Annals of Operations Research (2020) 286:119–146
https://doi.org/10.1007/s10479-019-03261-8

S.I.: CLAIO 2016

An analytics approach to the FIFA ranking procedure and the World Cup final draw

Sebastián Cea[1] · Guillermo Durán[2,3,4,5] · Mario Guajardo[6] · Denis Sauré[1] · Joaquín Siebert[1] · Gonzalo Zamorano[1]

Published online: 10 May 2019

Abstract
This paper analyzes the procedure used by FIFA up until 2018 to rank national football teams and define by random draw the groups for the initial phase of the World Cup finals. A predictive model is calibrated to form a reference ranking to evaluate the performance of a series of simple changes to that procedure. These proposed modifications are guided by a qualitative and statistical analysis of the FIFA ranking. We then analyze the use of this ranking to determine the groups for the World Cup finals. After enumerating a series of deficiencies in the group assignments for the 2014 World Cup, a mixed integer linear programming model is developed and used to balance the difficulty levels of the groups.

Keywords OR in sports · Analytics · Ranking · FIFA World Cup · Football

1 Introduction

The World Cup football competition is the most popular sporting event in the world. Organized by Fédération Internationale de Football Association (FIFA) in four-year cycles, it brings together the national teams of more than 200 countries. The event consists of two phases, the qualification phase and the tournament phase, the latter often known simply as the World Cup finals. In the qualifying phase, participating teams compete within their continental confederations for a certain number of berths in the final phase. Currently, there are six such regional bodies: the South American Football Confederation (CONMEBOL), the Union of European Football Associations (UEFA), the Asian Football Confederation (AFC), the Confederation of African Football (CAF), the Oceania Football Confederation (OFC) and the Confederation of North, Central American and Caribbean Association Football (CONCACAF). The 2014 World Cup finals, played in Brazil, saw 32 teams participate, 3.4 million spectators attended the matches at the stadiums and 3.2 billion people around the world watched on television (FIFA 2014a, b).

✉ Guillermo Durán
gduran@dm.uba.ar

Extended author information available on the last page of the article

As of the 1998 World Cup, 32 teams qualify for the finals, and these are divided into eight groups of four teams each as determined by a random draw.[1] The tournament begins with a group stage in which each team plays once against each of the other three teams in their group. The two teams that perform best in each group then advance to the knockout stage.

The draw defining the group stages has sustained various modifications throughout the last decade. Prior to the 2010 World Cup, teams were ranked according to past performance (measured as a combination of their FIFA rankings and performance in previous World Cup finals): the top seven teams (of the 32 that qualified) plus the host country's team were seeded into different groups, while the remaining teams were drawn at random from different pots, constructed so as to achieve a maximum geographic separation principle (which essentially aims at minimizing the number of teams from the same confederation placed on a same group). From 2010 to 2014 the procedure remained essentially the same, except for the fact that past performance was measured solely by the FIFA World Ranking at the time of the draw, eight months before the finals FIFA (2017) (ad-hoc rules ensuring geographic separation where implemented on each occasion). For the 2018 World Cup, draw pots were constructed so as to consider performance criteria as follows: teams were placed into four draw pots according to their ranking (artificially placing the host country on top of the ranking), so that the first eight teams in the ranking are on a first pot, the second eight teams on a second pot, and so on. At the draw event, pots were emptied sequentially, starting with the first pot; one by one, countries were drawn from each pot and placed into groups randomly while ensuring that no two countries from the same confederation were placed on the same group (with the exception of UEFA: in each group there has to be at least one, and no more than two UEFA countries). This procedure aimed at balancing the "quality" of each group while guaranteeing geographic diversity within each group. Arguably, today more than ever before, the FIFA Ranking has a significant impact on a team's chances at the final stage of the World Cup.

However reasonable the procedure above might appear, both the rankings and the draw have been subject to considerable criticism over the years. In the months leading to the 2014 World Cup, for example, many sports journalists and football fans noted with disapproval that teams like Colombia, Belgium and Switzerland had received high enough rankings to be seeded. Others pointed to what they perceived was the widely unbalanced composition (in terms of competitiveness) of the groups.

As for the academic community, the appearance in recent years of a number of scientific studies evidences a gradual increase of scholarly interest in the debates surrounding World Cup issues. Most of these works have dealt with predicting results (e.g., Maher 1982; Dixon and Robinson 1998; Rue and Øyvind Salvesen 2000; Dyte and Clarke 2000) while others have focused specifically on the FIFA rankings (McHale and Davies 2007; Suzuki et al. 2010; Lasek et al. 2013), but mostly studying their predictive power rather than proposing modifications. More recently, Lasek et al. (2016) elaborate strategies to improve a team's position in the FIFA ranking, based on choosing opponents for friendly games so as to maximize the probability of advancing in the ranking. Also recently, Alarcón et al. (2017) report the application of operations research to schedule the South American qualifiers to the 2018 World Cup, through an approach detailed in Durán et al. (2017). On the topic of defining the groups for the World Cup, the academic literature is almost non-existent. To the best of our knowledge, the only effort along this line is Guyon (2015), which discusses the deficiencies in the method used by FIFA and makes proposals for improvement.

In the present article we offer a critical analysis of the construction of the FIFA ranking and the group draw procedures, as used up until the 2018 World Cup. It is worth noting that

[1] Starting in 2026 the number of teams that qualify to the final stage is expected to increase to 48.

the FIFA ranking procedure has been changed after the 2018 Word Cup. We discuss these changes and relate them to our own recommendations. In turn, the drawing procedure was changed for the 2018 World Cup incorporating new rules that resemble one of the suggestions in Guyon (2015). Yet the drawing procedure ought to change for the 2026 World Cup, with modifications not announced as of the time of this writing, thus research efforts in this respect remain relevant.

We begin by analyzing the ranking. After revealing a series of apparent weaknesses in the ranking method used to seed teams for the 2014 World Cup, we specify a reference ranking for judging the performance of various possible simple modifications to it. The reference ranking is the expected result of a round-robin tournament between the teams that qualified to the 2014 World Cup finals, obtained using a Monte Carlo simulation (we use data for international matches played between 2009 and 2013). The main component of this approach is the calibration of a variation on the predictive model proposed by Maher (1982), a seminal work used as basis for most such models in the literature.

The proposed simple modifications to the FIFA ranking procedure are based on the results of a set of multinomial logistic regression models we construct in order to identify the variables that best explain the outcomes of matches between the national teams. The modifications are then evaluated using the reference ranking. This is followed by the application of a mixed integer linear programming model developed to define groups for the 2014 World Cup finals. The model takes into account geographic criteria while aiming to achieve a balance the "quality" across the various groups. Our results show that the methodologies used by FIFA for both the team rankings and group assignments could have been be improved upon considerably.

The remainder of this article is organized into four sections. Section 2 describes the construction of a reference ranking based on the predictive model due to Maher (1982). Section 3 analyses the FIFA ranking procedure used for the 2014 World Cup and identifies key variables for predicting match results. Section 4 develops proposals for simple modifications to the aforementioned ranking procedure and tests their performance using simulations and the reference ranking. Section 5 formulates a mixed integer linear programming model for group assignment and sets out the results. Section 6 presents our conclusions. Additional results and material are presented in "Appendices A and B".

2 Reference ranking and predictive model

Much of the negative media comment directed at FIFA's ranking relates to its use for seeding team for World Cup finals, has been prompted by the frequent inclusion in the top 10 of teams that have traditionally not performed well in major competitions (see McHale and Davies 2007). These criticisms are essentially subjective, however; what is needed is an objective way of assessing the quality of a ranking. This would be particularly useful for our work, as we provide a set of modifications to the rules for awarding points that are simple enough to be implemented in practice. Since each of these modifications would impact the team rankings, we must have a way of evaluating their performance.

Any criterion for comparing rankings presupposes the existence of an ideal ranking that reflects the quality of the teams such that when playing on neutral ground, the better-ranked side has a greater probability of winning than the worse-ranked side. This ideal ranking ought to be derived from historical data (and not only on the outcome of a single event competition, more on this later), a complicating factor considering that national team matches are normally played in international competitions which are held sporadically and are highly influenced by the tournaments' structures. The problem of inferring a ranking from paired comparison data goes back to Bradley and Terry (1952). In a recent survey of ranking systems for football by Lasek et al. (2013), a ranking's quality is associated with its predictive ability, which is inferred using a logistic regression model whose explanatory variables are the team ranking and the home-away status of the matches. This is the main quality criterion we adopt in this paper. An alternative view, also explored in our work, is to assess the quality of a ranking retrodictively, that is, by measuring how well it explains the past game outcomes (Coleman 2005; Martinich 2002).

2.1 Reference ranking

With the above in mind, and considering both the strong influence of the structure of a tournament on its outcome and the fact that the format which maximizes the correlation between the best team and the outcome is a round-robin tournament (Scarf and Yusof 2011), we base our reference ranking on the outcome of a round-robin tournament played in neutral ground. Because the main use of the ranking is to place teams into pots during the FIFA World Cup's group-stage draw, we restrict participation in such tournament to teams the qualified for the 2014 World Cup final stage. To estimate the outcome of said tournament, we calibrate a variation of the Maher (1982) model for a match, that assumes that the number of goals scored by the teams follow independent Poisson variables. Some works have proposed different approaches, such as the bivariate Poisson distribution (Karlis and Ntzoufras 2003) and a variant of it called Z-Poisson distribution (Lillestøl and Andersson 2011). Other works have focused on modeling the difference in goals scored instead of the number of goals of the teams (Karlis and Ntzoufras 2008; Van Haaren and Van den Broeck 2015). A discussion about these and other approaches can be found in Lasek et al. (2016). We adopt the most classic approach of assuming independent Poisson distributions and proceed to approximate the probability of each team winning a round-robin competition via a Monte Carlo simulation: we then derive our reference ranking from such approximate probabilities. In what follows we present our adaptation of the Maher (1982) model.

