TOP
https://doi.org/10.1007/s11750-020-00576-9

ORIGINAL PAPER

Sports scheduling and other topics in sports analytics: a survey with special reference to Latin America

Guillermo Durán[1,2]

Received: 31 March 2020 / Accepted: 28 June 2020

Abstract
Recent years have witnessed vigorous development in the application of mathematical and computational techniques to many aspects of the organization and planning of sports competitions. Known as sports scheduling, this field is one of the subjects of the present survey, which reviews different problems tackled in the associated literature, the techniques employed for solving them and the results of their implementation in real-world cases. Other topics in sports analytics such as player performance, result prediction, fantasy games and analysis of rankings are also examined. Special attention is given to applications in Latin America. The main challenges facing sports scheduling and other areas of sports analytics are also discussed.

Keywords Sports analytics · Sports scheduling · Travelling tournament problem · Integer linear programming · Heuristics

Mathematics Subject Classification 00A69 · 90C10 · 90C59 · 90C90

1 Introduction

Sports are not only one of the world's most popular recreations but also a global industry with a large economic footprint. The last Super Bowl, the annual championship game of professional American football (NFL), was a great sporting and artistic event. Held in Miami in February 2020, it featured a half-time show with Jennifer Lopez and Shakira, pulling in a worldwide television audience of some 100 million viewers for a single event that lasted more than 3 h. The Rio Olympics in 2016 brought together more than 11,000 athletes representing 207 nations, who competed in 41 disciplines from 28 different sports. The games were followed on television

✉ Guillermo Durán
gduran@dm.uba.ar

[1] Instituto de Cálculo y Departamento de Matemática, FCEN-UBA and CONICET, Buenos Aires, Argentina

[2] Departamento de Ingeniería Industrial, FCFM-Universidad de Chile, Santiago, Chile

Published online: 13 July 2020

by 5 billion viewers around the world and the cost of staging them was 13 billion US dollars. Football, the most popular sport in Latin America, has some 200 million registered players at all levels in more than 200 countries affiliated with FIFA, the game's international governing body. The organization's most recent World Cup, hosted by Russia in 2018, took place over an entire month, attracted a television audience estimated at more than 3.5 billion and required investments of about 14 US billion dollars, 2 billion more than the previous World Cup held in 2014 in Brazil.

These figures not only underline the importance of sports across the planet but also suggest a breadth of possibilities for the utilization of quantitative tools to enhance athlete performance and improve the organization, logistics, economics and overall functioning of sports activities generally. Such applications of mathematical tools are known in the literature as sports analytics, and make particularly heavy use of the techniques of optimization, statistics and data science.

One of the fields within sports analytics that has seen the greatest development in recent years is the efficient organization and planning of competitions known as sports scheduling. The present survey reviews the main problems addressed in the sports scheduling literature, the mathematical and computational techniques employed in their solution and the results they have generated in actual practice. Special emphasis is given to applications in Latin America, their implementation and impact.

The first theoretical results in the field of sports scheduling date back to the 1980s and the work of Dominique (De Werra 1980, 1981, 1982). In particular, the author proved certain results on the minimum number of breaks (consecutive matches with the same home-away status) in round-robin tournaments using concepts borrowed from graph theory.

Most sports leagues around the world use a round-robin schedule format in which each team plays every other team a fixed number of times. If they meet each other just once, the format is a single round robin; if twice, it is a double round robin, etc. According to Nemhauser and Trick (1998), there are two types of round-robin schedules: temporally constrained and temporally relaxed. In the former case, the number of available game slots (known as "rounds") is equal to the number of games each team must play, plus any necessary byes for leagues with an odd number of teams. Such schedules are said to be compact. Many leagues use this setup, including most professional football leagues in Europe and Latin America. In temporally relaxed schedules, on the other hand, the number of available rounds for each team is greater than the minimum necessary so each team will have multiple byes. This type of schedule is used by professional leagues in North America such as the National Basketball Association (NBA) (Bean and Birge 1980; Bao 2009) and the National Hockey League (NHL) (Fleurent and Ferland 1993; Costa 1995; Craig et al. 2009), numerous amateur sports leagues (Knust 2010), and cricket leagues in Australia, England and New Zealand (Willis and Terrill 1994; Wright 1994, 2005).

Interesting surveys of the sports scheduling literature can be found in Easton et al. (2004), Kendall et al. (2010) and Rasmussen and Trick (2008), but all of them were published at least a decade ago so an update covering the material published in the years since is now in order. Benchmarks of instances from some applications can be found in Nurmi et al. (2010) and Van Bulck et al. (2020).

As regards sports analytics problems other than sports scheduling, the range of possibilities is even greater. A number of these will be reviewed in this survey, including results prediction, fantasy games, rankings analysis and player performance, the best-known example of the latter being an application to baseball described in the book entitled Moneyball (Lewis 2003).

The remainder of this paper consists of three sections. Section 2 is devoted to sports scheduling and surveys problems of optimization and feasibility in the drawing up of sports schedules and efficient referee assignments, with special focus on Latin-American applications. A number of basic formulations using integer programming are also exhibited. Section 3 reviews some results on other topics in sports analytics, including player performance, results prediction, fantasy games and sports ranking analysis. Finally, Sect. 4 presents our conclusions and discusses a number of interesting possibilities for future work in sports scheduling and other areas of sports analytics.

2 Sports scheduling

This section addresses a number of different sports scheduling problems and develops, for three cases, their formulation as a mathematical model. The first subsection introduces the optimization problems addressed in sports scheduling, most of which are variations on the famous travelling tournament problem (TTP). The second subsection is devoted to feasibility problems, for which merely finding a feasible solution is a complex task given the set of constraints. Our focus will be on applications to football, which generally fall into this category. The third subsection deals with problems relating to the efficient assignment of referees in sports leagues, which are essentially derivations of the travelling umpire problem (TUP) and the referee assignment problem (RAP). Finally, the fourth subsection presents integer programming models for some of the problems formulated in the three previous subsections.

2.1 Optimization problems: the TTP and its variants

An excellent test bed for sports scheduling models, algorithms and methodological tools is the TTP as set out in Easton et al. (2001). Given a set of n teams and the distances between their home cities (using as a source of the first instances the US-based Major League Baseball), the TTP consists in scheduling a fictitious compact double round-robin tournament (each team plays every other team twice, once at home and once away) with $2(n-1)$ rounds in such a way that no team plays fewer than L or more than U consecutive games either at home or away, no team plays any other team in two consecutive rounds, and the total distance travelled is minimized. Also, no team returns home during a sequence of away games known as a road trip.

The TTP is very difficult to solve; indeed, small instances of just 12 teams have still not been solved to optimality (see the TTP website Trick 2001). This remains the case even for circular instances in which the team venues form a circle such that the distance between two consecutive team venues is always equal to 1 (Trick 2001).

By contrast, solving such instances of the famous travelling salesman problem (TSP) is trivial. However, it has also been proved in the case of the TTP that if all distances between any pair of teams are equal to 1, distance minimization and breaks maximization are equivalent and solving the instance is, therefore, straightforward (Urrutia and Ribeiro 2006).

The computational complexity of the TTP has not yet been fully determined apart from a few special cases of L and U. What is known is that the problem is polynomial for $L = 1$ and $U = 2$ (Easton et al. 2003), and NP-hard for $L = 1$ and $U = 3$ (Thielen and Westphal 2011) as well as for $L = 1$ and $U = \infty$ (Bhattacharyya 2016). For any other values of L and U, the computational complexity remains an open question.

Various techniques have been suggested for solving the TTP. A method proposed in Easton et al. (2003) combines integer programming with constraint programming (CP) while in Goerigk and Westphal (2012) integer programming is combined with local search. In Benoist et al. (2001), an algorithm mixes CP with Lagrangian relaxation and in Henz (2004) CP is used in conjunction with large neighborhood search. Different heuristic solution approaches have also been developed, such as simulated annealing (Anagnostopoulos et al. 2006; Van Hentenryck and Vergados 2006), tabu search (Cardemil and Durán 2002; Di Gaspero and Schaerf 2007) and genetic algorithms (Choubey 2010). Yet other proposals attempt to solve the TTP using approximation algorithms (Miyashiro 2008; Yamaguchi et al. 2009; Westphal and Noparlik 2012).

A number of authors have analyzed variants of the original TTP. One that arises quite naturally is the travelling tournament problem with predefined venues (TTPPV) in which each pair of teams plays a single match with a predetermined home-away status (Melo et al. 2006). The problem consists in finding a compact single round-robin schedule compatible with the preset venue definitions such that the total distance traveled by the teams is minimized and no team plays more than three consecutive games either at home or away. Three different formulations of the TTPPV are presented in Melo et al. (2006), one of which will be considered here in Sect. 2.4.1. The computational tests detailed in the cited study show that instances of the problem up to $n = 8$ can be solved to optimality with exact formulations but beginning with $n = 10$, recourse must be had to heuristic methods.

The mirrored version of the problem has been analyzed in two articles. A two-stage method that can solve to optimality instances of up to eight teams is proposed in Cheung (2008). A heuristic combining GRASP with an iterated local search presented in Ribeiro and Urrutia (2007) has obtained very good solutions for large instances including the Brazilian football league in 2003 which then had 24 teams. Another variant of the TTP, for application to multiple round robins (i.e., not just a double), is presented in Hoshino and Kawarabayashi (2011). The authors designed a method based on Dijkstra's well-known algorithm to obtain a multi-round schedule. They tested their approach in a case study of the Nippon Professional Baseball (NPB) in Japan, where 2 leagues of 6 teams play a total of 40 sets of three intra-league games over a series of 8 single round robins. The resulting schedules reduced the total travel distance of the NPB's manually defined 2010 season calendar by 25%.

A temporally relaxed variation called the time-relaxed travelling tournament problem (TRTTP) is developed in Bao (2009). In this problem, L and U are defined as in the TTP. The number of byes per team in the schedule is controlled by a parameter K. When $K = 0$, the problem just reduces to the TTP. In general, the computational complexity of the TRTTP is unknown. For the case where $U = 1$, the problem is trivial and the author presents an algorithm that proves it is polynomial for any K with $L = 1$ and $U = 2$. For other values of L, U and K it is conjectured that the problem is NP hard. Various instances generated for the TTP have also been studied for the TRTTP (Brandao and Pedroso 2014).

