Optimization in sports league scheduling: experiences from the Belgian Pro League soccer

Dries Goossens[0000−0003−0224−3412]

Ghent University, Faculty of Economics and Business Administration,
Tweekerkenstraat 2, 9000 Ghent, Belgium
dries.goossens@ugent.be

Abstract. Every sports competition needs a schedule of play, stating who will play whom, when, and where. A good schedule is important given its impact on the competition's fairness and outcome, public attendance, commercial interests, and cost of policing. This paper discusses experiences with scheduling the Belgian Pro League soccer competition, for which we develop the official schedule since 2006. We present methods that have proven their value in real-life sports scheduling, and discuss how they benefited from continuous improvement, in order to accommodate changing requirements. We discuss fairness issues, as well as a discrete choice experiment we carried out to estimate the schedule's impact on stadium attendance and TV viewership.

Keywords: sports scheduling, Belgian Pro League, TV broadcasting, soccer, stadium attendance, breaks, canonical schedule

1 Introduction

It is safe to say that millions of people, all over the world, are enthralled by sports, be it actively participating or as a fan or spectator. Not surprisingly, sports has become big business. For instance, in 2014, the sports market in North America alone was worth 60.5 billion dollar and is expected to reach 73.5 billion dollar by 2019 according to a sports industry report by PricewaterhouseCoopers[1]. Apart from the purely economic impact, sports is also relevant for society because of its contribution to well-being in general. Indeed, according to Forrest and McHale [19], the impact of sports on happiness scores collected from nearly 28,000 adults in the UK turned out to be comparable with the impact of having a job.

Sports offer ample of opportunities for optimization on all kinds of levels. What strategy should a team or player apply (e.g. [28])? What players should a manager hire (e.g. [6])? How should referees be assigned to matches (e.g. [3])? How should the competition owner organize its league (e.g. [23])? How should public funds be allocated to sports (e.g. [4])? In this paper, we focus on the question how the schedule of play should be constructed. Not much more than

---

[1] See https://www.pwc.com/us/en/industry/entertainment-media/ publications/assets/pwc-sports-outlook-north-america-2015.pdf

a decade ago, most professional sports competitions were scheduled by unskilled personnel, equipped with little more than pen and paper. In some competitions, and certainly in amateur leagues, this is still the case. A good schedule is important though, because it affects the outcome and fairness of the competition, public attendance, commercial interests, as well as the cost of policing. As league owners manage to close astonishing broadcasting deals, and in light of the growing financial interests in the sports industry, the importance of a good schedule has become more and more apparent.

Sport scheduling is about determining a suitable date and venue for each match that is to be played (according to a given competition format). The complexity of this problem depends on what should be taken into account. Generating a schedule where each team plays against each other team an equal number of times, such that each team plays at most once on any given matchday is a simple task (see e.g.[12]). However, as more constraints and considerations have to be taken into account, sports scheduling can quickly become a huge challenge. Nurmi et al. [39] provide a framework for sports scheduling problems, modeled from the requirements of various professional sports leagues. This framework includes constraints related to venue availability, teams that should not have simultaneous home games, bounds on the number of consecutive matches against teams from a given strength group, and many more.

We discuss practical applications involving the scheduling of the Belgian Pro League soccer, for which we develop the official schedule since 2006. Our contract with the Belgian Pro League currently involves 2 professional and 1 amateur division, each playing a different league format and play-off competitions. Although there have been other papers that discuss the involvement of academics in real-life soccer scheduling (e.g. [5], [14], [41]), we believe that our collaboration with the Belgian Pro League is exceptional in its length. It allows us to discuss how our solution method needed continuous improvement, in order to accommodate changes in the competition format, as well as new requirements. Furthermore, we discuss how we gradually got involved in determining these requirements, e.g. by carrying out a discrete choice experiment in order to estimate the impact of the schedule on TV viewership and stadium attendance.

This paper is organized as follows. Section 2 presents an overview of the terminology and main results in sports scheduling. In Section 3, we discuss the competition format currently used in the Belgian Pro League. We also give an overview of the most important stakeholders, and their wishes and aspirations with respect to the schedule. Section 4 reports on our solution approach, and the changes it underwent over the years. We provide and a discussion and some directions for future research in Section 5.

2 Terminology and fundamentals

Sports scheduling is a relatively young academic discipline: although the first substantial contributions were published in the early 80's, a considerable increase in sports scheduling articles appeared only in recent years, including publications

in top journals in operations research. Simultaneously, several academics reached out to the sports scheduling practitioners, offering new methods to deal with the increasingly more complex problem of scheduling a league. Excellent overviews of sports scheduling have been written by Easton et al. [15] and Rasmussen and Trick [40]. An extensive bibliography can be found on a website maintained by Knust[2] and in an annotated bibliography by Kendall et al. [30]. In this section, we explain the most important concepts discussed in the sports scheduling literature.

In a sports tournament, n teams play against each other over a period of time according to a given schedule. The teams belong to a league, which organizes matches (or games) between the teams. Each match consists of an ordered pair of teams, denoted i-j, where team i plays a home match - that is, uses its own venue (stadium) for a game - and team j plays away. In a so-called *round robin tournament* each team plays against every other team an equal number of times. Most sports leagues play a double round robin tournament, where each team faces each other team twice, once at home and once away. Matches are typically grouped into so-called rounds, which are played on one or more consecutive days (usually a weekend). Teams play at most once per round; if a team does not play on some round, we say it has a bye on that round. A schedule is called *compact* or *temporally constrained* when its number of rounds is minimal. When the number of teams is even, this means that each team plays on every round. When more rounds are used than needed, we say the schedule is *temporally relaxed*.