2.2 Predictive model

Let $X_{A,B}$ and $Y_{A,B}$ be the respective number of goals scored by teams A and B in a match where A plays at home. Following Maher (1982) we assume that $X_{A,B}$ and $Y_{A,B}$ are Poisson-distributed independent random variables with rates $\lambda_{A,B}$ and $\gamma_{A,B}$, respectively. We further assume that in a given match the rate at which a team scores depends on its offensive capacity, the opposing team's defensive capacity and the game's home-away status. More specifically, we assume that

$$ln(\lambda_{A,B}) = a_A - d_B + \rho_h, \tag{1a}$$

$$ln(\gamma_{A,B}) = a_B - d_A + \rho_a, \tag{1b}$$

where parameters a_i and d_i are respectively team i's offensive and defensive capacities, and variables ρ_h and ρ_a are corrections, not depending of the teams, for the rates that take into account home-away status (both variables omitted if a match is played on neutral ground). The probability of a match score being $\{X_{A,B} = m, Y_{A,B} = n\}$, where m are the goals by A and n the goals by B, is then given by

$$P(X_{A,B} = m, Y_{A,B} = n) = \frac{e^{-\lambda_{A,B}}(\lambda_{A,B})^m}{m!} \cdot \frac{e^{-\gamma_{A,B}}(\gamma_{A,B})^n}{n!}. \tag{2}$$

The above model has two parameters per team plus the two for home-away status. However, to avoid identification issues we set $a_i = 0$ for an arbitrary team. (This, because adding and subtracting a constant to all a and d parameters, respectively, results on the same values for the λ and γ parameter: we do not do the same for the ρ parameters as we implicitly assumed a null effect for matches on a neutral field.)

We estimate the model parameters via maximum likelihood estimation (MLE) using the outcomes of all official FIFA matches between January 2009 and October 2013. In particular, we performed a Poisson regression using the statistical software Stata, using (1a) and (1b) to predict goal rates and (arbitrarily) setting $a_{\text{Albania}} = 0$ as the reference level. The values of the a and d parameter estimates can be found in "Appendix B": for the home and away specific parameters we obtained $\rho_h = 0.167$ and $\rho_a = -0.191$.

To obtain our reference ranking we simulated a round-robin tournament 10^8 times (played on neutral ground) and approximated a team's probability of winning such a tournament with the percentage of simulations in which the team won the tournament.[2] In our simulation, each match granted two points to a win, one to a tie, and none to a loss. Table 1 depicts the (relative) reference ranking for the 32 teams participating in the 2014 World Cup finals.

In Table 1, teams winning the same number of (simulated) tournaments are ordered according to the mean number of points they receive during the tournaments. Considering that many competitions award three points for a win, we also computed a reference ranking under this modification: Table 1 shows how the ranking of each team changes (Rank. diff.) with this modification.

3 Analysis of FIFA ranking

Introduced in 2006, the FIFA ranking methodology is based on the ratings of its 211 member nations. Up until the latest (2018) World Cup, points were allocated to teams as follows: in any given match, P points were awarded to a team, where

$$P = M \cdot I \cdot T \cdot C. \tag{3}$$

Factor M is the number of points obtained from the match result: three points for a win, one for a draw, and zero for a loss. Factor I is the match status, that is, the importance of the match: 4 for a World Cup finals match, 3 for a confederation cup finals match (Copa America, Gold Cup, UEFA European Championship, etc.), 2.5 for a World Cup or confederation cup qualifier and 1 for a friendly (exhibition game) or minor confederation-level tournament match. Factor T is opponent strength, calculated as 200 less the opposing team's FIFA ranking. There are two exceptions: the value of this factor for a game against the top-ranked team would be 200; also, for opponents ranked below 150, the value is set at 50. Finally, factor C is the average strength of the confederations the two teams in the match belong to.[3] An example of this points system is given in Table 2 for a match between Honduras and Chile in the 2010 World Cup in which Chile won by a score of 1–0.

[2] In our simulations, the team with the highest number of points by the end of the tournament wins the competition. We use the goal-difference and goals scored criteria in a hierarchical manner to resolve ties.

[3] The confederation strength factor is based on the number of wins by all of the teams in a confederation in the last three World Cups. Before the 2014 World Cup, the values were 1 for CONMEBOL and UEFA, 0.88 for CONCACAF, 0.86 for the AFC and the CAF, and 0.85 for the OFC. For further details, see FIFA (2017).

Table 1 Reference ranking for teams in 2014 World Cup finals

Ranking	Country	P. champ.	Mean pts.	Rank. diff.
1	Brazil	46.5697	47.7700	–
2	Spain	36.9249	47.0040	–
3	Argentina	6.4035	42.0760	–
4	Netherlands	3.9351	41.1154	–
5	Germany	3.4002	40.4323	–
6	Colombia	1.0144	38.5166	–
7	England	0.6228	37.3053	–
8	Uruguay	0.5067	36.8033	–
9	Chile	0.1508	34.7379	–
10	Ecuador	0.1431	34.7912	–
11	Portugal	0.1025	34.4494	− 1
12	Mexico	0.0963	34.1847	+ 1
13	France	0.0873	34.2921	–
14	Russia	0.0158	31.8503	–
15	Italy	0.0112	31.0522	–
16	Ivory Coast	0.0066	30.1468	–
17	Switzerland	0.0042	29.9513	–
18	Croatia	0.0024	29.0754	–
19	Belgium	0.0004	27.0742	–
20	Ghana	0.0002	26.4549	− 1
21	Japan	0.0002	26.4432	+ 1
22	United-States	0.0002	26.0779	
23	Nigeria	0.0001	26.1404	–
24	Bosnia–Herzegovina	9.70E−05	25.2164	–
25	Korea Republic	4.80E−05	24.5138	–
26	Greece	1.70E−05	24.5556	–
27	Costa Rica	1.40E−05	23.6856	–
28	Iran	1.30E−05	23.6211	–1
29	Australia	3.00E−06	22.8116	+ 1
30	Cameroon	0	20.9119	–
31	Honduras	0	20.7207	–
32	Algeria	0	18.2170	–

For each participating team, the table depicts the ranking, probability of winning the round-robin tournament, average number of points obtained, and change in ranking when three points are granted per win

For a team i, we define $G_i := \{(j_k, t_k) : k = 1 \ldots\}$ as the set of matches played by team i, where j_k and t_k denote respectively the opponent and the date of match k. In addition, for $s = 1, \ldots, 4$, we define $G_{i,s}(t)$ as the set of indexes $(1, 2, 3 \ldots)$ for the matches played by team i between dates $t - s$ and $t - s + 1$. Let $\widehat{R_i}(t)$ be the rating of team i on date t. Under the procedure used so far to seed teams in the World Cup finals, this rating is calculated as the weighted sum of the point averages obtained over the last four years by the following formula:

Table 2 Example of FIFA ranking point system

	Honduras	vs.	Chile
Ranking position	38		22
M: Match result	0		3
I: Match status	4		4
T: Opponent strength	178		162
Confederation	CONCACAF		CONMEBOL
Strength of confederation	0.88		1
C: Confederation strength (average)		(0.88+1)/2 = 0.94	
$P = M \cdot I \cdot T \cdot C$	0		1827

$$\widehat{R}_i(t) := \sum_{s=1}^{4} \alpha_s \left(\frac{1}{\max\{|G_{i,s}(t)|, 5\}} \sum_{h \in G_{i,s}(t)} P_{i,h} \right), \tag{4}$$

where $P_{i,h}$ is the point total obtained by team i in match h. The max function in the denominator on the right-hand side is included to reflect the FIFA rule that each national team must play at least five matches per year; if fewer than five are played, the formula considers the remainder to be losses. The term α_s incorporates another element of FIFA procedure: points obtained are depreciated on an increasing scale according to how many years ago they were earned. The values used up to the 2018 World Cup were: $\alpha_1 = 1, \alpha_2 = 0.5, \alpha_3 = 0.3, \alpha_4 = 0.2$.

The FIFA ranking system orders the 211 associated teams every month by their ratings, assigning a rank of one to the highest rated team, two to the second-highest, and so on. The main purpose of the ranking is to identify the seeds for the draw defining the World Cup finals group stage.[4] As noted earlier, each of the seven top-rated teams plus the host country's national squad is seeded into one of the group stages' eight groups of teams. Such an arrangement favors these teams given that by being placed in separate groups, they do not face their strongest rivals in this initial stage of the finals.

3.1 Deficiencies of the FIFA ranking procedure

Although the methodology as just described does provide a relatively simple method for calculating the rankings, it also has a series of deficiencies. These are discussed in what follows.

Friendly matches A team's rating as given by (4) is a weighted sum of the point averages obtained by the team over the previous four years. Since little weight is placed on friendlies ($I = 1$), it can be easily shown that teams with a high rating have little incentive to participate in such matches. Take, for example, the case of Chile, which in August 2014 had an average over the preceding year of 673.61 points (see Table 3). Playing and winning a friendly against the world's top-rated team would give Chile 600 points ($3 \cdot 1 \cdot 200 \cdot 1 = 600$), the maximum possible for such a game but below its average as of that moment. Thus, a victory over the best team in the world would actually lower its rating, and probably worsens its ranking as well.

[4] The ranking is also used in setting up the World Cup qualifying tournaments and confederation cups in some confederations.