Two real-world applications of the TTP were developed in Latin America. One was for Argentina's professional volleyball league (Bonomo et al. 2012), the other for the country's professional basketball league (Durán et al. 2019a).

In the volleyball case, schedules were defined for the league's First Division using a variation on the TTP, which to our knowledge is the first application of this problem to a real-world league reported in the optimization literature. The division consists of 12 teams that play a double round robin. They are grouped into pairs known as couples. Matches are held on Thursdays and Saturdays (referred to as weekends) and in every pair of weekend games the two teams in each couple play against the two teams of another couple. One weekend in the season is reserved for games between the members of each couple. Since the teams' home venues are scattered widely across the country, a particularly important aspect of the scheduling is the minimization of travel distance.

The volleyball problem is thus a variation of the TTP in which road trips are undertaken by pairs of teams travelling together rather than each team travelling individually. This coupled format gives rise to two key problems: (a) how to pair off the teams into couples, and (b) how to schedule the matches. They are approached in two consecutive stages. The first stage is a procedure that takes into account travel distances and is modeled using integer linear programming. This model is solved to optimality in 10 min to 3 h and incorporates symmetry-breaking constraints. In the second stage, if the schedule is mirrored the solution is tackled by an integer programming model that finds a solution in anywhere from just minutes to 6 h while for non-mirrored formats a tabu search heuristic does the same in a matter of minutes (in this latter case, using integer programming resulted in very large optimality gaps after 10 h of execution time and feasible solutions of poorer quality than those found by the heuristic). The schedules generated by this application have been successfully implemented since the 2007–2008 season, reducing total travel distances by 15–30% while meeting all other requirements.

As for the professional basketball application, the league's schedules have been designed since the 2014–2015 season using a new format adopted by the top two divisions. Following the setup employed by the National Basketball Association (NBA) in North America, matches are held any day of the week and those played away are scheduled in road trips of one to four games. The main scheduling objective was to reduce the teams' total travel distance compared to previous season formats through the use of predetermined trips submitted by the teams. The mathematical form of the problem is therefore another variation on the TTP, without the couple format of the volleyball case and allowing only a subset of road trips as requested by the teams. This variant is called trip preferences—time-relaxed travelling tournament problem (TP-TRTTP).

The modeling in this case is again divided into two stages, the first one defining the sequences in which each team plays the other teams and the second one assigning the days on which each game is played. Both use integer programming models that incorporate a series of constraints reflecting criteria set by the Argentine Basketball Club Association. The second-stage model was solved in every instance in a matter of minutes. The first-stage model is much more difficult to solve, however, and in certain cases, the season has had to be decomposed into thirds or quarters, each of which is then solved separately with run times ranging from 5 min to 5 h. The objective function of the first-stage model seeks to maximize the number of games that are scheduled in accordance with the road trip preferences submitted by the teams. Since these preferences incorporate the teams' desire to reduce travel distances, formulating the problem in this manner automatically achieves such reductions. Implementation of the models reduced the average travel distance per away game under the previous manually designed couple format by more than 30%, with consequential benefits in lower travel costs and less player fatigue.

A number of other real-world applications of OR scheduling to professional leagues in these sports have been reported in the literature. In basketball, there are the cases of New Zealand (Wright 2006), the Czech Republic (Froncek 2001) (using the couples format) and Germany (Westphal 2014), this last work a finalist for the EURO Excellence in Practice Award (Vilnius, 2012) and the INFORMS Wagner prize (2013). In volleyball, an example is the Italian league (Cocchi et al. 2018). In all of these applications, however, travel distance is not a consideration so the problem to be solved is essentially one of feasibility. This aspect will be treated in the next subsection.

2.2 Feasibility problems: football applications

Unlike the cases discussed in the previous section, where optimization figures prominently because of the travel reduction objective, many sports scheduling problems have as their main objective the satisfaction of a series of constraints imposed by a governing body of the league to be scheduled. In such situations, the problem is essentially one of feasibility, and very often it is precisely the need to satisfy all of those restrictions that results in great computational complexity. Typically there will still be an objective function (to minimize breaks, minimize occurrences of multiple games in a given locality/venue in the same round, etc.) but in the search for a good solution it will not be the primary concern.

The majority of the world's football leagues fit this category given that their season formats are typically round robins with a single match per week, which effectively rules out the possibility of multi-game road trips and, therefore, the need to minimize travel distance. The prime focus of this subsection will thus be on feasibility problems, especially as they arise in football and with an emphasis on successful Latin American applications.

To the best of our knowledge, there are only two documented real-world cases of the application of OR scheduling to football in the last century. The first case was the 1989–1990 season of the Dutch league (Schreuder 1992) in which 18 teams competed in a single round robin. The underlying model considered various commercial, sporting, and organizational factors. The aim was to satisfy certain conditions by maximizing a weighted sum of soft requirements. The solution method was heuristic and 80% of the requirements were satisfied.

The other case, this one in Latin America, was the 1995 season of the Argentinean league. A simulated annealing approach was used to assign teams to a canonical schedule, the same predetermined template used by most leagues around the world where mathematical scheduling techniques are not employed. According to Eduardo Dubuc [reported in Paenza (2006)], the assignments considered a number of factors relating to television broadcasting of the games. The approach was dropped the following year, however.

In the following subsections, we survey various real applications implemented in Latin America, Europe and international tournaments held over the last 20 years.

2.2.1 Applications in Latin American football

The first lasting experience with OR scheduling in football was initiated in Latin America, where mathematical techniques have been used to generate schedules for Chile's professional leagues since 2005. Actual implementation is handled by the Chilean Professional Football Association (ANFP), which applies integer programming-based methods to decide what matches are played in each round of the season. Various objectives are taken into account such as keeping costs down and ensuring that the schedules are attractive for fans. The Association has so far defined more than 60 schedules for the professional-level First, Second and Third Divisions, the Women's League and the youth divisions. The direct economic impact of the improvements obtained through these OR techniques has been estimated at approximately US dollars 70 million, which includes reductions in television broadcaster operating costs, growth in pay-television subscriptions, increased ticket revenue, and lower travel costs for the teams (Alarcón et al. 2017).

As noted earlier, the problems encountered in these applications are essentially those of feasibility. Since solving the basic model directly in its integer linear formulation is only possible for small instances, these problems are typically addressed through the use of a three-stage approach, as was proposed in Nemhauser and Trick (1998). In the first stage, an integer programming model generates home-away patterns (HAP) for the teams. In the second stage, a different feasible HAP is assigned to each team (recall that in any feasible round-robin schedule, no two teams may have the same HAP). Finally, in the third stage another integer programming model is used to verify the remaining constraints reflecting conditions requested by the Association and the teams themselves. Using this approach, solutions are found for single round-robin instances of Chile's First Division in 3–8 h.

In certain cases, however, the problem has proved to be solvable in a single stage using constraint programming or a branch-and-cut approach. A detailed exposition of these models and their implementation is set out in Durán et al. (2007), Durán et al. (2012) and Noronha et al. (2007). A simplified version of the basic integer linear programming formulation is described below in Sect. 2.4.2.

These applications of OR techniques in Chilean professional football have also had significant non-economic impacts (Alarcón et al. 2017). For example, the incorporation of team requirements and various sporting criteria has improved schedule fairness and the transparency of the whole scheduling process, and has also increased Chilean football fans' interest in the leagues. The models and methods used in the scheduling have been disseminated widely, which has helped to promote OR as an effective tool for addressing practical problems. Our outreach activities on sports scheduling have involved thousands of high school and university students in four countries and a more general audience of millions of television viewers and Internet users.

The Chilean football scheduling project as a whole was a finalist for the 2016 Franz Edelman Award, the most important international operations research competition, organized by the US-based Institute for Operations Research and the Management Sciences (INFORMS). It was also a finalist for the EURO Excellence in Practice Award (Bonn, 2009).

Other reported real-world scheduling applications in Latin American football have been implemented by leagues in Brazil, Honduras, Ecuador and Argentina. Similar techniques are now being used by DIMAYOR, Colombia's top-tier division, in a project which has yet to be written up in the literature.

The OR scheduling of Brazil's highest division for the 2009 and 2010 seasons is presented in Ribeiro and Urrutia (2012). The league, organized by the Brazilian Soccer Confederation (CBF), had 20 teams in both years and the season format was a mirrored double round robin. The modeling of the problem had two objectives: the minimization of the number of breaks (a logical condition for schedule fairness) and the maximization of the number of attractive games between elite clubs that could be broadcast by the largest television network in the country, which is the league's major sponsor. The CBF also imposed a series of hard constraints, many of which were typical of such formats such as no breaks in the first two and last two matches of the season, crossed home-away patterns for certain pairs of teams (i.e., when one plays at home the other plays away) and no scheduling of the most attractive matches for the last rounds of the season.

The authors modeled the problem using the above-mentioned three-stage solution approach based on the method proposed in Nemhauser and Trick (1998). For each of the two objectives, bounds were readily calculated specifying, respectively, the minimum necessary number of breaks, and the maximum possible number of attractive games between elite clubs that could be broadcast on television. The problem was attacked in the third stage (that is, after generating the home-away patterns and assigning them to the teams) using a classic multi-objective optimization technique, imposing the minimum number of breaks as a constraint and maximizing the other (attractive games) objective. If the optimal value was the upper bound, the solution was said to be at the "ideal point" and is the one retained; otherwise, the two first stages are repeated. In practice, "ideal point" solutions considered optimal in relation to both objectives were obtained for every instance in less than 10 min of run time.

The proposed methodology was compared with the manual schedules drawn up by the CBF for the 2005 and 2006 seasons, revealing in both cases that not only were the manual game calendars far from the optimal solutions on both objectives but also that certain of the hard constraints were not satisfied.

