ORIGINAL PAPER

Kendall correlations and radar charts to include goals for and goals against in soccer rankings

Roy Cerqueti[1,2] · Raffaele Mattera[1] · Valerio Ficcadenti[3]

Abstract

This paper deals with the challenging themes of the way sporting teams and athletes are ranked in sports competitions. Starting from the paradigmatic case of soccer, we advance a new method for ranking teams in the official national championships through computational statistics methods based on Kendall correlations and radar charts. In detail, we consider the goals for and against the teams in the individual matches as a further source of score assignment beyond the usual win-tie-lose trichotomy. Our approach overcomes some biases in the scoring rules that are currently employed. The methodological proposal is tested over the relevant case of the Italian "Serie A" championships played during 1930–2023.

Keywords Ranking · Sport statistics · Radar charts · Football

1 Introduction

Sports competitions are often structured in official championships, where individual athletes or sporting teams compete to win. In such championships, one has a list of matches where the same teams/players play different games under a plethora of rules to be respected. In some cases, the team/player losing a match is eliminated from the competition—like in Grand Slam tennis tournaments, such as Wimbledon, the US Open, the Australian Open, and the French Open, where the winning cup is

✉ Raffaele Mattera
  raffaele.mattera@uniroma1.it

  Roy Cerqueti
  roy.cerqueti@uniroma1.it

  Valerio Ficcadenti
  ficcadv2@lsbu.ac.uk

1 Department of Social and Economic Sciences, Sapienza University of Rome, Rome, Italy

2 GRANEM, University of Angers, Angers, France

3 School of Business, London South Bank University, London, UK

assigned at the final game to one of the last two surviving players. In other cases, all the teams/players play the same number of matches, and the winner comes out from the outcomes of all the matches—this is the case of football, with the official championships of European countries like Serie A in Italy, Premier League in the UK, Ligue 1 in France and Campeonato Nacional de Liga de Primera División in Spain. In the former case, there is no real need to quantify the performance of the players to identify the winner and the other positions in the final ranking—the winner is the player who wins the final match, while her/his competitor takes a "silver medal". In the latter case, one has to state some rules that assign a score to the playing teams at the end of each match. The analysis of the existing scoring rules and the proposal of new criteria—more reasonable, from different perspectives—offers room to carry out scientific research at a methodological level but also in the context of applications, see Sziklai et al. (2022) for an overview on tournaments' efficacy.

In general, sports statistics is a widely acknowledged field of science (see, e.g. Albert and Koning 2007). One of the most famous papers in sports statistics is Reep and Benjamin (1968), where the authors analysed more than 2500 football matches and found that fewer passes are associated with a higher probability of a goal. This paper is the root of the football philosophy of the so-called "long ball system", for which the ball should be kicked over long distances to avoid a high number of passes. There was (and still is) a long debate on Reep and Benjamin's results regarding the presence of some biases in the analysis. From our point of view, debating the outcomes of the analysis does not affect the universal validity of Reep and Benjamin's research question. Intuitively, statistics in sports might be efficiently exploited to advance methods to predict the outcomes of different matches. On this, Baker and Scarf (2006) face the case of 20 annual sporting contexts by including the heterogeneity of the prediction criteria in their investigation. More recently, Mattera (2023) provides a forecasting exercise of football outcomes by employing score-driven models. Still, in the context of football, Heuer et al. (2010) advance a Poisson Process-based model for predicting the outcomes of football matches.

One can deal with the players' scores and performance from a perspective still related to the outcomes. In this respect, Volf (2009) provides a view of the scores in sports matches as the realizations of a point process based on the plethora of elements surrounding matches and players. Along the same line, Gabel and Redner (2012) deal with the scoring procedure of basketball games and elaborate on a random walk-type stochastic process behind the evolution of such a procedure over time. According to Volf (2009), Higham et al. (2014) identify the performance indicators for the case of rugby by highlighting their roles in the formation of the scores of the teams. Also, Boys and Philipson (2019) discuss the ranking procedures of sportsmen in the special context of cricket. A relevant contribution is Sandri et al. (2020), where the authors explore game performance variability through Markov switching models. In Ausloos (2024), the author offers a new perspective on the way the final ranking of cyclist rides should be carried out. The interested reader is also signposted to, e.g., Strauss and Arnold (1987), Merritt and Clauset (2014), Migliorati et al. (2023) and references therein contained.

This paper adds and contributes to the literature on scoring procedures with an application to the relevant case of football championships. Specifically, we propose

a novel method for assigning a score to the teams to identify the winner and the final ranking of every season of the considered championship. In doing so, we are close to several studies dealing with the analysis of the performance and the scoring procedures in football matches. We mention Ausloos et al. (2014), where the authors deal with the analysis of the structure of the ranking when considering UEFA and FIFA championships at a country level. Ausloos (2014) provides a view of the football rankings as unified frameworks through a rank-size analysis, with specific attention to the illustrative power of the Lavalette law. Ribeiro et al. (2010) build a model based on random walks for describing the scores of the soccer leagues. More recently and on the same line, Vernon-Carter et al. (2023) present soccer leagues as competitive complex systems where competitiveness can be measured through the scoring of individual soccer teams. We refer the interested reader also to Glickman and Stern (2005), Mendes et al. (2007), Thakkar and Shah (2021), Ficcadenti et al. (2023), Cefis and Carpita (2024). The methodological proposal is tested on the paradigmatic case of all the seasons of the Italian championship, Serie A.

Our starting point is the disappointing evidence that the current rule for Serie A admits the existence of special circumstances leading to the mathematical assignment of the winner's cup to a specific team, disregarding some matches to be played to officially end the season. That happens because of simple arithmetic consequences in the rule set. Indeed, the rigidity of the score on the basis of the trichotomy win-tie-lose makes recoveries be impossible when the distance in scores is large enough. This often implies a deteriorated level of game qualities in the matches played toward the end of the season, when the team that is mathematically the winner of the championship season starts losing matches against low-level teams. An example can be taken from the 2018–2019 championship. Juventus won the Serie A title with five games to spare. They clinched the title on April 20, 2019, after a 2–1 victory over Fiorentina, which put them 20 points clear of their closest challengers at the time, Napoli, with only 15 points left to play for. This significant points gap made it mathematically impossible for any other team to catch up with Juventus in the season's remaining fixtures.

Following this victory, the matches that Juventus played for the rest of the season lacked the same competitive edge, at least from their perspective, as the title was already secured. They tied against Internazionale, Torino and Atalanta and lost against Roma first and Sampdoria later. This situation illustrates the potential downside of having a team win the league so early: the intensity and competitive nature of their remaining matches can diminish, potentially affecting the overall quality of the league's competition towards the end of the season. While Juventus continued to compete professionally, the urgency and high stakes associated with their matches were notably reduced, aligning with the concerns expressed about the impact of early championship wins on the quality of the game.

We hypothesize a novel scoring rule for which scored and conceded goals play a relevant role in determining the final ranking of the considered championship season. In so doing, we are not far from Cerqueti et al. (2022), where there is an application on football data to rank teams according to their goals. The following approach is of data science-computational type. We consider the sample of all the Italian Serie A championship seasons, from 1929–1930 to 2022–2023, with specific

reference to the official final rankings of the teams. We then implement a four-step procedure based on a combination of Kendall τ and radar charts. The procedure leads to what we call *New Rankings* of the seasons, which are grounded on a different way to assign the scores to the teams also including the goals—hence, removing or reducing the cases of mathematical certainty of being the winner well-before the end of the championship. The approach is quite close to Gorgi et al. (2023), where the authors advance a pair of comparison methods for reconstructing the rankings of the football championships in the presence of forced interruption. This is particularly relevant in that there has been evidence of different cases of interrupted championship—the most recent one being linked to the COVID-19 pandemic. However, the quoted paper uses Kendall τ only to test the proposed framework's validity. Differently, we here base the analysis on such a statistical correlation measure along with a radar chart-based evaluation of the multidimensional performance of specific entities—the seasons of the considered championship, in our case.