Table 3 Chile's rating as of August 2014

Year	Annual average	α_t	Contribution
2014	673.61	1	673.61
2013	391.47	0.5	195.73
2012	469.7	0.3	140.91
2011	447.07	0.2	89.41
$\widehat{R}_{Aug,2014}$			1099.66

Table 4 FIFA ranking as of October 2013

Ranking	Country	Points	Ranking	Country	Points
1	Spain	1513	6	Uruguay	1164
2	Germany	1311	7	Switzerland	1138
3	Argentina	1266	8	Netherlands	1136
4	Colombia	1178	9	Italy	1136
5	Belgium	1175	10	England	1080

This circumstance was plainly in evidence when the seeds for the 2014 World Cup were chosen in October 2013. The 10 top-rated teams at that time, along with their point totals, are set out in Table 4. Four of the teams (Colombia, Belgium, Uruguay and Switzerland) attained their rankings in part by playing fewer friendlies between October 2012 and October 2013 than the other national sides. During that year, Switzerland had played only three and Colombia, Belgium and Uruguay just four while Netherlands and England had played five and Italy six. A simulation analysis revealed that some teams with lower ratings would have been seeded had they not played any friendlies during the year previous to the draw (e.g., Chile, which was ranked 12th when the seeding was decided, would have ranked second).

Point depreciation and scheduling The point-depreciation scheme used by FIFA favors certain confederations due to the scheduling of their cup tournaments. As previously shown, matches for the various confederation cups (Copa America, the Gold Cup, the UEFA European Championship, the Asian Cup, the Africa Cup of Nations) are worth more points than friendlies, but since the cup matches are played on different schedules, they are also depreciated differently. Thus, countries whose cup tournaments are scheduled closer to the World Cup benefit from the fact that their cup games are more recent. This favors European teams over South American teams, for example, given that the European tournament is played just two years before the following World Cup whereas the South American tournament is played three years before it (in the run-up to the 2014 World Cup, the Copa America was held in 2011, the UEFA Championship in 2012). At the same time, not all teams play the same number of "important" matches. For example, each CONMEBOL team plays 18 qualifying games, while UEFA teams play 10–12 games, which arguably favors CONMEBOL.

Match result points Although the value M assigned to match results follows the typical practice of awarding three points for a win, one for a draw, and zero for a loss, we posit that this leads to a disproportionate point difference between a draw with one of the best teams in the world and a win over a middle-ranked side in matches of the same status (i.e., both friendlies, qualifiers, or World Cup finals games). As an example, drawing in a friendly against the top-rated team earns a maximum of 200 points whereas beating a team ranked 100th will gain a minimum of 255 (the precise numbers will depend on the confederation). Quite apart from the relative merits of the two cases, what seems so perverse is that in certain situations, carefully arranging to play a middling team is more highly rewarded than the courage to face the world's top-rated one.

Home-away status and confederation In addition to the deficiencies of the FIFA ranking system used up to July 2018, there are other factors that are not currently part of the procedure but which are relevant to our analysis. One of these factors is a game's home-away status. Under this system, this plays no role in the point assignments, yet there is a general consensus that teams play better at home. This *home advantage* is recognized in the literature as a key factor in explaining sports event results. In the case of football, Downward and Jones (2007) attribute this effect to crowd pressure on the referees and the support of the fans, while Pollard (1986) emphasizes the home team's greater familiarity with the pitch.

Another issue that has been raised is the influence on ranking points of confederation strength. In our opinion, however, the impact of this factor, if any, is not obvious. What does seem to be important is the extension of the home-away concept from country of origin to confederation of origin. This is motivated by the fact that for the World Cup finals, of the 19 times they have been held either in Europe or the Americas, the winner on 17 occasions has been a team from the same continent (the only exceptions being Brazil in Sweden '58 and Germany in Brazil '14).

3.2 Home-away status

To estimate the importance of the home-away status and confederation strength factors in the FIFA ranking and make some concrete recommendations for improvement, we now examine the predictive ability of home-away status. More specifically, we directly estimate the probability of a match result in terms of the difference between the teams' rankings and the home-away status. The model we use here is a multinomial logistic regression (for an alternative approach, see Dyte and Clarke (2000); for more details on logistic regression and an interpretation of the model, see Anderson et al. 1992).

Let $Y_{i,k} \in \{\text{win}, \text{draw}, \text{loss}\}$ denote the result of i at match k, and l_k denote the index of match (i, t_k) in G_{j_k}. (In the sequel, we let $R_i(t)$ denote the ranking of team i at the time t, and j_k and t_k denote the rival of i and the time of match k, respectively.) Define

$$V_{i,k} := \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{home} \cdot H_{i,k} + \beta_{away} \cdot A_{i,k}, \tag{5}$$

where $H_{i,k}$ ($A_{i,k}$) takes a value of 1 if and only if team i (j_k) plays game k at home.

In our base model, which we refer to as MRH, for team i playing match k we have that

$$P(Y_{i,k} = \text{x}) = \begin{cases} \frac{e^{V_{i,k}}}{1+e^{V_{i,k}}+e^{V_{j_k,l_k}}} & \text{x} = \text{win} \\ \frac{1}{1+e^{V_{i,k}}+e^{V_{j_k,l_k}}} & \text{x} = \text{draw} \\ \frac{e^{V_{j_k,l_k}}}{1+e^{V_{i,k}}+e^{V_{j_k,l_k}}} & \text{x} = \text{loss}. \end{cases} \tag{6}$$

Thus, MRH incorporates home-away status. In addition to this base model we consider the following models that arise from alternative specifications for $V_{i,k}$.

Table 5 Senior men's national team matches October 2005–October 2013

	Friendlies	Qualifiers	Confederation cup finals	World Cup finals	Total
Has home-away status	2716	3079	189	10	5994
Neutral ground	1170	282	485	118	2055
Total	3886	3361	674	128	8049

Table 6 Regression models results

Variable	MR		MRH		MRHL		MRHS	
	Coefficient	Error	Coefficient	Error	Coefficient	Error	Coefficient	Error
β_0	0.3071	0.0274	0.3019	0.0519	0.1478	0.0895	0.1478	0.0895
β_r	0.0131	0.0003	0.014	0.0003	0.014	0.0003	0.014	0.0003
β_{home}			0.3463	0.0614	0.5005	0.0954	0.3601	0.0954
β_{away}			– 0.5058	0.0645	– 0.3517	0.0974	– 0.467	0.0974
β_c					0.2049	0.0975		
β_{AFC}							0.0611	0.06
β_{OFC}							0.477	0.2012
β_{CNMBL}							0.3136	0.1021
β_{UEFA}							– 0.1454	0.0631
β_{CAF}							0.1705	0.0633

- Model without home-away factor (MR):

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)).$$

- Model with confederation-level home-away factor (MRHL):

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{home} \cdot H_{i,k} + \beta_{away} \cdot A_{i,k} + \beta_c \cdot C_{i,k},$$

  where $C_{i,k}$ takes a value of 1 if and only if game k of team i is played on the confederation that i belongs to but i is not at home.
- Model with confederation strength factor (MRHS):

$$V_{i,k} = \beta_0 - \beta_r(R_i(t_k) - R_{j_k}(t_k)) + \beta_{home} \cdot H_{i,k} + \beta_{away} \cdot A_{i,k} + \beta_{c_i}$$

  where $c_i \in$ {AFC, OFC, CAF, CONMEBOL, UEFA, CONCACAF} denotes the confederation to which team i belongs to. (To avoid identification issues, we set $\beta_{CONCACAF} = 0$.)

The first of these three models (MR) is a benchmark for the predictive ability of the FIFA ranking system, the second model (MRHL) determines the influence of confederation-level home-away status, and the third model measures the impact of incorporating a team's confederation membership (recall this factor is used by the FIFA ranking).

We estimated the model parameters via MLE using data from every match played from October 2005 to October 2013 (8,049 games in total: Table 5 provides summary statistics on these games). In particular, we performed a multinomial logistic regression using the statistical software *R*, with the *mlogit* package. The outcome is depicted in Table 6.

With the MR model, the positive sign obtained for β_r indicates that a better-ranked team is associated with a greater probability of winning a match than its opponent. The sign and

Table 7 Summary of model results

Model	Log likelihood	No. of param.	AIC	BIC
MR	– 7.704	2	15.412	15.426
MRH	– 7.409	4	14.828	14.856
MRHL	– 7.407	5	14.825	14.860
MRHS	– 7.391	9	14.800	14.862

value of the intercept (β_0) show that with similarly ranked teams, it is more likely that one of the teams will win than that the two will draw.[5] Also, we observe that a team's probabilities of winning or drawing when playing in neutral ground equalize when its opponent is $\frac{\beta_0}{\beta_r} \approx 23$ positions higher in the rankings [this follows from (6) and setting $V_{ik} = 0$ in (5)].

From the MRH model results we may infer that all three variables are statistically significant (for all of the parameters a 95% confidence interval did not include 0). The sign of the β_{home} variable is positive, in line with previous results reported in the literature that have demonstrated the importance of this factor. Analogously, the β_{away} variable not only is negative but has an absolute value greater than that of the β_{home}. Considering the value of β_0, this means that, for example, when two similarly ranked teams play, it is more likely that the home team wins than that it ties, which is also more likely than that it loses.

We can use the MRH model to estimate how much difference in team rankings is the equivalent of the home advantage. This relationship is obtained by solving the following equation:

$$\beta_{home} - \beta_{away} = 2x\beta_r \quad \Rightarrow \quad x = 30.4. \tag{7}$$

(This follows from using (5) to set $V_{i,k} = V_{j_k,l_k}$.) This result indicates that the most closely matched games in senior men's national team football are those in which the strength of the home team is 30 units below that of its visiting opponent.

We compare the models using the Akaike (AIC) and Bayesian (BIC) information criteria.[6] The results for each of them are displayed in Table 7.

