INTERFACES
Vol. 00, No. 0, Xxxxx 2011, pp. 000–000
ISSN 0092-2102 | EISSN 1526-551X | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

Scheduling the Brazilian Soccer Tournament: Solution Approach and Practice

Celso C. Ribeiro
Universidade Federal Fluminense, Department of Computer Science,
Niterói, RJ 2431-240, Brazil.

Sebastián Urrutia
Universidade Federal de Minas Gerais, Department of Computer Science,
Belo Horizonte, MG 31270-010, Brazil.

Sports, with their massive investments in players and structures, have become a big business. Professional and amateur leagues face challenging problems, including logistics, revenue maximization, broadcast rights, fairness issues, game attractiveness, and security. The annual Brazilian soccer tournament is a compact, mirrored double round-robin tournament played by 20 teams in each of its two main divisions; it is possibly the world's most attractive soccer tournament because of the quality of the teams and players in the competition. With substantial revenue and community pride on the line, devising optimal schedules is crucial to players, teams, fans, sponsors, cities, and for security issues. Fair and balanced schedules for all teams are a major issue for ensuring attractiveness and confidence in the tournament outcome. The organizers seek schedules that satisfy a number of constraints. As often as possible, the most important games should be played in weekend rounds so that the open TV channels can broadcast many attractive games. We describe the integer programming formulation of the scheduling problem and the three-phase decomposition approach we proposed for solving it. We also report on the practical experience we observed after two years of running the system and the main results achieved during its successful history.

Key words: sports scheduling; soccer; football; tournament; integer programming.
History: This paper has been refereed.

Soccer, the most popular sport in Brazil, will surely see its status raised to unprecedented levels, both in Brazil and worldwide, thanks to the 2010 World Cup in South Africa, the 2014 World Cup to be hosted by Brazil, and the awarding of the 2016 Olympics to Rio de Janeiro.

Professional soccer leagues are big businesses and a major economic activity. Fans check newspapers, radio, television, and the Internet in their quest for information about their local and national teams, preferred players, and scores. Transfers of major soccer players amount to tens of millions of dollars per player. The sponsorship of national teams by sporting goods manufacturers involves huge contracts. Broadcast rights in some professional sports competitions amount to hundreds of millions of dollars. Television networks buy the rights to broadcast the games for huge amounts of money; in return, they want the most attractive games to be scheduled at specific times. Revenues from attendance and tournament attractiveness also depend strongly on the schedule of the games, which must satisfy a number of constraints. Geographical, technical, fairness, and security constraints impose intricate patterns of scheduling that are difficult to determine.

Teams and professional leagues do not want to waste their investments in players and structure due to poor schedules involving, for example, unattractive teams playing on prime dates or several important games played at the same time (resulting in loss of television rights since they cannot be broadcast simultaneously). National and international competitions played in parallel require strong coordination of travel and game schedules. Professional soccer leagues face challenging optimization problems. Efficient schedules are therefore of major interest for teams, leagues, sponsors, fans, and the media. Recent surveys by Rasmussen and Trick (2008) and Kendall et al. (2010) address the literature on sports scheduling.

The annual soccer tournament organized by the Brazilian Football Confederation (CBF) is Brazil's most important sporting event. Its major sponsor is TV Globo, the largest media group and television network in Brazil. Fair and balanced schedules for all teams are major issues for ensuring attractiveness and confidence in the tournament's outcome. Furthermore, TV sponsors condition their support based on schedules that make it possible for open (i.e., nonpaid) channels to broadcast the most important games. Large cities hosting two or more teams and the large number of fans impose additional security constraints to avoid clashes of fans before or after the games.

Nurmi et al. (2010) have noticed that, in spite of the relevance of good game schedules, very few professional leagues have adopted optimization models and software to date. This seems to be due both to the hardness of the problem and to some fuzzy preference restrictions and criteria that can be hard to describe, and also to the resistance of teams and leagues to using new tools that introduce modern techniques in sports management. Bartsch et al. (2006) applied heuristics and branch-and-bound techniques to schedule the professional soccer leagues of Austria and Germany. Della Croce and Oliveri (2006) adapted the integer programming decomposition approach of Nemhauser and Trick (1998) in scheduling the Italian football league. Durán et al. (2007, 2009) also used integer programming to schedule the Chilean soccer league. Improvements to their solution approach have been proposed by Noronha et al. (2007). Ribeiro and Urrutia (2007) presented a bicriteria integer programming approach for scheduling the annual first-division football Brazilian tournament, in which they sought a solution that minimizes the number of breaks and maximizes the number of games that open TV channels could broadcast. Goossens and Spieksma (2009) reported on the application of integer programming to schedule the Belgian soccer league for the 2006–2007 and 2007–2008 seasons. Rasmussen (2008) presented a solution approach using a logic-based Benders decomposition and column generation to solve a triple round-robin tournament for the Danish soccer league. Fiallos et al. (2010) developed an integer programming model solvable by CPLEX, which was used for the first time in 2010 to schedule the double round-robin professional soccer tournament of Honduras.