In the case of Honduras, the OR scheduling of the Central American country's professional league for the 2010 season is reported in Fiallos et al. (2010). A number of conditions were considered and assigned different levels of priority. The objective function minimized the cost of penalties for constraint violations. The 10 teams in the league played a mirrored double round robin (i.e., the second half of the season repeats the first except that the teams' home-away status in each match is reversed), followed by a playoff series. The integer linear program modeling the problem was solved to optimality in a matter of minutes.

The development of the Ecuadorian application began in early 2011, when the authors of Recalde et al. (2013) made a presentation to the Ecuadorian Football Federation (FEF by its Spanish initials) demonstrating that the design of feasible schedules using mathematical programming would bring significant benefits unattainable with manual scheduling methods. The cited paper presents two methods that were proposed for solving the problem, one an integer programming model and the other a heuristic. The resulting schedules got the approval of the FEF and were used from 2012 to 2018 by the league's first and second divisions as well as the country's youth leagues.

Late in 2018 the Liga Pro de Ecuador, which replaced the FEF as the organizer of the First and Second Divisions, contacted our group with a request to take over the mathematical programming of their schedules. Beginning with the 2019 season, we have designed the two divisions' match calendars using an ILP model that satisfies various conditions. The most important of these include maintaining a fair balance between the teams both in travel distances and match locations in high- and low-altitude regions of the country while also satisfying the special requirements of individual teams. The First Division, currently with 16 teams, plays a mirrored double round robin while the 10 Second Division teams play a quadruple round robin. Each of the two double round robins is formatted on the French system, in which the first round in the first round robin is the same as the last round of the second round robin while the rest of the matches in each round robin follow the same sequence. The third and fourth round robins simply repeat the first two. This French format has been used in the past by the first divisions of the French and Russian leagues, is the current arrangement in Luxembourg and the Czech Republic, and as we will see later, is also employed in the South American Qualifiers for the World Cup.

Also in 2018, OR scheduling techniques were adopted by the Superliga Argentina de Fútbol (SAF), responsible for organizing the season schedules of Argentina's First Division and its youth leagues. The First Division clubs all have teams in each of these youth leagues, classed by age as major divisions (Under-20, Under-18, Under-17) and minor divisions (Under-16, Under-15, Under-14). Regular season play in the youth circuits typically follows a single round-robin format, the minor divisions playing the same schedule as the majors but with the home-away status of the matches reversed. This setup can give rise to very significant differences in travel distances between a club's major and minor division teams, a frustrating situation for club officials, coaches and players alike but almost impossible to avoid with manual season scheduling techniques. Nor are manual methods able to take into account any number of other criteria that go into the design of a satisfactory match calendar.

Integer programming models have been implemented that simultaneously schedule these multiple youth divisions while also meeting a series of other desirable conditions. The central one was a better balance between the teams in travel distances, which was pursued through the application of two alternative solution methods. One of them was based on regional team clusters, treating the problem as one of feasibility, while the other was built around an explicit analysis of actual distances between the teams' home venues, an approach more closely resembling those described in the previous section (optimization problems). The solutions generated by these two methods have been used by the youth leagues to draw up their season schedules since 2018, and have produced a series of benefits for all stakeholders (Durán et al. 2020).

In 2018–2019, the application of these techniques was extended to the First Division of Argentina's professional football league (Durán et al. 2019c). The season consists of a single round-robin tournament. This can result in large differences between teams in travel distances over the course of the tournament, an important consideration in a league where an away match round trip can total more than 3000 km. In addition, since previous tournaments were scheduled manually, little care was taken to alternate the venue where a given pair of teams met. This meant that many such matchups were played at the same venue several seasons in a row. In the most extreme case, two particular teams were scheduled to play each other at the same venue 6 years running.

Among its other positive aspects, the model implemented for the Argentinean league has generated relatively even road trips for neighboring teams and changed the venues in every case where a pair of teams had played at the same location in three or more consecutive seasons. The classic three-phase solution proved to be ineffective in obtaining feasible solutions that were merely feasible, no doubt because of the large number of teams (26 in 2018–2019 and 24 in 2019–2020). The incorporation of geographical patterns (considering for each team, in addition to whether in a given round it plays at home or away, the region where each away match is played) was decisive in obtaining solution times of no longer than 1 h. And finally, the scheduling also included the definition of the exact day and kickoff time of each match within each round, results rarely attempted in football scheduling applications.

2.2.2 Applications in European football

Various real-world applications of OR scheduling in European football have been reported over the last couple of decades. In Bartsch et al. (2006), the authors describe a system used on various occasions by the German and Austrian professional leagues. It was applied in Germany to the 1997–1998 season when 18 teams played a mirrored double round robin. The objective function minimized the cost of penalizing the violated constraints and various hard or soft constraints representing attractiveness to fans, fairness, and organizational criteria were imposed. The model used a heuristic methodology to generate the schedules. In Austria, the same solution approach was utilized six times between 1997 and 2003 to define the schedules for the league's 10 teams playing a non-mirrored quadruple round robin.

A report on the scheduling of Denmark's professional league premier division for the 2006–2007 season using mathematical programming is given in Rasmussen (2008). The 12 Danish clubs played a triple round robin and the schedule requirements included geographical and top-team considerations expressed by both hard and soft constraints. The objective function minimized the penalties imposed for violating certain of the restrictions. The solution used Benders' decomposition and column-generating techniques, providing good solutions with short computation times.

The top-level Belgian soccer league has been scheduled since the 2007–2008 season using OR methods (Goossens and Spieksma 2009). The 18 league clubs play a mirrored double round robin and the central objective is the minimization of the number of breaks. A variety of other conditions are imposed relating to television broadcast requirements and what are known as carry-over effects (team i confers a unit of carry-over effect on team j if i plays the same third club in round k that j plays in round $k + 1$). The constraints are classified by priority level and penalties are levied for violations. The solution approach uses a mixed integer programming (MIP) model that satisfies the mirroring constraint and minimizes the number of breaks. Local search procedures are then used to improve the solution by minimizing the global penalty.

The scheduling of the Norwegian league is discussed in Flatberg et al. (2009). The 14 teams in the league play a non-mirrored double round robin. The approach is built around an integer programming model whose objective function incorporates carry-over effects.

Other cases of the application of optimization techniques to the scheduling of European football leagues that have not been reported in the literature but have been brought to the attention of the author include leagues in the Czech Republic and Poland, both instances dating back a number of years. A current example is the c.

The Spanish and English leagues, two of the world's most important circuits, do not use the canonical template, which is easily applied manually, nor have the scheduling methods they employ ever been reported in the literature. In Spain, the canonical schedule was used until some years ago (on this, see Goossens and Spieksma 2012, a highly interesting analysis of the numbers of teams and the schedule and season formats in Europe's top-tier leagues). As for England, the techniques utilized by the firm responsible for the scheduling are described briefly on the webpage of the Premier League (see https://www.premierleague.com/news/419020).

2.2.3 Applications in international football tournaments

Very little has been reported in the literature on the use of mathematical and computational techniques to draw up schedules for international football tournaments. To the best of our knowledge, the first case also originated in Latin America, with the application of OR scheduling to the South American qualifying stage of the 2018 World Cup.

Every 4 years, the ten national teams in the South American Football Confederation (CONMEBOL) compete in this qualifying process to secure one of the slots in the World Cup finals assigned to CONMEBOL members. The format for the qualifiers is a double round-robin tournament played over nine double rounds in which every team plays twice in each one. The entire tournament is spread over 2 years, the individual double rounds being scheduled several months apart.

The tournament schedule was based on the same mirrored schedule format for some 20 years until CONMEBOL, faced with growing complaints from its members, decided to look into a different arrangement for the 2018 qualifiers. Integer programming models were used to construct schedules that overcame the main drawbacks of the previous approach. After exploring numerous different design criteria, a schedule template based on the French format was proposed to the CONMEBOL members, the same one described here earlier for the Second Division of the Ecuadorian league.

The main characteristics of the template were that every team plays once at home and once away in each double round, each team's double round home-away sequences are well balanced (i.e., five double rounds start at home and four start away, or the other way around), and no team is scheduled to play against both of the strongest teams (Argentina and Brazil) in any given double round. The traditional mirrored format previously used was unable to satisfy all of these conditions.

This proposal was unanimously approved by the CONMEBOL members and was used for the qualifiers of the last (2018) World Cup. The design was presented at an OR congress in Valencia, Spain in 2018, where it was a finalist for the EURO Excellence in Practice Award. The implementation of the proposed models and their impact are detailed in Durán et al. (2017). They will be employed again for the 2022 World Cup qualification process, the sole modification being that the restriction barring teams from playing both Argentina and Brazil in the same double round has been dropped.

2.3 Referee assignment problem: the TUP and the RAP

Unlike the scheduling literature on assigning teams to matches, few works have been dedicated to the efficient assignment of matches to referees. As with the application of the TTP to match scheduling, the publications that do exist utilize two different formulations of the problem. In this case, the two are the Travelling Umpire Problem (TUP) (Trick et al. 2012) and the Referee Assignment Problem (RAP) (Duarte et al. 2007a). Both attempt to capture the most essential elements of an efficient referee assignment.

The TUP is formulated as follows. Given a set of n referees, $2n$ teams and a compact double round-robin schedule format (every pair of different teams i and j plays once in team i's venue and once in team j's), the number of rounds is $4n-2$. The problem consists in assigning a referee to each match at a given venue in each round (the assumed numbers of referees and teams implies all referees are needed in each round) such that the total cost of referee travel is minimized over the course of the tournament. There are five constraints in the problem:

1. Each match must be assigned to a referee.
2. Each referee must officiate at one match per round.
3. Each referee must officiate at each venue at least once over the season or tournament.
4. No referee may officiate more than once at the same venue within any window of $n-d_1$ consecutive rounds, where d_1 a parameter to be defined.
5. No referee may officiate at more than one match involving the same team within $\left\lfloor \frac{n}{2} \right\rfloor - d_2$ consecutive rounds, where d_2 is also a parameter to be defined.

Exact formulations of the problem using integer linear programming (one of which is presented below in Sect. 2.4.3) and constraint programming are discussed in Trick and Yildiz (2012), which also proposes a genetic algorithm heuristic. A simulated annealing (SA) heuristic is offered for the same problem in Trick et al. (2012). For large instances, the genetic algorithm generates better results than the exact methods and the SA heuristic, and does so in reasonably short execution times.

