



The Probabilistic Final Standing Calculator: a fair stochastic tool to handle abruptly stopped football seasons

Hans Van Eetvelde¹ · Lars Magnus Hvattum² · Christophe Ley¹

Received: 28 December 2020 / Accepted: 1 August 2021 / Published online: 11 August 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The COVID-19 pandemic has left its marks in the sports world, forcing the full stop of all sports-related activities in the first half of 2020. Football leagues were suddenly stopped, and each country was hesitating between a relaunch of the competition and a premature ending. Some opted for the latter option and took as the final standing of the season the ranking from the moment the competition got interrupted. This decision has been perceived as unfair, especially by those teams who had remaining matches against easier opponents. In this paper, we introduce a tool to calculate in a fairer way the final standings of domestic leagues that have to stop prematurely: our Probabilistic Final Standing Calculator (PFSC). It is based on a stochastic model taking into account the results of the matches played and simulating the remaining matches, yielding the probabilities for the various possible final rankings. We have compared our PFSC with state-of-the-art prediction models, using previous seasons which we pretend to stop at different points in time. We illustrate our PFSC by showing how a probabilistic ranking of the French Ligue 1 in the stopped 2019–2020 season could have led to alternative, potentially fairer, decisions on the final standing.

Keywords Bivariate Poisson · Plus–minus rating · Prediction · Ranking · (Tournament) Rank Probability Score

1 Introduction

The COVID-19 pandemic has left its marks in the world of sports, forcing in February–March 2020 the full stop of nearly all sports-related activities around the globe. This implied that the national and international sports leagues were abruptly

✉ Hans Van Eetvelde
Hans.VanEetvelde@UGent.be

¹ Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium

² Molde University College, Faculty of Logistics, Molde, Norway

stopped, which had dramatic impacts especially in the most popular sport, football (or soccer). Indeed, the absence of regular income through ticket selling, television money and merchandising around live matches entailed that a large majority of professional clubs no longer were able to pay their players and other employees (Sky Sports 2020; Kicker 2020). Professional football having become such a big business, this also had an impact on other people whose main income was related to football matches (BBC Sport 2020). Moreover, the fans were also much looking forward to having their favourite sport resume. However, the intense and uncertain times made it a tough choice for decision-makers to take the risk of letting competitions go on, especially as it would be without stadium attendance for the first weeks or possibly months. Consequently, some leagues never resumed. The Dutch Eredivisie was the first to declare that it would not be continued, on 25 April 2020, followed by the French Ligue 1 on 28 April and the Belgian Jupiler Pro League on 15 May. The German Bundesliga was the first professional league to get restarted, followed by the English Premier League, Spanish Primera Division and Italian Serie A.

The leagues that decided to not resume playing were facing a tough question: how to evaluate the current season? Should it simply be declared void, should the current ranking be taken as final ranking, or should only a part of the season be evaluated? Either decision would have an enormous financial impact given the huge amount of money at stake. Which team would be declared champion? Which teams would be allowed to represent their country at the European level and hence earn a lot of money through these international games, especially in the Champions League? Which teams would be relegated and, as a potential consequence, be forced to dismiss several employees? Moreover, the broadcasting revenue is allocated on the basis of final standings. Different countries reacted differently: the Eredivisie had no champion nor relegated teams, but the teams qualified for the European competitions were decided based on the ranking of 8 March. The Ligue 1 and the Jupiler Pro League, on the other hand, declared also a champion and relegated teams on the basis of the ranking from the moment their seasons got stopped. However, the Jupiler Pro League had to come back on this decision, and the relegation was nullified. The Ligue 1 based their final standing on the ratio of points earned per game, since not all teams had played an equal amount of games when the season got stopped. Obviously, several teams were not pleased by such decisions, considering them to be unfair (Holroyd 2020) because, inevitably, this favoured those teams that had still to play the strongest opponents in the remaining matches over those that were looking forward to a rather light end-of-season.

This naturally raises the question to find a more balanced, scientifically sound way to evaluate the final standing of abruptly ended seasons. It also represents a bigger challenge than evaluating in a fair way an abruptly stopped single game, where for instance cricket has adopted a mathematical formula known as the Duckworth–Lewis–Stern method (Duckworth and Lewis 1998).

The literature addressing the challenging question of how a stopped season should be evaluated in the most objective way is fuelled by proposals from after the outbreak of COVID-19. Ley et al. (2019) used current-strength based rankings, a statistical model based on the bivariate Poisson distribution and weighted maximum likelihood, to provide all possible final standings of the Belgian Jupiler Pro

League together with their associated probabilities. Their results were summarized in a mainstream journal (Het Laatste Nieuws 2020). Guyon (2020) proposed an Elo-based approach applied to the English Premier League. Lambers and Spieksma (2020) suggested an eigenvector-based approach, while Csató (2020) discussed general criteria that a fair ranking should fulfil in such a situation and proposed the generalized row sum method. Recently, Gorgi et al. (2020) used a statistical model to determine a ranking based on the expected total number of points per team.

In this paper, we investigate the extent to which a relatively simple stochastic model can serve the purpose of producing fair final standings for prematurely stopped round-robin type football competitions. Our original approach goes as follows. We construct a stochastic soccer model that is fitted on the played matches and then is used to simulate the remainder of the competition a large number of times, thus yielding for every team the probabilities to reach each possible final rank. This output is much richer in terms of information than giving only the most likely or the expected ranking. This also explains the terminology for our model, namely Probabilistic Final Standing Calculator, which we abbreviate as PFSC. In order to assess its predictive strength, we compare our PFSC with two benchmark prediction models. The first is the best performing model of Ley et al. (2019), which uses a similar stochastic model to estimate the current strength of a team based on its matches played in the last two years, rather than just the current season. The second is the plus-minus ratings approach of Pantuso and Hvattum (2020), where match outcome predictions are based on the ratings of the individual players making up the teams. In the latter method, player ratings are based on data from several past seasons. Thus, the benchmarks are chosen by selecting a high-quality team rating model and a state-of-the-art player rating model. Furthermore, both benchmarks use significantly more data than the PFSC. Next to these two advanced models, we compare our PFSC to two simpler models, to see if these models could serve as an alternative for our method. The first alternative is simply using the current ranking at the time that the season is halted as the final ranking. The second alternative resembles our PFSC, but does not take into account the number of goals scored in a match.