This paper describes the formulation, implementation, and practical use of the optimization software developed by the authors in partnership with CBF to determine good schedules for the first (Series A) and second (Series B) divisions of the Brazilian soccer tournament, extending previous work published in Ribeiro and Urrutia (2007). In the *Tournament Structure* section, we summarize the main characteristics of the tournament, including its structure and the participating teams. The *Schedule Requirements* section discusses the requirements that the schedule must meet. The associated integer programming model is presented in the *Integer Programming Model* section. *Solution Approach* describes the solution approach based on a three-phase decomposition strategy. *Development and Practical Experience* presents some statistics related to software development and algorithm performance, and includes a report on the practical experience resulting from using the proposed interactive software to schedule the 2009 and 2010 editions of the tournament.

We note that we use the terms *football* and *soccer* to refer to the sport regulated by the Fédération Internationale de Football Association (FIFA).

Tournament Structure

In this paper, we consider a tournament played by an even number n of teams. A *round-robin* tournament is one in which every team plays against every other team a fixed number of times. Each team faces each other team exactly once (respectively twice) in a single (respectively double) round-robin tournament. The tournament is said to be *compact* if the number of rounds is minimum and, consequently, each team plays exactly one game in every round. Each team has its own venue in its home city and each game is played at the venue of one of the two teams in confrontation. The team that plays at its own venue is called the home team and is said to play a *home game*; the other is called the away team and is said to play an *away game*.

Double round-robin tournaments are often partitioned into two phases. Each game occurs exactly once in each phase, but with different home rights. In the case of a *mirrored* schedule, games in the second phase are scheduled in the same order as in the first, but with exchanged venues.

A tournament schedule establishes the round and the venue in which each game will be played. A *home-away pattern* (HAP) determines in which condition (home or away) each team plays in each round. For any given schedule, one says that there is a home (respectively away) *break* whenever a team plays two consecutive home (respectively away) games.

The annual Brazilian soccer tournament lasts for seven months, from May to December. Each of its main divisions (Series A and Series B) is structured as a compact, mirrored double round-robin tournament played by $n = 20$ teams. There are $2 \cdot (n-1) = 38$ rounds. Games of *weekend rounds* are played on Saturdays and Sundays, while those of *midweek rounds* are played on Wednesdays and Thursdays. The dates available for game playing change from year to year and must be coordinated with other competitions, such as the state tournaments, Brazil's Cup, South America's Cup, and the Santander Libertadores Cup. If the game between teams A and B is played in the first phase at the venue of A in some round $k = 1, \ldots, n-1$, then the second game between A and B will be played in round k of the second phase (or in the overall round $k+n-1$), but at the venue of B. Some specific games are required to be played only on weekends. For any $r = 1, \ldots, n-1$ such that both rounds r and $r+n-1$ are weekend rounds, we say that r is a *double weekend round*.

Each year, the last four qualified teams in Series A are downgraded to play Series B in the next year, while the four best-qualified teams in Series B are upgraded to play Series A. TV rights, marketing revenues, and gate attendance are much larger for teams in Series A than for those in Series B.

Twelve teams belong to the so-called *Group of Twelve* (G12), formed by the strongest founding teams of the league. The only advantage of these teams is their larger broadcast rights. Not all of them necessarily play in Series A every year.

Teams playing in the same Series (A or B) are organized by pairs (as determined by CBF) with complementary home-away patterns of game playing. Usually, teams in the same pair are based in the same home city. Table 1 displays the names of the 20 teams participating in the 2010 edition of the tournament, along with their home cities, the states in which their home cities are located, and an indication of whether a team belongs to G12. Figure 1 shows the locations of the home cities of the participating teams. Santos, an important city with a well-known team, is very close to the city of São Paulo. For simplicity, whenever we refer to the teams of the city of São Paulo, we are also including Santos.