In the RAP, on the other hand, the number of referees exceeds what is needed per round and the number of matches at which it is desired to have each referee officiate is known. The problem is to find an assignment that minimizes the sum over all referees of the absolute value of the differences between the desired number and the number actually officiated at. If a match has multiple refereeing positions, multiple referees can be assigned to it. In football, for example, there are typically four: a head referee, two assistant referees who patrol the touchlines and a so-called fourth official who provides general assistance to the other three. The constraints on the problem are as follows:

1. All refereeing positions must be filled for all matches.
2. No referee may be assigned to officiate at two matches in quick succession (i.e., with insufficient time to travel from the first one to the second).
3. No referee may be assigned to officiate at a match for which they are unavailable (due to injury, assignment to some other tournament, etc.)
4. Every referee must have the minimum necessary qualifications for the position assigned to.
5. No referee may officiate at more than a specified maximum number of matches.

Instances of the TUP are available on a website (Toffolo et al. 2015). In Oliveira et al. (2015) it is shown that the TUP is NP-complete for $0 \leq d_1 \leq \left\lfloor \frac{n}{2} \right\rfloor$ and

$d_2 = \left\lfloor \frac{n}{2} \right\rfloor - 1$, while in Duarte et al. (2007a) it is demonstrated that the RAP is NP-complete also.

Although both problems capture the general aspects of referee assignment, in real cases other ad hoc conditions must be added to reflect the particularities of the application in question. So far, there have been very few real-world applications of referee assignment using these techniques, but those that have been reported include the following: cricket in England (Wright 1991), Major League Baseball in the US (Trick et al. 2012), and in Latin America, football in Chile (Alarcón et al. 2014) and basketball in Argentina (Durán et al. 2019b). The latter two cases will be described in more detail below. Other publications found in the literature on the subject of referee assignment are confined to certain methodological aspects or computational experiments (see, for example, Oliveira et al. 2016; Toffolo et al. 2016; Xue et al. 2015; Duarte et al. 2007b).

It should be noted that unlike match scheduling, where solutions typically relate to complete seasons or tournaments, referee assignments are produced for short periods of time. The reason is that situations requiring assignment changes often arise as a season progresses such as injuries, travel to other tournaments, or conflicts with certain teams. When this happens, new data must be incorporated into the problem so that updated solutions can be generated.

For Chilean football, integer linear programming models were developed for referee assignment at various league levels (Alarcón et al. 2014). A number of criteria for enhancing the transparency and objectivity of the assignment process were considered. As well as better balances in the number of matches each referee had to officiate at, the frequency each was assigned to a given team, and the distances each had to travel over the course of a season, these enhancements included the generation of assignments that align referee skill and experience with the level of importance of the matches they are assigned to.

The problem is solved using two approaches, one traditional and the other a pattern-based approach inspired by the home-away patterns that have been successfully employed for scheduling the season match calendars of sports leagues around the world. The two approaches have been implemented for real instances of the problem, obtaining results that have significantly improved upon the manual assignments. Also, the pattern-based approach has achieved major reductions in execution times, solving real instances to optimality in a matter of seconds. By contrast, the traditional formulation takes anywhere from several minutes to an hour or more to find a solution.

The models initially assign referees for a complete tournament but are rerun for each round or every two or three rounds, incorporating new and relevant information as it arises. They were used a decade ago by Chile's football leagues at both the professional and youth levels.

As for Argentinean basketball, a referee assignment application employed by the league that to some extent resembles the TUP, mainly in that both attempt to minimize referee travel costs, is reported in Durán et al. (2019b). The approach developed for the purpose solves the problem using an integer linear programming model. Simultaneously with travel cost minimization, the objective is to satisfy a series of other conditions. The problem is broken down into a series of relatively small subproblems representing successive periods of the season, the solution of each one being part of the input to the one following.

Upon its development this approach was tested by applying it to the First Division's 2015–2016 season, the last one before the model was adopted when referees were still assigned using manual methods. The travel costs simulated by the model turned out to be 26% lower than those actually incurred under the manual definitions, and unlike the latter, all of the different restrictions that had been requested by the league authorities were satisfied. The tool has been used by the First and Second Divisions since the 2016–2017 season.

2.4 Integer programming models

In what follows, we set out for illustrative purposes the basic integer linear programming (ILP) models for a variant of the travelling tournament problem (the TTPPV), the Chilean football problem and the Travelling Umpire Problem.

2.4.1 An ILP model for the TTPPV

This formulation is modeled with $O(n^3)$ variables and $O(n^4)$ constraints as presented in Melo et al. (2006). Let G be the set of matches represented by ordered pairs of teams determined by the predefined home-away status of each match. The game between teams i and j is represented by an ordered pair that is either (i, j) or (j, i), and is played in the former case at the team i venue and in the latter case at the team j venue. Thus, for every pair of different teams i and j, either $(i,j) \in G$, or $(j, i) \in G$.

2.4.1.1 Variables The following two families of binary variables are needed to formulate the model:

$$z_{tjk} = \begin{cases} 1 & \text{if team t plays at home against team j in round k.} \\ 0 & \text{otherwise.} \end{cases}$$

$$y_{tij} = \begin{cases} 1 & \text{if team t travels from the team i venue to the team j venue .} \\ 0 & \text{otherwise.} \end{cases}$$

The z variables are typically found in this type of formulation. The y variables represent the trip performed by a team between two cities (recall that a journey between the cities of two different teams is performed at most once by each team).

2.4.1.2 Objective function The objective function minimizes the travel distances over all of the teams.

$$\mathrm{m}\acute{\pi}\mathrm{n} \sum_{t=1}^{n} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij} \cdot y_{tij}.$$

2.4.1.3 Constraints

1. Each game in G occurs exactly once.

$$\sum_{q=1}^{n-1} z_{tjq} = 1, \quad \forall (t,j) \in G.$$

2. The variables corresponding to games between teams t and j at the team t venue take the value of zero if the game is predetermined to occur at the team j venue.

$$\sum_{q=1}^{n-1} z_{tjq} = 0, \quad \forall (j,t) \in G.$$

3. Each team plays one game in each round.

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} \left(z_{tjk} + z_{jtk} \right) = 1, \quad t = 1, \dots, n \quad k = 1, \dots, n-1.$$

Constraints 4 through 8 below define the logical relationship between the y and z variables.

4. Team t makes a trip from the home city of team i to that of team j if it plays two consecutive away games against teams i and j, in that order.

$$y_{tij} \geq z_{it,k-1} + z_{jtk} - 1, \quad t,i,j = 1, \dots, n, \text{ with } t \neq i, t \neq j, i \neq j, \quad k = 2, \dots, n-1.$$

5. Team t makes a trip from the home city of team i to its own home city if it has an away game against the latter followed by a home game in the succeeding round.

$$y_{tit} \geq z_{it,k-1} + \sum_{\substack{j=1 \\ j \neq t}}^{n} z_{tjk} - 1, \quad t,i = 1, \dots, n, \text{ with } t \neq i, \quad k = 2, \dots, n-1.$$

6. Team t travels from its home city to that of team i to play away against the latter following a home game in the previous round.

$$y_{tti} \geq \sum_{\substack{j=1 \\ j \neq t}}^{n} z_{tj,k-1} + z_{itk} - 1, \quad t,i = 1, \dots, n, \text{ with } t \neq i, \quad k = 2, \dots, n-1.$$

7. Team t travels from its home city to the home city of team i if it plays away against the latter in the first round.

$$y_{tti} \geq z_{it1}, \quad t,i = 1, \dots, n \text{ with } t \neq i.$$

8. Team t returns from the home city of team i to its own home city if it plays away against team i in the last round.

$$y_{tit} \geq z_{it,n-1}, \quad t, i = 1, \dots, n \text{ with } t \neq i.$$

9. In any 4 consecutive games team t cannot play more than three away, implying that the team cannot play more than three consecutive away games.

$$\sum_{q=k}^{k+3} \sum_{\substack{j=1 \\ j \neq t}}^{n} z_{jtq} \leq 3, \quad t, i = 1, \dots, n \quad k = 1, \dots, n-4.$$

10. In any four consecutive games team t must play at least one away, implying that the team cannot play more than three consecutive games at home.

$$\sum_{q=k}^{k+3} \sum_{\substack{j=1 \\ j \neq t}}^{n} z_{jtq} \geq 1, \quad t = 1, \dots, n \quad k = 1, \dots, n-4.$$

11. The z and y variables are binary.

$$z_{tjk} \in \{0, 1\}, \quad t, j = 1, \dots, n \quad k = 1, \dots, n-1$$
$$y_{tij} \in \{0, 1\}, \quad t, i, j = 1, \dots, n.$$

2.4.2 An ILP model for the Chilean first division football league

Below we present a model for a 20-team league that plays a single round robin (i.e., 19 rounds). The teams are grouped into four zones of five teams with separate standings and point totals for each zone. The two top teams in each group at season's end participate in a playoff system to determine the champion. The model is a simplified version of the basic integer linear programming model used by the Chilean football league. The complete version may be found in Durán et al. (2007).

2.4.2.1 Variables To determine the games to be held in each round, we define $\forall i \neq j \in I$ (the set of teams) and $\forall k \in K$ (the set of rounds) a family of binary variables as follows:

$$x_{ijk} = \begin{cases} 1 & \text{if team } i \text{ plays at home against team } j \text{ in round } k. \\ 0 & \text{otherwise.} \end{cases}$$

To represent home and away game sequence constraints (that is, to model the "breaks"), we define $\forall i \in I$ and $\forall k = 1, \dots, 18$ the following auxiliary variables, also binary:

$$y_{ik} = \begin{cases} 1 & \text{if team } i \text{ plays at home in rounds } k \text{ and } k+1. \\ 0 & \text{otherwise.} \end{cases}$$

$$w_{ik} = \begin{cases} 1 & \text{if team } i \text{ plays away in rounds } k \text{ and } k+1. \\ 0 & \text{otherwise.} \end{cases}$$

For the instance, considered here, the model generates approximately 8000 variables.