The paper is organized as follows. Section 2 describes the considered dataset. Section 3 presents the methodology used for dealing with the data, outlining the four-steps procedure used to obtain the New Rankings. Section 4 collects the main results of the analysis, along with some related discussions. The last section offers some conclusive remarks and traces lines of future research.

2 Data

This study utilises a comprehensive dataset encompassing the outcomes of football matches from the Italian Serie A championship, spanning from its inception, 1929–1930, to the present day, amounting to $M = 90$ seasons.[1] For brevity, we refer to the seasons by mentioning only the related last year, so that e.g. 1929–1930 becomes 1930 for us.

The summary statistics of the dataset in Table A1 offers a glimpse into the volume and nature of the data analysed. These statistics encompass various metrics to understand football dynamics, including the number of goals scored by home and away teams ("Goals For" identified with GF and "Goals Against" with GA), points accumulated throughout the season, and the number of wins, draws, and losses. The dataset comprises records from the 1930 season of Serie A, totalling 34 matches, to the 2023 season, with 38 matches played. For each match, the dataset records the date, the teams involved, the goals scored by each team, and the final result (win, draw, or loss), allowing for a detailed examination of team performances at the season level and the evolution of the championship over time.

In addition to the standard metrics, we have developed rankings based on GF and GA each season, identified by the variables GF_r and GA_r, respectively. These rankings provide an alternative perspective on team performance at the end of the season, emphasizing, for example, offensive and defensive capabilities beyond the traditional league standings.

---

[1] A simple counting of the years between 1929 and 2023 (last ended season) would not work because of the Championship suspensions occurred during the second world war and the COVID-19 pandemic.

In preparing the data for analysis, several preprocessing steps were undertaken to ensure the dataset was fit for purpose. These steps included verifying and correcting match outcomes, normalising team names to account for historical changes, and identifying and treating any missing or incomplete records. Finally, only official rankings (including penalties applied by authorities), GF and GA were employed to serve as the primary basis for our analysis. Developing rankings based on GF and GA for each season involved taking the total number of goals scored by and against each team, allowing us to derive GF_r and GA_r. This approach offers a nuanced view of team strategy and performance, fitting the objective of our study and complementing the official league standings with metrics highlighting each team's offensive and defensive strengths.

3 Methodology

This section illustrates the four steps of the procedure for achieving the New Rankings for the Italian Serie A championship seasons. First, we consider the unofficial rankings with teams ordered on the basis of the scored goals or, simply, *Goals For* (in decreasing order, so that rank=1 is associated with the team with the highest number of scored goals in the championship season) and conceded goals or, simply, *Goals Against* (in increasing order, so that rank=1 is assigned to the team with the lowest number of conceded goals). Thus, we have three rankings for each season on the same set of teams. Second, we compute the Kendall τ of all the possible couples of rankings, hence obtaining three values of the Kendall τ for each season. Third, we build a radar chart for each season, whose axes are associated with the three Kendall values. Therefore, a triangle describes each championship season. We compute the area of the obtained triangles, and then we suitably normalise it so that the areas range from 0 (case of all Kendall τ equals − 1) to 1 (case of all Kendall τ equals + 1). The areas of the triangles represent the target (normalised) Kendall τ. Fourth, we detect rankings with a target Kendall τ correlation with the official one. The obtained rankings are the *New Rankings*. As we will see soon, the New Rankings are often far from the official ones.

We enter the details.

3.1 Kendall τ correlation analysis

The association between official team rankings and goal metrics (GF and GA) is achieved through Kendall τ correlation analysis. The τ correlation coefficient measures the strength and direction of the association between two ranked variables. It is defined as:

$$\tau = \frac{2}{n(n-1)} \sum_{i<j} \text{sign}(x_i - x_j) \cdot \text{sign}(y_i - y_j) \tag{1}$$

where n is the number of observations, x_i and x_j are the ranks of the i-th and j-th observations for the first variable, and y_i and y_j are the ranks of i and j for the second

variable. The sign function, sign($\cdot$), returns -1, 0, or 1 depending on the sign of its argument.

In our case, we use the τ_b variant, Kendall (1945). Such a coefficient is a measure of association based on the ranks of the data and is adjusted for ties. We point out that ties occur when two or more items have the same rank. The formula defines it:

$$\tau_b = \frac{P - Q}{\sqrt{(P + Q + T)(P + Q + U)}} \tag{2}$$

being P and Q the number of concordant and discordant pairs, respectively and T and U the number of ties only in x and y, respectively. If a tie is registered for the same couple in x and y, such a tie is not considered for the value of T and U.

The τ_b coefficient accounts for ties by adjusting the denominator to reflect the number of tied ranks, which can affect the distribution of concordant and discordant pairs. This adjustment makes τ_b a more accurate reflection of the association between two variables when ties are present in the data. In datasets where ties are common—as in our case[2]—τ_b offers a more reliable correlation estimate than the standard Kendall τ coefficient, which does not adjust for ties.

This coefficient is computed for each season to analyse the relationships between the official rankings (*Rank*) and the rankings based on goals for (GF_r) and goals against (GA_r), as well as the relationship between GF_r and GA_r themselves. It is worth recalling that in our data, the ties can be met only in GF_r and GA_r, as the official ranking is built on a set of rules that avoid the presence of ties. Table 1 reports a summary of the instances considered, and Fig. 1 shows correlations over time.

The Kendall correlation coefficients are computed using Python, leveraging the `scipy.stats.kendalltau` function for correlation analysis. As we will appreciate when we introduce the radar charts, a normalization process is needed to use the correlations smoothly to form the areas. Such a normalization procedure adjusts the correlation values to a [0, 1] scale centred at 0.5, allowing for a consistent geometric interpretation across different years. The normalisation formula applied to each tau correlation coefficient, denoted as τ_b, is defined as follows:

$$\tau_{b;N} = \frac{\tau_b + 1}{2} \tag{3}$$

Here, $\tau_{b;N}$ represents the normalised correlation coefficient. By adding 1 to the original correlation coefficient τ_b, the new range starts from 0 (previously -1) to 2 (previously 1). Dividing this result by 2 adjusts the scale to range from 0 to 1. If the quantity $\tau_{b;N}$ in (3) has a value of 0.5, then we do not have correlation; values closer to 1 indicate a strong positive correlation, while values closer to 0 suggest a strong negative correlation.

To facilitate the understanding of the steps, we report in Table 2 a snapshot of the 1939 and 2023 cases, being chosen as explicative instances; in Table 3 one can see the rank correlations and their respective normalisation.

---

[2] We use the Dense Ranking Method, namely elements with the score receive the same rank.