In general, it can be seen that models that incorporate Home-Away effects perform better (lower AIC and BIC) than MR. The MRHL estimates show that the β_c variable is significant and positive, meaning that a team playing on its home confederation has more chances of winning than its opponent if the latter is from a different confederation. We also note, however, that the confederation-level home-away effect is weaker than the national one. Furthermore, although β_c is significant, the increase in log likelihood of model MRHL does not clearly offset having an extra parameter compared to MRH (whether one model is better than the other will depend on which criteria for model selection one uses). As for the MRHS model, its estimates show that β_{AFC} is not significant, that is, it is not statistically different from 0. One can conclude then that countries from both the AFC and CONCACAF behave similarly, in average. Additionally, the CONMEBOL, OFC and CAF countries do better than CONCACAF while the UEFA nations do worse.[7]

[5] This is explained as in our data from October 2005 to October 2013 there are 1951 games that end up tied, among a total of 8049 games. Thus the estimated model will favor having some side winning with higher probability.

[6] Both criteria are estimators of the relative quality of a statistical model for given data, and are typically used for model selection: for further details, see, for example, Hastie et al. (2001).

[7] Although some of the UEFA countries have great achievements in the history of football (such as Germany and Italy), the confederation in total gathers 55 teams including some that consistently rank among the worst in the world (such as San Marino, Andorra, Malta, and Liechtenstein) which help explain this result.

Overall, MRHS and MRH achieve the lowest AIC and BIC, respectively. In particular, we conclude that models that include Home-Away effects are those that best explain the data. Given how decisive is the home-away factor in match results, its inclusion in the factors determining the number of points a match contributes to a team's rating would be desirable.

3.3 Connection to new and old Elo-inspired rankings

Starting in August 2018, FIFA changed its rating procedure, using a formula inspired by the Elo method. Under this formula, the rating of team i after its kth game is

$$\begin{aligned}\widehat{R}_i(t_k) = \ & \widehat{R}_i(t_k-) + I \cdot (M - M_e)1 \\ & \times\{M > M_e \vee \text{any game not belonging to a knock-out phase}\}, \end{aligned} \tag{8}$$

where I denotes the importance of the match (now taking one out of nine possible values, ranging from 2.5 to 30), M denotes the outcome of the match (2 points for a win, 1.5 for a win decided in penalty shoot-outs, 1 for a draw or a loss in penalty shoot-outs, and 0 for a loss), and $M_e = 2/(1 + 10^{(\widehat{R}_{j_k}(t_k-)-\widehat{R}_i(t_k-))/600})$ represents the "expected result of the match." Here, $1\{\}$ denotes the indicator function, thus the increment in the rating of a team after a match might be negative, unless the game corresponds to the knock-out phase of a competition.

It is worth noting that the formula above differs in key aspects from the well-known World Football Elo Ranking adaptation Elo Ratings (2018), which we use as a benchmark later in this paper. In this later adaptation of the Elo method, the rating of team i after its kth game is given by

$$\widehat{R}_i(t_k) = \widehat{R}_i(t_k-) + I \cdot G \cdot (M - M_e), \tag{9}$$

where I denotes the importance of the match (using a different scale), G is a factor that depends on the goal difference in the final score of the game, M denotes the outcome of the match (as in the FIFA formula above), and

$$M_e = 1/(1 + 10^{(\widehat{R}_{j_k}(t_k-)+100\cdot 1\{j_k \text{ plays at home}\}-\widehat{R}_i(t_k-)-1001\{i \text{ plays at home}\})/400})$$

represents the "expected result of the match."

Both rating systems above move away from the ranking used by FIFA so far. On the positive side, the fact that the rating is the cumulative sum of match points (possibly with some negative terms) should eliminate the incentives to avoid friendly games, and issues related to point depreciation. Also, they exclude direct influence of teams conferences, and adopt the 2–1–0 pointing systems for each game. Note, however, that unlike the newest FIFA formula, equation (9) includes a correction to account for the home-away status, and for the goal difference. According to the analysis earlier in this section, the exclusion of the home-away status ought to significantly hamper the predictive ability of the new FIFA rating system.

A common feature of both methods above is that the marginal change of points before and after a game depends on the difference between the outcome of the game and its "expectation." In this regard, both formulas for M_e can be interpreted as attempting to adjust a Logistic model using the rating of the teams to predict the outcome of a game. Nonetheless, note that such a prediction does not account explicitly for the possibility of a tie (which happens in about 1/4 of the games in out data set), which ultimately distorts point assignment for tight games.

For the purpose of this paper, we consider modifications to the rating system used by FIFA between 2006 and 2018 because: it is the only system used so far in drawing teams for the World Cup finals; the use of the new system depends on its predictive ability as of 2022, which depends non-trivially on the initialization of the ratings in August 2018 (this, in a nutshell, considered equally spaced rating according to the post 2018 Wold Cup (old) FIFA ranking positions, which introduces a transient effect that has not been properly studied); and, finally, a move to a new system ought to consider, based on our work, components that are especially relevant to football, such as those included in the World Football Elo Ranking (e.g. home-away status) but not necessarily restricted to them (we include a more detailed discussion on the subject later on the paper).

4 Proposals for improvement and results

The reference ranking we constructed in Sect. 2 is meant to be used only as a baseline, as the complexity of its construction would probably rule out its application in practice. Next, we suggest a series of simple modifications to the formula employed by FIFA so far, informed by the discussion advanced in the previous section and that can be easily interpreted by football fans.

4.1 Improvement proposals for the ranking procedure

In light of the deficiencies revealed above and the results obtained for the models presented in the previous section, we develop below a number of proposals for improvement.

P1 *Include home-away factor and omit confederation strength factor* The analysis in the previous section demonstrates the major importance of the national home-away factor variable and the relatively minor importance of the confederation variable. To reflect these two effects, we propose that the C variable be dropped from the match ranking point formula (3) and that the opponent strength factor T be adjusted as prescribed by (7) above. Thus, the formula becomes $P_{i,k} = M_{i,k} \cdot I_{i,k} \cdot T_{i,k}$, where

$$T_{i,k} = \begin{cases} \max\{(186 - R_{j_k}(t_k), 50\} & \text{if } i \text{ plays at home,} \\ \max\{(216 - R_{j_k}(t_k), 50\} & \text{if } i \text{ plays away,} \\ \max\{(201 - R_{j_k}(t_k), 50\} & \text{if } i \text{ plays on neutral ground.} \end{cases}$$

This formula aims at imposing a penalty equivalent to 30 ranking positions to reflect the advantage accruing to the team that plays at home.

P2 *Omit friendly matches* To mitigate the negative effects of friendlies, we propose that they simply be dropped from the ratings calculation. This would eliminate the incentive to avoid these games, which play a very useful role for the teams in that they allow them to test new players and experiment with new strategies (note that the incentives against experimenting persist even when points are not discounted). In concrete terms, we suggest that $G_{i,s}(t)$ in (4) be redefined as the set of *non-friendly* matches played by a team between dates $t - s$ and $t - s + 1$ for all s.

P3 *Change the number of points awarded for a win* To reduce the wide gap between the points earned for drawing with a top-ranked team and the points gained for beating a middle- or low-ranked one, we propose that only two points be awarded for a victory instead of three. This would mean that a draw with a top-ten team would gain a number of points similar to or even more than a win over a middle-ranked side. In concrete terms, we propose that match result points be given by

$$M_{i,k} = \begin{cases} 2 & \text{if } Y_{i,k} = \text{win}, \\ 1 & \text{if } Y_{i,k} = \text{draw}, \\ 0 & \sim . \end{cases}$$

Davidson (1970) shows that when points are assigned as above, the expected ranking coincides with that obtained by applying a variation of the model in Bradley and Terry (1952), which derives rankings from pairwise comparisons.

P4 *Eliminate point depreciation by year* To correct the problem of imbalances in the depreciation of points earned in past games played in certain confederation tournaments due to differences in scheduling, we propose that the depreciation parameters α_s in (4) be replaced as follows:

$$\alpha_s = \begin{cases} 1 & s \in \{1, 2, 3, 4\}, \\ 0 & \sim . \end{cases}$$

By thus eliminating the depreciation of points depending on the year earned, a team's point total at any given moment reflects its performance over the previous four years, embracing all possible confederation tournaments a national team may have participated. Note that this proposal, in conjunction with **P2**, ought to also mitigate the negative effect of having different qualifying tournament formats per confederation (e.g. different number of matches per tournaments).

The modification introduced in **P1** above is akin to that made in the computation of the expected outcome in the World Elo football rating. Similarly, **P3** and **P4** above made the rating a cumulative sum of points earned in the previous four years. Note however, that unlike Elo-based ratings, each term is said sum is non-negative (a key feature of the Elo method). We consider that including a term that represents and expected outcome would constitute a mayor departure from the ranking system, and thus we did not consider it. Finally, note that **P2** is not considered by Elo-based ratings, which in our opinion, reduces the incentives to use friendly games as laboratories to try out new formations and tactics.

4.2 Results

We implemented all combinations of the modifications suggested in the previous section to compute all possible alternatives to the FIFA ranking. In particular, using historical data of official FIFA matches we computed these alternative rankings for each month between January 2009 and May 2014 (right before the 2014 World Cup). In what follows, for a set of recommendations $\{i_1, \ldots, i_k\} \subseteq \{1, 2, 3, 4\}$ we denote by $Pi_1 \ldots i_k$ the ranking that arises from implementing recommendations $i_1, \ldots, i_k$ (for example, ranking $P23$ implements modifications 2 and 3). We then generated fifteen proposals of new rankings.