2.4.2.2 Objective function The model maximizes an objective function that measures the concentration of games between teams in the same group (known as "attractive games") played towards the final rounds of the tournament. The OF takes the following form:

$$\text{máx} \sum_{1 \le k \le 19} \sum_{e} \sum_{i \in t(e)} \sum_{j \in t(e)} k \cdot x_{ijk},$$

where

$$t(e) \text{ denotes the set of teams in group } e, \; e \in E = \{1, 2, 3, 4\}.$$

In certain cases other games may be included in the "attractive games" category, such as the "classic rivalries" or matches between teams fighting to avoid relegation to the Second Division.

2.4.2.3 Constraints The constraints formulated below are the basic ones required to define a single round robin plus a few examples of restrictions peculiar to the Chilean tournament. In the complete version of the model, the constraints number about 3000.

1. Each team plays every other team once over the course of the 19 rounds in the tournament.

$$\sum_{k} \left[x_{ijk} + x_{jik}\right] = 1 \qquad \forall \, i,j \in I.$$

2. Each team plays each round either at home or away.

$$\sum_{j} \left[x_{ijk} + x_{jik}\right] = 1 \qquad \forall \, i \in I, \; k \in K.$$

3. Each team plays at least nine rounds, but not more than ten, at home.

$$9 \le \sum_{j} \sum_{k} x_{ijk} \le 10 \qquad \forall \, i \in I.$$

4. The following constraint is required to define the logical relationship between the x and y variables.

$$2 \cdot y_{ik} \le \sum_{j} [x_{ijk} + x_{ij(k+1)}] \le 1 + y_{ik} \qquad \forall \, i \in I, \; k < 19.$$

5. No more than 1 home break per team.

$$\sum_{k<19} y_{ik} \leq 1 \qquad \forall\, i \in I.$$

6. The following constraint is required to define the logical relationship between the x and w variables.

$$2 \cdot w_{ik} \leq \sum_{j} [x_{jik} + x_{ji(k+1)}] \leq 1 + w_{ik} \qquad \forall\, i \in I,\ k < 19.$$

7. No more than 1 away break per team.

$$\sum_{k<19} w_{ik} \leq 1 \qquad \forall\, i \in I.$$

8. Let A be a set of rounds such that if a team plays at home (away) in a round belonging to A, it must play away (at home) in the following round. Usually, $A = \{1, 18\}$. The objective is to balance the teams' home and away games between the early and late stages of the tournament.

$$\sum_{j} \left[x_{ijk} + x_{ij(k+1)}\right] = 1 \qquad \forall\, i \in I,\ k \in A.$$

9. Each team must play two of its four group opponents at home and the other two away.

$$\sum_{k} \sum_{j\, \in\, t(e)} x_{ijk} = 2 \qquad \forall e \in E,\ i \in t(e).$$

10. When a northern (southern) team plays two consecutive away games, neither of them will be played in the South (North).

$$\sum_{i \in South} [x_{ijk} + x_{ij(k+1)}] \leq 1 - w_{jk} \qquad \forall j \in North,\ k < 19$$

$$\sum_{i \in North} [x_{ijk} + x_{ij(k+1)}] \leq 1 - w_{jk} \qquad \forall j \in South,\ k < 19.$$

11. The three strong teams (UC, CC, UCH) play each other in three classic matchups, each team playing one of the matches at home.

$$\sum_{k} \left[x_{hik} + x_{jik}\right] = \sum_{k} \left[x_{hjk} + x_{ijk}\right] \qquad h = UC, i = CC, j = UCH.$$

12. No team may play two consecutive games against a strong team.

$$\sum_{j \in \text{Strong Teams}} \left[x_{ijk} + x_{jik} + x_{ij(k+1)} + x_{ji(k+1)}\right] \leq 1 \qquad \forall\, i \in I,\ k < 19.$$

2.4.3 An ILP model for the TUP

Set forth in what follows is the exact formulation of the travelling umpire problem using an ILP as presented in Trick and Yildiz (2012). This specification has $O(n^4)$ variables and constraints.

2.4.3.1 Input data and definitions

- $\mathbb{S}, \mathbb{T}, \mathbb{U}$: Set of rounds, teams and referees, respectively.
- $\mathrm{OPP}(s, i) = \begin{cases} j & \text{if team } \mathbf{i} \text{ plays at home in round } \mathbf{s} \text{ against team } \mathbf{j} \\ -j & \text{if team } \mathbf{i} \text{ plays away in round } \mathbf{s} \text{ against } \mathbf{j} \end{cases}$
- d_{ij} : kilometers between the home venues of teams **i** and **j**
- $n_1 = n - d_1 - 1$
- $n_2 = \left\lfloor \frac{n}{2} \right\rfloor - d_2 - 1$
- $N_1 = \{0, \dots, n_1\}$
- $N_2 = \{0, \dots, n_2\}$.

2.4.3.2 Variables The following two families of binary variables are needed to formulate the model:

$$x_{isu} = \begin{cases} 1 & \text{if a match at venue } \mathbf{i}\text{, in round } \mathbf{s}\text{, is assigned to referee } \mathbf{u}. \\ 0 & \text{otherwise.} \end{cases}$$

$$z_{ijsu} = \begin{cases} 1 & \text{if referee } \mathbf{u} \text{ officiates at a game at venue } \mathbf{i} \text{ in round } \mathbf{s} \text{ and then a game at venue } \mathbf{j} \text{ in round } \mathbf{s}+1. \\ 0 & \text{otherwise.} \end{cases}$$

2.4.3.3 Objective function The model minimizes an objective function that measures the total number of kilometers travelled by the set of referees.

$$\text{mín} \sum_{i \in \mathbb{T}} \sum_{j \in \mathbb{T}} \sum_{u \in \mathbb{U}} \sum_{s \in \mathbb{S} : s < |S|} d_{ij} \cdot z_{ijsu}.$$

2.4.3.4 Constraints

1. Each game must have an assigned referee:

$$\sum_{u \in \mathbb{U}} x_{isu} = 1, \quad \forall i \in \mathbb{T}, s \in \mathbb{S} \,:\, \mathrm{OPP}(t, i) > 0.$$

2. Each referee is assigned exactly one game per round:

$$\sum_{i \in \mathbb{T}\,:\,\mathrm{OPP}(s,i)>0} x_{isu} = 1, \quad \forall s \in \mathbb{S}, u \in \mathbb{U}.$$

3. Each referee officiates at least once at a match involving each team at home:

$$\sum_{s \in \mathbb{S}\,:\,\mathrm{OPP}(s,i)>0} x_{isu} \geq 1, \quad \forall i \in \mathbb{T}, u \in \mathbb{U}.$$

4. Each referee officiates no more than once at any given venue every $n - d_1$ rounds:

$$\sum_{s_1 \in N_1} x_{i(s+s_1)u} \leq 1, \quad \forall i \in \mathbb{T}, u \in \mathbb{U}, s \in \mathbb{S} \,:\, s \leq |S| - n_1.$$

5. Each referee officiates at a match involving any given team no more than once every $\left\lfloor \frac{n}{2} \right\rfloor - d_2$ rounds:

$$\sum_{s_2 \in N_2} \left(x_{i(s+s_2)u} + \sum_{k \in T\,:\,\mathrm{OPP}(s+s_2,k)>0} x_{k(s+s_2)u} \right) \leq 1, \quad \forall i \in \mathbb{T}, u \in \mathbb{U}, s \in \mathbb{S} \,:\, s \leq |S| - n_2.$$

6. The following constraint is required to define the logical relationship between the x and z variables, thus allowing the referee trips to be modeled:

$$x_{isu} + x_{j(s+1)u} - z_{ijsu} \leq 1, \quad \forall i,j \in \mathbb{T}, u \in \mathbb{U}, s \in \mathbb{S} \,:\, s < |S|.$$

3 Other topics in sports analytics

In this section, we survey a few areas of research in sports analytics and the related literature. In the area of sports business, Fried and Mumcu (2016) is worthy of note. Player performance has been the subject of a number of books in recent years. Two examples of these works are Alamar (2013) and Miller (2015).

Of cases involving a particular sport, the best known is the one described in the book "Moneyball" (Lewis 2003), mentioned in the Introduction. It is based on the true story of Billy Beane, general manager of the Major League Baseball team the Oakland Athletics. Beane used data science to determine the team's roster and improve its performance, achieving great success with much smaller budgets than the big clubs in the league. These techniques are now widely practiced in basketball, not just for choosing players but also for game play decisions (Zuccolotto and Manisera 2020), especially among NBA teams.

The following subsections focus on results prediction, fantasy games and rankings analysis, three lines of investigation in sports analytics that make particularly heavy use of statistical tools and data science.

3.1 Result prediction

Sports results prediction has seen steady growth in recent years. The most influential Internet site for this activity is fivethirtyeight.com, which focuses on politics and economics as well as sports. Taking its name from the number of electors in the US electoral college, the site was created by the American writer and statistician Nate Silver in 2008. Silver had already made his name in the early 2000s with the development of a statistical system that projects the future performance of baseball players. Another website, this one featuring predictions for cricket, basketball and football using artificial intelligence tools, is predict22.com.

For predictions relating exclusively to football, a Bayesian methodology was proposed in Suzuki et al. (2010). Its performance was tested on the results of the 2006 World Cup in Germany. The pioneer website for football-only predicting is 851.cl. This site was launched for the 2015 Copa América (formerly the South American Football Championship), was utilized again for the 2016 edition of the tournament, and has been applied in various seasons of the Chile's First Division.

The model behind 851.cl is based on a predictive model devised by Dixon and Coles (1997) that assumes goal-scoring follows a Poisson distribution. It has two specific coefficients per team denoted "offensive strength" and "defensive strength", and two general ones (the same for each team) capturing the home advantage denoted "extra at-home offensive strength", and "extra at-home defensive strength". In any given match with a home team H and an away team A, goals by H behave according to a Poisson variable with parameter λ while goals by A behave according to a Poisson variable with parameter μ, where λ = "offensive strength H"—"defensive strength A" + "extra at-home offensive strength" and μ = "offensive strength A"—"defensive strength H"—"extra at-home defensive strength".