Table 1 Summary of the pairwise rank-correlation analyses

Analysis pair	Explanation
GF_r versus GA_r	Correlation between ranks based on goals for (GF_r) and goals against (GA_r)
Rank versus GF_r	Official rank correlation with ranks based on goals for (GF_r)
Rank versus GA_r	Official rank correlation with ranks based on goals against (GA_r)
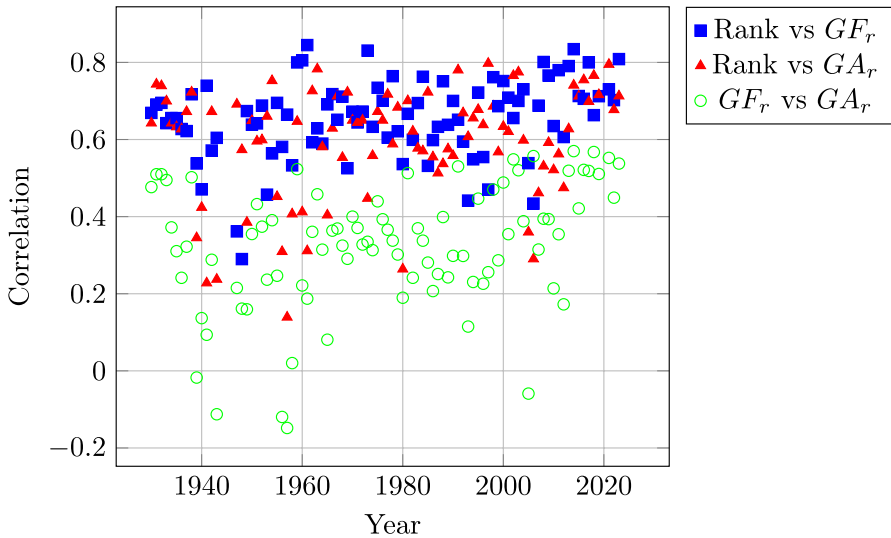
Fig. 1 The different correlation analyses are reported over the seasons. We use the non normalised version of the Kendall tau in formula (2). *Rank* is the official ranking, GA_r is the ranking when "Goals Against" is considered, and GF_r indicates the ranking when "Goals For" is considered

3.2 Mapping correlations into radar charts

In the analysis, the normalised correlation coefficients for each year are represented geometrically, forming triangles that encapsulate the relationship among different performance metrics. The triangles are constructed as radar charts, by plotting points on axes that extend from a central point, with each axis representing one of the analysis types. The process is described as follows:

1. For each year under analysis, the normalised correlation coefficients ($\tau_{b;N}$) for the selected metrics are retrieved. These coefficients range from 0 to 1, and are centred at 0.5.
2. Angles for the vertices of the triangles are calculated to distribute the analysed couples of metrics around a centre evenly. This is achieved using the formula:

$$\theta_h = 2\pi \frac{h}{n} \tag{4}$$

Table 2 Table of the values of the considered variables and rankings for the championship's seasons ended in 1939 and 2023

Data	Rank	Team	Pt	GF	GA	GF_r	GA_r
2023	1	Napoli	90	77	28	1	1
2023	2	Lazio	74	60	30	5	2
2023	3	Inter	72	71	42	2	6
2023	4	Milan	70	64	43	4	7
2023	5	Atalanta	64	66	48	3	9
2023	6	Roma	63	50	38	8	4
2023	7	Juventus	62	56	33	6	3
2023	8	Fiorentina	56	53	43	7	7
2023	9	Bologna	54	53	49	7	10
2023	10	Torino	53	42	41	11	5
2023	11	Monza	52	48	52	9	11
2023	12	Udinese	46	47	48	10	9
2023	13	Sassuolo	45	47	61	10	13
2023	14	Empoli	43	37	49	12	10
2023	15	Salernitana	42	48	62	9	14
2023	16	Lecce	36	33	46	14	8
2023	17	Verona	31	31	59	15	12
2023	18	Spezia	31	31	62	15	14
2023	19	Cremonese	27	36	69	13	15
2023	20	Sampdoria	19	24	71	16	16
1939	1	Bologna	42	53	31	2	3
1939	2	Torino	38	45	34	3	5
1939	3	Inter	37	55	37	1	7
1939	4	Genova	35	53	30	2	2
1939	5	Roma	31	39	35	5	6
1939	6	Liguria	31	35	34	7	5
1939	7	Napoli	31	30	35	11	6
1939	8	Juventus	29	28	34	12	5
1939	9	Milano	28	36	34	6	5
1939	10	Lazio	28	33	40	8	8
1939	11	Bari	27	33	46	8	9
1939	12	Novara	26	27	32	13	4
1939	13	Modena	25	32	40	9	8
1939	14	Triestina	24	23	28	14	1
1939	15	Livorno	24	40	49	4	10
1939	16	Lucchese	24	31	54	10	11

where θ_h represents the angle for the h-th vertex, n is the total number of analysis types, and h ranges from 0 to $n-1$. In our context, the three angles in radiants are 0, 2.09 and 4.18.

3. Each correlation coefficient is then used as a point on its respective axis, determined by the corresponding angle θ_h.

Table 3 Table of the normalised and non-normalised correlations for the analysis of the championship's seasons ended in 1939 and 2023

Year	Analysis type	τ_b	$\tau_{b;N}$
1939	Rank versus GF_r	0.538	0.769
1939	Rank versus GA_r	0.345	0.673
1939	GF_r versus GA_r	−0.017	0.491
2023	Rank versus GF_r	0.809	0.904
2023	Rank versus GA_r	0.713	0.856
2023	GF_r versus GA_r	0.538	0.769

4. The points are connected in sequence, forming a closed shape that, in the context of this analysis, is a triangle; one gets three vertexes thanks to the three types of analyses considered. See Fig. 2, where the examples of 1939 and 2023 cases are reported on the basis of the data presented in Table 3.

3.3 Area calculation from the resulting triangles and correlation target

The area of each triangle described above is calculated to quantify the combined strength of the correlations. Given the vertices positioned at angles θ_1, θ_2, and θ_3 with their respective normalised correlation coefficients, the Cartesian coordinates for each vertex are determined by:

$$x_h = \tau_{b;N}^{(h)} \cos(\theta_h) + \text{shift}_x \tag{5}$$

$$y_h = \tau_{b;N}^{(h)} \sin(\theta_h) + \text{shift}_y \tag{6}$$

where $\tau_{b;N}^{(h)}$ is the normalised correlation coefficient for vertex h, and $\text{shift}_x = \text{shift}_y = 10$ and used to ensure all points are positioned in the positive quadrant to simplify the area calculation.

The area of a triangle formed by the three points representing the considered normalised correlations is calculated, to provide a geometric representation of these correlations. Given the vertices coordinates (x_1, y_1), (x_2, y_2), and (x_3, y_3), the area (A) of the triangle is given by:

$$A = \frac{1}{2}\left|x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)\right| \tag{7}$$

The procedure of computing A is implemented to each of the $M = 90$ seasons considered in our dataset. An example of the calculations can be found in Table 4 for the cases 1930 and 2023.

A time series version of the calculated areas can be found in Fig. 3, where there is a clear view of the time-evolution of the considered areas.

We then hypothesise that the area of the triangles represents the (normalised) Kendall correlation targets of the New Rankings with the official ranks. So, considering goals for and against would give a ranking of the teams whose correlation with the official one is the area of the triangle of the related radar chart.
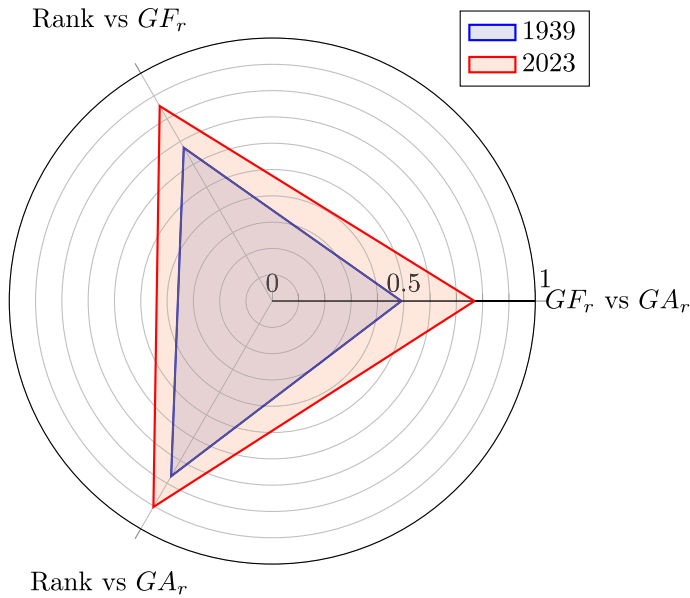
Fig. 2 The vertices are suitably labelled, and the area of the triangle is shaded to visually represent the correlations' magnitude. The radial grid lines are set at intervals of 0.1 to indicate the scale of the normalised correlation (according to Formula 3), ranging from negative correlation (zero in the graph) to perfect positive correlation (one in the graph). The shaded areas, calculated from the triangles formed by these correlations, provide a quantitative measure of the combined correlation strength among attributes for each year, in the graph being 1939 and 2023