In the following analysis we include two additional benchmark rankings. First, a modified FIFA ranking resulting from placing Brazil in first place (this is a natural modification, as Brazil was artificially down in the FIFA ranking due to the fact that it did not play qualifying games). Second, we consider the World Football Elo Ranking from October 2013 Elo Ratings (2018). While construction of this second ranking follows from mayor changes to the procedure used by FIFA between 2006 and 2018, we include it because of its (documented) high predictive power Lasek et al. (2013) and because of its similarity to the ranking system to ought to be used for the 2022 World Cup. Table 8 compares the 16 top positions for the FIFA, the modified FIFA, the Elo, the Reference and the proposal $P1234$ rankings, only

Table 8 Comparison of the different rankings, for October 2013

Ranking	Ref. ranking	P1234	FIFA	FIFA (mod)	Elo
1	Brazil	Brazil	Spain	Brazil	Brazil
2	Spain	Spain	Germany	Spain	Spain
3	Argentina	Argentina	Argentina	Germany	Germany
4	Netherlands	Germany	Colombia	Argentina	Argentina
5	Germany	Uruguay	Belgium	Colombia	Netherlands
6	Colombia	Netherlands	Uruguay	Belgium	England
7	England	Chile	Switzerland	Uruguay	Italy
8	Uruguay	Australia	Netherlands	Switzerland	Uruguay
9	Chile	Ivory Coast	Italy	Netherlands	Colombia
10	Ecuador	Italy	England	Italy	Portugal
11	Portugal	Japan	Brazil	England	Chile
12	Mexico	United States	Chile	Chile	Belgium
13	France	England	United States	United States	France
14	Russia	Colombia	Portugal	Portugal	United States
15	Italy	South Korea	Greece	Greece	Russia
16	Ivory Coast	Greece	Bosnia	Bosnia	Switzerland

considering the teams participating in the 2014 World Cup finals, for October 13, 2013, the date used by FIFA to select the seeded teams for the 2014 World Cup.

Opinions will, of course, vary on which of these rankings is the *best* one, and there is little chance of overcoming such differences. An objective appreciation can nevertheless be obtained by comparing the distances separating these rankings from the reference ranking developed in Sect. 2. We consider various metrics for this comparison.

- First, we consider a weighted mean square error (w-MSE) between a ranking and the reference baseline. For a ranking proposal S, define

$$\text{w-MSE}_S := \sum_{i \in N} w_{R_i^*} \left(R_i^S(Oct, 2013) - R_i^*(Oct, 2013) \right)^2, \tag{10}$$

  where ${R^S}_i(\cdot)$ denotes the ranking of team i under proposal S, R^* denotes the reference ranking, N the set of countries in the 2014 World Cup finals sorted by its positions in the Reference ranking, and $\{w_j : j = 1, \ldots, N\} > 0$ is a set of weights such that $\sum_{j=1}^{|N|} w_j = 1$ and $w_j = c\, e^{-\gamma\, j}$, $j = 1, \ldots, |N|$, with $\gamma = 0.1$. Here, constant $c \approx 0.1$ is such that the sum of all the weights is equal to 1. These weights allow to assign more importance to the differences in the higher positions of the rankings. The election of $\gamma = 0.1$ has the objective of a smoother transition in the importance of top places in the rankings. Thus arguably, the lower the w-MSE of the ranking, the better the proposal.
- Second, considering that underlying ratings might provide a better notion of the distance between teams, we consider a rating-equivalent to w-MSE, which we denote by r-MSE, in which we simply replace the ranking R_i^S of a team by its underlying rating $\widehat{R}_i^S$ in (10).[8] For the case of the modified FIFA ranking, we consider that Brazil has the same

[8] In this case, because a higher rating translate into a lower ranking, we set $w_r = c\,(1 - e^{-r})$ where r denotes the rating of a team and c is a normalizing constant.

Table 9 Performance of all the proposals under the three metrics at October 2013

Proposal S	w-MSE$_S$	r-MSE$_S$	τ_S
FIFA ranking	35.612	0.436	0.573
P1 (+ home-away factor, − conf.)	50.168	0.337	0.488
P2 (+ excl. friendlies)	31.288	0.252	0.593
P3 (+ change in points for win)	38.347	0.326	0.573
P4 (− point depreciation)	27.179	0.240	0.645
P12	50.316	0.341	0.484
P13	48.683	0.336	0.492
P14	31.230	0.250	0.625
P23	23.047	0.228	0.645
P24	36.038	0.172	0.581
P34	33.431	0.258	0.629
P123	45.732	0.341	0.508
P124	43.900	0.240	0.500
P134	28.901	0.150	0.637
P234	30.392	0.167	0.605
P1234	36.662	0.223	0.573
Elo	11.212	0.131	0.758
Modified FIFA ranking	24.969	0.347	0.613

rating than Spain. For Elo, both FIFA and our proposed rankings, we normalize (4) so that $\max\{\widehat{R}_i^S(t) : i \in N\} = 1$ and $\min\{\widehat{R}_i^S(t) : i \in N\} = 0$; and for the case of the reference ranking we consider the (normalized) mean points obtained by a team in the tournaments that define the ranking. This is, we consider

$$\widehat{R}_i^*(t) = (\text{Mean pts.}(i) - \text{Min mean})/(\max\{\text{Mean pts.}(j) - \text{Min mean} : j \in N\},$$

where Mean pts.(i) denotes the mean points obtained by team i in the round-robin tournaments used to construct the reference ranking, and Min mean := $\min\{\text{Mean pts.}(j) : j \in N\}$.
Again, the lower the r-MSE of the ranking, the better the proposal.

- Third, we compute the *Kendall's tau* distance Kendall (1938) between each proposal and the reference baseline. For a proposal S the metric at time t is given by

$$\tau_S := \frac{2}{|N|(|N|-1)} \sum_{i \in N} \sum_{j \in N : j > i} \text{sgn}(R_i^S(t) - R_j^S(t)) \cdot \text{sgn}(R_i^*(t) - R_j^*(t)),$$

where the function sgn(·) returns the sign of its argument. This metric is a measure of ranking correlation, it is equal to 1 when both rankings coincide, and it is − 1 when one ranking is the reverse of the other. So, values of τ_S closer to 1 mean that the proposal S is better.

Table 9 depicts the performance of all the rankings under these three metrics. There, we observe that our proposed rankings improve upon FIFA's ranking in terms of w-MSE in seven of fifteen cases. However, when comparisons are draw in terms of ratings, all proposals improve upon the incumbent. In terms of the third metric, eight of our fifteen proposals strictly outperform the FIFA ranking. Note that seven proposals improve upon the FIFA ranking in terms of the three metrics. On the other hand, the modified FIFA ranking outperforms the FIFA ranking in the three metrics. If we compare our proposals with the modified FIFA ranking we observe that our rankings outperform this ranking in one case in terms of the first metric, in all the cases in terms of the second one and in five cases in terms of the third one. Note that proposal 23 is better than modified FIFA ranking in all the metrics. Recall that this ranking arises from artificially placing Brazil on top of the ranking for October 2013.

It is worth noting that the Elo ranking improves upon both the FIFA ranking and all our proposals in each metric. This is consisting with the analysis in the previous section: while being a mayor departure from the FIFA ranking used between 2006 and 2018, the method incorporates in some sense all our recommendations, and is in spirit closer to the reference ranking in the sense that it attempts to predict match outcomes based on the teams past performance. This speaks of the potential for FIFA from fully embracing the Elo method.

In addition to measuring a ranking's performance in terms of the distance to the reference ranking, we can also measure quality in terms of how well it can explain the outcome of past games. For this, we consider a rather retrodictive approach and construct the following score: for a ranking S, define

$$S_S := 2\left(\sum_{j\in N}\sum_{(j_k,t_k)\in G_i} 1\{R_i^S(t_k) < R_{j_k}^S(t_k)\}1\{Y_{i,k}=\text{win}\}\right)\Big/\left(\sum_{j\in N}|G_i|\right),$$

where $1\{\cdot\}$ is the indicator function. Hence, S_s corresponds to the fraction of games won by the team with the best ranking. Because of our prescription pertaining the home-away effect (see 7), we also consider the alternative score,

$$\widehat{S}_S := 2\left(\sum_{j\in N}\sum_{(j_k,t_k)\in G_i} 1\{R_i^S(t_k) + E_{i,k} < R_{j_k}^S(t_k)\}1\{Y_{i,k}=\text{win}\}\right)\Big/\left(\sum_{j\in N}|G_i|\right).$$

where

$$E_{i,k} := \begin{cases} 30 & \text{if } i \text{ plays at home in game } k, \\ -30 & \text{if } i \text{ plays away in game } k, \\ 0 & \sim . \end{cases}$$

The idea behind the corrected score is to adjust for the fact that teams that play at home have an advantage equivalent (in average) to roughly 30 spots in the FIFA ranking. Table 10 depicts the performance of the FIFA and proposed rankings under these additional metrics considering all the 4684 games played between Jan 2009 and May 2014. Note that 1119 of them (23.9%) are draws and so, they cannot be predicted under this analysis.[9]

We observe that, although the obtained percentages for metric S_S are similar, the FIFA ranking performs equal to or worse than all our proposals. The best performance is achieved by P134, which improves the FIFA ranking score by 1.4%. This means that the P134 ranking is consistent with the scores of 65 more matches than the FIFA ranking. As for the metric $\widehat{S}_S$, the performances of the different rankings comes closer. The FIFA ranking improves its performance by 2.5%. However, it is outperformed by 10 of our proposals. The best one is P14, with 0.5% or 23 more consistent games outcomes than the FIFA ranking. Again P34

[9] Because of the dynamic nature of this score, we did not compute it for the cases of the Reference ranking, the modified FIFA ranking and the World Football Elo ranking.