These "extra" coefficients are not necessarily positive, and for matches played at a neutral venue they are omitted from the calculations, in which case H and A are taken simply as the identifiers of the two teams. All the coefficients are estimated by generalized linear regression (calculated with the *R* software package) using data on the teams' past matches. The probabilities attached to the different possible results for each match are calculated as the product of the two Poisson variables (i.e., the probability that H scores m goals and A scores n goals is the probability of the former multiplied by that of the latter given the corresponding λ and μ coefficients for H and A, respectively). Using these probabilities, future scheduled matches are simulated to obtain each team's probability of ending the season at a given place in the standings or advancing to a certain round in a cup tournament.

A broadly similar model on the website 301060.exactas.uba.ar. incorporates additional home-away factors specific to each team with a view to achieving more realistic formulations of them than a single general specification used for all teams alike. This extension of the model was used to predict the outcomes of the 2018 World Cup, the 2019 Copa America, and the 2018–2019 and 2019–2020 Superliga Argentina (i.e., First Division) seasons. As an indication of this approach's accuracy, we note that in the case of the 2018 World Cup, the score in 30 out of the 48 group stage matches was precisely the one the model found most likely, while the same was true for 11 out of the 16 matches in the knockout stage.

Analogous models have been implemented for basketball (at 280777.exactas.uba.ar) and rugby (at 190665.exactas.uba.ar). In the basketball case, each team's zero score was set at 40 points, which was then subtracted from the actual scores in the database of past games and the remainder divided by 5. Calculating the different strength factors on this basis gave results that corresponded closely to the actual game scores. Thus, the points attributed to team A playing against B are calculated as $5 \times P + 40$, where P is the Poisson variable similar to the football case that takes into account offensive strength, defensive strength and home-away status.

The model was used to make predictions for the 2019 Basketball World Cup, which were uploaded to the 280,777 website. In 82.6% of the tournament games, the team that won in the tournament was the team to which the model had assigned the greatest probability of winning. Furthermore, the second highest probability (25.23%) of winning the Cup was attributed to the team (Spain) that actually won. The same model has also been used to predict game results for Argentina's professional basketball league.

The modeling for rugby had to incorporate the four different ways of scoring: penalty goals, drop goals, tries and conversions, which are worth 3, 3, 5, and 2 points, respectively. Unlike the other sports, however, the final score is not the only factor that counts because the points awarded for a game in the team standings include bonuses for scoring 4 or more tries or losing by a score difference of no more than 7 points. In this case, therefore, we decided to formulate three models, one each for penalty goals/drop goals, tries, and conversions. Penalty goals and drop goals were considered together given that they are both worth three points and drops are relatively uncommon. To calculate the probabilities for them and for tries, the models used were similar to the one formulated for goals in football. For conversions, on the other hand, since they are attempted after a try is scored, we assumed their distribution is binomial(t,p), where "t" is the number of tries converted in a match and "p" is calculated using the maximum likelihood estimator. The models were run for the 2019 Rugby World Cup and the results uploaded to the 190,665 website. For 89% of the matches, the models assigned the greatest probability of winning to the team that did in fact win, suggesting that rugby is more predictable than basketball, which in turn is more predictable than football.

The ideas incorporated in the above-described models will surely be extended to other sports in the future. The models themselves have been made available through the above-cited Internet webpages, which also contain brief explanations of the techniques underlying them and allow visitors to submit questions by e-mail. This should be an effective way of taking advantage of the popularity of sports to create greater awareness among the general public of the power of mathematics and quantitative methods generally.

3.2 Fantasy games

Building sports results prediction models has a number of aspects in common with the development of quantitative tools for fantasy sports games. In football, the first fantasy game was "Fantacalcio" (http://www.fantacalcio.kataweb.it), which was designed for Italy's Serie A professional soccer league. A creation of Riccardo Albini in the late 1980s, Fantacalcio is run by a major national newspaper. Albini was himself inspired by the fantasy game of North America's Major League Baseball (MLB) (mlb.mlb.com/mlb/fantasy).

Various other fantasy games have been created for sports around the world, such as English football's Fantasy Premier League (fantasy.premierleague.com/), Argentina's football "Gran DT" (grandt.com.ar) and the fantasy basketball game for the NBA in North America (http://www.nba.com/fantasy/). Similar games for virtual soccer like Hattrick (http://www.hattrick.org) and Xpert (http://www.xperteleven.com) have also become very popular. But especially interesting is the growing use of fantasy games in recent years to improve the teaching of mathematics and stimulate student motivation for studying the subject at all education levels (see http://www.fantasysportsmath.com).

In the case of Gran DT in Argentina, two mathematical programming models were developed to act as virtual coaches for choosing a virtual team line-up for each round of the real Argentinian football league. The a priori design created a competitive team for each match, while the a posteriori model determined what its optimal line-up would have been for each match had all the results for the entire season been known at the time. The a priori model was entered in the fantasy game on a number of occasions, achieving results that placed it among the highest scoring participants (generally among the top 1%). The details of the models and the results are discussed in Bonomo et al. (2014).

In Gupta (2017) an approach is proposed for a fantasy version of the English Premier League (football) using a hybrid of autoregressive integrated moving average (ARIMA) and recurrent neural networks (RNNs) that makes time-series predictions of the points scored by a team's individual players and then maximizes the point total across the entire team. It uses linear programming and incorporates various constraints to capture the characteristics of the problem.

An optimization model for a popular cycling fantasy game ("The Gigabyke game", http://www.gigabyke.be) is presented in Beliën et al. (2011). The model contains certain features of knapsack problems, multiperiod inventory problems and logical constraint modeling, making it particularly suitable as a comprehensive case study for an undergraduate course in operations research and management science. But just as fantasy games models are ideal pedagogical tools for statistics and operations research, the concepts they involve also provide useful tools for supporting decision-making by coaches or managers of real sports.

3.3 Rankings

Another area that has generated considerable interest in the literature is sports rankings. There are a number of works that study the predictive power of the football rankings published by FIFA, which attempt to classify the world's national sides. Two examples are McHale and Davies (2007) and Lasek et al. (2013).

In Lasek et al. (2016), the authors propose strategies for teams to improve their ranking while Cea et al. (2020) presents a discussion of what are considered to be deficiencies in the ranking method used by FIFA up until the 2018 World Cup and makes proposals for its improvement. An alternative predictive model is calibrated to provide a reference ranking for evaluating the performance of various simple changes to that procedure. It should be noted, however, that since 2018 FIFA has used a new ranking procedure whose formula takes its inspiration from the Elo rating system, which was originally created by the Hungarian physicist Arpad Elo to rank chess players but has since been applied to various sports. No study of the predictive abilities of the new procedure has yet been published. The use of the FIFA rankings and the draw format for the World Cup finals are discussed in Cea et al. (2020) and Guyon (2015). Certain ideas proposed in the latter work were adopted for the 2018 World Cup draw.

Similar studies have been done on rankings and performance analyses in professional tennis. In Radicchi (2011), Radicchi proposed an algorithm for identifying the best tennis player in history. In an article published the following year (Dingle et al. 2012), the authors explored the relationship between official rankings of professional tennis players and rankings computed using a variant of the PageRank algorithm, originally generated by the creators of Google to rank pages (see Page et al. 1998; Brin and Page 1998).

They show that Radicchi's equations constitute a direct application of the PageRank algorithm and present up-to-date comparisons of official rankings with PageRank-based rankings for both the Association of Tennis Professionals (ATP) and Women's Tennis Association (WTA) tours. For top-ranked players the two rankings are broadly in line but for those near the bottom there is wide variation, raising questions about how well the official ranking mechanism reflects true player ability. For a 390-day sample of tennis matches, PageRank-based rankings were found to be better predictors of match outcome than the official rankings. A later adaptation of the authors' ideas presented in Aronson (2015) arrives at similar conclusions.

4 Conclusions and future challenges

This survey has reviewed the main problems addressed in the sports scheduling literature, the mathematical and computational techniques used to solve them and their practical results. Much effort has been devoted to the study of these problems in recent years and as we have just seen, many interesting real-world applications have been implemented across a range of sports in countries around the globe. In particular, a number of successful applications in football, volleyball and basketball have been reported over the last 15 years from Latin America, many of which have been described here above.

There are, however, numerous challenges still to be tackled. One of them is the parallel scheduling of multiple leagues, as is done with the six interrelated divisions of the Argentinean youth leagues discussed here earlier. Other cases found in the literature are an amateur softball league in the US (Grabau 2012), a regional rugby league in New Zealand (Burrows and Tuffley 2015) and an amateur table tennis league in Germany (Schönberger 2017). Multi-league scheduling is approached from a more methodological standpoint in Davari et al. (2020). The article addresses the problem of simultaneously scheduling multiple sports leagues with interdependencies arising from the fact that clubs have teams in different sports, each one playing in one of the leagues being scheduled. All of a given club's teams share the same venue so the schedules must ensure the venue is available each time one of the team's is slated to play at home. Other multi-league challenges are encountered in the so-called "sports teams grouping problem" where given a large number of teams, it must be decided which team is assigned to which tournament such that each one plays with a round-robin format and the objective is to minimize total travel distances. This is studied in Toffolo et al. (2019), where various integer linear programming models are formulated to solve the problem.

Another topic that will no doubt arise in the future is the design of practical applications of relaxed round robins where it is desired to minimize differences or mismatches in the rest times a pair of teams have enjoyed as of the day they are scheduled to meet. This is a sensitive issue for team managers, coaches and players that has not been dealt with in the literature on real-world cases. The so-called "rest mismatches problem" is examined in Atan and Cavdaroglu (2018).

The "stable tournament problem" formulated in Guajardo and Jörnsten (2017) is yet another situation for which practical applications could be developed in the future. Proceeding on an analogy with Gale and Shapley's "stable marriage problem" (Gale and Shapley 1962), the authors analyze round-robin schedules that consider the preferences of the teams as regards the order in which they meet their rivals. They define a concept of schedule stability and propose an integer programming model that generates schedules satisfying that definition.