For each model, the probabilistic final standing of a not yet ended season is obtained by simulating the remaining matches 100,000 times, which gives us for every team the probabilities of reaching each possible place in the final standing. It is not appropriate to compare the predictions of these models on the 2019–2020 competitions which were resumed after the break, since those matches were played under different circumstances, including the absence of fans. It has been shown (Fischer and Haucap 2020) that these changed conditions could influence team performances, by lowering the effect of the home advantage. Therefore, we rather compare the three models on the basis of the three preceding seasons of the five most important European football leagues (England, Spain, Germany, Italy and France), which we stopped artificially after every match day. Our evaluation of the models' performance is done in two ways: by means of the Rank Probability Score (RPS) (Epstein 1969) and the Tournament Rank Probability Score (TRPS) (Ekstrøm et al. 2020), see Sect. 2.5 for their definition. From this comparison, we can see at which point in time the PFSC is able to catch up with the two high-performing but more complicated prediction models. The reader may now wonder why we do not use any of

these more elaborate models as PFSC; the reason is that we wish to propose a handy tool that sports professionals can indeed use without the need of too long computation time or large data collections. In the same vein, we will also make the PFSC freely available in the form of an R-package (R Core Team 2020). The package can be installed from <https://github.com/vaneetvelde/PFSC>.

The remainder of the paper is organized as follows. In Sect. 2, we describe our PFSC along with the two benchmark models and the alternatives models, as well as the two model performance evaluation metrics. Section 3 then presents the results of this broad comparison, illustrates the advantages of our PFSC by analyzing the French Ligue 1 season 2019–2020 and considers how fairer decisions could be obtained on the basis of our PFSC. We conclude the paper with final comments in Sect. 4.

We finish this introduction with a historical remark. The problem treated here, namely finding a fair way to determine final league standings if one cannot continue playing, goes back to the very roots of probability theory. The French writer Antoine Gombaud (1607–1684), famously known as Chevalier de Méré, was a gambler interested by the “problem of the points”: if two players play a pre-determined series of games, say 13, and they get interrupted at a score of, say, 5–2, and cannot resume the games, how should the stake be divided among them? The Chevalier de Méré posed this problem around 1654 to the mathematical community, and the two famous mathematicians Blaise Pascal (1623–1662) and Pierre de Fermat (1607–1665) accepted the challenge. They addressed the problem in an exchange of letters that established the basis of our modern probability theory (Devlin 2010).

2 Methods

In this section, we start by explaining the PFSC (Sect. 2.1) and the two benchmark models (Sects. 2.2 and 2.3), followed by a description of the two evaluation measures for comparison (Sect. 2.5). In what follows, we suppose to have a total of n teams competing in a round-robin type tournament of M matches.

2.1 The PFSC: a bivariate Poisson-based model

For modelling football matches, the PFSC will make use of the bivariate Poisson distribution. Building on the original idea of (Maher 1982) to model football match outcomes via Poisson distributions, the bivariate Poisson distribution has been popularized by Karlis and Ntzoufras (2003). Let Y_{ijm} stand for the number of goals scored by team i against team j ($i, j \in \{1, \dots, n\}$) in match m (where $m \in \{1, \dots, M\}$) and let $\lambda_{ijm} \geq 0$ resp. $\lambda_{jim} \geq 0$ be the expected number of goals for team i resp. j in this match. The joint probability function of the home and away score is then given by the bivariate Poisson probability mass function

$$P(Y_{ijm} = x, Y_{jim} = y) = \frac{\lambda_{ijm}^x \lambda_{jim}^y}{x!y!} \exp(-(\lambda_{ijm} + \lambda_{jim} + \lambda_C)) \sum_{k=0}^{\min(x,y)} \binom{x}{k} \binom{y}{k} k! \left(\frac{\lambda_C}{\lambda_{ijm} \lambda_{jim}} \right)^k,$$

where $\lambda_C \geq 0$ is a covariance parameter representing the interaction between both teams. This parameter is kept constant over all matches, as suggested in Ley et al. (2019), who mentioned that models where this parameter depends on the teams at play perform worse. Note that, $\lambda_C = 0$ yields the independent Poisson model. The expected goals λ_{ijm} are expressed in terms of the strengths of team i and team j , which we denote r_i and r_j , respectively, in the following way: $\log(\lambda_{ijm}) = \beta_0 + (r_i - r_j) + h \cdot \mathbf{I}(\text{team } i \text{ playing at home})$, where h is a real-valued parameter representing the home effect and is only added if team i plays at home, and β_0 is a real-valued intercept indicating the expected number of goals e^{β_0} if both teams are equally strong and play on a neutral ground. The strengths r_1, \dots, r_n can take both positive and negative real values and are subject to the identification constraint $\sum_{i=1}^n r_i = 0$. Over a period of M matches (which are assumed to be independent), this leads to the likelihood function

$$L = \prod_{m=1}^M \mathbf{P}(Y_{ijm} = y_{ijm}, Y_{jim} = y_{jim}), \quad (1)$$

where y_{ijm} and y_{jim} stand for the actual number of goals scored by teams i and j in match m . The unknown values of the strength parameters r_1, \dots, r_n are then computed numerically as maximum likelihood estimates, that is, in such a way that they best fit a set of observed match results.

Ley et al. (2019) established that the bivariate Poisson model and its Independent counterpart are the best-performing maximum likelihood-based models for predicting football matches. They evaluated the match-based predictive performance for various Poisson models, as well as for Bradley–Terry and Thurstone–Mosteller models where the outcome (win/draw/loss) is modelled directly instead of as a function of the goals scored. The evaluation data consisted of ten seasons of the English Premier league and ten years of national team matches.