São Paulo is the richest state in Brazil and has a number of strong teams both in its capital and in some small but economically active cities. In contrast, the participating teams from all other states have their home cities exclusively in the state capitals.

We conclude this section by two additional definitions. *Regional games* involve two opposing teams whose home cities are located in the same state. *Classic* games are those that involve two opponents based in the same home city, that is, teams based in Rio de Janeiro, or São Paulo (including Santos), Porto Alegre, Belo Horizonte or Goiânia, for the 2010 edition of the tournament. These games have a long tradition of rivalry. They are usually the most important games, and attract the largest attendance and more interest from the fans and the media. All classic games are also regional games.

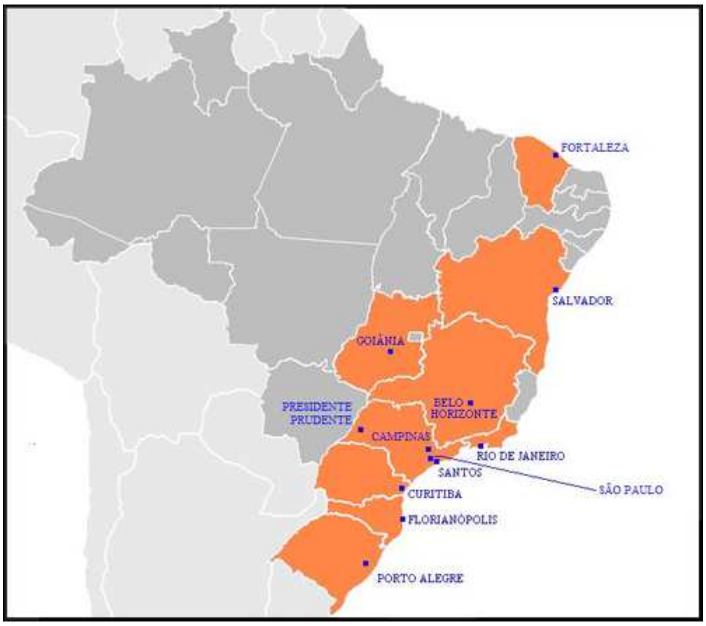
Schedule Requirements

The tournament schedule should satisfy a number of constraints, ranging from fairness to security issues, and from technical to broadcasting criteria. Most reflect strategies for maximizing revenues

Table 1 **The table shows participating teams in Series A of the 2010 edition of the Brazilian soccer tournament; it shows their home cities and home states, and indicates if a team belongs to G12. Teams in the same pairs appear in consecutive rows and are determined by CBF. Teams in odd rows play at home in the first round of the tournament.**

Index	Team	City	State	G12?
1	Atlético Goianiense	Goiânia	Goiás	no
2	Goiás	Goiânia	Goiás	no
3	Atlético Mineiro	Belo Horizonte	Minas Gerais	yes
4	Cruzeiro	Belo Horizonte	Minas Gerais	yes
5	Avaí	Florianópolis	Santa Catarina	no
6	Atlético Paranaense	Curitiba	Paraná	no
7	Botafogo	Rio de Janeiro	Rio de Janeiro	yes
8	Fluminense	Rio de Janeiro	Rio de Janeiro	yes
9	Ceará	Fortaleza	Ceará	no
10	Vitória	Salvador	Bahia	no
11	Corinthians	São Paulo	São Paulo	yes
12	São Paulo	São Paulo	São Paulo	yes
13	Flamengo	Rio de Janeiro	Rio de Janeiro	yes
14	Vasco da Gama	Rio de Janeiro	Rio de Janeiro	yes
15	Internacional	Porto Alegre	Rio Grande do Sul	yes
16	Grêmio	Porto Alegre	Rio Grande do Sul	yes
17	Guarani	Campinas	São Paulo	no
18	Grêmio Prudente	Presidente Prudente	São Paulo	no
19	Palmeiras	São Paulo	São Paulo	yes
20	Santos	Santos	São Paulo	yes

Figure 1 **The map shows the locations of the home cities of the 20 participating teams in Series A of the 2010 edition of the Brazilian soccer tournament.**



and tournament attractiveness; others attempt to avoid unfair situations that could benefit one team by scheduling a more convenient sequence of games. These requirements, which have been discussed and established over the years by teams, federations, city administrators, security agencies, and sponsors, can be grouped into the following five classes:

Round-robin constraints

(A.1) Each team must play each other team twice, once at home and once away (double round-robin).

