



Forecasting a winner for Malaysian Cup 2013 using soccer simulation model

Muhammad Mat Yusof, Mohd Soffian Omar Fauzee, and Rozita Abdul Latif

Citation: [AIP Conference Proceedings](#) **1605**, 1153 (2014); doi: 10.1063/1.4887753

View online: <http://dx.doi.org/10.1063/1.4887753>

View Table of Contents: <http://scitation.aip.org/content/aip/proceeding/aipcp/1605?ver=pdfcov>

Published by the [AIP Publishing](#)

Articles you may be interested in

[Forecasting performance of denoising signal by Wavelet and Fourier Transforms using SARIMA model](#)

AIP Conf. Proc. **1605**, 961 (2014); 10.1063/1.4887720

[Large eddy simulation of wind turbine wake dynamics in the stable boundary layer using the Weather Research and Forecasting Model](#)

J. Renewable Sustainable Energy **6**, 033137 (2014); 10.1063/1.4885111

[Modeling the Malaysian motor insurance claim using artificial neural network and adaptive NeuroFuzzy inference system](#)

AIP Conf. Proc. **1522**, 1431 (2013); 10.1063/1.4801297

[Forecasting Diffusion of Technology by using Bass Model](#)

AIP Conf. Proc. **1148**, 149 (2009); 10.1063/1.3225258

[Modelling Set Up for ThroughProcess Simulation of Aluminium Cup Production](#)

AIP Conf. Proc. **712**, 388 (2004); 10.1063/1.1766555

Forecasting A Winner For Malaysian Cup 2013 Using Soccer Simulation Model

Muhammad Mat Yusof^a, Mohd Soffian Omar Fauzee^b and Rozita Abdul Latif^c

^a*School of Quantitative Sciences, UUM College of Arts and Sciences,
Universiti Utara Malaysia, 06010, Sintok, Kedah*

^b*School of Education and Modern Languages, UUM College of Arts and Sciences,
Universiti Utara Malaysia, 06010, Sintok, Kedah*

^c*Faculty of Sport Science and Recreation, Universiti Teknologi MARA, 40450 Shah Alam, Selangor*

Abstract. This paper investigates through soccer simulation the calculation of the probability for each team winning Malaysia Cup 2013. Our methodology used here is we predict the outcomes of individual matches and then we simulate the Malaysia Cup 2013 tournament 5000 times. As match outcomes are always a matter of uncertainty, statistical model, in particular a double Poisson model is used to predict the number of goals scored and conceded for each team. Maximum likelihood estimation is used to measure the attacking strength and defensive weakness for each team. Based on our simulation result, LionXII has a higher probability in becoming the winner, followed by Selangor, ATM, JDT and Kelantan. Meanwhile, T-Team, Negeri Sembilan and Felda United have lower probabilities to win Malaysia Cup 2013. In summary, we find that the probability for each team becoming a winner is small, indicating that the level of competitive balance in Malaysia Cup 2013 is quite high.

Keywords: soccer simulation model, Malaysia Cup 2013, Poisson model.

PACS: 02.50.-r

INTRODUCTION

Soccer (or football) is a popular sport that is enjoyed across many cultures and countries. Currently, there are more than 240 million registered players around the world and millions of fans are watching, thinking and speculating about the result of soccer matches. In professional soccer events like the World Cup or English Premier League (EPL) in England, the interests for this kind of sport is caused by two reasons, firstly the quality of play and secondly the uncertainty about match outcome [1].

In Malaysia, soccer is considered a popular type of team sport. Many professional soccer events have been organized in Malaysia such as the Malaysian Cup, Malaysian Super League, Malaysian Premier League, Malaysian FA Cup, FAM League and President Cup. These tournaments are organized by The Football Association of Malaysia (FAM), which is the governing body for football in Malaysia. Based on the hierarchy of those tournaments, Malaysian Cup is the prestigious tournament and considered the highest competition in Malaysia.