Using the bivariate Poisson model in the final standing, prediction works as follows. The parameters λ_C , β_0 , h and the strength parameters r_1, \dots, r_n are estimated using the matches played so far in the current season. Next, these parameters are used to simulate 100,000 times the remaining matches, by sampling the number of goals for each team in each match from the corresponding bivariate Poisson distribution. For each simulated end of season, a final standing is created based on the played and simulated matches, taking into account the specific rules of the leagues. The probabilistic final standing is then calculated by averaging the results over all 100,000 simulations, giving each team a probability to reach every possible rank. Note that, Gorgi et al. (2020) also used the bivariate Poisson distribution as their statistical model, but they only calculate expected ranks and not the complete probabilistic picture as we do.

This model is relatively simple compared to the benchmark models that we describe below, but it has some nice properties that make it perfectly suited for determining the final standing of a prematurely stopped competition. First, the PFSC only takes into account match results, so data requirements are benign. Second, the PFSC only takes into account matches of the current season, so there is no bias to teams that performed well in the previous season(s). Third, each played game has

the same weight in the estimation of the team strengths. These three properties make this method a fair way to evaluate an unfinished football season. On top of this, the code for the model can easily be executed in a short time.

2.2 Current-strength-based team ratings

The first benchmark model is an extension of the previous model. The idea of Ley et al. (2019) was to use a weighted maximum likelihood, where the weight is a time depreciation factor $w_{time,m} > 0$ for match m , resulting in

$$L = \prod_{m=1}^M \left(P(Y_{ijm} = y_{ijm}, Y_{jim} = y_{jim}) \right)^{w_{time,m}}.$$

The exponentially decreasing time decay function is defined as follows: a match played x_m days back gets a weight of

$$w_{time,m}(x_m) = \left(\frac{1}{2} \right)^{\frac{x_m}{\text{Half period}}}.$$

In other words, a match played *Half period* days ago only contributes half as much as a match played today and a match played $3 \times \text{Half period}$ days ago contributes 12.5 % of a match played today. This weighting scheme gives more importance to recent matches and leads to a so-called current-strength ranking based on the estimated strength parameters of the teams.

Another difference is that this model uses two years of past matches to estimate the team strengths. The half period is set to 390 days, as this was found to be the optimal half period by Ley et al. (2019) when evaluated on ten seasons of the Premier League. The predicted probabilities for each rank in the final standing are obtained in the same way as in the PFSC.

2.3 Plus–minus ratings

Plus–minus ratings, the second benchmark model, are based on the idea of distributing credit for the performance of a team onto the players of the team. We consider the variant of plus–minus proposed by Pantuso and Hvattum (2020). Each football match is partitioned into segments of time, with new segments starting whenever a player is sent off or a substitution is made. For each segment, the set of players appearing on the pitch does not change, and a goal difference is observed from the perspective of the home team, equal to the number of goals scored by the home team during the segment minus the number of goals scored by the away team. The main principle of the plus–minus ratings considered is to find ratings such that the sum of the player ratings of the home team minus the sum of the player ratings of the away team is as close as possible to the observed goal difference.

Let S be the set of segments, $P_{h(s)}$, respectively, $P_{a(s)}$ the set of players on the pitch for the home, respectively, away team during segment $s \in S$. Denote by $g(s)$ the goal difference in the segment as seen from the perspective of the home

team. If a real-valued parameter β_j is used to denote the rating of player j , the identification of ratings can be expressed as minimizing

$$\sum_{s \in S} \left(\sum_{j \in P_{h(s)}} \beta_j - \sum_{j \in P_{a(s)}} \beta_j - g(s) \right)^2,$$

the squared difference between observed goal differences and goal differences implied by the ratings of players involved. To derive more reasonable player ratings, Pantuso and Hvattum (2020) considered a range of additional factors, which we also consider here: (1) Segments have different durations, so the ratings in each segment are scaled to correspond with the length of the segment. (2) The home team has a natural advantage, which is added using a separate parameter. (3) Some segments have missing players, either due to players being sent off by the referee or due to injuries happening after all allowed substitutions have been made. These situations are represented using additional variables corresponding to the missing players, while remaining player ratings are scaled so that their sum corresponds to an average rating for a full team. (4) The player ratings are not assumed to be constant over the whole data set, but rather to follow a curve that is a function of the age of players. This curve is modelled as a piece-wise linear function which is estimated together with the ratings by introducing corresponding age adjustment variables. (5) Each segment is further weighted by factors that depend on the recency of the segment and the game state. A complete mathematical formulation of the plus-minus rating system was provided by Pantuso and Hvattum (2020).

To move from plus-minus player ratings to match predictions, an ordered logit regression model is used. This model derives probabilities for a home win, a draw, and an away win based on a single value associated with each match: the sum of the ratings of the players in the starting line-up for the home team, minus the sum of the ratings of the players in the starting line-up for the away team, plus the home field advantage of the corresponding competition.

As with the previous benchmark, the remaining matches of a league are simulated. However, some slight differences can be observed. Following Sæbø and Hvattum (2019), the starting line-ups of the teams are also simulated, based on the players available in the squads. Each player has a 10 % chance of being injured or otherwise inadmissible for a given match. Subject to these random unavailable players, the best possible starting line-up is found, consisting of exactly one goalkeeper, and at least three defenders, three midfielders, and one forward. Based on this, probabilities of a home win, draw and away win are derived using the ordered logit regression model. Since this does not provide a goal difference, but just a result, the simulation further assumes that losing teams score 0 goals and drawing teams score one goal, whereas the number of goals for winning teams is selected at random from 1 to 3.

2.4 Alternative models for the PFSC

In our experiments, we also consider two simpler benchmarks. They both meet the requirement of solely depending on the match results of the current season. The first one is using the official standing at the time when the season is halted. We thus assign the probability of 100 % to the current position of the team and 0 % to every other position.