(A.2) Each team must play exactly once in each round, either home or away (compact schedule).

(A.3) Each team must play each other team exactly once in the first (respectively second) phase, along the $n-1$ initial (respectively last) rounds. Games in the second phase are played exactly in the same order as in the first, but with interchanged venues. Consequently, the schedule of the second phase is directly determined by that of the first phase (i.e., a mirrored schedule).

Home-away patterns of game playing

(B.1) Teams in the same pair always have complementary home-away patterns of game playing: whenever one plays at home, the other plays away. When a game is played between two teams from the same home city, only one acts as the home team with home rights.

(B.2) Teams playing at home in the first round are determined beforehand by CBF, depending on historical data (teams that played away in the first round of the previous year's tournament and teams that have been upgraded from Series B should, as much as possible, play at home in the first round).

(B.3) Teams alternate their playing condition (home or away) in every round during the first four rounds of the first phase and during the last two rounds of the second phase. That is, there are no breaks in the first four rounds or in the last round of the tournament.

(B.4) Teams playing at home in the first round necessarily play away in the last, and vice versa.

(B.5) All teams have the same number of home breaks and the same number of away breaks in each of the two phases of the tournament.

(B.6) The number of breaks should be kept as low as possible. In the *Solution Approach* section, we show that feasible schedules with a minimum number of breaks must include one home break and one away break for each team in each of the two phases of the tournament.

(B.7) Not counting the game with the other team in its pair, each team plays nine games at home and nine games away in each phase of the schedule.

Classic and regional games

(C.1) As many classic games as possible should be played in double weekend rounds. Fans and the general public have more free time on weekends to spend on leisure activities, such as attending the games at the stadiums or watching them on TV. This requirement enables larger attendance and audiences for classic games, which are often characterized by historic local rivalries, and will be handled by the objective function of the model presented in the *Integer Programming Model* section.

(C.2) No regional games can be played during the first three rounds (early in the competition they do not draw large audiences because the fans are less motivated) or in the last four rounds (to avoid travel schedules that are advantageous or disadvantageous to one team in a city that hosts multiple teams) of the tournament.

(C.3) Whenever a classic game is played at São Paulo in a particular round, there are no games between G12 teams of Rio de Janeiro and São Paulo in this same round; this avoids competition that could divide the interest of the public.

(C.4) Whenever a classic game is played at Rio de Janeiro in a particular round, there are no games between G12 teams of Rio de Janeiro and São Paulo in this same round; this avoids competition that could divide the interest of the public.

(C.5) There can be at most one classic game played in each city in any round. This constraint is imposed because of the logistics that are necessary to avoid security and circulation problems.

(C.6) No team can play two classic games in consecutive rounds, because they are usually harder games for which the teams request more interplay time. Losing two classic games in a row would strongly affect a team's motivation and the interest of its fans.

(C.7) Some games may have to be scheduled at predefined given rounds, according to the will of teams, sponsors, and organizers.

Geographical and G12 constraints

(D.1) No team whose home city is located in the state of São Paulo can play five or more consecutive games in that state; otherwise, this would be considered as an advantage by the other teams because such cities are geographically close to each other.

(D.2) No team whose home city is located in a state other than São Paulo can play four or more consecutive games in that state; this avoids the advantage of a team staying at home for a long period, while other teams are traveling across the country.

(D.3) Each team must play a game outside the state in which its home city is located in either one of the first two rounds. This requirement is automatically implied by (B.3) and (C.2).

(D.4) There are at least two and at most four games involving only G12 teams in every round; this generates an even distribution of games between strong teams throughout the tournament. It also ensures the existence of multiple games between strong teams in any round, offering more choices for broadcasting by open and paid TV channels.