Table 1. A compact double round robin schedule for 6 teams

R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
A-B	B-E	B-D	C-B	B-F	B-A	E-B	F-A	B-C	F-B
C-D	D-A	A-F	E-A	D-E	D-C	A-D	D-B	A-E	E-D
E-F	F-C	E-C	F-D	A-C	F-E	C-F	C-E	D-F	C-A

Table 1 presents an example of a compact schedule for a double round robin competition with 6 teams (A-F). If a team plays two home matches or two away matches in two consecutive rounds, it is said to have a *break*. For instance, team E in Table 1 has an away break in rounds 8-9. It is easy to see that any round robin schedule for an even number of teams will feature at most 2 teams that have a perfect alternation of home and away matches. Consequently, any schedule for a single round robin tournament with an even number of teams will have at least $n - 2$ breaks [12]. Goossens and Spieksma [22] introduced the so-called *generalized break*, which arises if a team has two home (away) games in a given pair of arbitrary (i.e. not necessarily consecutive) rounds, and present a number of theoretical results on this topic.

---

[2] See `http://www.inf.uos.de/knust/sportssched/sportlit_class`

Table 2. The timetable (left) and home-away pattern set (right) corresponding to the schedule in Figure 1

Team	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
A	B	D	F	E	C	B	D	F	E	C
B	A	E	D	C	F	A	E	D	C	F
C	D	F	E	B	A	D	F	E	B	A
D	C	A	B	F	E	C	A	B	F	E
E	F	B	C	A	D	F	B	C	A	D
F	E	C	A	D	B	E	C	A	D	B

Team	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
A	H	A	H	A	H	A	H	A	H	A
B	A	H	H	A	H	H	A	A	H	A
C	H	A	A	H	A	A	H	H	A	H
D	A	H	A	A	H	H	A	H	H	A
E	H	A	H	H	A	A	H	A	A	H
F	A	H	A	H	A	H	A	H	A	H

We can see a schedule as the combination of a timetable (e.g. Table 2 – left) and a set of home-away patterns (e.g. Table 2 – right). The timetable (sometimes also referred to as an opponent schedule) specifies for each round which opponent each team faces. The sequence of (H)ome and (A)way matches (and possibly also (B)yes) according to which a team plays is called the *home-away pattern* (HAP) of this team. A pair of home-away patterns is said to be complementary if the patterns never feature two home games (or two away games) on the same round. Clearly, a timetable and a home-away pattern set need to be compatible before a schedule can emerge: for each match in the timetable, the corresponding home-away patterns need to give one team the home advantage, and designate an away match for the other team. Seeing a schedule as a merging of a timetable and a pattern set, leads to two fundamental problems: the pattern set feasibility problem and the break minimization problem. The pattern set feasibility problem departs from a set of home away patterns and asks the question whether there exists a corresponding timetable (such that a given tournament format, e.g. single round robin, can be played), in which case we say the pattern set is feasible. It is still an open question whether this problem is NP-complete or not. Miyashiro et al. [36] present a necessary condition for HAP sets to be feasible. Moreover, for HAP sets having a minimum number of breaks, they show how this condition can be checked in polynomial time and conjecture it to be sufficient. In the break minimization problem, a timetable is given and the question is to determine the home assignment (i.e. the home away patterns) such that the number of breaks is minimized. To the best of our knowledge, the complexity status of this problem is also still unknown. However, Miyashiro and Matsui [37] proved a conjecture by Elf et al. [16], stating that deciding whether a feasible home-away pattern set with $n - 2$ breaks exists for a given timetable, can be solved in polynomial time.

There is a strong link between sports scheduling and graph theory (see e.g. [11], [13] and [29]). For instance, a schedule for a single round robin tournament with an even number of teams can be seen as a one-factorization of K_n, the complete graph with n nodes. The nodes in this graph correspond to the teams, and each edge between two nodes represents a match between the two corresponding teams. A one-factorization of K_n is a partitioning into edge-disjoint one-factors. A one-factor is a set of edges such that each node in the graph is incident to

exactly one of these edges (i.e. a perfect matching), and corresponds to a round. One particular one-factorization, the so-called canonical one-factorization, can be traced back to a contribution by Kirkman in the 19th century [31], and has received special attention in sports scheduling. Indeed, when looking at the schedules used in practice, in particular for soccer (see e.g. [24]), the majority of the schedules are canonical schedules, i.e. based on the canonical one-factorization. Their success may be explained by the fact that there is an easy way to generate them (i.e. the circle method, also known as the clock method or the polygon method), and that they can be constructed using the minimum number of breaks [12]. Note however that apart from minimizing breaks, many other constraints typically play a role in sports scheduling, which may heavily complicate the problem. For instance, if costs corresponding to each possible combination of a match and a round are given and the objective is to minimize the sum of these costs for the scheduled matches, the resulting sports scheduling is NP-hard [7].

Of the many fairness issues that can arise when constructing a schedule, the carry-over effect as introduced by Russell [42] has probably been studied most intensively. Russell defines team A to receive a *carry-over effect* from team B if some team X plays against team B in one round, and against team A in the next round. The carry-over effect is perceived to be particularly relevant in physical sports. For instance, if team B is very strong, and tough-playing, one can imagine that its opponent, team X, is weakened by injuries, suspensions, fatigue or lowered morale, which could be an advantage for its next opponent, team A. The opposite may be true if team B is a weak team. As carry-over effects unavoidably are present in any schedule, one can only strive to spread these carry-over effects as evenly as possible over all teams to increase overall fairness. Let c_{ij} be the number of times team j receives a carry-over effect from team i in the previous round in a schedule. A schedule's degree of imbalance with respect to carry-over effects is given by the so-called carry-over effect value, which is defined as $\sum_{ij} c_{ij}^2$ [42]. A lower bound for the carry-over effect value of a single round robin tournament is $n(n-1)$; if this bound is attained, we say the schedule is *balanced*. Russell also presents an algorithm that results in a balanced schedule when n is a power of two. For other values of n, the best known results are by Anderson [2] and Guedes and Ribeiro [27]. Lambrechts et al. [33] have shown that the canonical schedule maximizes the carry-over effect value.