The second alternative model is a Thurstone–Mosteller model (Glenn and David 1960), another maximum-likelihood model which, in contrast to our PFSC, only takes into account the outcome of a game (win/draw/loss) rather than the number of goals scored by each team. The Thurstone–Mosteller model considers latent normally distributed variables $Y_{i,m}$ which stand for the performance of team i in match m . When the performance of team i is much better than the performance of team j in match m , say $Y_{i,m} - Y_{j,m} > d$ for some positive parameter d , then team i beats team j in that match. If the difference in their performances is lower than d , i.e. $|Y_{i,m} - Y_{j,m}| < d$, then the game will end in a draw. The expected performance $\mathbb{E}[Y_{i,m}]$ is the strength of the team, denoted by r_i , possibly adjusted for the home advantage by adding a parameter h . The variance σ^2 of the performances is assumed to be constant and can be chosen arbitrarily. If we call $P_{H_{ijm}}$, the probability of a home win in match m , $P_{D_{ijm}}$ the probability of a draw in match m and $P_{A_{ijm}}$ the probability of an away win in match m , then the outcome probabilities in this model are

$$P_{H_{ijm}} = P(Y_{i,m} - Y_{j,m} > d) = \Phi\left(\frac{(r_i + h) - r_j - d}{\sigma\sqrt{2}}\right);$$

$$P_{A_{ijm}} = P(Y_{j,m} - Y_{i,m} > d) = \Phi\left(\frac{r_j - (r_i + h) - d}{\sigma\sqrt{2}}\right);$$

$$P_{D_{ijm}} = 1 - P_{H_{ijm}} - P_{A_{ijm}},$$

where Φ denotes the cumulative distribution function of the standard normal distribution. These probabilities are then used to build a maximum likelihood function, which allows us to estimate the included parameters, based on the played matches of the current season.

2.5 Metrics to evaluate and compare the three models

We have provided three proposals for predicting the final standings of abruptly stopped football seasons. These are evaluated by predicting, for several completed seasons from different top leagues, the remaining matches after artificially stopping each season after every match. The evaluation of their predictive abilities is done at two levels: single match outcomes and final season standings. For the former, we use the Rank Probability Score as metric, for the latter the Tournament Rank Probability Score.

The Rank Probability Score (RPS) is a proper scoring rule that preserves the ordering of the ranks and places a smaller penalty on predictions that are closer

to the observed data than predictions that are further away from the observed data (Epstein 1969; Gneiting and Raftery 2007; Constantinou and Fenton 2012). The RPS is defined for a single match as

$$RPS = \frac{1}{R-1} \sum_{r=1}^R \left(\sum_{j=1}^r (o_j - x_j) \right)^2$$

where R is the number of possible outcomes, o_j the empirical probability of outcome j (which is either 1 or 0), and x_j the forecasted probability of outcome j . The smaller the RPS, the better the prediction. The RPS is similar to the Brier score, but measures the accuracy of a prediction differently when there are more than two ordered categories, by using the cumulative predictions in order to be sensitive to the distance. Let us give some further intuition about this metric. Let 1 stand for home win, 2 for draw and 3 for away win, so obviously $R = 3$. The formula of the RPS can be simplified to

$$\frac{1}{2}((o_1 - x_1)^2 + (o_1 + o_2 - x_1 - x_2)^2 + (1 - 1)^2) = \frac{1}{2}((o_1 - x_1)^2 + (o_3 - x_3)^2),$$

which shows, for instance, that a home win predicted as draw is less severely penalized than would be a predicted away win in such a case.

Ekstrøm et al. (2020) extended the RPS to final tournament or league standings, and consequently termed it TRPS for Tournament RPS. The idea is very similar to the RPS, as the TRPS compares the cumulative prediction X_{rt} that team t will reach at least rank r (with lower values of r signifying a better ranking) to the corresponding empirical cumulative probability O_{rt} . The latter also only attains two different values: a column t in O_{rt} is 0 until the rank which team t obtained in the tournament, after which it is 1. Consequently, the TRPS is defined as

$$TRPS = \frac{1}{T} \sum_{t=1}^T \frac{1}{R-1} \sum_{r=1}^{R-1} (O_{rt} - X_{rt})^2,$$

where T is the number of teams and R is the total number of possible ranks in a tournament or league. A perfect prediction will yield a TRPS of 0 while the TRPS increases when the prediction worsens. The TRPS is a proper scoring rule, very flexible and handles partial rankings. It retains for league predictions the desirable properties of the RPS, and as such assigns lower values to predictions that are almost right than to predictions that are clearly wrong.

3 Results

3.1 Performance comparison based on artificially stopped previous seasons

The PFSC, the current-strength team ratings and the plus-minus player ratings are evaluated in terms of correctly predicting match outcomes and the final league table.

The evaluation is conducted on the top leagues of England, France, Germany, Italy and Spain, for the 2016–2017, 2017–2018 and 2018–2019 seasons. Each season and league is halted after each match day, and the outcomes of the remaining matches, as well as the final league tables, are predicted.

Figure 1 shows the mean RPS of the remaining matches, given the current match day. The figure illustrates that the performances of the current-strength team ratings and the plus–minus player ratings are similar throughout. The PFSC is a much simpler model, and only uses data from the current season. Therefore, its performance is relatively bad in the beginning of the season. However, in most cases, the PFSC converges towards the performance of the benchmark methods after around 10 match days, and in all but one of fifteen league-seasons it has caught up after 25 match days. The exception is the 2017–2018 season of the Italian Serie A, where the PFSC leads to worse predictions than the benchmarks throughout. The results for the Thurstone–Mosteller model indicate that if we do not take into account the number of goals, the performance is worse, even after 25 or more match days.

When nearing the end of the season, the RPS is calculated over few matches. Therefore, the mean RPS behaves erratically and sometimes increases sharply. This is because a single upset in one of the final rounds can have a large effect on the calculated RPS values. However, as the plots in Fig. 1 show, all methods follow each other closely, indicating that the results that are difficult to predict are equally hard for all methods.

In Fig. 2, the mean TRPS is shown for each league and season. As the final table becomes easier to predict the more matches have been played, the TRPS converges to zero. Therefore, the difference in performance among the three methods also converges to zero. However, the PFSC has a similar prediction quality as the two advanced benchmark methods already somewhere between 10 and 25 match days into the season. Even for the Italian Serie A in 2017–2018, the TRPS is similar for all methods after 30 match days, although the RPS indicated that the predictions of the PFSC are worse for the remaining matches in that particular season. As for the RPS, the performances of the current-strength team ratings and the plus–minus player ratings are very similar. On the other hand, both the use of the current ranking and the Thurstone–Mosteller model shows worse and very volatile results.