The Malaysian Cup began in 1921 and the Malaysian Cup 2013 is the 87th edition organized by FAM. Currently, this tournament involves 16 teams, contested by the top 10 teams from Malaysian Super League 2013 (the top-tier in FAM Malaysian soccer league), the top 4 teams from Malaysian Premier League 2013 (the second-tier in FAM Malaysian soccer league) and 2 winning teams in playoffs round (There are 4 teams competing in playoffs matches comprising of 2 bottom teams in Malaysian Super League 2013 and teams who finished 5th and 6th places in Malaysian Premier League 2013). The detailed explanation on the design of Malaysian Cup 2013 is shown in the next section.

In this paper, we will use statistical modeling and soccer simulation model to calculate the probability for each team winning Malaysian Cup 2013. We argue that match outcome in soccer are always a matter of uncertainty, therefore a statistical model, in particular a Poisson model is useful to predict the number of goals scored and conceded for each team. To do this, we will develop a soccer simulation model that uses Monte Carlo technique. Here, we use Monte Carlo technique to generate both match outcomes, that is, we use simulated values from one of the statistical methods for match prediction models to represent match scores. In literature we find that the use of Monte Carlo technique in sport is widespread. For example, Koning et. al. [2] and Goddard [3] applied Monte Carlo simulation to predict result in a soccer (football) tournament and Norman and Clark [4] used Monte Carlo simulation to predict the outcomes in cricket tournament.

Proceedings of the 21st National Symposium on Mathematical Sciences (SKSM21)

AIP Conf. Proc. 1605, 1153-1159 (2014); doi: 10.1063/1.4887753

© 2014 AIP Publishing LLC 978-0-7354-1241-5/\$30.00

The remainder of the paper proceeds as follow. In the next section, the dataset for measuring the parameters of attacking and defensive strength for each team is briefly described, followed by a description of the modeling match outcome approach. The results of the model fitting are then presented. Finally, we concluded with a discussion of the limitations of our works, and opportunities for further development.

METHODOLOGY

The fundamental idea of this paper is to simulate Malaysian Cup 2013. In order to duplicate features, appearance and characteristics of Malaysian Cup 2013, we firstly need to know how the Football Association of Malaysia (FAM) design a Malaysian Cup 2013. In this paper, we do not attempt to describe in detail all the characteristics of Malaysian Cup 2013 but merely seek to describe in general three main components of tournament design used in Malaysian Cup 2013. Interested reader who wantsto know in depth the design of Malaysian Cup 2013 can refer to this website: http://en.wikipedia.org/wiki/2013_Malaysia_Cup.

In general, the design of Malaysian Cup 2013 can be divided into three main components; they are tournament structure, a seeding policy and a progression rule. Here, we describe tournament structure as the arrangement of individual matches without specifying which them will be paired, the seeding policy is the assignment of teams into individual matches, and progression rules are the rules used to describe how each teamwill be advanced or eliminated from the tournament.

When we look at the point of view of tournament structure, Malaysian Cup 2013 usedone-group round robin followed byknockout structure. In group round, a 2-leg round-robin is used in which each team will play two times, once at home and once away. Meanwhile, in knockout stages, the quarter and semi-final used a knockout (2-leg round-robin) matches, and in the final round, knockout 1-leg matches is used.

The placement of teams in groups is based on seeding policy. The sixteen teams that qualified are firstly divided into four pots. Pot 1 is reserved for LionXII, Selangor, JDT and Kelantan. Pot 2 is for Pahang, ATM, Perak and PKNS. Pot 3 is for Terengganu, T-Team, Sarawak and Sime-Darby and the final pot, Pot 4 is reserved for Kedah, Johor, Felda United and Negeri Sembilan. After that, one team from each pot is chosen at random to group 1, group 2, group 3 and group 4respectively. In Figure 1, we show the allocations of teams in each group and demonstrate the three-diagram of matches for quarter finals, semi finals and final.