Table 4 Angles, normalised correlations, coordinates of the vertices of the triangles and resulting areas for the seasons 1939 and 2023

	1939	2023
θ_1	0.000	0.000
θ_2	2.094	2.094
θ_3	4.189	4.189
GF_r versus GA_r	0.491	0.769
Rank versus GA_r	0.673	0.856
Rank versus GF_r	0.769	0.904
x1	10.491	10.769
x2	9.664	9.572
x3	9.616	9.548
y1	10.000	10.000
y2	10.582	10.742
y3	9.334	9.217
Area	0.531	0.921
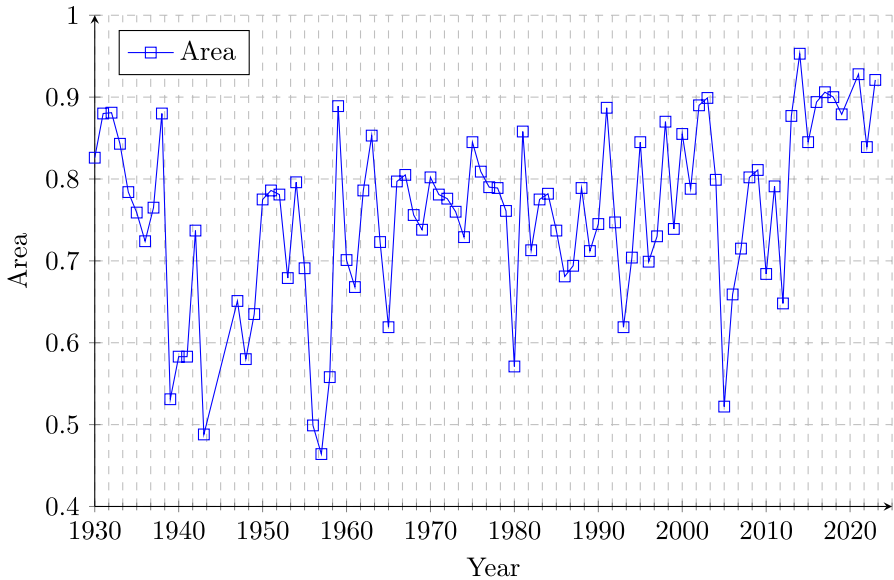
Fig. 3 The areas calculated with Eq. (7). Each point represents a season

3.4 Finding the new rankings

This section contains the computational strategy devised to discover alternative ranking systems that may more accurately reflect football championships' dynamics and a performance-centric nature in the final rankings. The goal is to pinpoint permutations of team rankings yielding Kendall τ_b correlations that match the geometric areas previously calculated with Eq. (7), suggesting a ranking better-representing team performance throughout the seasons.

3.4.1 Generating permutations and calculating Kendall's tau correlations

For a given number of teams, n, in the football championship (indicated in Table A1 as "N. Teams"), we embark on a systematic generation of permutations to simulate the possible teams' positions in the final ranking, therefore simulating various possible seasons' outcomes. Owing to computational constraints and the $n!$ increasing number of permutations with n, our exploration is confined to a select subset of permutations. In this way, we can still show here to what extent the official rankings are affected by the partially missed account of "Goal For" and "Goal Against". Specifically, we run the first 362,880 permutations for the 2023 case. Such a threshold is based on system capabilities and the aspiration to encompass a broad spectrum of potential rankings.

Within each permutation, we compute Kendall τ_b, with respect to the original ranking sequence using Eq. (2). As already said, Kendall's tau is a measure used to ascertain the ordinal association between two quantities.

One should use the original sequence of numbers between 1 and n for each season against the permuted sequence of the same set of numbers. The permuted series should be ordered according to the lexicographic criterion to simplify the process. This computation yields a distribution of τ_b values formed by the correlations associated with all the possible permutations. The target τ_b associated with a given permutation illustrates the degree of correlation of such a permutation with the original team ranking. Such a permutation of the original ranking can be viewed as the outcome of a championship where the teams are ranked according to the permuted ranking. To illustrate this statement, we refer to Fig. 4, where 362,880 permutations are evaluated with different values of n. This exemplifies the idea of having championships of n teams whose rankings are shuffled and compared to the original one, which is assumed to be $(1, \dots, n)$.

3.4.2 Identifying the optimal permutations

At the heart of our analysis lies the quest for permutations that realise Kendall correlation aligned with the target τ_b values stemming from the geometric correlation analysis. For each season, we may have different $n \in 16, 18, 20, 21$, as can be grasped from Table A1. The procedure is as follows:

1. We extract the target τ_b values from the preceding geometric analysis for the year as the area of the triangle/radar chart. We transform the various As in a τ_bs.
2. We calculate the absolute difference between each τ_b target and the various $\tau_b^{(j)}$ resulting from comparing each permutation with the target, where j is the index of permutation.
3. We isolate permutations whose $\tau_b^{(j)}$ values are nearest to the target τ_b, implementing a tolerance threshold to facilitate a significant comparison. This tolerance is derived from the rounded (to the third decimal digit) interval between consecutive lexicographically ordered permutations $\tau_b^{(1)} - \tau_b^{(2)}$ values in our permutation analysis, accommodating the inherent variability in the dataset; in fact, in this way, it depends on the $n!$ possible operations.

This methodology empowers us to single out some permutations (i.e., hypothetical rankings accounting for GA and GF) that most accurately conform to the theoretical ideals elucidated by our prior analysis. These optimal permutations shed light on alternative ranking methodologies that more faithfully mirror team performances and the competitive dynamics across the football season.
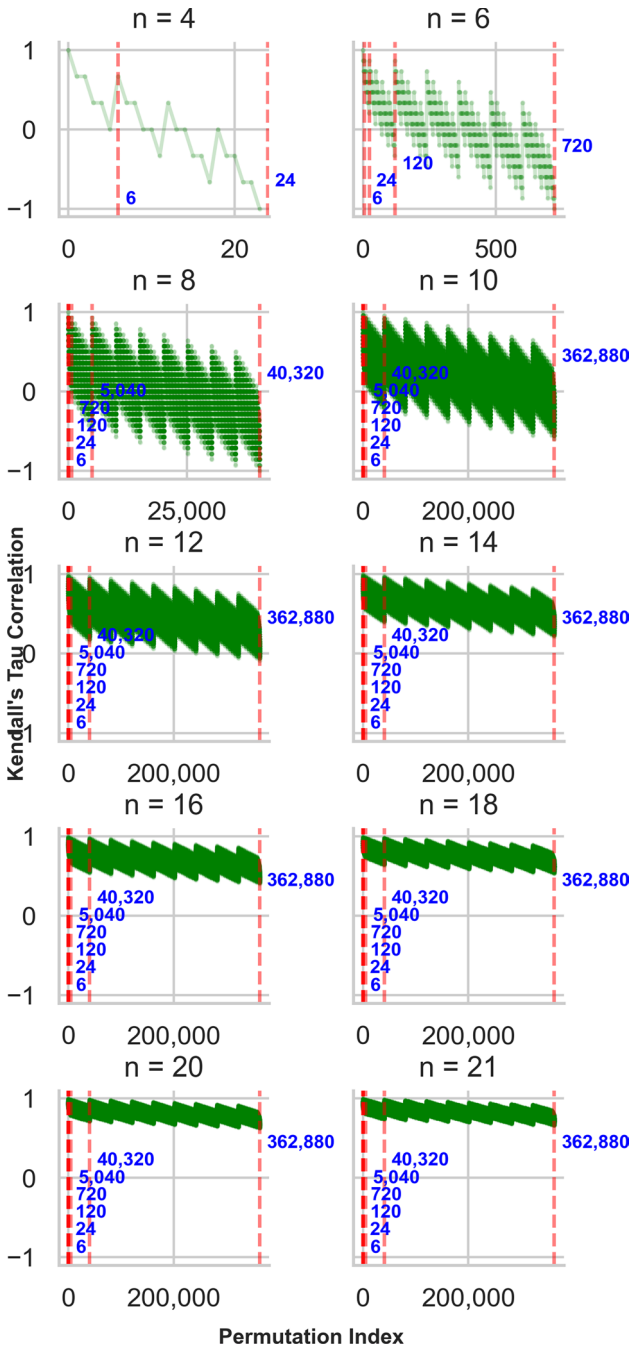
3.4.3 Implementation and challenges