3   The Belgian Pro League Soccer

We first discuss the current league format used by the Belgian Pro League (Section 3.1), followed by an overview of the various stakeholders and their (often conflicting) wishes and interests with respect to the schedule (Section 3.2).

3.1   League format

Since 2016–2017, the Belgian Pro League consists of 2 professional soccer divisions: a first division (1A) with 16 teams, which is also known as the Jupiler

Pro League and a second division (1B) with 8 teams known as the Proximus League. Both division play a so-called regular stage, which involves a double round robin tournament for 1A and a quadruple round robin tournament for 1B, completed in 30 and 28 rounds respectively. The regular stage is followed by a post-competition, inaptly called *play-offs* by the Belgian Pro League. Indeed, in general, play-offs involve a direct knock-out competition, where teams are tied in pairs and the loser of each match is eliminated. This is not the case with the play-off stage in Belgian soccer: depending on their result in the regular season, the teams are divided into one of 4 play-off groups, each of which plays a double round robin tournament. The most awaited play-off group includes the six best teams of 1A: they keep half of the points they collected in the regular stage and play for the league title, as well as access to the (qualification stage of the) Champions League or the Europa League. Two other play-off competitions include 6 teams, collected from the other teams in 1A and the top half of 1B. These teams start with a clean sheet and play for the final ticket for the Europa League qualifiers. A final play-off competition includes the bottom half of 1B; these 4 teams retain half of the points from the regular stage and play to avoid being relegated to the amateur competition. An overview of the league format is given in Figure 1.
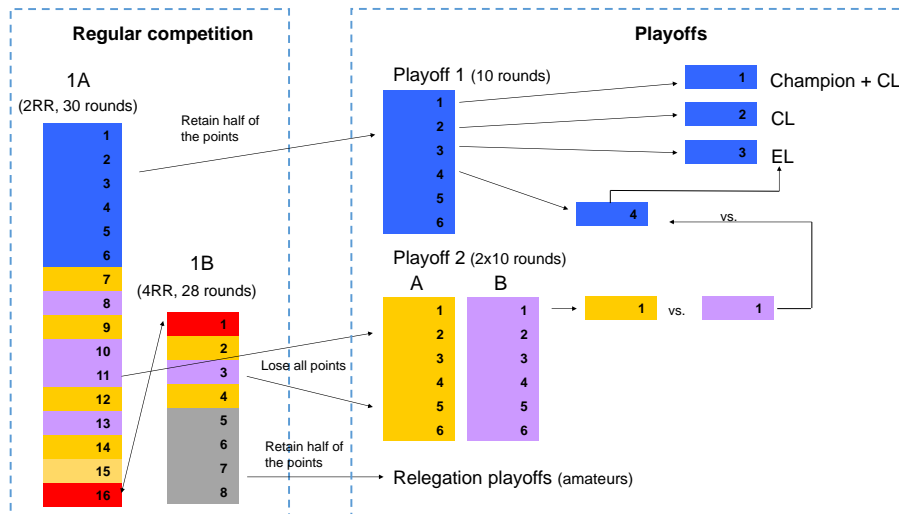
Fig. 1. The competition format for the Belgian Pro League.

Note that 2 teams do not take part in the play-off stage: the winner of the regular stage of 1B and the last team after the regular stage of 1A see an early end to their season. Certainly for the winner of 1B, this is peculiar, as this team does not get the opportunity to battle for access to the Europa League, as opposed to teams that performed worse in the regular stage of 1B. This competition format

is quite rare in soccer; it bears resemblance to the format used in Cyprus, which also has a double round robin post-competition, after the teams have been split into two groups depending on their performance in the regular stage [24]. A similar format, including the rule that teams transfer half of their points from the regular stage to the post-competition stage has been in use in the Maltese Premier League, but it was recently abandoned in favor of a classic single stage, double round robin competition format. The main motivation for the Belgian Pro League competition format is the fact that it allows for more top matches, without drastically reducing the number of teams in the top division. This format should increase the level of the top teams, allowing them to perform better in the European competitions, and at the same time generate more revenue from TV broadcasting rights, ticketing, and sponsorship.

3.2 Stakeholders and their requirements

The Belgian Pro League Being the owner of the competition, the Belgian Pro League obviously has a large say in how the competition is organized and what the schedule should look like. The Pro League has appointed a so-called *schedule manager*, who gathers the various wishes and requirements from the other stakeholders and assesses the relative importance of the scheduling requests. The schedule manager is also responsible for rescheduling postponed matches and handling complaints.

Themainconcern of the Pro League is the fairness of the competition. Breaks play an important role in this. For instance, no team should start or end the season with two home games (or two away games), as it is assumed that this would give them an advantage (disadvantage, respectively). Similarly, as a succession of away games is seen as the path towards a performance slump, teams cannot have more than two consecutive away matches. In fact, breaks in general should be minimized.