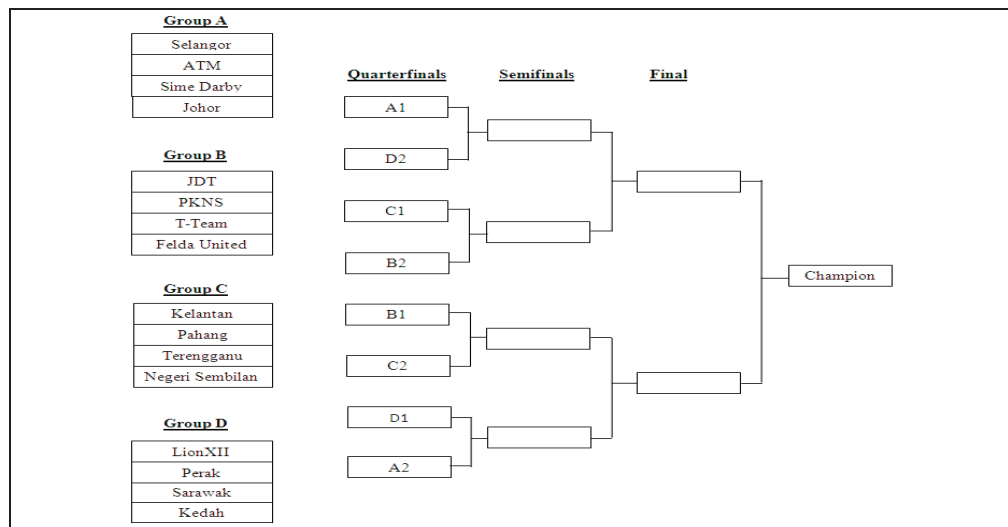


FIGURE 1.The team allocations in each group and the structure used in Malaysian Cup 2013. Note that the A1 refers to the champion for group A, D2 refers to second runner-up for group D, etc.

Data Description

In order to simulate Malaysian Cup 2013, we need to simulate the individual matches. To simulate individual matches, we need a model that predicts the outcome of the individual matches. In this paper, we apply a double Poisson model to modeling match outcome. Note that in sport contest like soccer, the probability that ball possession will result in a goal is small but the number of ball possession events is large. In particular, if we assume a constant ability over time of a team to score goals, the Poisson distribution seems a plausible model [5,6,7]. In order to use the model, we firstly need to estimate the attacking strength and defensive weaknesses for all teams contested in Malaysian Cup 2013.

We estimate the attacking strength and defensive weakness based on historical match outcome. Our data were obtained from two sources: (1) Sokernet website (<http://ligasuper.sokernet.com/keputusan/keputusan-liga-super-2013>; accessed on 25.6.13) and (2) Sportstat website (<http://www.sportstats.com/soccer/malaysia/premier-league/results/>; assessed on 25.6.13). Since Malaysian Cup 2013 tournament involved teams in two different leagues, we need to use two different datasets. The first dataset refers to historical match outcomes for 2013 Malaysian Super League. Meanwhile, the second dataset refers to historical match outcomes for 2013 Malaysian Premier League. In total, we have used 132 match results in 2013 Malaysian Super League and 48 match results in 2013 Malaysian Premier League.

By using the maximum likelihood estimation, Table 1 shows the attacking strength and defensive weakness for sixteen teams playing in Malaysian Cup 2013.

TABLE 1. Attacking strength and defensive weakness for teams participating in Malaysian Cup 2013. Home advantage is 1.3234.

