ratings

[13] transfermarkt.de, "transfermarkt.de - Marktwerte," http://www.transfermarkt.de/de/default/marktwert/basics.html. [Online]. Available: http://www.transfermarkt.de/de/default/marktwert/basics.html

[14] FIFA, "The history of all FIFA World Cups from the official website of the FIFA," http://www.fifa.com/tournaments/archive/worldcup/index.html. [Online]. Available: http://www.fifa.com/tournaments/archive/worldcup/index.html

[15] World Elo Ratings, "The current ELO ratings of most football teams." http://www.eloratings.net/. [Online]. Available: http://www.eloratings.net/

[16] FIFA, "FIFA/Coca-Cola World Ranking - Ranking Table," http://www.fifa.com/worldranking/rankingtable/index.html. [Online]. Available: http://www.fifa.com/worldranking/rankingtable/index.html

[17] ESPN, "Soccer Power Index: FAQ," http://espnfc.com/world-cup/story/_/id/4504383/ce/us/faq. [Online]. Available: http://espnfc.com/world-cup/story/_/id/4504383/ce/us/faq

[18] transfermarkt.de, "transfermarkt.de - Market values and general information about players," http://www.transfermarkt.de/. [Online]. Available: http://www.transfermarkt.de/de/

[19] P. van der Kruit, "Home advantage and tied games in soccer," http://www.astro.rug.nl/~vdkruit/jea3/homepage/voetbal.pdf, 2006.

[20] E. Bittner, A. Nußbaumer, W. Janke, and M. Weigel, "Football fever: goal distributions and non-gaussian statistics," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 67, no. 3, pp. 459–471, 2009. [Online]. Available: http://EconPapers.repec.org/RePEc:spr:eurphb:v:67:y:2009:i:3:p:459-471

[21] JQuery Team, "jQuery," http://jquery.com/. [Online]. Available: http://jquery.com/

[22] N. Downie, "Chart.js," http://www.chartjs.org/. [Online]. Available: http://www.chartjs.org/

[23] The Bootstrap Contributors, "Bootstrap," http://getbootstrap.com/. [Online]. Available: http://getbootstrap.com/

[24] The QLess-py Contributors, "Python Bindings for qless)," https://github.com/seomoz/qless-py. [Online]. Available: https://github.com/seomoz/qless-py

[25] The redis Contributors, "Redis Homepage," http://redis.io/. [Online]. Available: http://redis.io/

[26] J. W. Wilkinson, "JSON Spirit," http://www.codeproject.com/Articles/20027/JSON-Spirit-A-C-JSON-Parser-Generator-Implemented. [Online]. Available: http://www.codeproject.com/Articles/20027/JSON-Spirit-A-C-JSON-Parser-Generator-Implemented

[27] MIT, "MIT License," http://opensource.org/licenses/MIT. [Online]. Available: http://opensource.org/licenses/MIT

[28] C. Leitner, A. Zeileis, and K. Hornik, "Forecasting sports tournaments by ratings of (prob)abilities: A comparison for the euroÂ 2008," *International Journal of Forecasting*, vol. 26, no. 3, pp. 471 – 481, 2010, sports Forecasting. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169207009001459