Given the large number of rounds that come with the league format discussed in Section 3.1, midweek rounds are inevitable. However, teams generally do not appreciate a home game on a midweek round, as the revenue from ticketing and catering tends to be lower than on weekend days. Consequently, the Pro League prefers a balance between home and away matches on midweek rounds: if a team plays at home on one midweek round, they should not have a home game on the next midweek round. Note that, as midweek rounds need not be consecutive, we can see this as avoiding a generalized break.

Another fairness issue is the carry-over effect. At some point, carry-over effects have been suggested in the Belgian media as the cause of a team's relegation (see e.g. [26]). Even though it was shown that there is no statistical evidence that carry-over effects have any meaningful impact on the outcome of soccer matches in the Belgian Pro League [25], the feeling still remains that unbalanced carry-over effects are to be avoided.

Finally, the Pro League strives for a balanced start of the season for all teams, avoiding that any team faces a disproportionate number of strong opponents in

the first rounds. Exceptions are made for those teams that are involved in qualification rounds for Champions League or Europa League. In order to maximize their chances of success at the European stage, these teams are offered a start of the season without big domestic clashes.

The TV broadcasters TV broadcasters aim to maximize their revenue by attracting as many viewers as possible. Over the years, we have had many different scheduling requests from various TV broadcasters. In the beginning, each round needed to feature at least one (and preferably two) of four top teams playing an away match. The underlying motivation was that a top team's home matches are less attractive, as the top team tends to win these games without much opposition. Soon after that, we received a request to have at most one top match per round, and a balanced spread of top matches over the season. Later, the TV broadcaster wanted to have so-called Super Sundays, i.e. a Sunday with (at least) two top matches. Given this diverse set of requests, at some point we started to wonder whether the TV broadcasters actually knew what kind of schedule would maximize their viewing rates. The number of papers on the determinants for TV viewership of soccer matches is surprisingly limited. Most of them focus on the relation between viewer ratings and match outcome uncertainty (e.g. [1] and [8]); Forrest, Simmons, and Buraimo [17] apply their model to support the broadcasters' choice of which matches to televise in the English Premier League. However, none of these contributions show how the schedule should be constructed in order to maximize TV viewership.

On a typical weekend round in the regular stage of the competition, the following kick-off times are used for matches in 1A: Friday 20:30, Saturday 18:00, Saturday 20:00 (2 matches), Saturday 20:30, Sunday 14:30, Sunday 18:00, and Sunday 20:00. Essentially, the schedule determines which opponents will face each other in which part of the season and at what kickoff time. To obtain a better understanding of the impact of these match characteristics (month, kickoff time, and opponent strength) on TV viewership, we conducted a discrete choice experiment using an online survey questionnaire distributed to chart the preferences of Belgian soccer fans in watching the league matches [44]. Our results show that fans on average dislike matches scheduled for Wednesday 20:30, Saturday 18:00, Sunday 14:30, and Sunday 20:30. In particular, midweek matches are to be avoided, although there is quite some heterogeneity in the individual preferences for the Wednesday 20:30 kick-off time (probably due to the difference between employees who have to go to work the next morning and fans without a job). If midweek matches are inevitable, they are best planned in October or March to ensure a reasonably sized audience. We found that it does not seem beneficial to start the season late in July or even early in August, as a considerable proportion of the fans take holidays in this period. Surprisingly, fans do not prefer to see their team play against a top team regardless of the month and kick-off time: matches against teams of average strength can attract more viewers than matches against top teams if carefully scheduled on attractive rounds and kick-off times. Our model can also be used to predict audience ratings

for each (type of) match, depending on the planned month and kickoff time, which in turn can be used as input parameter for optimizing the schedule.

The police and the tax payer Soccer matches in the Belgian Pro League are events that can attract up to 30,000 fans. Moreover, some matches, labeled *risk matches*, are known to have an increased risk of hooliganism. Hence, it is clear that the police is an important stakeholder in professional soccer. Moreover, it is important to realize that a mayor has the right to forbid a match being played if he or she judges that public order and safety cannot be guaranteed.

Months before the start of the new season, various local police forces are queried about their concerns and requests with respect to the schedule. Most requests deal with risk matches, which should be avoided on dates when other events are scheduled in the city that require the attention of the local police (e.g. a festival, summit, or other sports event). Sometimes, the request only concerns the kick-off time, as some matches can be managed more easily during daylight. The impact on traffic of scheduling a match on a specific date and time is also regularly pointed out by the police. A few police zones contain the venues of two (or even more) teams, and in this case some coordination is needed: preferably, teams from the same police zone do not play a home game on the same weekend. For sure, no pair of teams from the same zone should host risk games simultaneously.

Ultimately, the tax payer bears the vast majority of the policing cost for soccer matches. Clearly, the schedule has an impact on the size of the policeforce needed, as well as the costs involved. For instance, scheduling a risk game in the beginning of the season will allow the police to deploy a smaller force compared to when this game is decisive for e.g. relegation. Furthermore, matches scheduled e.g. in the Christmas period result in a higher man-hour rate than midweek matches.

The clubs and their fans Clubs have a wide variety of wishes. Teams sharing a venue have a strong reason for not playing home games in the same round. Although Belgium currently does not have multi-purpose stadia (i.e. stadia designed to host multiple types of events, such as pop concerts and various types of sports contests), in some cases, the parking facilities are shared with a shopping mall or another sports club, resulting in a list of days on which home games are to be avoided. Most teams believe they can increase their chances of playing a good season by starting with a home game, ideally against an easy opponent. Champions League or Europa League contenders typically request a home game after their European match, as they believe a home game renders them less vulnerable to the effects of fatigue resulting from this extra midweek match (and trip). Teams with cash flow problems prefer a steady number of home matches per month. Occasionally, a team requests to delay the first home game by starting the season with two consecutive away games, in order to have enough time to finish venue construction or renovation works.