3.2 Evaluating the French Ligue 1 2019–2020 season with the PFSC

We have shown in the previous section that our simple PFSC is comparable in terms of predictive performance at match and final standing levels to the two benchmark models that are more computationally demanding, more time-consuming and require more input data. This establishes the PFSC as a very good candidate for obtaining fair final standings.

We will now illustrate how our PFSC can be used in practice by decision-makers to reach fairer decisions on how an abruptly stopped season should be evaluated. To this end, we will show how the French Ligue 1 could have been settled after it was abruptly stopped in the 2019–2020 season. The match results were downloaded from the website <https://www.football-data.co.uk/>. The code for the PFSC applied

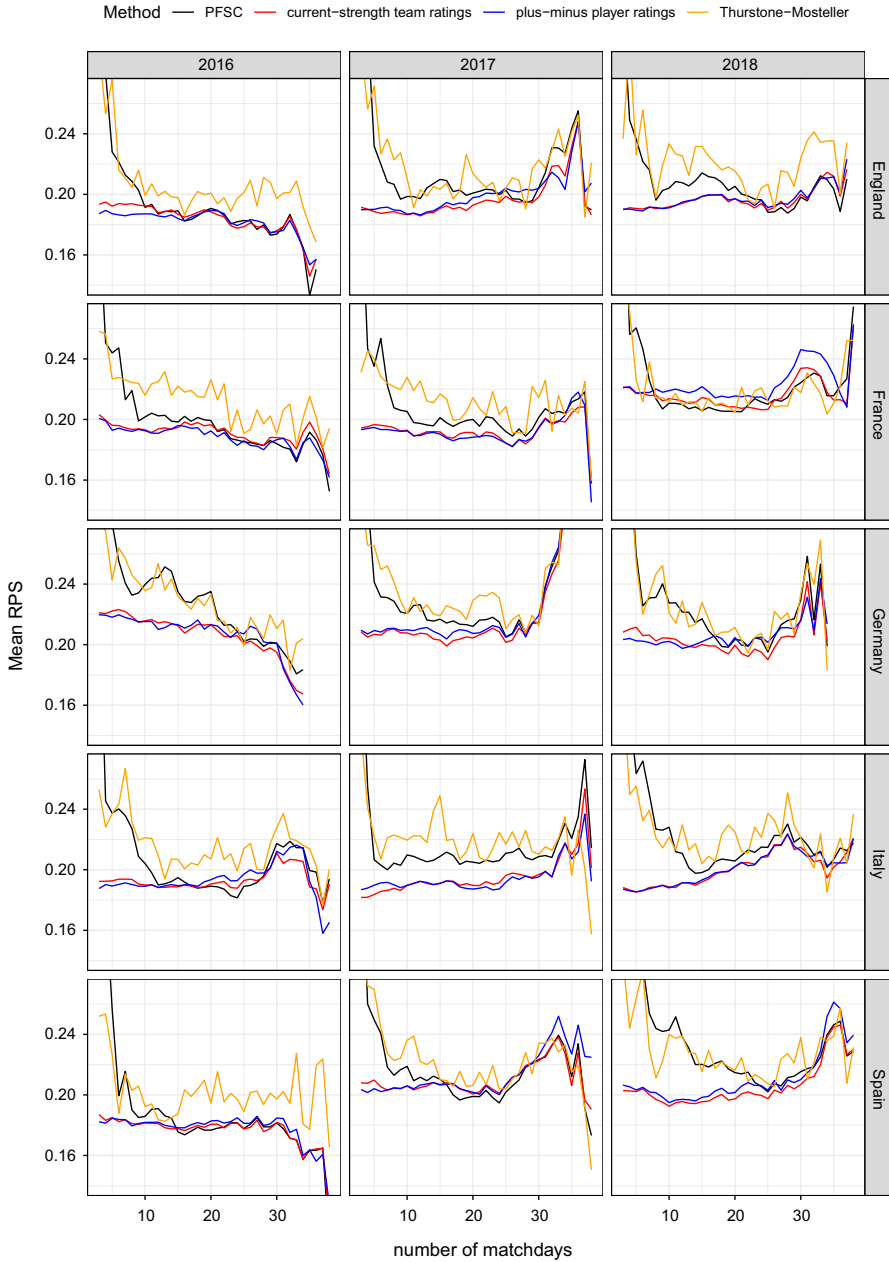


Fig. 1 Mean RPS (Rank Probability Score) values over all remaining matches, calculated after each match day for five leagues and three seasons