Table 10 Retrodictive power of FIFA and proposed rankings

Proposal S	S_S (%)	$\widehat{S}_S$ (%)
FIFA Ranking	54.1	56.6
P1 (+ home-away factor, − conf.)	54.2	56.7
P2 (+ excl. friendlies)	54.6	56.2
P3 (+ change in points for win)	54.1	56.2
P4 (− point depreciation)	55.1	57.0
P12	54.2	56.2
P13	54.2	56.7
P14	55.4	57.1
P23	54.7	56.4
P24	54.9	56.7
P34	55.1	56.8
P123	54.4	56.2
P124	54.8	56.7
P134	55.5	56.9
P234	55.1	56.9
P1234	55.0	56.7

and P134 outperform the FIFA ranking. Note that in all cases, $\widehat{S}_S$ is greater than S_S, which strongly supports the correction by the home-away effect.

Finally, a similar analysis is conducted but only restricted to the 64 matches of the 2014 World Cup, without taking into account extra time and penalties. We grant a point when a match's winner is the country with a better ranking and 0.5 before each draw. The results are presented in Table 11.

In this case, the highest predictive power is in the FIFA ranking, which scores 49.5. That is, in 43 of the matches where there was a winner, it was the one with a better FIFA ranking (remember that the 13 draws of the 64 matches played, bring 6.5 points). Of our proposals, those with the best performance are P14 and P3 both with a score of 46.5, that is, in 40 of 51 games the country with a better ranking was the winner. In this regard, note that the results of Table 10 follow from analyzing 4684 games, against only 64 of the World Cup. Also, consider that World Cup games are biased by the FIFA ranking (seeded countries are determined by this ranking). The fact that the reference ranking (which is arguably the best predictor) has a performance of 43.5 speaks of the unpredictability of a tournament with a small number of games.

Overall, we can conclude after the analysis of this section that there exists ample space for improving the FIFA ranking used in previous World Cups using simple modifications, with our proposals constituting a reasonable starting point for the discussion.

5 Definition of groups in World Cup finals draw

The main purpose of the FIFA ranking is to determine the seeded teams for the draw that defines the group members for the group stage of the World Cup finals. Recall that in this stage, the teams only play against other members of the same group. Below we analyze the

Table 11 Comparison using 2014 World Cup Matches

Proposal S	Score$_S$	Percentage$_S$
FIFA Ranking (October 2013)	49.5	77.3
P1 (+ home-away factor, − conf.)	42.5	66.4
P2 (+ excl. friendlies)	44.5	69.5
P3 (+ change in points for win)	46.5	72.7
P4 (− point depreciation)	45.5	71.1
P12	40.5	63.3
P13	41.5	64.8
P14	46.5	72.7
P23	44.5	69.5
P24	40.5	63.3
P34	45.5	71.1
P123	40.5	63.3
P124	37.5	58.6
P134	45.5	71.1
P234	41.5	64.8
P1234	38.5	60.2
Elo	44.5	69.5
Modified FIFA ranking (Oct 2013)	48.5	75.8
Reference ranking	43.5	68.0

draw procedures for creating these groups and propose a more balanced method using an integer programming model.

5.1 Deficiencies of the 2014 World Cup draw system

Under the FIFA system for the 2014 World Cup finals, the groups into which the teams are sorted for the group stage are formed by a random draw. The 32 teams in the World Cup finals are first divided up into four pots. Pot one contains the eight seeded teams, which include the seven top-ranked teams at the time of the draw plus the team from the host country. The other three pots contain the rest of the teams distributed by geographic regions. At the 2014 World Cup, pot two contained the two unseeded South American teams plus five African teams for a total of seven, pot three consisted of 8 North American and Asian teams, and pot four had the nine non-seeded European teams. To level the pots at eight teams each, one of the European teams (Italy) was transferred by draw to pot two. The resulting composition of the four pots is shown in Table 12.

The groups are then formed by drawing one of the eight teams from each of the four pots into each of the eight groups, each group thus ending up with four teams. Additional rules were imposed to prevent three European teams or two South American teams being drawn into the same group. This procedure promotes geographic diversity but is unfair to the higher-ranked teams within a given pot, which have a high probability of being placed in the more difficult groups. For example, in 2014 the United States was negatively affected in that the team began in pot three with much weaker teams, which under the procedure automatically meant it would not be in the same group with any of them and thus was that much more likely in the group stage to be up against relatively strong sides. In particular,

Table 12 Composition of pots and FIFA ranking for the 2014 World Cup finals draw (number in parentheses is relative rank)

Pot 1	Pot 2	Pot 3	Pot 4
11 Brazil (11)	9 Italy (9)	13 USA (13)	8 Netherlands (8)
1 Spain (1)	12 Chile (12)	24 Mexico (23)	10 England (10)
2 Germany (2)	17 Ivory Cost (17)	31 Costa Rica (24)	14 Portugal (14)
3 Argentina (3)	22 Ecuador (21)	34 Honduras (27)	15 Greece (15)
4 Colombia (4)	23 Ghana (22)	44 Japan (28)	16 Bosnia (16)
5 Belgium (5)	32 Algeria (25)	49 Iran (29)	18 Croatia (18)
6 Uruguay (6)	33 Nigeria (26)	56 Korea (30)	19 Russia (19)
7 Switzerland (7)	59 Cameroon (32)	57 Australia (31)	21 France (20)

Table 13 2014 World Cup group assignments

Group	Teams			
A	11 Brazil (11)	18 Croatia (18)	24 Mexico (23)	59 Cameroon (32)
B	1 Spain (1)	8 Netherlands (8)	12 Chile (12)	57 Australia (31)
C	4 Colombia (4)	15 Greece (15)	17 Ivory Coast (17)	44 Japan (28)
D	6 Uruguay (6)	31 Costa Rica (24)	10 England (10)	9 Italy (9)
E	7 Switzerland (7)	22 Ecuador (21)	21 France (20)	34 Honduras (27)
F	3 Argentina (3)	16 Bosnia (16)	49 Iran (29)	33 Nigeria (26)
G	2 Germany (2)	14 Portugal (14)	23 Ghana (22)	13 USA (13)
H	5 Belgium (5)	32 Algeria (25)	19 Russia (19)	56 Korea (30)

it would mean that the US will face at least two and possibly three of the top-20 seeded teams. Another team disadvantaged by the procedure was the UEFA team transferred from pot four into pot two: such a team (which turned out to be Italy) was also guaranteed to face at least two and possibly three top-20 seeded teams. The final group assignments for the 2014 World Cup are shown in Table 13. Note that for each team, the preceding number is its FIFA ranking as of October 2013 when the seeded teams were named while the number following (in parentheses) is its relative ranking among the 32 teams that qualified for the finals.

To measure the overall difficulty of a group, we summed the rankings of its teams; the lower the resulting total, the better the group's teams (in average). In addition, to account for the heterogeneity within a group, we computed the range of the rankings of the teams within the group (the difference between the maximum and the minimum ranking): the lower the range, the more competitive the group. The results of these metrics for the 2014 World Cup are depicted in Table 14, revealing major disparities between the eight groups. According these measures groups D and G are arguably the most competitive, as they have high overall qualities and small ranges, i.e. they include rather good and homogeneous teams. Groups A, F and H, on the other hand, have low overall qualities and high ranges. Note that Italy and the United States, our two examples in the previous paragraph, ended up in rather competitive groups. (Italy was eliminated in the group stage while the United States came in second, thus advancing to the knockout stage.)

Although these deficiencies have been actively discussed in popular media (e.g. see Guyon 2014), the academic literature is rather scarce. The exception is Guyon (2015) who performs

Table 14 Ranking sum and range for 2014 World Cup groups, using the FIFA ranking

Group	A	B	C	D	E	F	G	H
Metric								
Ranking Sum	112 (84)	78 (52)	80 (64)	56 (49)	84 (75)	101 (74)	52 (51)	112 (79)
Range	48 (21)	56 (30)	40 (24)	25 (18)	27 (20)	46 (26)	21 (20)	51 (25)

In parenthesis, the performance measured using the relative ranking

a rigorous analysis on the probabilities of being allocated to a competitive group. The analysis reveals great unfairness in this respect, concluding that the procedure was seriously biased against teams such as Chile and USA. Guyon (2015) then discusses alternative procedures attempting to keep the practicalities of the draw, while improving its outcome. These alternative procedures include the sequential list all the acceptable groups, draw continent first and then the teams, and adding and S-curve-type constraint. The author provides an example of a draw that considerably improves FIFA's outcome. To our knowledge, this was to date the only academic effort proposing improvements to the FIFA procedure. In what follows, we develop an alternative proposal based on an integer linear programming model aimed at correcting the imbalances in the different groups' difficulty levels.

5.2 Proposals for improvement and results

In brief terms, we first formulate a model whose main set of variables indicate the group each of the participating teams is assigned to, as well as a set of logical (all groups must have four teams, each team may be assigned only to one group, etc.) constraints, as well as those that enforce meeting FIFA's geographical separation principle. The objective function of said model is to minimize the difference between the maximum and the minimum ranking sums of each group's members. Then, we formulate a second model that imposes that the difference between the maximum and minimum ranking sums must be that found by the first model (by adding an additional constraint), and focuses on minimizing the difference between the maximum and minimum range across all groups. A formal presentation of this second model (which generalizes the first model) is given in the "Appendix A".

The model was implemented in the Julia language (Bezanson et al. 2017), using the package JuMP (Dunning et al. 2017), in conjunction with Gurobi (Gurobi LLC 2018). It was solved in a few seconds for all of the instances considered. The groups obtained using the FIFA and the relative rankings are shown in Tables 15 and 16, respectively.