In general, teams want to maximize the number of fans attending their home games, as this is closely related to their revenue. Various models have been developed to analyze the factors that determine stadium attendance (see e.g. [20], [43], [9], [34], [10]). Compared to TV broadcasting, more factors can influence a soccer fan's decision whether or not to travel to the stadium: distance from the stadium, weather, availability of car parking, availability of tickets, safety, service, catering, etc. Here, we limit ourselves to those factors that can be impacted by the schedule. Forrest and Simmons [18] show that consecutive home games negatively impact attendance. Hence, teams generally prefer a schedule in which breaks are minimized. Wang et al. [44] also studied the impact of month, kick-off time, and opponent strength on stadium attendance for Belgian soccer fans. They find that, contrary to the results with respect to TV viewership, January is not a popular month for attending a soccer match in the stadium (presumably because of the cold). Apparently, August's nice summer weather is neutralized by the fact that many people leave on holiday. Wednesday 20:30 and Sunday 20:30 are deplorable kick-off times with respect to stadium attendance, as the next day is a working day. In contrast, Saturday 18:00 and 20:00 and Sunday 14:30 and 18:00 all work very well.

4   Solution approach

Before 2006, the Pro League's highest division consisted of 18 teams, playing a double round robin tournament without play-offs; the schedule was developed manually. The starting point was a so-called basic match schedule, i.e. a schedule that specifies the round and the home advantage for all matches, however using placeholders instead of real teams. For instance, the schedule in Table 3 shows the first 7 rounds of the basic match schedule that has been used by the Pro League for decades. Clearly, when each team is coupled with a placeholder (1–18), a schedule follows. Although the origin of this basic match schedule is not clear, it is a canonical schedule and it has many advantages. For instance, it consists of two halves, such that each team faces each other team exactly once in each half, and each half has the minimum number of breaks. Furthermore, no team plays three home games or away games in a row, and no team starts or ends the season with two home games or two away games. Moreover, it condenses the whole scheduling problem to a relatively easy question: which placeholder should be matched with which team? The way in which this should be done, however, was less transparent. The schedule manager did this assignment without the use of any optimization tool, till the the point where he found no further improvement by manually swapping the placeholder assignment of a pair of teams. Not surprisingly, the resulting schedules typically satisfied merely a small fraction of the constraints. As many teams found their wishes brushed aside without understanding how the schedule was constructed, this lead to several accusations of favoritism and lack of transparency in the media (see e.g. [32]).

When we got involved, our main challenge was to automate and optimize the assignment of teams to placeholders in the basic match schedule. This, however,

Table 3. A basic match schedule for a double round robin tournament with 18 teams, rounds 1-7

R1	R2	R3	R4	R5	R6	R7
1-3	2-4	1-7	2-8	1-11	2-12	1-15
4-17	3-18	3-5	4-6	3-9	4-10	3-13
6-15	5-1	6-2	5-18	5-7	6-8	5-11
8-13	7-16	8-17	7-3	8-4	7-18	7-9
10-11	9-14	10-15	9-1	10-2	9-5	10-6
12-9	11-12	12-13	11-16	12-17	11-3	12-4
14-7	13-10	14-11	13-14	14-15	13-1	14-2
16-5	15-8	16-9	15-12	16-13	15-16	16-17
18-2	17-6	18-4	17-10	18-6	17-14	18-8

could not be done without explicitly stating all requirements and their relative importance. The latter turned out problematic: a committee with representatives from some of the clubs, had to come to an agreement on the scheduling priorities, whereas they were used to deal with the schedule manager individually and informally. Eventually, we agreed to classify each requirement into one of 5 priority classes, where each class has a certain weight (it took some time to convince them that putting all requirements in the highest priority class is not the way to obtain the best schedule). The optimization model was based on the assignment model. We used a binary decision variable $x_{i,p}$ which is 1 if team i is assigned to placeholder p and 0 otherwise. Furthermore, y_c is 1 if constraint c is not satisfied, and 0 otherwise. We use parameters q_c for the weight associated with the class to which constraint c belongs, and $a_{i,p,c}$ and b_c as parameters to model the various requirements (M is a sufficiently large number). The problem formulation is then as follows:

$$\mathbf{Min} \sum_c q_c \cdot y_c \tag{1}$$

subject to

$$\sum_p x_{i,p} = 1 \quad \forall i \tag{2}$$

$$\sum_i x_{i,p} = 1 \quad \forall p \tag{3}$$

$$\sum_i \sum_p a_{i,p,c} \cdot x_{i,p} \leqslant b_c + M \cdot y_c \quad \forall c \tag{4}$$

$$x_{i,p} \in \{0,1\} \quad \forall i, p \tag{5}$$

$$y_c \in \{0,1\} \quad \forall c. \tag{6}$$

The objective function minimizes the total weight of violated constraints. Constraints (2) and (3) ensure that each team is assigned to a placeholder and

vice versa. The third set of constraints can be used to model all requirements stakeholders may have by selecting appropriate values for $a_{i,p}$ and b_c.

In the beginning of our collaboration, the Pro League insisted that we used their basic match schedule. The main reasons for this were that they were highly familiar with it, and that lower divisions – which to some extent tune their schedules based on the schedule of the highest division – also used it. However, we soon became the victim of our early success: as we were able to handle many more constraints than the manual approach, stakeholders gradually came up with more wishes and requirements and higher expectations in the next seasons. At some point, it became clear that no satisfactory solution existed, if we restricted the solution space to schedules resulting from the Pro League's basic match schedule. Hence, the next step was to use a phased approach, where we first assign each team to one of the home-away patterns from the basic match schedule, and subsequently decide on the opponents for each team in each round, given the home-away pattern assignment. We call this the two-phase approach; it is a so-called *first break, then schedule* approach, based on a decomposition introduced by Nemhauser and Trick [38].