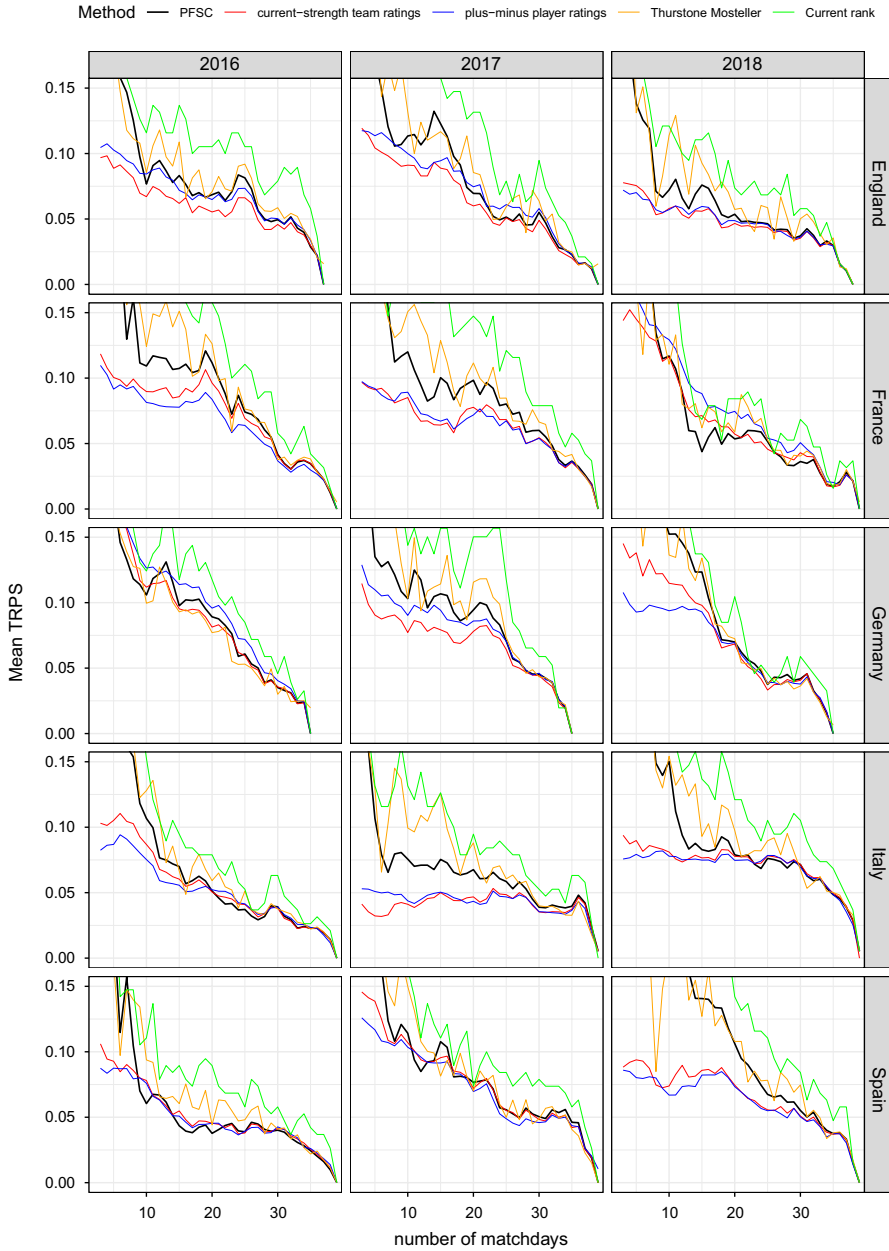


Fig. 2 Mean TRPS (Tournament Rank Probable Score) values calculated after each match day for five leagues and three seasons

Table 1 The official ranking of the French Ligue 1 in the 2019–2020 season

	Team	Points	Win	Draw	Loss	Goals	Goals against	Goal difference	Matches	Points per match
1	PSG	68	22	2	3	75	24	51	27	2.52
2	Marseille	56	16	8	4	41	29	12	28	2.00
3	Rennes	50	15	5	8	38	24	14	28	1.79
4	Lille	49	15	4	9	35	27	8	28	1.75
5	Nice	41	11	8	9	41	38	3	28	1.46
6	Reims	41	10	11	7	26	21	5	28	1.46
7	Lyon	40	11	7	10	42	27	15	28	1.43
8	Montpellier	40	11	7	10	35	34	1	28	1.43
9	Monaco	40	11	7	10	44	44	0	28	1.43
10	Strasbourg	38	11	5	11	32	32	0	27	1.41
11	Angers	39	11	6	11	28	33	-5	28	1.39
12	Bordeaux	37	9	10	9	40	34	6	28	1.32
13	Nantes	37	11	4	13	28	31	-3	28	1.32
14	Brest	34	8	10	10	34	37	-3	28	1.21
15	Metz	34	8	10	10	27	35	-8	28	1.21
16	Dijon	30	7	9	12	27	37	-10	28	1.07
17	St. Etienne	30	8	6	14	29	45	-16	28	1.07
18	Nîmes	27	7	6	15	29	44	-15	28	0.96
19	Amiens	23	4	11	13	31	50	-19	28	0.82
20	Toulouse	13	3	4	21	22	58	-36	28	0.46

The order of the teams is determined by the number of points earned per match

on the present case study is available in Online Resource 1. The official ranking of the Ligue 1 is given in Table 1. The probabilities provided by the two advanced benchmark models are given in Online Resource 2.

At the time of stopping the competition, each team had played at least 27 matches. From the findings of the previous section, we know that after this number of match days, our PFSC is a competitive model for predicting the remaining matches and the final ranking. Based on the played matches, the teams strengths were estimated, which resulted in the strengths reported in Table 2. We can see that Paris Saint-Germain (PSG) is by far considered as the strongest team in the league. Surprisingly, Olympique Lyon comes out as the second strongest team, while only standing on the 7th place in the official ranking at that time. This could indicate that Lyon did have some bad luck during the season. Looking at their match results, we could see that in almost all their lost matches, they lost with a goal difference of 1 goal. Only PSG at home managed to get a margin of two goals against Lyon. At the bottom of the table, we find that Toulouse was the weakest team in the league, followed by Amiens, St. Etienne and Nîmes. This is in agreement with the official ranking, up to a slightly different ordering of the teams.

Using these strengths, we have simulated the remainder of the season 100,000 times and by taking the mean over these simulations, we have calculated the probabilities for each team to reach each possible position, which is summarized in Table 3. We can see that PSG would win the league with approximately 100 %

Table 2 The estimated ratings r_i , $i = 1, \dots, 20$, of the teams in the French Ligue 1, based on the played matches in the 2019–2020 season, obtained via the bivariate Poisson model in our PFSC

	Teams	Estimated strengths
1	PSG	0.85
2	Lyon	0.28
3	Rennes	0.24
4	Marseille	0.19
5	Lille	0.15
6	Bordeaux	0.15
7	Reims	0.08
8	Nice	0.03
9	Montpellier	0.02
10	Monaco	0.01
11	Strasbourg	-0.01
12	Nantes	-0.02
13	Brest	-0.09
14	Angers	-0.10
15	Metz	-0.16
16	Dijon	-0.17
17	Nîmes	-0.27
18	St. Etienne	-0.29
19	Amiens	-0.32
20	Toulouse	-0.59