Teams	Attacking	Defensive	Teams	Attacking	Defensive
ATM FA	0.8860	1.4266	Negeri Sembilan FA	0.5164	0.9330
Sime Darby FC	0.6916	0.9742	Pahang FA	1.3699	1.2896
Johor FA	0.5069	0.9142	Kelantan FA	0.9654	1.0096
Selangor FA	0.9862	0.8169	Terengganu FA	0.9787	1.1662
PKNS FA	1.4698	1.1798	Sarawak FA	0.8795	0.3845
T-Team FC	0.8968	1.0805	Perak FA	1.0264	0.8589
Johor Darul Takzim	1.0890	1.1387	Lion XII	1.0493	0.7042
Felda United FC	0.8159	0.9961	Kedah FA	0.6571	0.4507

Modelling Match Outcome

The prediction of match outcome is an important concern of research for a number of reasons. These are, for example, so that we can: measure the implications of changing the design from the point of view of the best team winning [8,9]; measure changes in competitive balance [2]; measure efficiency in the sports betting markets [3]; measure the effects of design changes [10]; and measure the importance of an individual match in a tournament [11,12].

To date, two types of probabilistic models for match prediction exist in the literature. These are (i) modeling match outcome (direct approach) and (ii) modeling match scores (indirect approach). The first approach utilizes ordinal regression models to forecast the ordered response variable, win/draw/loss. It typically uses ordered logit or ordered probit to calculate the probabilities. This approach is often used by econometricians. The indirect approach models the distribution of scores by each team, either independently or dependently. It typically employs an independent Poisson model [8], bivariate Poisson model or variation of that model, such as a diagonally inflated bivariate Poisson model [5]. This is often the approach taken in the applied statistics literature.

Goddard [3] through his relative test performance state that both approaches can be used to predict match outcome. However, McHale and Davies [13] argue that the indirect approach has an advantage not only of match outcome prediction, but the varieties of exact scores are assigned probabilities. Thus, it supplies a much richer outcome description. Therefore, this paper will use the second approach which is the modeling match scores. In particular, Malaysian Cup 2013 uses a 2-leg round-robin. Thus we argue that modeling match scores are useful in distinguishing the winner or progression team based on total number of goals scored and conceded for each team.

By letting X_{ij} be the number of goals scored by team i and Y_{ij} be the number of goals scored by team j , the match outcome model is given by

$$\begin{aligned} (X_{ij}, Y_{ij}) &\sim \text{Poisson}(\alpha_i \beta_j \gamma, \alpha_j \beta_i) \text{ for team } i \text{ playing at home against team } j \\ (X_{ij}, Y_{ij}) &\sim \text{Poisson}(\alpha_i \beta_j, \alpha_j \beta_i \gamma) \text{ for team } i \text{ playing at away against team } j \end{aligned} \quad (1)$$

where α_i and β_i are the attacking strength and defensive weakness of team i respectively, and γ denotes the home advantage parameter.

RESULTS

In producing the probability for each team win Malaysian Cup 2013, we have simulated the tournaments 5000 times. All the tournament simulation is written in MathCAD software. Before running the full simulation, we firstly conducted an exercise to test our tournament simulation code. In the first test, we consider a situation in which the variation in relative strength and weakness among the contestants is small. Therefore we expect the probability of each team becoming the winner in group stage is nearly 0.25 and the probability for each becoming champion in Malaysian Cup 2013 is nearly 0.06. Therefore, based on our simulation test, we find that each team has a probability near to that value. Thus, we consider our simulation code is fit to run for full simulation.

In Figure 2 and Figure 3, we show part of our screenshot from MathCAD software that demonstrates our simulation of Malaysia Cup 2013.