The implementation was conducted using Python, with the assistance of libraries such as `itertools` for permutation generation, `scipy` for statistical computations, and `pandas` for data handling. This computational framework facilitated a thorough exploration of ranking permutations against predefined criteria, offering a fresh perspective on the assessment of football championships.

A significant challenge in this process is the identification of all the permutations that match Kendall τ_b target correlations, resulting in the target areas obtained from the triangle's geometric analysis, driving the problem to something computationally challenging. In fact, to find the best permutations that meet the case of $n = 21$, one has to explore potentially 21! possibilities, which are evidently complex and expensive.

4 Results and discussion

This section elaborates on the outcomes derived from our methodological framework and discusses their implications within the realm of football analytics, specifically in the context of Serie A. The analysis provides insightful revelations about team performance dynamics and proposes a novel perspective on ranking methodologies.

Our geometric representation of team performances allows a different view of the Serie A history and is illustrated in Fig. 3. It unveils a trend where the last ten seasons exhibit a distinguishable pattern. This observation suggests a shift towards more balanced team strategies, aiming to optimise both offensive and defensive plays (see Okada and Takagi 2008, on the impact of various strategies on GA and GF). The gap between the calculated areas and the ideal scenario (Area = 1) quantifies the extent to which the official rankings might overlook the intricate balance between goals scored (GF) and goals conceded (GA) when accounting for more granular teams' performance in forming the final ranking.

The transformation of these areas into target Kendall τ_b coefficients provides a foundation for empirical analysis. As depicted in Fig. 5, the majority of seasons align positively with the official rankings, indicating a generally robust system but not completely accounting for GA and GF. Anomalies identified in the negative range (1943, 1956, 1957) call for a closer examination of those particular seasons and potentially underline the need for a refined ranking mechanism that better captures team performance nuances.

Our exploration into the New Rankings, facilitated by the examination of permutations and Kendall τ_b correlations, highlights the potential for alternative standings that deviate significantly from the official rankings. As shown in Fig. 6 for 2023, incorporating goals scored and conceded into the rankings can result in substantial shifts in team positions. This variability underscores the impact of evaluating team performances beyond mere wins, draws, and losses, advocating for a more granular approach to ranking that acknowledges the multifaceted nature of football competitions. The case presented in Fig. 6 is already meaningful even if the number of iterations tested stops at 362,880, and to complete the exercise, one should have gone to 20!, as 20 were the team competing. Another way to observe the impact of targeting a level of Kendall τ_b that ensures capturing a more comprehensive set of
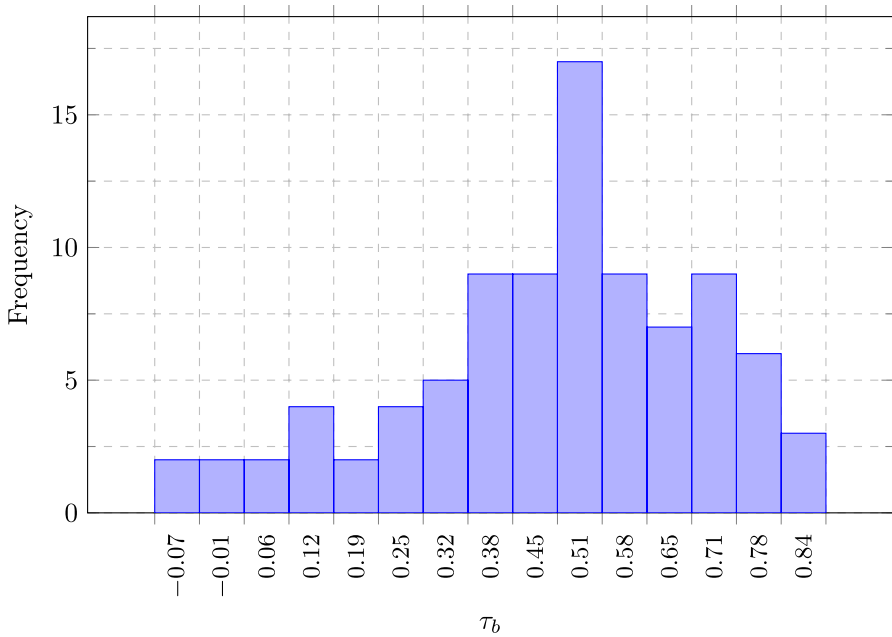
Fig. 5 Histogram of the resulting τ_b obtained from mapping the areas back to the $[-1,1]$ correlation range along the Italian Serie A history, 1930–2023

features into the ranking is proved by Fig. 7. Such a figure contains the Kendall τ_b in the cases of inversion of the first f elements as well as of the l last elements of the rankings. We take n ranging from 10 to 30 to include the cases of interest for the analysed championships. For example, if $f = 5$ and $n = 10$, we have the "reverted" ranking $(5, 4, 3, 2, 1, 6, 7, 8, 9, 10)$ and when $l = 5$ and $n = 10$, we have $(1, 2, 3, 4, 5, 10, 9, 8, 7, 6)$. One can notice that in the case of $n = 16$ teams playing in a championship, the inversion of the ranking for the last $l = 4$ teams in the ranking means pursuing a target Kendall τ_b of 0.9, indicating that with very little variations on the ranking system, teams may or may not face a relegation. This explains the impact of the inversion well when GF and GA are taken into full consideration.