Table 3 The Probabilistic final standing (in percentages) of the Ligue 1 in the 2019–2020 season, according to our PFSC method

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
PSG																				
Marseille		79	17	4																
Rennes		11	42	37	8	2	1													
Lille		10	38	41	8	2	1													
Lyon			2	10	34	19	12	8	6	4	3	2	1							
Reims				2	12	16	15	14	12	10	8	6	3	1						
Montpellier				2	11	14	14	13	12	10	9	7	5	2	1					
Bordeaux			1	7	12	13	13	13	12	11	10	9	7	3	1					
Nice			1	6	10	12	12	13	13	13	12	10	7	3	1					
Strasbourg			1	5	9	9	11	11	12	12	11	9	5	2	1					
Nantes				4	7	7	9	10	11	12	13	13	12	6	3	1				
Monaco			1	3	6	6	8	10	12	13	14	13	11	6	2	1				
Angers				2	3	3	5	7	8	11	13	16	17	11	5	2				
Metz								1	1	2	3	6	11	23	23	17	9	3		
Brest							1	1	1	2	3	5	10	21	24	18	10	4		
Dijon											1	2	4	10	17	24	24	15	4	
St-Etienne												1	2	7	14	22	26	20	7	
Nîmes													1	2	6	12	23	38	18	
Amiens														2	1	2	7	19	70	1
Toulouse															1	2	7	19	70	1

Probabilities are rounded to the nearest 1 percent

probability, thanks to the big lead they had and the high team strength. Marseille had a 79 % chance of keeping the second position, with also a certain chance of becoming third (17 %), or even fourth (4 %). Furthermore, we see that Lyon, thanks to their high estimated strength, had the highest probability to be ranked as fifth (34 %). Their frustration with respect to the decision as it was taken officially by the Ligue 1 is thus understandable (Holroyd 2020). In the bottom of the standing, we see that Toulouse was doomed to be relegated, with almost no chance of not ending at the 19th or 20th place in the league. Amiens had still about 29 % chance of staying in the first league.

Now, how could this table be used by decision-makers to handle the discontinued season? One has to decide which team will become the champion, which teams will play in the Champions League and Europe League and which teams will be relegated to the second division.

For the first answer, some leagues nowadays have entered a rule stating that if enough matches are played, the current leader of the season would be considered as the champion. However, this does not take into account the gap between the first and the second in the standing. We would recommend changing the rule, in the sense that a team can only be declared champion if it has more than $C\%$ chance to win the league according to the PFSC (C could, e.g. be 80, but this decision of course has to be made by the leagues). For our example, there is little doubt. PSG was expected to be the winner of the league with an estimated chance of 100%, so they should be considered as the champions of the Ligue 1. A similar strategy can be adopted regarding which teams should be relegated to the second division.

For the participation in the Champions League and Europe League, the leagues need a determined final standing instead of a probabilistic final standing. We will next show how we can get a determined final standing using our PFSC, and how we can use the PFSC to help to determine financial compensations.

3.3 Determined final standing and financial compensations for the French Ligue 1 via the PFSC

Following up on the results of Sect. 3.2, we make a determined final standing by calculating the expected rank, using the probabilities. This results in the standing shown in Table 4. In the example of the French League, we see that PSG gets the direct ticket for the Champions League, while Marseille and Rennes get the tickets for the CL qualification rounds. Lille and Lyon would have received the tickets for the group stage of the Europe League and Reims the ticket for the qualifications. This shows that Nice was one of the teams that got an advantage from the decision of the French league to halt the season.

However, transforming our probabilistic standing to a determined standing causes a number of teams to be (dis)advantaged. For example, in Table 4, we can see that the expected rank of Rennes is 3.51, which is the third-highest expected rank. Assigning Rennes to the third rank is therefore an advantageous outcome. Lille, on the other hand, has an expected rank of 3.58, which is only the fourth-best expected rank. Lille is therefore at a disadvantage when being assigned to rank 4.

Table 4 Determined final standing, using the PFSC probabilities. This standing could be used to decide which teams will play in the Champions League and Europe League

Rank	Team	Expected rank
1	Paris SG	1.00
2	Marseille	2.26
3	Rennes	3.51
4	Lille	3.58
5	Lyon	6.31
6	Reims	8.12
7	Montpellier	8.43
8	Bordeaux	8.96
9	Nice	9.26
10	Strasbourg	9.56
11	Nantes	10.11
12	Monaco	10.17
13	Angers	11.23
14	Metz	14.44
15	Brest	14.63
16	Dijon	16.05
17	St Etienne	16.44
18	Nimes	17.38
19	Amiens	18.57
20	Toulouse	19.99

This issue could be solved by using a compensation fund. Assume that the expected profit (in particular prize money for the league placement and starting and prize money from Champions League and Europe League) of a team ending in rank i is equal to P_i . The expected profit for, e.g. Marseille would be $0.79 * P_2 + 0.17 * P_3 + 0.04 * P_4$. In the determined ranking, they end as second, so they will receive P_2 . Actually, they receive too much, since they had no chance of ranking higher than second, but they had a reasonable chance to become third or even fourth. To compensate for this, they should hand over $P_2 - (0.79 * P_2 + 0.17 * P_3 + 0.04 * P_4) = 0.17 * (P_2 - P_3) + 0.04 * (P_2 - P_4)$ to the compensation fund. This will then be used to compensate teams that are disadvantaged by the establishing of a determined ranking. There will still be the difficulty of estimating the expected profit from reaching a certain rank (e.g. a team reaching the Europe League will have further merchandising advantages besides the profit mentioned above as compared to the team classified just outside of these ranks), but we believe that this tool could be very useful for decision-makers in determining which teams have received an advantage or disadvantage from an early stop of the league, and how to compensate for this.

When a league is played to completion, the difference between ranks i and $i + 1$ could be large (ten points or more) or small (same number of points, slightly better goal difference), but the final financial rewards would not differentiate between these cases. The suggested compensation fund also does not depend on these differences. That is, each simulation of the league arrives at a potential final table, and the ranks

in that table are determined in the same way as the ranks of a completed league. Thus, the proposed distribution of rewards only compensates for the uncertainty in the final ranking, not in the actual gaps between ranks.