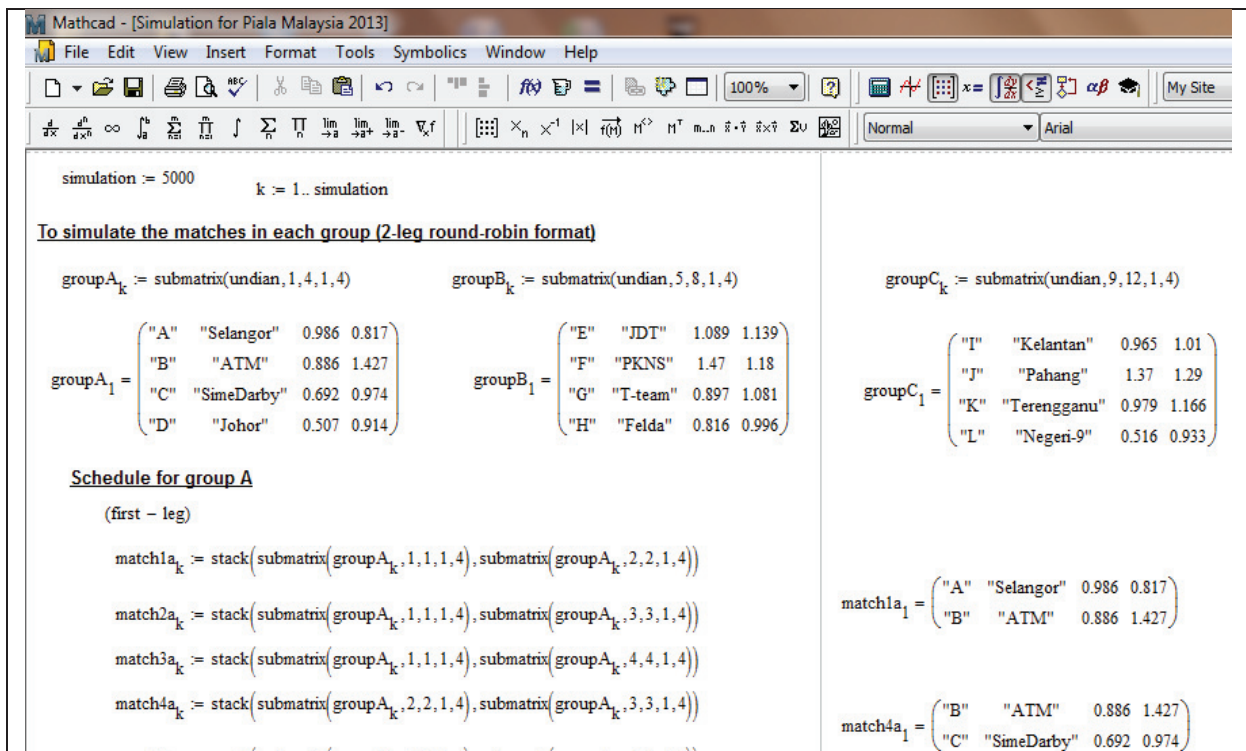


FIGURE 2. Screenshot taken from MathCAD software that illustrates the structure of matches in groups A.