The first phase can in fact also be modeled using formulation (1)-(6), however this time the binary decision variable $x_{i,p}$ is 1 if team i is assigned to home-away pattern p and 0 otherwise. Notice that constraints (4) now only include constraints that relate to the home advantage; no constraints involving specific opponents are included in this model. The second phase uses a binary decision variable $x_{i,j,k}$ which is 1 if team i plays a home game against team j on round k, and 0 otherwise. We use $a_{i,j,k,c}$ and b_c as parameters to model the various wishes from the stakeholders related to opponents. Note that we enforce that both halves of the schedule are mirrored, i.e. the second half of the schedule is identical to the first half, except the home advantage, which is inverted. Mirroring simplifies the model because we only need to schedule the first half of the season: constraints related to the second half can be modeled in terms of first-half constraints. The formulation is as follows:

$$\textbf{Min} \sum_c q_c \cdot y_c \tag{7}$$

subject to

$$\sum_j (x_{i,j,k} + x_{j,i,k}) = 1 \quad \forall i, k \tag{8}$$

$$\sum_k (x_{i,j,k} + x_{j,i,k}) = 1 \quad \forall i, j : i \neq j \tag{9}$$

$$\sum_i \sum_j \sum_k a_{i,j,k,c} \cdot x_{i,j,k} \leqslant b_c + M \cdot y_c \quad \forall c \tag{10}$$

$$x_{i,j,k} \in \{0, 1\} \quad \forall i, j, k \text{ allowed by the HAPs of team } i \text{ and } j \tag{11}$$

$$y_c \in \{0, 1\} \quad \forall c. \tag{12}$$

The objective function again minimizes the total weight of violated constraints. Constraints (8) ensure that each team plays exactly once on each round, and constraints (9) enforce that each team meets each other team once in the first half of the season. Constraints (10) are used to model all opponent-related stakeholder requirements. This approach keeps all properties of the basic match schedule related to home advantage and breaks, but drastically increases the search space, as the order of the opponents is no longer fixed. An important argument to abandon the basic match schedule was the carry-over effect. Indeed, since the basic match schedule, being a canonical schedule, has maximally unbalanced carry-over effects [33], any change results in a positive impact on the carry-over effect value.

Solving the two phases sequentially, however, does not guarantee an optimal solution and typically leaves a lot of room for improvement. Hence, we developed a tabu search algorithm around the first phase. This algorithm finds the best move in a neighborhood around some current solution; this neighborhood consists of all solutions that can be reached by swapping the home-away patterns of two teams in this solution. Computationally, computing the consequences of a swap with respect to the objective function in phase 1 is trivial. The evaluation of phase 2 can be more time-consuming, although it frequently happens that there is no need to solve phase 2, since the result after solving phase 1 is already worse than the best solution so far in this neighborhood. We maintain a tabu list to make sure that moves are not immediately reversed.

As new requirements kept coming, it did not take long before we again found no suitable solution in the solution space we considered. Hence we argued to drop the tradition of having the second half mirrored from the first half. Indeed, this created for each requirement connected to the first half of the season a counterpart in the second half, and vice versa, effectively doubling the number of constraints. It took quite some effort to convince the Pro League to allow non-mirrored schedules; in the end, the fact that a non-mirrored schedule had been played before in the Belgian Pro League, albeit a long time ago and for reasons no one could recollect, was a final nudge in the right direction. The fact that the home-away patterns where still mirrored resulted in a schedule that was still about 80% mirrored and at least did not preclude a mirrored schedule provided that this was feasible given the requirements.

The introduction of the play-offs in the season 2009–2010 entailed a number of changes with respect to the scheduling process. Until then, the exact kick-off times were not communicated in the beginning of the season: the broadcaster had the right to delay that decision until one month before the match. This allowed them to take into account more recent information on e.g. which teams qualify for what stages of other competitions (domestic cup, Europa League, Champions League) and which teams perform well in the domestic league. This is still the case for the regular stage of the season, however, for the play-off stage, all kick-off times are to be determined by the schedule manager before the play-offs start. Furthermore, interdependencies between the various play-off competitions required a totally different scheduling approach, in which all play-

off competitions are scheduled simultaneously. On the other hand, each play-off involves at most six teams, such that an IP model based on variables $x_{i,j,k,t} = 1$ if team i plays against team j on round k using kick-off time t (and 0 otherwise), and $y_{i,k} = 1$ if team i has a break on round k (and 0 otherwise) can be solved in a reasonable computation time with e.g. IBM Ilog Cplex.

Finally, we also moved away from the set of home-away patterns that follow from the basic match schedule. Interesting properties of these HA patterns are that (i) they don't begin or end with a break, (ii) they don't have consecutive breaks, (iii) they have at most 3 breaks, and (iv) they allow a mirrored schedule. It is easy to see that for e.g. a double round robin tournament with 16 teams, there are 26 such home-away patterns (13 complementary pairs), i.e. 10 more on top of the 16 HAPs implied by the basic match schedule. Hence, we found that among the 1287 possibilities to pick 8 pairs of complementary HAPs out of 13, 46 HAP sets lead to a feasible schedule. Again, this greatly increased the search space. Currently, we are also including HAPs with more than 3 breaks that do not allow a mirrored schedule, but have other properties that fit well with the requirements of that season.