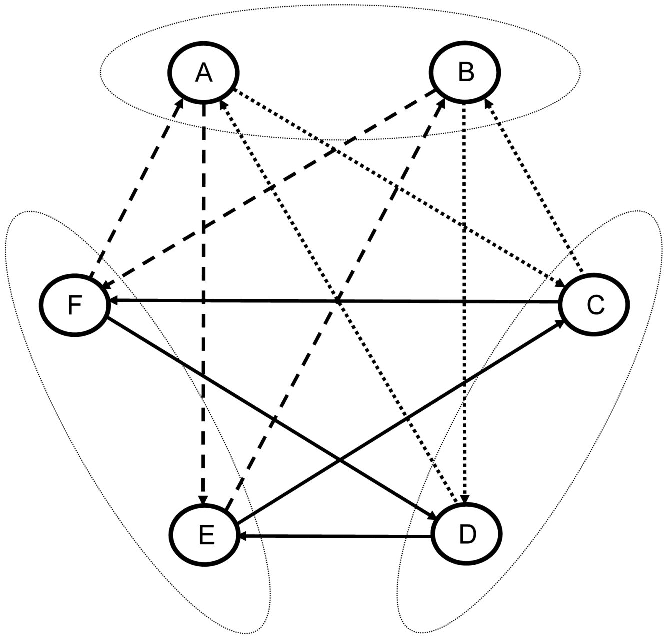
(D.5) No team can play more than five consecutive games against G12 teams, to avoid a long series of hard games.

Perfect matching of paired teams

(E.1) In the quest for a fair and balanced schedule, CBF imposes additional constraints to enforce a tight equilibrium to any two teams belonging to the same pair. Let A and B be two teams belonging to the same pair and C and D two other teams belonging to any other pair. If A plays with C at home (respectively away) in the first phase, then it plays away (respectively at home) with D in this phase. Consequently, B will play away (respectively at home) with C and at home (respectively away) with D in this phase. The same constraints are automatically implied for the second phase, since all its games have interchanged venues with respect to those in the first phase. Such constraints lead to a strong equilibrium between teams in the same cities and regions and are considered by CBF officials as among the most important to be enforced. Figure 2 illustrates a possible perfect matching of six paired teams, which applies to every subset of four teams organized in two pairs.

The maximization of gate attendance and TV audiences is the major issue in the schedule of the Brazilian tournament. Major revenues earned by the teams come from broadcast and merchandising rights paid by the sponsors, who request good schedules that draw large audiences. Fair and balanced schedules for all teams are also a major issue for the attractiveness of the tournament

Figure 2 In this example of a perfect matching of six teams paired into three pairs, an arrow in the arc from team i to team j means that this game is played at the venue of team j.



and for the confidence in its outcome, and plays a major role in the success of the competition. Consequently, determining a good schedule is mainly an issue of finding a solution that maximizes gate attendance and TV audiences, while satisfying requirements that impose intricate patterns of game playing that are hard to compute. To maximize gate attendance and TV audiences, we seek a schedule with a maximum number of classic games played in double weekend rounds. Most of the constraints handle fairness and equilibrium requirements.

Integer Programming Model

The complete schedule of a mirrored double round-robin tournament is fully determined by that of its first phase, because the games in the second phase are played in the same order as those in the first, but with interchanged venues. Therefore, the home-away patterns of the teams playing a mirrored double round-robin tournament can always be seen as formed by two symmetric halves, with the second completely determined by the first. Consequently, an optimal schedule may be obtained by considering only the home-away assignments corresponding to the single round-robin tournament associated with the first phase. The integer programming model and the solution approach proposed in this paper use this property, because all assignments corresponding to games in the second phase are directly implied by those in the first. This strategy automatically handles requirement (A.3).

Let $T = \{1, \ldots, n\}$ be the set of participating teams, with those indexed by odd numbers playing at home in the first round according to the recommendations made by CBF via requirement (B.2). Teams indexed by $2e-1$ and $2e$ belong to the same pair for $e = 1, \ldots, n/2$. Furthermore, let $H = \{1, \ldots, N\}$ be the set representing all feasible home-away patterns for the first phase of the tournament that satisfy constraints (B.3) to (B.6) and start with a home game. For every pattern $\ell = 1, \ldots, N$ and every round $k = 1, \ldots, n-1$, let $h(\ell, k) = 1$ if the team associated with the feasible pattern indexed by ℓ plays at home in round k; let $h(\ell, k) = 0$ otherwise.