Yet another scheduling problem stems from the fact that due to unforeseen events such as bad weather conditions, league schedules are not necessarily played as they were announced at the beginning of the season. The study presented in Yi et al. (2020) reveals that in many cases from different football leagues, matches that were rescheduled to another date had a profound impact on the quality of the overall schedule. The authors propose the insertion in the schedule of so-called catch-up rounds as buffers, and offer various suggestions on how rescheduling can be done without unduly affecting overall schedule quality. An idea in the same vein is presented in Durán et al. (2019a), which incorporates rest weeks at various points in the season that can be used for rescheduling matches, but the authors also point out that in practice this is done manually. An interesting line of research for such cases involving non-compact seasons where the schedules include road trips would be to develop an exact or heuristic method that allows reschedulings, whether or not for rest weeks, while maintaining the minimization of team travel distances.

An issue that is still very much an open topic in sports scheduling relates to tournaments, both compact and relaxed, that involve a large number of teams and/or matches. Such is the case with the National Basketball Association in North America (Bao 2009; Bean and Birge 1980) and professional football in Argentina (Durán et al. 2019c). In these situations, the usual scheduling methods tend not to be very successful and techniques of problem decomposition such as those surveyed here must be resorted to, but much work in this area remains to be done.

Combining in a single model, the generation of league schedules and referee assignment for a practical application has yet to be achieved. An attempt to meet this double challenge is proposed in Linfati et al. (2019). The formulation of the problem includes the usual objective functions and restrictions for the two subproblems to address such considerations as travel distance reduction and minimum referee qualifications for certain matches. The proposed method is tested on two Latin American leagues (the first divisions of Chile's football and basketball leagues) and the World Volleyball League, the latter a competition between participating countries' national teams, with very promising results.

As regards other areas in sports analytics, this survey reviewed the use of quantitative tools for results prediction, fantasy games and ranking analysis in various sports. Also mentioned was the study of player performance, team improvement, efficient use of team budgets, etc., with applications primarily to baseball, basketball and American football. The use of these latter tools in football (soccer) is considerably less common but will no doubt grow significantly in the years to come.

There no doubt exist many other aspects of sports analytics that are amenable to mathematical techniques. For example, we are currently working on the application of statistical techniques to interesting problems that arise in football such as the moments at which goals are scored in the most important leagues and the generation of a league inequality index based on the Gini Index used in economics (for more details on this metric, see, for example, Ceriani and Verme 2012). The inequality analysis could be used to measure the differences between the teams of a given league in their financial resources and how these relate to the teams' respective point totals in a given tournament.

In conclusion, the application of mathematical and computational techniques to a variety of problems arising in sports scheduling and other areas of sports analytics over a range of different sports has been extensively studied and reported in the literature. However, there remains ample room for broadening of the field to address new questions and challenges. The coming years thus promise to be very productive.

Acknowledgements The author is grateful to both anonymous referees and one of the editors in chief for their comments and suggestions, which were of great help in improving the final version of this article. Thanks are also due to the Asociación Nacional de Fútbol Profesional de Chile (ANFP); the Superliga Argentina de Fútbol (SAF); the Liga Pro in Ecuador; the Confederación Sudamericana de Fútbol (CONMEBOL); the Asociación de Clubes de Básquetbol de la Argentina (AdC); and the Asociación de Clubes Liga Argentina de Voleibol (ACLAV) for their collaboration on various of the projects discussed in this survey. The author also wants to thank to Mario Guajardo, Gonzalo Zamorano, Javier Marenco, Pablo Rey, Facundo Gutiérrez, Iván Monardo, Alejandro Alvarez, Florencia Fernández, Flavia Bonomo, Kenneth Rivkin, Manuel Durán, Marcelo Mariosa and Claudia Zelzman for their helpful comments on this survey. Partial funding for the preparation of this article was provided by the Department of Industrial Engineering at the University of Chile, the Instituto Sistemas Complejos de Ingeniería (ISCI) in Santiago, Chile (CONICYT PIA/BASAL AFB180003), and Grants nos. UBACyT 20020170100495BA (UBA, Argentina) and ANPCyT PICT 2015-2218 (Mincyt, Argentina). Finally, the author owes a special debt of gratitude to Rafael Epstein and Andrés Weintraub, who taught him the supreme art of applying operations research to real-world problems.

References

Alamar B (2013) Sports analytics: a guide for coaches, managers, and other decision makers. Columbia University Press, New York

Alarcón F, Durán G, Guajardo M (2014) Referee assignment in the Chilean football league using integer programming and patterns. Int Trans Oper Res 21(3):415–438

Alarcón F, Durán G, Guajardo M, Miranda J, Muñoz H, Ramírez L, Ramírez M, Sauré D, Siebert M, Souyris S, Weintraub A, Wolf-Yadlin R, Zamorano G (2017) Operations research transforms scheduling of Chilean soccer leagues and South American world cup qualifiers. Interfaces 47(1):52–69

Anagnostopoulos A, Michel L, Van Hentenryck P, Vergados Y (2006) A simulated annealing approach to the traveling tournament problem. J Sched 9(2):177–193

Aronson A (2015) TenisRank: Un nuevo ranking de jugadores de tenis basado en PageRank (in Spanish). Master thesis in computer science, University of Buenos Aires

Atan T, Cavdaroglu B (2018) Minimization of rest mismatches in round robin tournaments. Comput Oper Res 99:78–89

Bao R (2009) time relaxed round robin tournament and the NBA scheduling problem. Ph.D thesis, Cleveland State University

Bartsch T, Drexl A, Kröger S (2006) Scheduling the professional soccer leagues of Austria and Germany. Comput Oper Res 33(7):1907–1937

Bean J, Birge J (1980) Reducing travelling costs and player fatigue in the National Basketball Association. Interfaces 10:98–102

Beliën J, Goossens D, Van Reeth D, De Boeck L (2011) Using mixed integer programming to win a cycling game. INFORMS Trans Educ 11(3):93–99

Benoist T, Laburthe L, Rottembourg B (2001) Lagrange relaxation and constraint programming collaborative schemes for traveling tournament problems. In: Proceedings of the 3rd international workshop on the integration of AI and OR techniques (CP-AI-OR), pp 15–26

Bhattacharyya R (2016) Complexity of the unconstrained traveling tournament problem. Oper Res Lett 44(5):649–654

Bonomo F, Cardemil A, Durán G, Marenco J, Saban D (2012) An application of the traveling tournament problem: the Argentine volleyball league. Interfaces 42(3):245–259

Bonomo F, Durán G, Marenco J (2014) Mathematical programming as a tool for virtual soccer coaches: a case study of a fantasy sport game. Int Trans Oper Res 21(3):399–414

Brandao F, Pedroso JP (2014) A complete search method for the relaxed traveling tournament problem. EURO J Comput Optim 2:77–86

Brin S, Page L (1998) The anatomy of a large-scale hypertextual web search engine. Comput Netw ISDN Syst 30(1–7):107–117

Burrows W, Tuffley C (2015) Maximizing common fixtures in a round robin tournament with two divisions. Australas J Comb 63(1):153–169

Cardemil A, Durán G (2002) Un algoritmo tabú search para el traveling tournament problem (in Spanish). Revista Ingeniería de Sistemas (Universidad de Chile) 18:95–115

Cea S, Durán G, Guajardo M, Sauré D, Siebert J, Zamorano G (2020) An analytics approach to the FIFA ranking procedure and the World Cup final draw. Ann Oper Res 286(1):119–146

Ceriani L, Verme P (2012) The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. J Econ Inequal 10:421–443

Cheung K (2008) Solving mirrored traveling tournament problem benchmark instances with eight teams. Discret Optim 5(1):138–143

Choubey N (2010) A novel encoding scheme for traveling tournament problem using genetic algorithm. IJCA Spec Issue Evolut Comput 2:79–82

Cocchi G, Galligari A, Picca NF, Piccialli V, Schoen F, Sciandrone M (2018) Scheduling the Italian National Volleyball Tournament. Interfaces 48(3):271–284

Costa D (1995) An evolutionary Tabu search algorithm and the NHL scheduling problem. INFOR Inf Syst Opera Res 33(3):161–178

Craig S, While L, Barone L (2009) Scheduling for the National Hockey League using a multi-objective evolutionary algorithm. In: Proceedings of the Australasian joint conference on artificial intelligence, pp 381–390

Davari M, Goossens D, Belien J, Lambers R, Spieksma F (2020) The multi-league sports scheduling problem, or how to schedule thousands of matches. Oper Res Lett 48(2):180–187

De Werra D (1980) Geography, games and graphs. Discret Appl Math 2:327–337

De Werra D (1981) Scheduling in sports. N-Holl Math Stud 11:381–395

De Werra D (1982) Minimizing irregularities in sports schedules using graph theory. Discret Appl Math 4:217–226

Di Gaspero L, Schaerf A (2007) A composite-neighborhood tabu search approach to the traveling tournament problem. J Heuristics 13:189–207

Dingle N, Knottenbelt W, Spanias D (2012) On the pageranking of professional tennis players. Lect Notes Comput Sci 7587:237–247

Dixon M, Coles S (1997) Modeling association football scores and inefficiencies in the football betting market. Appl Stat 46(2):265–280

Duarte A, Ribeiro CC, Urrutia S (2007) Referee assignment in sports tournaments. Lect Notes Comput Sci 3867:158–173

Duarte A, Ribeiro CC, Urrutia S (2007) A hybrid ILS heuristic to the referee assignment problem with an embedded MIP strategy. Lect Notes Comput Sci 4771:82–95

Durán G, Miranda J, Guajardo M, Sauré D, Souyris S, Weintraub A, Wolf R (2007) Scheduling the Chilean soccer league by integer programming. Interfaces 37:539–552

Durán G, Guajardo M, Wolf YR (2012) Operations research techniques for scheduling Chile's second division soccer league. Interfaces 42(3):273–285

Durán G, Guajardo M, Sauré D (2017) Scheduling the South American qualifiers to the 2018 FIFA World Cup by integer programming. Eur J Oper Res 262(3):1109–1115

Durán G, Durán S, Marenco J, Mascialino F, Rey P (2019) Scheduling Argentina's professional basketball leagues: a variation on the relaxed travelling tournament problem. Eur J Oper Res 275(3):1126–1138