4 Conclusion

In this paper, we proposed a novel tool, the Probabilistic Final Standing Calculator, to determine the most likely outcome of an abruptly stopped football season. Unlike other recent proposals, we provide probabilities for every team to reach each possible rank, which is more informative than only providing the single most likely or expected final ranking. We have shown the strength of our PFSC by comparing it to two benchmark models that are based on much more information, and yet our PFSC is exhibiting similar performances except when a season would get stopped extremely early, which however was anyway more a theoretical than a practical concern (a season stopped after less than a third of the games played would certainly be declared void). Our evaluation has been done at both the match-level (via the RPS) and the final standing level (via the TRPS).

We have shown on the concrete example of the 2019–2020 season of the French Ligue 1 how our PFSC can be used, also for a fair division of the money shares. We hope that our PFSC will help decision-makers in the future to reach fairer decisions that will not lead to the same level of dissatisfaction and controversies that one could observe in various countries in the 2019–2020 season. The idea of the PFSC can also be applied, up to minor modifications, to several other types of sports tournaments.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10182-021-00416-6>.

Acknowledgements The authors would like to thank the Special Issue editors and two anonymous reviewers for helpful comments that led to an improvement of our paper.

Funding The authors have no funding to report.

Availability of data and material The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors have no conflict of interest to declare.

Code availability The code for the application part is available in Online Resource 1.

References

BBC Sport Arsenal plan to make 55 staff redundant because of coronavirus pandemic. (2020) <https://www.bbc.co.uk/sport/football/53668802>. Accessed 14 December 2020

- Constantinou, A., Fenton, N.: Solving the problem of inadequate scoring rules for assessing probabilistic football forecast models. *J. Quant. Anal. Sports* **8**(1), 1 (2012)
- Csató, L.: Coronavirus and sports leagues: how to obtain a fair ranking if the season cannot resume? [arXiv:2005.02280](https://arxiv.org/abs/2005.02280). Accessed 23 December (2020)
- Devlin, K.: *The Unfinished Game: Pascal, Fermat, and the Seventeenth-Century Letter that Made the World Modern*. Basic Books (2010)
- Duckworth, F., Lewis, A.: A fair method for resetting the target in interrupted one-day cricket matches. *J. Oper. Res. Soc.* **49**, 220–227 (1998)
- Ekstrøm, C., Van Eetvelde, H., Ley, C., Brefeld, U.: Evaluating one-shot tournament predictions. *J. Sports Anal.* (in press) (2020)
- Epstein, E.: A scoring system for probability forecasts of ranked categories. *J. Appl. Meteorol.* **8**(6), 985–987 (1969)
- Fischer, K., Haucap, J.: Does crowd support drive the home advantage in professional soccer? Evidence from German ghost games during the COVID-19 pandemic (2020). <http://hdl.handle.net/10419/222278>. Accessed 23 December 2020
- Glenn, W., David, H.: Ties in paired-comparison experiments using a modified Thurstone–Mosteller model. *Biometrics* **16**(1), 86–109 (1960)
- Gneiting, T., Raftery, A.: Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **102**(477), 359–378 (2007)
- Gorgi, P., Koopman, S., Lit, R.: Estimation of final standings in football competitions with premature ending: the case of COVID-19 (2020). <https://papers.tinbergen.nl/20070.pdf>. Accessed 23 December 2020
- Guyon, J.: The model to determine Premier League standings (2020). <https://www.thetimes.co.uk/article/the-model-to-determine-premier-league-standings-ttt8tmldd>. Accessed 14 December 2020
- Nieuws, H.: Wat als corona geen spelbreker was? Statistici: Club kampioen, Europese tickets volgens stand na speeldag 29 en geen Europees voor Anderlecht (2020). <https://www.hln.be/belgisch-voetbal/wat-als-corona-geen-spelbreker-was-statistici-club-kampioen-europese-tickets-volgens-stand-na-speeldag-29-en-geen-europees-voor-anderlecht-aeb9e387/>. Accessed 14 December 2020
- Holroyd, M.: Lyon among French football teams upset over league position decision (2020). <https://www.euronews.com/2020/05/01/french-ligue-1-abandoned-clubs-to-consider-appealing-decision>. Accessed 14 December 2020
- Karlis, D., Ntzoufras, I.: Analysis of sports data by using bivariate Poisson models. *J. R. Stat. Soc.: Ser. D (The Statistician)* **52**, 381–393 (2003)
- Kicker: 13 der 36 Profiklubs droht Insolvenz - noch in dieser Saison (2020). https://www.kicker.de/13_der_36_profiklubs_droht_insolv enz_noch_in_dieser_saison-773385/artikel. Accessed 14 December 2020
- Lambers, R., Spieksma, F.: True rankings (2020). <https://www.euro-online.org/websites/orinsports/wp-content/uploads/sites/10/2020/05/TrueRanking.pdf>. Accessed 14 December 2020
- Ley, C., Van de Wiele, T., Van Eetvelde, H.: Ranking soccer teams on the basis of their current strength: a comparison of maximum likelihood approaches. *Stat. Model.* **19**, 55–73 (2019)
- Maher, M.: Modelling association football scores. *Statistica Neerlandica* **36**, 109–118 (1982)
- Pantuso, G., Hvattum, L.: Maximizing performance with an eye on the finances: a chance-constrained model for football transfer market decisions. *TOP* Forthcoming. (2020)
- R Core Team R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, (2020). <https://www.R-project.org/>
- Sæbø, O., Hvattum, L.: Modelling the financial contribution of soccer players to their clubs. *J. Sports Anal.* **5**, 23–34 (2019)
- Sky Sports: Coronavirus: Football clubs at risk of going bust without government aid, open letter warns (2020). <https://www.skysports.com/football/news/11095/12083809/coronavirus-football-clubs-at-risk-of-going-bust-without-government-aid-open-letter-warns>. Accessed 14 December 2020