By arranging team permutations in lexicographical order, we systematically explore variations from the initial ranking, incrementally adjusting team positions. In Fig. 4, the permutation index j is plotted along the x-axis, and the corresponding Kendall's tau correlation coefficient, $\tau_b^{(j)}$, is plotted along the y-axis. This arrangement reveals a pattern of regular fluctuations in $\tau_b^{(j)}$ values, manifesting as seasonal cycles across the permutation index. We argue that the number of solutions depends on the target correlation. The extreme cases of $\tau_b = -1$ or $\tau_b = 1$ are associated with singular solutions to the problem – for $\tau = -1$ being the complete reversion of the ranking, while for $\tau_b = 1$ being the original series itself. More specifically, $\tau_b = 1$ is the perfect agreement with the official ranking, representing a scenario where the permutation does not alter the original team order, highlighting the unique case where the equivalence class contains only the official ranking itself. If one slightly
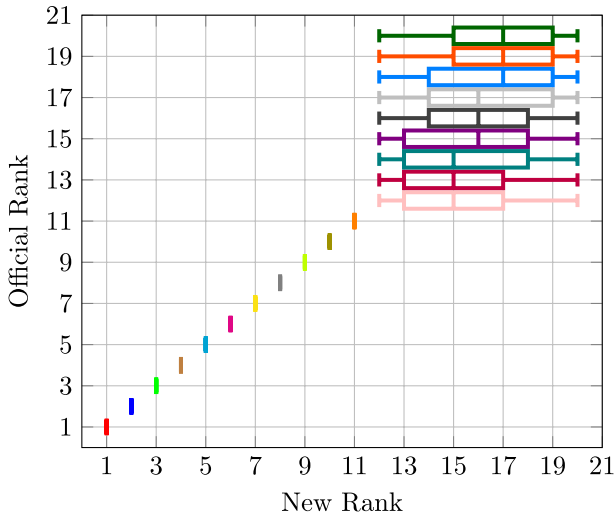
Fig. 6 In this box plot, the results for the 2023 season are reported. The target correlation is $\tau_b = 0.842962$ and the optimal correlation is $\tau_b^j = 0.842105$ with j taking some opportune values in the set $\{0, \dots, 362880\}$, being 362,880 the number of permutation tested in this case when the generated permutations are stored in lexicographic order. The whiskers of each box represent respectively min and max ranking obtained in the considered permutations to meet the target correlations. The vertical line that splits the box in two is the median obtained from the raking obtained for that position (y-axis) and the left and right sides of the box are the 25-th and 75-th percentile of ranks assigned to that ranking position. When there is a single bar, like for cases $1, \dots, 11$, it means that no changes has been recorded. It is worth recalling that here 362880 permutations are considered over 20! that should have been explored to complete the plot

modifies these two corner cases, the number of solutions that lead to the target case increases as one approaches a target $\tau_b = 0$ for the definition of Kendall correlation.

The cycles presented in Fig. 4 reflect the comprehensive range of permutations explored, with the length of each cycle corresponding to the total number of teams involved; for instance, with 18 teams, the permutation space, and hence the cycle length, expands to 18!. Due to computational constraints, the analysis presented in the figure samples only a fraction of the total permutation space.

This cyclical pattern underscores that multiple permutations yield identical Kendall τ_b correlations, indicating that multiple rankings could feasibly represent the data with equivalent statistical validity. Thus, as already said above, a specific τ_b value may correspond to a set of rankings rather than a single, unique order. This set forms an equivalence class, each member sharing the same Kendall τ_b correlation with the official ranking. The diversity within these classes illustrates the potential for alternative interpretations of team performance and rankings.

5 Concluding remarks

This study embarked on a novel exploration of Serie A football championship rankings by introducing a comprehensive methodology that integrates geometric analysis with Kendall's τ_b correlation coefficients. Through this approach, we scrutinised the
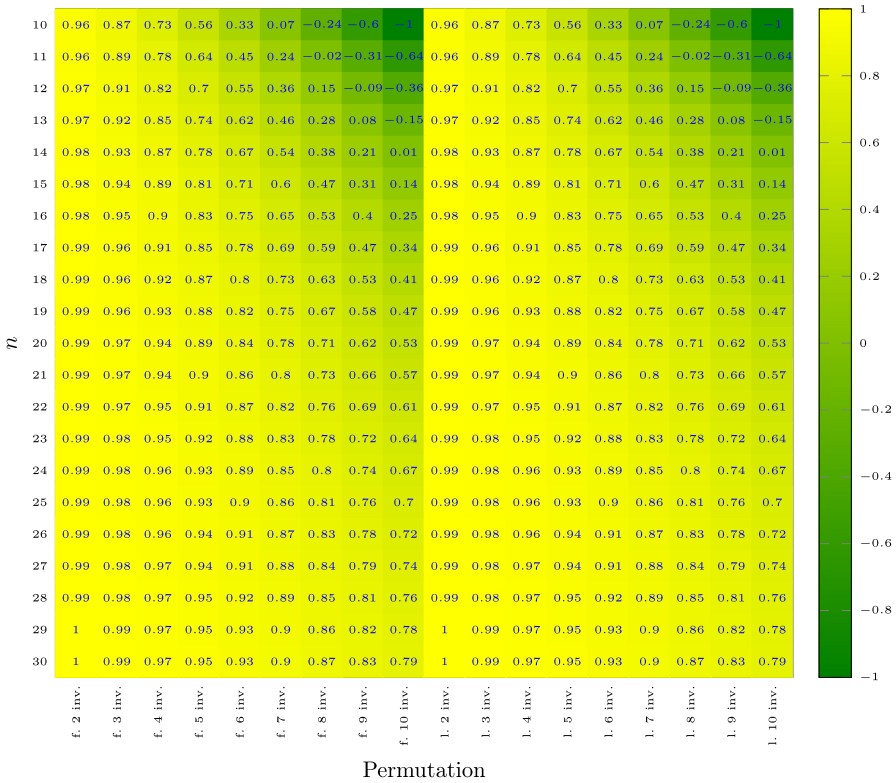
Fig. 7 The x-axis has ticks indicating that the Kendal Tau correlation reported in the cells is calculated comparing the original series 1,2,...,n (y-axis) with the series where the first ('f.') or the last ('l') k elements have been permuted, inverting their order. For example, when $n = 10$, for "f. 2 inv.", the value 0.96 is obtained by applying Formula (2) to the series (1,2,3,4,5,6,7,8,9,10) and (2,1,3,4,5,6,7,8,9,10)

alignment between official team rankings and those derived from goals scored and conceded, as well as with the official rankings, unveiling the potential for alternative rankings that might offer a deeper insight into team performance dynamics.

Our findings reveal that while the performance metrics-based rankings (here being Goals For and Against, GF and GA) broadly align with the official ranking system, there are distinct seasons where alternative ranking methodologies could provide a more nuanced understanding of team capabilities. Introducing a geometric representation to visualise the relationship between different ranking metrics not only enriches our analytical toolkit but also highlights the multifaceted nature of football competitions, where outcomes are influenced by a complex interplay of offensive and defensive strategies mirrored in the GA and GF-based rankings, here indicated with GA_r and GF_r respectively.

Furthermore, identifying equivalence classes among rankings underscores the notion that multiple valid perspectives can exist regarding team performance, challenging the singular narrative often presented by official standings. This observation invites a broader discussion on the criteria and metrics used to assess and compare

teams, suggesting that there is room for innovation in ranking methodologies that more accurately reflect the competitive landscape of Serie A football.

5.1 Implications for stakeholders

For stakeholders in the football community-ranging from team management and coaches to analysts and fans-our study offers a fresh lens through which to evaluate team performance. By considering alternative rankings, stakeholders can better understand a team's strengths and weaknesses, guiding strategic decisions from player development to game tactics. That certainly does not come to question the official outcome of a season winner, but more to award different teams because of their outstanding defensive tactics or aggressive ones manifested in GA or GF, also addressing some challenges described in Sziklai et al. (2022).

5.2 Limitations and future research

While our study contributes valuable insights into football ranking systems, it is not without limitations. The computational complexity of analysing all possible permutations of team rankings poses a challenge, necessitating further methodological innovations to explore the full permutation space efficiently. In particular, the application of suitably defined heuristics for reducing the cardinality of the set of permutations as n grows is a possible way to let the problem be tractable, no doubt. This opens the gate to more operational research-oriented studies—that are out of the scope of the present paper.

Additionally, our focus on Serie A limits the generalizability of our findings. Future research could extend this methodology to other leagues and sports, examining the universality of our observations across different competitive contexts and combining it with existing methods such as Sum of Ranking Differences, presented in Sziklai and Héberger (2020).