Durán G, Guajardo M, Gutiérrez F (2019) Efficient referee assignment in Argentina's professional basketball leagues using operations research methods. **(submitted)**

Durán G, Guajardo M, López A, Marenco J, Zamorano G (2020) Scheduling multiple sports leagues with travel distance fairness: an application to Argentinean youth football. Int J Appl Anal **(in press)**

Durán G, Guajardo M, Zamorano G (2019) Scheduling the Argentina's football superliga. In: Proceedings of the 30th. European conference on operational research, Dublin, Ireland

Easton K, Nemhauser G, Trick M (2001) The traveling tournament problem: description and benchmarks. Lect Notes Comput Sci 2239:580–584

Easton K, Nemhauser G, Trick M (2003) Solving the travelling tournament problem: a combined integer programming and constraint programming approach. Lect Notes Comput Sci 2740:100–109

Easton K, Nemhauser G, Trick M (2004) Sports scheduling. In: Leung J (ed) Handbook of scheduling, vol 52. CRC Press, Boca Raton, pp 1–52

Fiallos J, Pérez J, Sabillón F, Licona M (2010) Scheduling soccer league of Honduras using integer programming. In: Johnson A, Miller J (eds) Proceedings of the (2010) industrial engineering research conference, Cancún, Mexico

Flatberg T, Nilssen E, Stølevik M (2009) Scheduling the topmost football leagues of Norway. In: 23rd European conference on operational research, book of abstracts, Bonn, Germany

Fleurent C, Ferland J (1993) Allocating games for the NHL using integer programming. Oper Res 41:649–654

Fried G, Mumcu C (2016) Sport analytics: a data-driven approach to sport business and management. Routledge, London

Froncek D (2001) Scheduling the Czech national basketball league. Congr Numer 153:5–24

Gale D, Shapley L (1962) College admissions and the stability of marriage. Am Math Mon 69(1):9–14

Goerigk M, Westphal S (2012) A combined local search and integer programming approach to the traveling tournament problem. In: Proceedings of the practice and theory of automated timetabling (PATAT 2012), pp 29–31

Goossens D, Spieksma F (2009) Scheduling the Belgian soccer league. Interfaces 39(2):109–118

Goossens D, Spieksma F (2012) Soccer schedules in Europe: an overview. J Sched 15:641–651

Grabau M (2012) Softball scheduling as easy as 1-2-3 (strikes you're out). Interfaces 42(3):310–319

Guajardo M, Jörnsten K (2017) The stable tournament problem: matching sports schedules with preferences. Oper Res Lett 45(5):461–466

Gupta A (2017) Time series modeling for dream team in fantasy premier league. In: Proceedings of the international conference on sports engineering ICSE-2017, Jaipur, India

Guyon J (2015) Rethinking the FIFA World Cup final draw. J Quant Anal Sports 11(3):169–182

Henz M (2004) Playing with constraint programming and large neighborhood search for traveling tournaments. In: Proceedings of the 5th international conference on the practice and theory of automated timetabling (PATAT, 2004), pp 23–32

Hoshino R, Kawarabayashi K (2011) A multi-round generalization of the traveling tournament problem and its application to Japanese baseball. Eur J Oper Res 215:481–497

Kendall G, Knust S, Ribeiro C, Urrutia S (2010) Scheduling in sports: an annotated bibliography. Comput Oper Res 37(1):1–19

Knust S (2010) Scheduling non-professional table-tennis leagues. Eur J Oper Res 200(2):358–367

Lasek J, Szlávik Z, Bhulai S (2013) The predictive power of ranking systems in association football. Int J Appl Pattern Recognit 1(1):27–46

Lasek J, Szlávik Z, Gagolewski M, Bhulai S (2016) How to improve a team's position in the FIFA ranking? A simulation study. J Appl Stat 43(7):1349–1368

Lewis M (2003) Moneyball: the art of winning an unfair game. Norton & Company, London

Linfati R, Gatica G, Escobar J (2019) A flexible mathematical model for the planning and designing of a sporting fixture by considering the assignment of referees. Int J Ind Eng Comput 10:281–294

McHale I, Davies S (2007) Statistical analysis of the effectiveness of the FIFA world rankings. In: Albert J, Koning RH (eds) Statistical thinking in sports. Chapman & Hall CRC, Boca Raton, pp 77–89

Melo R, Urrutia S, Ribeiro C (2006) The traveling tournament problem with predefined venues. J Sched 12(6):607–622

Miller T (2015) Sports analytics and data science: winning the game with methods and models. Pearson Education, New York

Miyashiro R, Matsui T, Imahori S (2008) An approximation algorithm for the traveling tournament problem. In: Proceedings of the 7th international conference on the practice and theory of automated timetabling (PATAT, 2008)

Nemhauser G, Trick M (1998) Scheduling a major college basketball conference. Oper Res 46:1–8

Noronha T, Ribeiro C, Durán G, Souyris S, Weintraub A (2007) A branch-and-cut algorithm for scheduling the highly-constrained Chilean soccer tournament. Lect Notes Comput Sci 3867:174–186

Nurmi K, Goossens D, Bartsch T, Bonomo F, Briskorn D, Durán G, Kyngäs J, Marenco J, Ribeiro C, Spieksma F, Urrutia S, Wolf R (2010) A framework for a highly constrained sports scheduling problems. In: Proceedings of the 2010 IAENG international conference on operations research (ICOR at IMECS), Hong Kong

Oliveira L, Souza C, Yunes T (2015) On the complexity of the traveling umpire problem. Theor Comput Sci 562:101–111

Oliveira L, Souza C, Yunes T (2016) Lower bounds for large traveling umpire instances. Comput Oper Res 72(C):147–159

Paenza A (2006) Matemática... Estás Ahí? Episodio 2 (in Spanish). Siglo XXI, Buenos Aires

Page L, Brin S, Motwani R, Winograd T (1998) The pagerank citation ranking: bringing order to the web. In: Proceedings of the 7th international World Wide Web conference, Brisbane, Australia

Radicchi F (2011) Who is the best player ever? A complex network analysis of the history of professional tennis. PLoS One 6(2):e17249

Rasmussen R (2008) Scheduling a triple round robin tournament for the best Danish soccer league. Eur J Oper Res 185(2):795–810

Rasmussen R, Trick M (2008) Round robin scheduling—a survey. Eur J Oper Res 188:617–636

Recalde D, Torres R, Vaca P (2013) Scheduling the professional Ecuadorian football league by integer programming. Comput Oper Res 40(10):2478–2484

Ribeiro C, Urrutia S (2007) Heuristics for the mirrored traveling tournament problem. Eur J Oper Res 179:775–787

Ribeiro C, Urrutia S (2012) Scheduling the Brazilian soccer tournament: solution approach and practice. Interfaces 42(3):260–272

Schönberger J (2017) The championship timetabling problem-construction and justification of test cases. In: Proceedings of MathSport international (2017) conference, Italy, Padua, p 330

Schreuder J (1992) Combinatorial aspects of construction of competition Dutch professional football leagues. Discret Appl Math 35(3):301–312

Suzuki K, Salasar B, Leite G, Louzada-Neto F (2010) A Bayesian approach for predicting match outcomes: the 2006 (Association) Football World Cup. J Oper Res Soc 61(10):1530–1539

Thielen C, Westphal S (2011) Complexity of the traveling tournament problem. Theor Comput Sci 412:345–351

Toffolo T, Wauters T, Trick M (2015) An automated benchmark website for the traveling umpire problem. http://www.gent.cs.kuleuven.be/tup. Accessed 28 Feb 2020

Toffolo T, Wauters T, Van Malderen S, Vanden BG (2016) Branch-and-bound with decomposition-based lower bounds for the traveling umpire problem. Eur J Oper Res 250(3):737–744

Toffolo T, Christiaens J, Spieksma F, Vanden BG (2019) The sport teams grouping problem. Ann Oper Res 275:223–243

Trick M (2001) Challenge traveling tournament instances. http://www.mat.tepper.cmu.edu/TOURN/. Accessed 28 Feb 2020

Trick M, Yildiz H (2012) Locally optimized crossover for the traveling umpire problem. Eur J Oper Res 216:286–292

Trick M, Yildiz H, Yunes T (2012) Scheduling major league baseball umpires and the traveling umpire problem. Interfaces 42(3):232–244

Urrutia S, Ribeiro C (2006) Maximizing breaks and bounding solutions to the mirrored traveling tournament problem. Discret Appl Math 154:1932–1938

Van Bulck D, Goossens D, Schönberger J, Guajardo M (2020) RobinX: a three-field classification and unified data format for round-robin sports timetabling. Eur J Oper Res 280(2):568–580

Van Hentenryck P, Vergados Y (2006) Traveling tournament scheduling: a systematic evaluation of simulated annealing. Lect Notes Comput Sci 3990:228–243

Westphal S (2014) Scheduling the German basketball league. Interfaces 44:498–508

Westphal S, Noparlik K (2012) A 5.875-approximation for the traveling tournament problem. Ann Oper Res 218:347–360

Willis R, Terrill B (1994) Scheduling the Australian state cricket season using simulated annealing. J Oper Res Soc 45(3):276–280

Wright M (1991) Scheduling English cricket umpires. J Oper Res Soc 42(6):447–452

Wright M (1994) Timetabling county cricket fixtures using a form of tabu search. J Oper Res Soc 45(7):758–770

Wright M (2005) Scheduling fixtures for New Zealand cricket. IMA J Manag Math 16:99–112

Wright M (2006) Scheduling fixtures for basketball New Zealand. Comput Oper Res 3:1875–1893

Xue L, Luo Z, Lim A (2015) Two exact algorithms for the traveling umpire problem. Eur J Oper Res 243(3):932–943

Yamaguchi D, Imahori S, Miyashiro R, Matsui T (2009) An improved approximation algorithm for the traveling tournament problem. Lect Notes Comput Sci 5878:679–688

Yi X, Goossens D, Talla NF (2020) Proactive and reactive strategies for football league timetabling. Eur J Oper Res 282(2):772–785

Zuccolotto P, Manisera M (2020) Basketball data science: with applications in R. Chapman & Hall, Boca Raton

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.