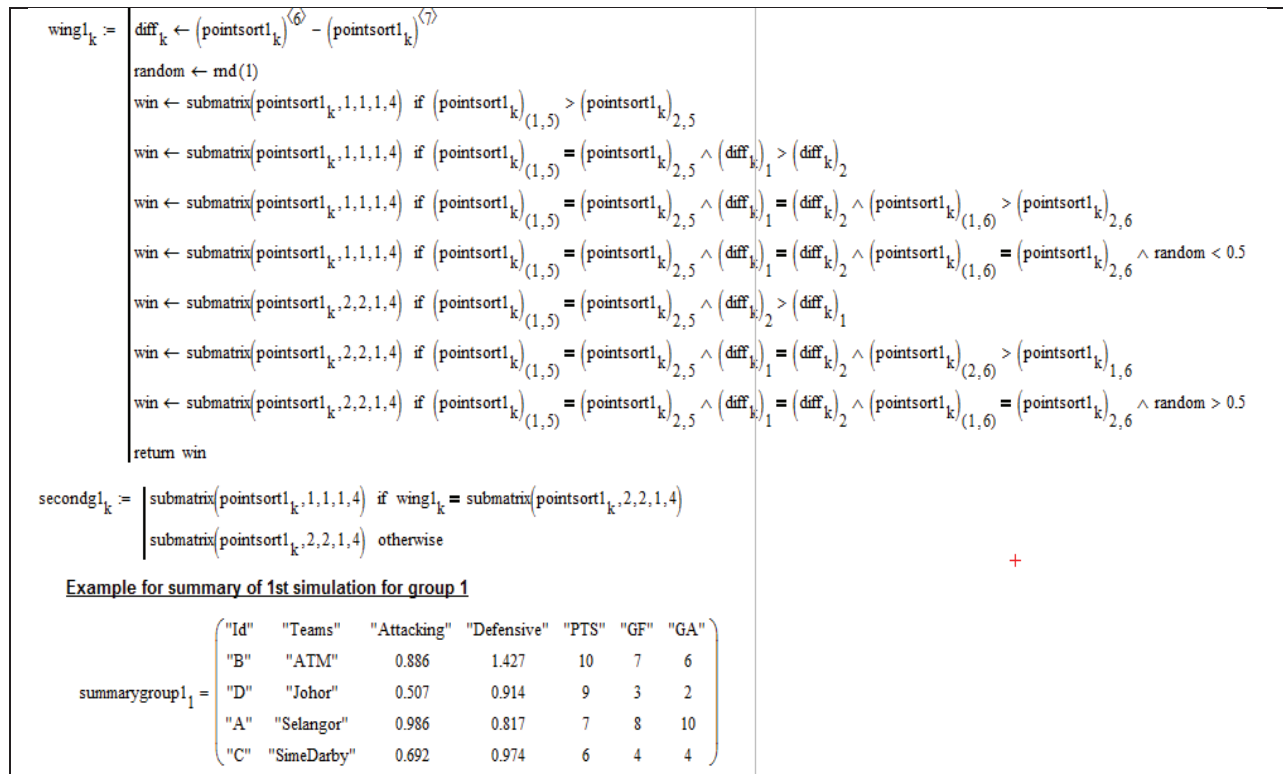


FIGURE 3. Screenshot taken from MathCAD software that illustrate the rule to determine a league table for Group A.

In Table 2, we show from our simulation result the probability for each team winning in group round. As we can see, in group A for example, the probability for ATM becoming a group winner is 0.275, the probability for Sime Darby winning is 0.256, the probability for Selangor winning is 0.248 and the probability for Johor becoming the group winner is 0.221.

TABLE 2. A probability of winning for each team winning in group round.

Teams	Prob.	Rank order in our simulation model	Actual rank order	Teams	Prob.	Rank order in our simulation model	Actual rank order
GROUP A				GROUP C			
ATM	0.275	1 st	1 st	Negeri Sembilan	0.228	4 th	3 rd
Sime Darby	0.256	2 nd	2 nd	Pahang	0.260	2 nd	2 nd
Johor	0.221	4 th	3 rd	Kelantan	0.267	1 st	1 st
Selangor	0.248	3 rd	4 th	Terengganu	0.246	3 rd	4 th
GROUP B				GROUP D			
PKNS	0.263	1 st	1 st	Sarawak	0.233	3 rd	1 st
T-Team	0.247	3 rd	3 rd	Perak	0.268	2 nd	3 rd
Johor DT	0.256	2 nd	2 nd	LionXII	0.269	1 st	2 nd
Felda United	0.234	4 th	4 th	Kedah	0.23	4 th	4 th

Table 2 also shows the comparison of rank order between our simulation in group stage result and the actual group stage result for Malaysian Cup 2013. As seen in table 2, our simulation result accurately predicts the group champion for group A, group B and group C. Meanwhile for group D, we expect LionXII will be the champion for group D but in reality, Sarawak became champion.

Finally in Table 3, we show the probability for all teams winning Malaysian Cup 2013. Based on Table 3, we find that LionXII has greater chances to become winner for Malaysian Cup 2013, followed by Selangor, ATM, Johor DarulTakzim and Kelantan. Meanwhile T-Team, Negeri Sembilan and Felda United have lower probability to win Malaysian Cup 2013.