Moreover, incorporating other performance metrics (in fact, the axis on the radar charts can be more than three), such as aggregated player statistics or situational variables (e.g., weather conditions during matches), could enhance the robustness and relevance of alternative ranking systems.

5.3 Final thoughts

In conclusion, our study enlightens the potential for alternative perspectives in evaluating football team performances, inviting a reconsideration of conventional ranking systems. As we continue to navigate the rich and evolving landscape of sports analytics, the pursuit of more sophisticated and representative methodologies for assessing team success remains a compelling and worthwhile endeavour.

Appendix

See Table A1.

Table A1 Overview of the dataset by using the main statistical indicators

Date	N. Games	N. Teams	GA μ	GA σ	GA Max	GA Min	GF μ	GF σ	GF Max	GF Min	Pt μ	Pt σ	Pt Max	Pt Min
2023	38	20	48.700	12.244	71	28	48.700	14.298	77	24	51.500	17.993	90	19
2022	38	20	54.450	14.551	78	31	54.450	16.031	84	27	52.100	18.462	86	27
2021	38	20	58.150	14.858	92	35	58.150	16.916	90	39	52.200	20.629	91	20
2019	38	20	50.950	12.664	75	30	50.950	14.144	77	25	51.450	18.357	90	17
2018	38	20	50.850	16.362	84	24	50.850	18.196	89	29	52.850	20.551	95	21
2017	38	20	56.150	13.789	81	27	56.150	19.082	94	29	53.000	20.561	91	18
2016	38	20	48.950	12.717	76	20	48.950	14.989	83	33	52.250	17.247	91	28
2015	38	20	51.200	13.045	75	24	51.200	12.344	72	28	50.650	16.096	87	19
2014	38	20	51.750	13.924	77	23	51.750	15.096	80	28	52.500	19.449	102	25
2013	38	20	50.150	11.595	84	24	50.150	14.254	73	27	51.700	17.107	87	22
2012	38	20	48.600	11.546	69	20	48.600	12.775	74	24	51.150	14.809	84	22
2011	38	20	47.750	9.095	66	24	47.750	12.345	69	27	52.000	14.984	82	24
2010	38	20	49.600	8.852	67	34	49.600	11.385	75	27	51.900	14.736	82	29
2009	38	20	49.400	9.976	67	32	49.400	12.647	70	30	52.250	15.563	84	30
2008	38	20	48.500	9.259	62	26	48.500	13.073	72	29	51.400	15.642	85	30
2007	38	20	48.450	11.436	69	31	48.450	13.713	80	27	49.400	16.340	97	26
2006	38	20	49.550	11.218	65	24	49.550	15.683	85	24	42.550	16.330	76	0
2005	38	20	48.000	12.561	73	27	48.000	11.164	67	32	50.750	13.913	86	35
2004	34	18	45.333	12.305	70	19	45.333	14.336	68	21	46.000	17.283	82	13
2003	34	18	43.833	9.805	62	29	43.833	12.282	64	23	45.889	14.196	72	21
2002	34	18	44.778	11.477	63	23	44.778	10.184	64	29	46.167	14.940	71	18
2001	34	18	46.944	11.059	68	27	46.944	10.898	68	31	46.111	14.262	75	20
2000	34	18	42.444	9.141	60	20	42.444	13.522	65	19	45.778	14.953	72	21
1999	34	18	46.944	8.875	63	31	46.944	12.056	69	26	46.389	12.368	70	20

Table A1 (continued)

Date	N. Games	N. Teams	GA				GF				Pt			
			μ	σ	Max	Min	μ	σ	Max	Min	μ	σ	Max	Min
1998	34	18	47.056	14.379	76	27	47.056	14.957	67	25	46.167	15.768	74	14
1997	34	18	44.889	11.601	67	24	44.889	8.210	60	28	45.333	11.911	65	19
1996	34	18	44.722	13.851	79	24	44.722	11.641	66	28	46.556	14.047	73	24
1995	34	18	42.944	11.212	65	25	42.944	12.891	69	18	46.722	15.215	73	12
1994	34	18	41.167	12.931	72	15	41.167	10.176	64	28	34.000	9.267	50	11
1993	34	18	47.667	12.276	75	32	47.667	9.146	65	36	34.000	8.000	50	17
1992	34	18	38.611	13.639	68	20	38.611	13.347	74	24	34.000	10.477	56	14
1991	34	18	39.000	12.098	63	19	39.000	9.786	57	20	34.000	9.184	51	18
1990	34	18	38.000	7.356	51	26	38.000	11.906	57	20	34.000	9.127	51	21
1989	34	18	35.833	8.583	49	19	35.833	15.026	67	17	34.000	9.381	58	22
1988	30	16	31.500	6.623	44	14	31.500	10.263	55	19	29.688	7.364	45	20
1987	30	16	28.875	7.247	41	17	28.875	8.732	42	13	29.438	7.668	42	15
1986	30	16	30.938	9.448	55	17	30.938	7.767	51	18	30.000	7.339	45	16
1985	30	16	31.500	9.026	48	19	31.500	9.750	48	16	30.000	8.383	43	15
1984	30	16	35.812	7.985	55	23	35.812	11.280	57	14	30.000	7.501	43	12
1983	30	16	31.562	7.924	56	23	31.562	8.602	49	21	30.000	7.137	43	13
1982	30	16	29.625	8.245	42	14	29.625	8.123	48	18	30.000	7.958	46	17
1981	30	16	28.688	7.552	46	15	28.688	8.972	46	18	29.062	7.801	44	16
1980	30	16	28.250	7.443	44	15	28.250	7.904	44	18	26.188	11.923	41	0
1979	30	16	28.438	7.589	42	16	28.438	8.718	46	14	30.000	7.448	44	15
1978	30	16	32.000	7.155	44	17	32.000	8.278	50	21	30.000	7.062	44	17
1977	30	16	33.250	9.074	48	14	33.250	8.466	51	22	29.938	9.327	51	14
1976	30	16	33.875	7.658	52	22	33.875	8.679	49	19	30.000	7.992	45	19
1975	30	16	29.188	7.591	42	15	29.188	10.925	50	14	30.000	8.042	43	17

Table A1 (continued)

Date	N. Games	N. Teams	GA				GF				Pt			
			μ	σ	Max	Min	μ	σ	Max	Min	μ	σ	Max	Min
1974	30	16	31.062	4.837	37	23	31.062	9.657	50	16	27.875	10.481	43	0
1973	30	16	28.062	6.757	41	16	28.062	13.849	65	13	30.000	9.018	45	16
1972	30	16	30.188	7.687	43	17	30.188	10.154	49	17	30.000	9.121	43	13
1971	30	16	31.562	6.703	43	19	31.562	9.508	54	18	30.000	7.348	46	21
1970	30	16	29.000	8.540	45	11	29.000	9.640	43	11	30.000	7.598	45	19
1969	30	16	31.062	10.994	49	12	31.062	9.219	55	20	30.000	8.025	45	19
1968	30	16	31.500	6.044	42	23	31.500	10.752	53	13	30.000	7.109	46	17
1967	34	18	34.056	12.119	57	17	34.056	11.588	59	20	34.000	9.756	49	17
1966	34	18	37.111	10.671	62	22	37.111	13.087	70	22	34.000	9.127	50	15
1965	34	18	35.556	8.002	51	23	35.556	13.712	68	19	34.000	8.931	54	21
1964	34	18	35.944	9.187	50	18	35.944	11.700	58	20	34.000	9.756	54	22
1963	34	18	41.167	11.195	59	20	41.167	10.875	58	18	34.000	7.806	49	20
1962	34	18	42.778	8.579	63	31	42.778	15.152	83	29	34.000	9.876	53	17
1961	34	18	45.222	6.924	63	34	45.222	15.292	80	27	34.000	7.746	49	18
1960	34	18	44.000	6.508	54	31	44.000	17.167	92	21	33.000	11.225	55	0
1959	34	18	49.833	9.871	72	32	49.833	19.704	95	29	34.000	8.478	52	23
1958	34	18	48.889	8.811	65	36	48.889	11.999	77	29	34.000	6.269	51	26
1957	34	18	46.944	7.025	63	39	46.944	10.524	65	32	34.000	5.605	48	22
1956	34	18	46.556	13.111	87	20	46.556	12.253	70	27	34.000	7.396	53	15
1955	34	18	46.222	5.786	55	35	46.222	13.761	81	24	34.000	7.088	48	21
1954	34	18	45.833	10.274	64	27	45.833	10.557	67	33	34.000	8.395	51	25
1953	34	18	46.111	9.615	67	24	46.111	10.471	73	31	34.000	6.544	47	22
1952	38	20	55.050	13.040	85	34	55.050	17.034	98	37	38.000	9.171	60	17
1951	38	20	59.600	11.614	76	39	59.600	20.666	107	44	38.000	10.063	60	27