We define the following decision variables:

$$x_{ijk} = \begin{cases} 1, & \text{if team } i \in T \text{ plays at home against team } j \in T \text{ in round } k = 1, \ldots, n-1, \\ 0, & \text{otherwise;} \end{cases}$$

and

$$y_{e\ell} = \begin{cases} 1, & \text{if team } 2e-1 \text{ follows pattern } \ell \in H \text{ and team } 2e \text{ follows its complement,} \\ & \text{for every pair } e = 1, \ldots, n/2, \\ 0, & \text{otherwise.} \end{cases}$$

Let C be the set formed by all cities hosting two or more teams and $T(c) \subset T$ be the set of teams whose home city is $c \in C$, with the cities indexed by 1 corresponding to São Paulo and indexed by 2 corresponding to Rio de Janeiro; Belo Horizonte is indexed by 3, Porto Alegre by 4, and Goiânia by 5. We also define SP as the set of teams whose home cities are located in the state of São Paulo. Furthermore, let $\bar{R} \subset \{1, \ldots, n-1\}$ be the set of rounds in which regional games cannot be played according to requirement (C.2) and $D \subset \{1, \ldots, n-1\} \setminus \bar{R}$ be the set of double weekend rounds in which classic games should be played whenever possible. We denote by $G12 \subset \{1, \ldots, n\}$ the set of teams of the G12 that play in the current edition of the tournament.

In particular, $C = \{\text{São Paulo}, \text{Rio de Janeiro}, \text{Belo Horizonte}, \text{Porto Alegre}, \text{Goiânia}\}$, $T(1) = \{11, 12, 19, 20\}$, $T(2) = \{7, 8, 13, 14\}$, $T(3) = \{3, 4\}$, $T(4) = \{15, 16\}$, $T(5) = \{1, 2\}$, $SP = T(1) \cup \{17, 18\}$, $\bar{R} = \{1, 2, 3, 16, 17, 18, 19\}$, $D = \{11, 12, 13, 15\}$, and $G12 = \{3, 4, 7, 8, 11, 12, 13, 14, 15, 16, 19, 20\}$ for the 2010 edition of the tournament (see Table 1). We recall that with the exception of the state of São Paulo, which has several teams in different cities, in all other states all teams are based in their capitals. This is unlikely to change; São Paulo is by far the most powerful and economically developed state; it is the only state that has small cities with enough resources to host a First Division team.

The objective function (1) handled by the model is to maximize the number of classic games that can be played in double weekend rounds, leading to larger gate attendance and TV audiences as determined by requirement (C.1):

$$\max \sum_{k \in D} \sum_{c \in C} \sum_{i \in T(c)} \sum_{\substack{j \in T(c) \\ j \neq i}} (x_{ijk} + x_{jik}). \tag{1}$$

Constraints (2) enforce that every team play against every other team exactly once in each phase of the tournament, corresponding to requirement (A.1):

$$\sum_{k=1}^{n-1} (x_{ijk} + x_{jik}) = 1, \qquad \forall i, j \in T : i < j. \tag{2}$$

Because requirement (B.6) imposes that each team must have one home break and one away break in each phase, teams who play at home in the first round will play 10 home games and 9 away games in the first phase of the tournament. Constraints (3) guarantee that the game between the two teams in each pair will be played at the home of the odd indexed team in the first phase of the tournament (and in the home of the other team in the second phase). Therefore, each team in the pair will play nine games at home and nine games away against the other teams, in accordance with requirement (B.7):

$$\sum_{k=1}^{n-1} x_{2e-1, 2e, k} = 1, \qquad e = 1, \ldots, n/2. \tag{3}$$

Constraints (4) impose requirement (C.2):

$$\sum_{\substack{c\in C\\ c\neq 1}}\sum_{i\in T(c)}\sum_{\substack{j\in T(c)\\ j\neq i}}(x_{ijk}+x_{jik})+\sum_{i\in SP}\sum_{\substack{j\in SP\\ j\neq i}}(x_{ijk}+x_{jik})=0, \qquad \forall k\in \bar{R}. \tag{4}$$

Constraints (5) (respectively (6)) handle requirements (C.3), (C.4), and (C.5), ensuring that there is at most one classic game played in São Paulo (respectively Rio de Janeiro) in any round and that whenever there is one such classic games, there are no games between G12 teams of Rio de Janeiro and São Paulo in the same round:

$$\sum_{i\in T(1)}\sum_{j\in T(2)}(x_{ijk}+x_{jik})+4\cdot\sum_{i\in T(1)}\sum_{\substack{j\in T(1)\\ j\neq i}}(x_{ijk}+x_{jik})\leq 4, \qquad k=1,\ldots,n-1, \tag{5}$$

$$\sum_{i\in T(2)}\sum_{j\in T(1)}(x_{ijk}+x_{jik})+4\cdot\sum_{i\in T(2)}\sum_{\substack{j\in T(2)\\ j\neq i}}(x_{ijk}+x_{jik})\leq 4, \qquad k=1,\ldots,n-1. \tag{6}$$