TABLE 3. The probability for each team winning a Malaysian Cup 2013

Teams	Prob. of winning	Teams	Prob. of winning
LionXII	0.122	Sime Darby	0.045
Selangor	0.108	Terengganu	0.045
ATM	0.106	Perak	0.043
Johor DT	0.094	Johor	0.037
Kelantan	0.082	Kedah	0.036
Sarawak	0.072	T-Team	0.034
PKNS	0.065	Negeri Sembilan	0.034
Pahang	0.050	Felda United	0.027

DISCUSSION AND FUTURE WORK

This paper demonstrates using simulation and statistical modeling to predict the probability for all teams participating in Malaysian Cup 2013 to become winner. We calculate the probability for each team win using a double Poisson model. Note that our result in this paper is fundamentally based on three main assumptions. Firstly, the attacking strength and defensive weaknesses for all teams can be measured quantitatively, which is based on historical match outcomes in Malaysian Super league 2013 and Malaysian Premier League 2013. Secondly, we assume the strength is fixed through the tournament. Here, we do not consider other factors affecting attacking strength and defensive weaknesses such as injuries of main players, resignation of team coach or corruption claim occurring throughout Malaysian Cup 2013. Thirdly, we assume that the match outcome is independent.

Based on the result in our soccer simulation model, LionXII has greater ability to win Malaysian Cup 2013, followed by Selangor, ATM, Johor DT and Kelantan. In group round, we find that six teams have greater chances to qualify to knockout stage. They are ATM, Sime Darby, Selangor PKNS, Johor DarulTakzim, Kelantan, Pahang, LionXII and Perak. Comparative studies with actual results show that seven out of eight teams qualify to second round. It shows that the double Poisson model we used is able to accurately predict the match result.

In future research, we aim to extend this soccer simulation model to investigate other characteristics of tournament design such as the effects of changing point system, the effect of changing tournament structure and seeding policy and changing of progression rule. Based on match prediction model, we aim to use other match prediction models and checking their match outcome accuracy. Several considerations should be included in measuring team strength such as injuries of main player, the dependent of result throughout the tournament and the tactical strategies used by team coach.

ACKNOWLEDGMENTS

We would like to take this opportunity to thank Prof Phil. Scarf from University of Salford, United Kingdom for his ideas and guidance related to the uses of simulation and statistical modeling in the area of sport tournament design.

REFERENCES

1. R. Koning, *DE Econs.*, **157**(2), 229-246(2009).
2. R. Koning, and M. Koolhaus, and G. Renes, and G. Ridder, *European J. Oper. Res.*, **148** (2), 268-276 (2003).
3. J. Goddard, *Int. J. Forecasting*, **21** (2), 331-340 (2005).
4. J. M. Norman, and S. R. Clarke, *J. Oper. Res. Soc.*, **61**, 980-986 (2010).
5. D. Karlis and I. Ntzoufras, *J. Stat. Softw.*, **14**, 1-32 (2003).
6. A. J. Lee, *Chance*, **10**, 15-19 (1997).
7. M. J. Maher, *Stat. Neerl.*, **36** (3)109-118 (1982).

8. T. McGarry, "On the design of sport tournaments" in *Statistics in Sport*, edited by J. Bennett, London: Edward Arnold, 1998, pp.199-217.
9. T. McGarry, and R. W. Schutz, *J. Oper. Res. Soc.* **48** (1), 65-74 (1997).
10. P. A Scarf, M. Bilbao, and M. M. Yusof, *Eur. J. Oper. Res.* **198** (1), 190-198 (2009).
11. D. K. Goossens and C. R. Spieksma, *Ann. Oper. Res.* **194**, 223-240 (2012).
12. P.A. Scarf and X. Shi, *IMA J. Manag. Math.* **16** (2), 161-178 (2005).
13. I. McHale and S. Davies, "Statistical analysis of the effectiveness of the FIFA world rankings" in *Statistical Thinking in Sports*, edited by J. Albert and R. Koning, CRC Press, 2007, pp. 77-89.