Table A1 (continued)

Date	N. Games	N. Teams	GA				GF				Pt			
			μ	σ	Max	Min	μ	σ	Max	Min	μ	σ	Max	Min
1950	38	20	63.250	11.562	89	43	63.250	22.396	118	25	38.000	10.285	62	16
1949	38	20	55.500	9.822	74	34	55.500	15.195	85	30	38.000	8.844	60	26
1948	40	21	57.143	12.920	82	33	57.143	19.487	125	31	38.381	11.948	65	0
1947	38	20	53.200	12.164	79	24	53.200	17.246	104	32	38.000	9.862	63	18
1943	30	16	46.375	10.105	66	31	46.375	14.431	75	24	30.000	6.673	44	21
1942	30	16	42.938	11.404	59	21	42.938	11.364	60	23	30.000	6.723	42	19
1941	30	16	45.812	11.374	84	34	45.812	9.042	60	31	30.000	4.747	39	17
1940	30	16	39.062	7.793	48	23	39.062	9.909	56	25	30.000	6.408	44	22
1939	30	16	37.062	7.150	54	28	37.062	9.788	55	23	30.000	5.489	42	24
1938	30	16	38.562	12.474	60	22	38.562	9.423	57	22	30.000	8.509	41	15
1937	30	16	39.938	11.710	67	25	39.938	9.391	56	23	30.000	6.928	42	18
1936	30	16	38.125	8.516	50	20	38.125	10.494	61	21	30.000	6.398	40	16
1935	30	16	39.312	10.025	54	22	39.312	11.943	63	21	30.000	7.220	44	15
1934	34	18	46.889	15.751	91	24	46.889	13.455	88	32	34.000	9.068	53	17
1933	34	18	52.278	14.187	75	23	52.278	15.350	83	28	34.000	8.630	54	21
1932	34	18	52.722	13.033	87	33	52.722	16.374	89	31	34.000	8.738	54	22
1931	34	18	52.889	12.583	71	31	52.889	17.098	87	30	34.000	10.825	55	19
1930	34	18	53.833	15.147	83	31	53.833	11.794	85	31	34.000	8.196	50	16

References

Albert J, Koning RH Eds. (2007) Statistical thinking in sports. CrC Press

Ausloos M, Cloots R, Gadomski A, Vitanov NK (2014) Ranking structures and rank-rank correlations of countries: the FIFA and UEFA cases. Int J Mod Phys C 25(11):1450060

Ausloos M (2014) Intrinsic classes in the Union of European Football Associations soccer team ranking. Cent Eur J Phys 12:773–779

Ausloos M (2024) Hierarchy selection: new team ranking indicators for cyclist multi-stage races. Eur J Oper Res 314(2):807–816

Baker R, Scarf P (2006) Predicting the outcomes of annual sporting contests. J R Stat Soc: Ser C: Appl Stat 55(2):225–239

Boys RJ, Philipson PM (2019) On the ranking of test match batsmen. J R Stat Soc: Ser C: Appl Stat 68(1):161–179

Cefis M, Carpita M (2024) The higher-order PLS-SEM confirmatory approach for composite indicators of football performance quality. Comput Stat 39(1):93–116

Cerqueti R, D'Urso P, De Giovanni L, Mattera R, Vitale V (2022) INGARCH-based fuzzy clustering of count time series with a football application. Mach Learn Appl 10:100417

Ficcadenti V, Cerqueti R, Varde'i CH (2023) A rank-size approach to analyse soccer competitions and teams: the case of the Italian football league "Serie A". Ann Oper Res 325(1):85–113

Gabel A, Redner S (2012) Random walk picture of basketball scoring. J Quant Anal Sports 8(1):6

Gorgi P, Koopman SJ, Lit R (2023) Estimation of final standings in football competitions with a premature ending: the case of COVID-19. AStA Adv Stat Anal 107(1):233–250

Glickman ME, Stern HS (2005) A state-space model for National Football League scores. In Anthology of statistics in sports. Society for Industrial and Applied Mathematics, pp 23–33

Heuer A, Mueller C, Rubner O (2010) Soccer: is scoring goals a predictable Poissonian process? Europhys Lett 89(3):38007

Higham DG, Hopkins WG, Pyne DB, Anson JM (2014) Performance indicators related to points scoring and winning in international rugby sevens. J Sports Sci Med 13(2):358

Kendall MG (1945) The treatment of ties in ranking problems. Biometrika 33(3):239–251

Mattera R (2023) Forecasting binary outcomes in soccer. Ann Oper Res 325(1):115–134

Mendes RS, Malacarne LC, Anteneodo C (2007) Statistics of football dynamics. Eur Phys J B 57:357–363

Merritt S, Clauset A (2014) Scoring dynamics across professional team sports: tempo, balance and predictability. EPJ Data Sci 3:1–21

Migliorati M, Manisera M, Zuccolotto P (2023) Integration of model-based recursive partitioning with bias reduction estimation: a case study assessing the impact of Oliver's four factors on the probability of winning a basketball game. AStA Adv Stat Anal 107(1):271–293

Okada H, Takagi T (2008) Evaluation of multi-objective genetic algorithm for RoboCupSoccer team evolution. In SICE annual conference, Chofu, Japan 2008, pp 151–154. https://doi.org/10.1109/SICE.2008.4654639

Reep C, Benjamin B (1968) Skill and chance in association football. J R Stat Soc Ser A 131(4):581–585

Ribeiro HV, Mendes RS, Malacarne LC, Picoli S, Santoro PA (2010) Dynamics of tournaments: the soccer case: a random walk approach modeling soccer leagues. Eur Phys J B 75:327–334

Sandri M, Zuccolotto P, Manisera M (2020) Markov switching modelling of shooting performance variability and teammate interactions in basketball. J R Stat Soc Ser C Appl Stat 69(5):1337–1356

Strauss D, Arnold BC (1987) The rating of players in racquetball tournaments. J R Stat Soc Ser C Appl Stat 36(2):163–173

Sziklai BR, Biró P, Csató L (2022) The efficacy of tournament designs. Comput Oper Res 144:105821

Sziklai BR, Héberger K (2020) Apportionment and districting by sum of ranking differences. PLoS ONE 15(3):e0229209

Thakkar P, Shah M (2021) An assessment of football through the lens of data science. Ann Data Sci 8:823–836

Vernon-Carter EJ, Ochoa-Tapia JA, Alvarez-Ramirez J (2023) Singular value decomposition entropy of the standing matrix for quantifying competitiveness of soccer leagues. Phys A 625:129007

Volf P (2009) A random point process model for the score in sport matches. IMA J Manag Math 20(2):121–131