5   Discussion

At the start of our collaboration with the Pro League in the season 2006–2007, the TV broadcasting rights for the highest division were sold for a price of 36 million euros per season. The current broadcasting contract has been marketed for 70 million euros per season, which is almost double as much (but still a trifle compared to the amounts paid in e.g. the English Premier League). In the seven seasons before our involvement, stadium attendance per season for the first division amounted to on average 2,954,311, which corresponds to an average of 9,654 fans per match. In the ten seasons for which we provided the schedule, stadium attendance equalled 3,451,945 per season (11,491 per match) on average, which corresponds to an increase of 17%. At the same time, the number of man-hours performed by the police decreased from 28,613 per season in the three seasons right before our involvement to 21,780 per season since we create the official schedule (i.e. a decrease of about 24%). This decrease is remarkable, because since the introduction of the play-offs more matches are played, and many of these additional matches are risk matches. Obviously, there are many factors that play a role in these changes, for instance the introduction of the play-offs has undoubtedly impacted the attractiveness of the competition, and the fanbase of the team that promotes to and the team that is relegated from the first division certainly plays a role. Nevertheless, we like to believe that our scheduling approach also has contributed considerably to these positive trends. Equally important, it has improved the transparency of the process. Although, obviously, the stakeholders do not have a real understanding of the optimization method itself, it is seen as neutral. Moreover, it is now clear to everyone what the various requirements are, and what weight they get. Setting these weights is not an exact science. Therefore, we present the Pro League with

several good schedules to choose from, instead of just the best. Occasionally, the Pro League eventually selects a schedule which is not optimal according to our approach, which is a motivation for us to reevaluate the weights that some of the requirements receive. Our method offers additional flexibility through what-if analyses, which can be used to assess the consequences of honoring certain wishes. As a result, the yearly tradition of stakeholders having a big fight in the media when the new schedule is announced has now disappeared.

Although we have made a lot of progress with the Pro League over the past 10 years, there are still a number of opportunities for further research. We believe that these opportunities not only apply to our case with the Belgian soccer league, but to the sports scheduling community in general. Phased approaches are very popular, and they perform well provided that the most important requirements can be solved in earlier phases. Nevertheless, there is still room for improvement with respect to solution methods. We are not able to solve highly-constrained sports scheduling problems optimally for league sizes larger than 12. We have little insight in what makes a HAPset feasible, particularly if the home-away patterns are not symmetrical and feature several breaks. Many practical sports scheduling problems require exactly these type of home-away patterns. Scheduling multiple leagues simultaneously is another major challenge for future research. Indeed, according to our experience, the focus is shifting from scheduling one important league to scheduling a multitude of leagues with interdependencies. This is even more apparent in amateur of youth sport leagues. Finally, we would like to point out the importance of sports analytics as a way to determine what is important in sports scheduling. We noticed that stakeholders often don't have a good view on the impact of the schedule on fairness, gate revenue, TV viewership, or other issues they care about. Hence, the sports scheduling community can contribute by clarifying these relationships, and assist practitioners with providing the right parameters for their sports scheduling problem.

References

1. Alavy, K., Gaskell, A., Leach, S., Szymanski, S.: On the edge of your seat: Demand for football on television and the uncertainty of outcome hypothesis. International Journal of Sport Finance 5, 75-95 (2010).
2. Anderson, I.: Balancing carry-over effects in tournaments. In: Combinatorial designs and their applications, edited by F. Holroyd, K. Quinn, C. Rowley, and B. Webb, pp. 1–16 (1997).
3. Alarcón, F., Duran, G., Guajardo, M.: Referee assignment in the Chilean football league using integer programming and patterns. International Transactions in Operational Research 21, 415–438 (2014).
4. Angulo-Meza, L., Soares de Mello, J.C.C.B, Valério, R.P.: Assessing the Efficiency of Sports in Using Financial Resources with DEA Models? Procedia Computer Science 55, 1151–1159 (2015).
5. Bartsch, T., Drexl, A., Kröger, S.: Scheduling the professional soccer leagues of Austria and Germany. Computers and Operations Research 33, 1907–1937 (2006).