In both cases the group's overall quality are considerably more similar than those formed by the FIFA draw. With the FIFA ranking, the model obtains an optimal objective value indicating a difference between the highest and lowest ranking sums of $f^* = 1$ and standard deviation 0.5, whereas with the groups defined by the FIFA draw, the difference was 60 and the standard deviation 21.6. Guyon (2015) provide an example of a draw using one of his suggestions, obtaining a difference of 26 and standard deviation 9.6. With the relative ranking, the optimal objective value obtained by our model was $f^* = 0$ and standard deviation 0, whereas with the groups defined by the FIFA draw, the difference was 35 and the standard deviation 13.0, and in Guyon's example the difference was 14 and the standard deviation was 4.0. Note that when the relative ranking is used, the ranges of the groups differ in at most 2 (clearly when the FIFA ranking is used this is not possible). It is important to note, however, that the draw of the groups is traditionally a central part of the preliminary activities

Table 15 2014 World Cup groups obtained with model using FIFA ranking

Group	A	B	C	D	E	F	G	H
	Brazil	Spain	Argentina	Germany	Netherlands	Uruguay	Chile	Ecuador
	England	USA	Croatia	Ivory Coast	Colombia	France	Italy	Mexico
	Russia	Korea	Costa Rica	Bosnia	Portugal	Ghana	Switzerland	Belgium
	Japan	Greece	Algeria	Iran	Cameroon	Honduras	Australia	Nigeria
Rnk. sum	84	85	84	84	85	84	85	84
Range	34	55	29	47	55	28	50	28

Table 16 2014 World Cup groups obtained with model using relative ranking

Group	A	B	C	D	E	F	G	H
	Brazil	Chile	Spain	Netherlands	Germany	Argentina	Russia	Colombia
	Portugal	Switzerland	Ecuador	Uruguay	Mexico	Ivory Coast	Belgium	England
	Italy	Bosnia	France	Ghana	Nigeria	Croatia	USA	Algeria
	Cameroon	Australia	Costa Rica	Korea	Greece	Japan	Iran	Honduras
Rnk. sum	66	66	66	66	66	66	66	66
Range	23	24	23	24	24	25	24	23

of the World Cup and FIFA organizes it as a TV show broadcast worldwide. The procedures by FIFA and Guyon (2015) are designed as to conserve the randomness and practicalities of drawing the groups in such a TV show. Whereas the implementation of an optimization model might seem far from such a tradition, the model could be used to identify a certain number of best solutions and then perform the draw among such solutions.

6 Conclusions

This paper reveals a number of major deficiencies in the FIFA ranking methodology used between 2006 and 2018, used for drawing teams in the group phase of the World Cup finals, and offers a set of simple proposals for improving it based on the results of mathematical models developed for this study and applied to a range of international match data. As regards the deficiencies, our empirical investigation shows that an important feature that ought to be considered is the home-away status. More importantly, said feature is not considered in current modifications to the ranking system, which ought to influence the draw for the next World Cup. Additionally, up until the 2018 World Cup, teams were allowed to gain scheduling advantages in the initial stage of the World Cup finals by avoiding friendly games in the year previous to the definition of the schedules. This is so because of the relatively few points granted for friendlies and the manner in which the points for all matches are averaged. Our work shows that in the 2014 World Cup, teams such as Belgium, Switzerland and Colombia benefited significantly from these rules.

To determine the importance of the different variables in national team game outcomes, we devised multinomial logistical regression models which were applied intensively to a broad match data bank. The results pointed up clearly how fundamental was the home-away factor due to the great advantage enjoyed by the team that plays at home. This, together with the difference in ranking between the two sides in a game, were found to be the pair of factors that best fit the historical match data. It was calculated that home advantage was comparable to having a superior world ranking of 30 positions. Other factors, such as the average strength of the confederation a team belongs to or playing on one's "home" continent, proved to have much less predictive ability.

As for improvements recommendations, our findings form the basis for a number of simple proposals. These consist in incorporating the home-away factor, eliminating the confederation strength variable, dropping friendly matches entirely, restructuring the points awarded for match results by reducing the value of a win from three points to two, and doing away with the point depreciation system based on how long previously a game was played.

To test the performance of our proposals and compare them with the ranking system used by FIFA between 2006 and 2018, we use three metrics of distance between a ranking and a reference ranking created using simulations. The result of the testing was that, for each metric, several of our proposals improve upon the FIFA ranking procedure, while a handful (seven) of proposals improve upon said procedure in all metrics at the same time.

Starting in August 2018, FIFA is using an Elo-based procedure for computing its ranking. Said formula addresses some deficiencies of the ranking system in ways comparable to the proposals included in this work. Although, in our opinion, this modification should improve the performance of the FIFA ranking, it falls short on a number of aspects. For example, it fails to consider the home-away factor, which has been shown to be quite important in predicting the outcome of a match. In addition, unlike the modern Elo ranking systems, the new formula does not consider goal difference in assigning points to a match. While there are others critiques to the new formula, their impact will ultimately depend on FIFA's implementation which we are yet to see. We tested numerically the performance of the World Football Elo Raking, which arguably should overperform FIFA's implementation (as it considers factors neglected by FIFA, e.g. home-away status), and found that it consistently outperformed our proposals in all metrics, which speaks of the potential benefits for FIFA from fully embracing the Elo method.

We also developed a mixed integer linear programming model to generate groups of teams for the World Cup finals group stage that maintain FIFA's geographic criteria while improving the balance between the different groups' relative strength or level of difficulty. Applying the model to the results of the 2014 World Cup, and defining the overall quality for each group as the sum of the rankings of its individual team members, we were able to create groups whose relative strength was much better balanced than that of the groups defined by FIFA. All this, while also reducing the range of the ranking of teams within a group, for all groups. As mentioned above, FIFA modified the formation of pots to draw the groups for the 2018 World Cup Russia according to their ranking: the seven best ranked teams plus the host country were put in the first pot; the other three pots contained the teams ranked 9–16, 17–24, and 25–32. We see this as an improvement, although it might still lead to unbalanced groups, due to the deviations from the ideal solution that can inherently occur in the traditional random draw.

Overall, we foresee the academic debate may play an important role in future improvements for both the draw and the ranking. These problems turn even more relevant under the increased importance of the FIFA ranking in the pots formation for the World Cup draws and other confederation-level tournaments (e.g. Copa America 2019).

Acknowledgements We would like to sincerely thank two reviewers for their valuable suggestions that allowed us to considerably improve a preliminary version of this work. We also thank ISCI, Chile (CONICYT

PIA FB0816) for its support. The second author was partially financed by ANPCyT PICT Grant 2015-2218 (Argentina) and UBACyT Grant 20020170100495BA (Argentina).

Appendix A: Mixed integer linear programming model

Sets

- G: groups.
- C: confederations.
- I: teams.
- S: seeded teams ($S \subset I$).
- J_c: teams in confederation c ($J_c \subset I, c \in C$).

Parameters

- R_i: ranking of team i ($i \in I$).
- L_c: minimum number of teams of confederation c in each group ($c \in C$).
- U_c: maximum number of teams of confederation c in each group ($c \in C$).
- SR: minimum accepted sum ranking difference.
- $\hat{R}$: upper bound on the ranking of any team.

Decision variables

$$x_{ig} = \begin{cases} 1 \text{ if team } i \text{ is assigned to group } g \\ 0 \text{ otherwise} \end{cases}$$

$$x_{ig}^{\max} = \begin{cases} 1 \text{ if team } i \text{ is assigned to group } g \text{ has the highest ranked team} \\ 0 \text{ otherwise} \end{cases}$$

$$x_{ig}^{\min} = \begin{cases} 1 \text{ if team } i \text{ is assigned to group } g \text{ has the lowest ranked team} \\ 0 \text{ otherwise} \end{cases}$$

- w_{min}: ranking sum of the group with the lowest ranking sum value.
- w_{max}: ranking sum of the group with the highest ranking sum value.
- z_{min}: range of the group with the lowest range.
- z_{max}: range of the group with the highest range.

Objective function

$$\min f = z_{max} - z_{min} \tag{A.1}$$

Constraints

$$\sum_{g \in G} x_{ig} = 1 \quad \forall\, i \in I \tag{A.2}$$

$$\sum_{i \in I} x_{ig} = 4 \quad \forall\, g \in G \tag{A.3}$$

$$\sum_{i \in S} x_{ig} = 1 \quad \forall\, g \in G \tag{A.4}$$

$$\sum_{i \in J_c} x_{ig} \geq L_c \quad \forall\, g \in G, c \in C \tag{A.5}$$

$$\sum_{i \in J_c} x_{ig} \leq U_c \quad \forall\, g \in G, c \in C \tag{A.6}$$

$$w_{min} \leq \sum_{i \in I} R_i x_{ig} \tag{A.7}$$

$$w_{max} \geq \sum_{i \in I} R_i x_{ig} \tag{A.8}$$

$$w_{max} - w_{min} \leq SR \tag{A.9}$$

$$x_{ig}^{\min} + x_{ig}^{\max} \leq x_{ig} \quad \forall i \in I, g \in G \tag{A.10}$$

$$\sum_j R_j x_{jg}^{\max} \geq R_i \, x_{ig} \quad \forall i \in I, g \in G \tag{A.11}$$

$$\sum_j (\bar{R} - R_j) \, x_{jg}^{\min} \geq (\bar{R} - R_i) \, x_{ig} \quad \forall i \in I, g \in G \tag{A.12}$$

$$\sum_{i \in I} x_{ig}^{\max} = 1 \quad \forall \, g \in \, G \tag{A.13}$$

$$\sum_{i \in I} x_{ig}^{\min} = 1 \quad \forall \, g \in \, G \tag{A.14}$$

$$z_{min} \leq \sum_{i \in I} R_i (x_{ig}^{\max} - x_{ig}^{\min}) \tag{A.15}$$

$$z_{max} \geq \sum_{i \in I} R_i (x_{ig}^{\max} - x_{ig}^{\min}) \tag{A.16}$$

$$x_{ig}, \, x_{ig}^{\max}, \, x_{ig}^{\min} \in \{0, 1\} \quad \forall i \in I, g \in G \tag{A.17}$$

$$w_{min}, \, w_{max}, \, z_{min}, \, z_{max} \geq 0 \tag{A.18}$$

Constraints (A.2) ensure that every team is assigned to exactly one group. Constraints (A.3) specify that each group contains exactly four teams while constraints (A.4) require that one of the four teams is a seed. Constraints (A.5) and (A.6) impose upper and lower bounds on the number of teams from a single confederation assigned to a single group (in the 2014 World Cup, only one team from each confederation was allowed in a group except for the European confederation, in which case the limit was two; and at least one UEFA team per group is necessary). Constraints (A.7)–(A.8) aid in calculating the group's minimum and maximum ranking sums, while constraint (A.9) imposes a bound on the difference of such values. Constraints (A.10)–(A.14) compute the range of each group, and constraints (A.15)–(A.16) aid in calculating the group's minimum and maximum ranges. Constraints (A.18) defines the nature of the variables, and finally, the objective function (A.1) minimizes the difference between the maximum and minimum ranges.