6. Boon, B., Sierksma, G.: Team formation: Matching quality supply and quality demand. European Journal of Operational Research 148 (2), 277–292 (2003).
7. Briskorn, D., Drexl, A., Spieksma, F.C.R.: Round tobin tournaments and three index assignments. 4OR 8, 365–374 (2010).
8. Buraimo, B.: Stadium attendance and television audience demand in English league football. Managerial and Decision Economics 29, 513-523 (2008).
9. Buraimo, B., Forrest, D., Simmons, R.: Insights for clubs from modeling match attendance in football. Journal of Operational Research Society 60(2), 147–155 (2009).
10. de Carvalho M., Boen F., Sarmento J., Scheerder J.: What brings youngsters into the stadium? Sociopsychological predictors of soccer attendance among Belgian and Portuguese young fans. Portuguese Journal of Sports Sciences (1), 21–40 (2015).
11. de Werra, D.: Geography, games and graphs, Discrete Applied Mathematics 2(4), 327–337 (1980).
12. de Werra, D.: Scheduling in sports, In: P. Hansen (ed.), Studies on Graphs and Discrete Programming, 381–395 (1981).
13. Drexl, A., Knust, S.: Sports league scheduling: graph- and resource-based models. Omega 35, 465–471 (2007).
14. Duran, G., Guajardo, M., Miranda, J., Saure, D., Souyris, S., Weintraub, A., Wolf, R.: Scheduling the Chilean soccer league by integer programming. Interfaces 37, 539–552 (2007).
15. Easton, K., Nemhauser, G., Trick M.: Sports scheduling. In: Leung (ed.), Handbook of Scheduling, CRC Press, Florida, USA, pp. 52.1–52.19 (2004).
16. Elf, M., Jünger, M., Rinaldi G.: Minimizing breaks by maximizing cuts. Operations Research Letters 31, 343-349 (2003).
17. Forrest, D., Simmons, R., Buraimo, B.: Outcome uncertainty and the couch potato audience. Scottish Journal of Political Economy 52, 641-661 (2005).
18. Forrest, D., Simmons, R.: New issues in attendance demand: The case of the English football league, Journal of Sports Economics 7(3), 247–266 (2006).
19. Forrest, D., McHale, I.G.: Subjective well-being and engagement in sport: Evidence from England. In: Rodrguez, P., Kesenne, S., Humphreys, B.R. (eds.) The Economics Of Sport, Health And Happiness: The Promotion of Well-being through Sporting Activities, Edward Elgar Publishing, pp. 184–199 (2011).
20. Garcia, J., Rodriguez, P.: The determinants of football match attendance revisited: Empirical evidence from the Spanish football league. Journal of Sports Economics 3, 18-38 (2002).
21. Goossens, D., Spieksma, F.C.R.: Scheduling the Belgian soccer league. Interfaces 39(2), 109–118 (2009).
22. Goossens, D., Spieksma, F.C.R.: Breaks, cuts, and patterns. Operations Research Letters 39(6), 428–432 (2011).
23. Goossens, D., Belien, J., Spieksma, F.C.R.: Comparing league formats with respect to match importance in Belgian football. Annals of Operations Research 194, 223–240 (2012).
24. Goossens, D., Spieksma, F.C.R.: Soccer schedules in Europe: an overview. Journal of Scheduling 15(5), 641–651 (2012).
25. Goossens, D., Spieksma, F.C.R.: The carryover effect does not influence football results. Journal of Sports Economics 13(3), 288–305 (2012).
26. Geril, J.: Ons budget voor transfers? Nul komma nul euro [in Dutch]. Het Nieuwsblad (VUM), February 2nd (2007).

27. Guedes, A.C.B., Ribeiro, C.C.: A heuristic for minimizing weighted carry-over effects in round robin tournaments. Journal of Scheduling, 14(6), 655–667 (2011).

28. Hirotsu, N., Wright, M.B.: Using a Markov process model of an association football match to determine the optimal timing of substitution and tactical decisions. Journal of Operational Research Society 53(1), 88-96 (2002).

29. Januario, T., Urrutia, S., Ribeiro, C.C., de Werra, D.: Edge coloring: A natural model for sports scheduling. European Journal of Operational Research 254, 1–8 (2016).

30. Kendall, G., Knust, S., Ribeiro, C.C., Urrutia, S.: Scheduling in Sports: An annotated bibliography. Computers and Operations Research 37, 1–19 (2010).

31. Kirkman, T.: On a problem in combinations, Cambridge and Dublin Mathematical Journal 2, 191–204 (1847).

32. Reunes, M.: Club boos op Charleroi [in Dutch]. Het Nieuwsblad (VUM), September 29th (2005).

33. Lambrechts, E., Ficker, A.M.C., Goossens, D., Spieksma, F.C.R.: Round-robin tournaments generated by the circle method have maximum carry-over. Mathematical Programming, to appear. https://doi.org/10.1007/s10107-017-1115-x. (2017).

34. Madalozzo, R., Villar, R.B.: Brazilian football: What brings fans to the game? Journal of Sports Economics 10, 639-650 (2009).

35. Mendelsohn, E., Rosa, A.: One-factorizations of the complete graph – a survey. Journal of Graph Theory 9, 43–65 (1985).

36. Miyashiro, R. Iwasaki, H., Matsui, T.: Characterizing feasible pattern sets with a minimum number of breaks. In: Lecture Notes in Computer Science, vol. 2740, 78-99 (2003).

37. Miyashiro, R., Matsui T.: A polynomial-time algorithm to find an equitable home-away assignment. Operations Research Letters 33 235-241 (2005).

38. Nemhauser, G.L., Trick, M.A.: Scheduling a major college basketball conference. Operations Research 46, 1-8 (1998).

39. Nurmi, K., Goossens, D., Bartsch, T., Bonomo, F., Briskorn, D., Duran, G., Kyngs, J., Marenco, J., Ribeiro, C.C., Spieksma, F., Urrutia, S., Wolf-Yadlin, R.: A Framework for Scheduling Professional Sports Leagues. In Ao, Sio-Iong (ed.): IAENG Transactions on Engineering Technologies 5, 14–28 (2010).

40. Rasmussen, P., Trick, M.A.: Round robin scheduling – A survey. European Journal of Operational Research 188, 617–636 (2008).

41. Recalde, D., Torres, R., Vaca, P.: Scheduling the professional Ecuadorian football league by integer programming. Computers & Operations Research 40, 2478–2484 (2013).

42. Russell, K.G.: Balancing carry-over effects in round robin tournaments. Biometrika 67(1), 127-131 (1980).

43. Scarf, P.A., Shi, X. The importance of a match in a tournament. Computers and Operations Research 35(7), 2406–2418 (2008).

44. Wang, C., Goossens, D., Vandebroek, M.: The impact of the soccer schedule on TV viewership and stadium attendance: Evidence from the Belgian Pro League. Journal of Sports Economics, to appear. https://doi.org/10.1177/1527002515612875. (2017).