Appendix B: Estimates for the predictive model of Section 2

See Table 17.

Table 17 Poisson regression outcome for predictive model in Sect. 2

Team	Attack coef.	Std. error	z-score	Defense coef.	Std. error	z-score
Algeria	0.33	0.26	1.23	− 0.06	0.23	− 0.26
Argentina	1.35	0.23	6.09	− 0.44	0.20	− 2.27
Australia	0.71	0.24	2.94	− 0.029	0.20	− 0.14
Belgium	0.85	0.23	3.64	− 0.14	0.20	− 0.67
Bosnia–Herzegovina	0.91	0.24	3.77	0.01	0.20	0.07
Brazil	1.38	0.22	6.39	− 0.83	0.21	− 4.02
Cameroon	0.40	0.25	1.57	− 0.16	0.22	− 0.75
Chile	1.10	0.22	4.90	− 0.30	0.19	− 1.57
Colombia	0.94	0.24	3.93	− 0.74	0.23	− 3.26
Costa Rica	0.54	0.23	2.30	− 0.21	0.19	− 1.11
Croatia	0.83	0.24	3.47	− 0.26	0.22	− 1.21
Ecuador	0.99	0.24	4.22	− 0.42	0.21	− 2.03
England	1.10	0.22	4.91	− 0.45	0.21	− 2.13
France	0.78	0.23	3.37	− 0.63	0.21	− 2.98
Ghana	0.74	0.24	3.12	− 0.20	0.21	− 0.97
Germany	1.44	0.22	6.62	− 0.22	0.19	− 1.15
Greece	0.41	0.26	1.62	− 0.36	0.22	− 1.65
Honduras	0.45	0.23	1.91	− 0.12	0.19	− 0.60
Iran	0.40	0.26	1.53	− 0.32	0.25	− 1.28
Italy	0.86	0.23	3.80	− 0.34	0.19	− 1.76
Ivory Coast	1.02	0.23	4.35	− 0.13	0.21	− 0.63
Japan	0.78	0.23	3.39	− 0.16	0.20	− 0.82
Korea	0.76	0.23	3.23	− 0.08	0.20	− 0.40
Mexico	0.95	0.22	4.30	− 0.43	0.19	− 2.23
Netherlands	1.19	0.22	5.45	− 0.59	0.21	− 2.78
Nigeria	0.72	0.24	3.00	− 0.20	0.21	− 0.91
Portugal	0.85	0.23	3.69	− 0.56	0.22	− 2.58
Russia	0.68	0.24	2.85	− 0.57	0.24	− 2.37
Spain	1.39	0.22	6.44	− 0.75	0.21	− 3.58
Switzerland	0.72	0.24	3.00	− 0.41	0.23	− 1.82
United states	0.94	0.22	4.26	0.00	0.18	0.01
Uruguay	1.17	0.22	5.21	− 0.34	0.19	− 1.74

References

Alarcón, F., Durán, G., Guajardo, M., Miranda, J., Muñoz, H., Ramírez, L., et al. (2017). Operations research transforms the scheduling of Chilean soccer leagues and South American World Cup qualifiers. *Interfaces*, *47*(1), 52–69.

Anderson, S. P., de Palma, A., & Thisee, J.-F. (1992). *Discrete choice theory of product differentiation*. Cambridge, MA: MIT Press.

Bezanson, Jeff, Edelman, Alan, Karpinski, Stefan, & Shah, Viral B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika*, *39*(3/4), 324–345.

Coleman, B. J. (2005). Minimizing game score violations in college football rankings. *Interfaces*, *35*(6), 483–496.
Davidson, R. R. (1970). On extending the Bradley–Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, *65*(329), 317–328.
Dixon, M., & Robinson, M. (1998). A birth process model for association football matches. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(3), 523–538.
Downward, P., & Jones, M. (2007). Effects of crowd size on referee decisions: Analysis of the FA Cup. *Journal of Sports Sciences*, *25*(14), 1541–1545.
Durán, G., Guajardo, M., & Sauré, D. (2017). Scheduling the South American Qualifiers to the 2018 FIFA World Cup by integer programming. *European Journal of Operational Research*, *262*(3), 1109–1115.
Dyte, D., & Clarke, S. R. (2000). A ratings based Poisson model for World Cup soccer simulation. *The Journal of the Operational Research Society*, *51*(8), 993–998.
FIFA. (2014a). 2014 FIFA World Cup Brazil technical report and statistics.
FIFA. (2014b). 2014 FIFA World Cup Brazil television audience report.
FIFA. (2017). FIFA ranking. http://www.fifa.com/fifa-world-ranking/. Accessed March 07, 2017.
FIFA. (2018). Revision of the FIFA/Coca-Cola World Ranking. https://img.fifa.com/image/upload/edbm045h0udbwkqew35a.pdf. Accessed April 24, 2019.
Gurobi Optimizer Reference Manual. http://www.gurobi.com. Accessed April 24, 2019.
Guyon, J. (2014). The World Cup draw is unfair. Here's a better way. The New York Times. https://www.nytimes.com/2014/06/05/upshot/the-world-cup-draw-is-unfair-heres-a-better-way.html. Accessed Feb 05, 2019.
Guyon, J. (2015). Rethinking the FIFA World Cup final draw. *Journal of Quantitatie Analysis in Sports*, *11*(3), 169–182.
Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning*. New York: Springer.
Huchette, Joey, & Lubin, Miles. (2017). JuMP: A modeling language for mathematical optimization. *SIAM Review*, *59*(2), 295–320.
Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *52*(3), 381–393.
Karlis, D., & Ntzoufras, I. (2008). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. *IMA Journal of Management Mathematics*, *20*(2), 133–145.
Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30*(1/2), 81–93.
Lasek, J., Szlávik, Z., & Bhulai, S. (2013). The predictive power of ranking systems in association football. *International Journal of Applied Pattern Recognition*, *1*(1), 27–46.
Lasek, J., Szlávik, Z., Gagolewski, M., & Bhulai, S. (2016). How to improve a team's position in the FIFA ranking? A simulation study. *Journal of Applied Statistics*, *43*(7), 1349–1368.
Lillestøl, J., & Andersson, J. (2011). The Z-Poisson distribution with application to the modelling of soccer score probabilities. *Statistical Modelling*, *11*(6), 507–522.
Maher, M. J. (1982). Modelling association football scores. *Statistica Neerlandica*, *36*(3), 109–118.
Martinich, J. (2002). College football rankings: Do the computers know best? *Interfaces*, *32*(5), 85–94.
McHale, I., & Davies, S. (2007). Statistical analysis of the effectiveness of the FIFA world rankings. In *Statistical thinking in sports* (pp. 77–89). CRC Press.
Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of Sports Sciences*, *4*(3), 237–248.
Rue, H., & Salvesen, O. (2000). Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *49*(3), 399–418.
Scarf, P. A., & Yusof, M. M. (2011). A numerical study of tournament structure and seeding policy for the soccer World Cup Finals. *Statistica Neerlandica*, *65*(1), 43–57.
Suzuki, K. A., Salasar, B. L. E., Leite, G. J., & Louzada-Neto, F. (2010). A Bayesian approach for predicting match outcomes: The 2006 (Association) Football World Cup. *Journal of the Operational Research Society*, *61*(10), 1530–1539.
Van Haaren, J., & Van den Broeck, G. (2015). Relational learning for football-related predictions. In *Latest advances in inductive logic programming* (pp. 237–244).
World Football Elo Ratings. http://www.eloratings.net. Accessed April 24, 2019.

Affiliations

Sebastián Cea[1] · Guillermo Durán[2,3,4,5] · Mario Guajardo[6] · Denis Sauré[1] · Joaquín Siebert[1] · Gonzalo Zamorano[1]

Sebastián Cea
secea@ing.uchile.cl

Mario Guajardo
mario.guajardo@nhh.no

Denis Sauré
dsaure@dii.uchile.cl

Joaquín Siebert
joaquin.siebert@ug.uchile.cl

Gonzalo Zamorano
gonzaloz@dii.uchile.cl

1 Department of Industrial Engineering, University of Chile, Beauchef 851, Santiago, Chile

2 Department of Mathematics and Calculus Institute, FCEN, University of Buenos Aires, Buenos Aires, Argentina

3 CONICET, Buenos Aires, Argentina

4 Department of Industrial Engineering, University of Chile, Santiago, Chile

5 Ciudad Universitaria, Pabellón II, C1428EGA Buenos Aires, Argentina

6 Department of Business and Management Science, NHH Norwegian School of Economics, Helleveien 30, 5045 Bergen, Norway