For any city hosting three or more teams, constraints (7) ensure that there is at least one round in between two classic games involving the same team, as imposed by requirement (C.6):

$$\sum_{\substack{i\in T(c)\\ i\neq j}}(x_{ijk}+x_{jik}+x_{ij,k+1}+x_{ji,k+1})\leq 1, \qquad \forall c\in C:|T(c)|\geq 3, \quad \forall j\in T(c), \quad k=1,\ldots,n-2. \tag{7}$$

Constraints (8) and (9) enforce that any team whose home city is located in the state of São Paulo will play away at least once outside the state in any sequence of five consecutive games. Constraints (8) enforce requirement (D.1) for the first phase, while constraints (9) apply to the second phase:

$$\sum_{i\notin SP}(x_{ijk}+x_{ij,k+1}+x_{ij,k+2}+x_{ij,k+3}+x_{ij,k+4})\geq 1, \qquad \forall j\in SP, \quad k=1,\ldots,n-5, \tag{8}$$

$$\sum_{i\notin SP}(x_{jik}+x_{ji,k+1}+x_{ji,k+2}+x_{ji,k+3}+x_{ji,k+4})\geq 1, \qquad \forall j\in SP, \quad k=1,\ldots,n-5. \tag{9}$$

Similarly, constraints (10) and (11) enforce that any team whose home city is not located in the state of São Paulo will play away at least once outside the state where it is located in any sequence of four consecutive games. Constraints (10) enforce requirement (D.2) for the first phase, while constraints (11) apply to the second phase:

$$\sum_{i\notin T(c)}(x_{ijk}+x_{ij,k+1}+x_{ij,k+2}+x_{ij,k+3})\geq 1, \qquad \forall c\in C\setminus\{1\}, \quad \forall j\in T(c), \quad k=1,\ldots,n-4, \tag{10}$$

$$\sum_{i\notin T(c)}(x_{jik}+x_{ji,k+1}+x_{ji,k+2}+x_{ji,k+3})\geq 1, \qquad \forall c\in C\setminus\{1\}, \quad \forall j\in T(c), \quad k=1,\ldots,n-4. \tag{11}$$

Constraints (12) state that there are at least two and at most four games involving only G12 teams in every round, as determined by requirement (D.4):

$$2\leq \sum_{i\in G12}\sum_{\substack{j\in G12\\ j\neq i}}x_{ijk}\leq 4, \qquad k=1,\ldots,n-1. \tag{12}$$

Constraints (13) enforce requirement (D.5), imposing that no team can play more than five consecutive games against G12 teams:

$$\sum_{\substack{j\notin G12\\ j\neq i}} (x_{ij,k} + x_{ij,f(k+1)} + x_{ij,f(k+2)} + x_{ij,f(k+3)} + x_{ij,f(k+4)} + x_{ij,f(k+5)} + x_{ji,k} + x_{ji,f(k+1)} + \\ + x_{ji,f(k+2)} + x_{ji,f(k+3)} + x_{ji,f(k+4)} + x_{ji,f(k+5)}) \geq 1, \qquad \forall i \in T, \quad k = 1,\dots,n-1, \quad (13)$$

with the function

$$f(a) = \begin{cases} a, & \text{if } a \leq n-1, \\ a-(n-1), & \text{otherwise,} \end{cases}$$

used to handle sequences of consecutive games that start in the first phase and finish in the second.

Constraints (14) and (15) enforce that each odd indexed team follows a different home-away pattern starting with a home game:

$$\sum_{\ell\in H} y_{e\ell} = 1, \qquad e = 1,\dots,n/2, \tag{14}$$

$$\sum_{e=1}^{n/2} y_{e\ell} \leq 1, \qquad \forall \ell \in H. \tag{15}$$

Constraints (16) and (17) determine whether each odd indexed team plays at home or away in each round and imply the satisfaction of requirement (A.2) for such teams:

$$\sum_{\substack{i\in T\\ i\neq 2e-1}} x_{2e-1,ik} = \sum_{\ell\in H} h(\ell,k)\cdot y_{e\ell}, \qquad \forall e = 1,\dots,n/2, \quad k = 1,\dots,n-1, \tag{16}$$

$$\sum_{\substack{i\in T\\ i\neq 2e-1}} x_{i,2e-1,k} = 1 - \sum_{\ell\in H} h(\ell,k)\cdot y_{e\ell}, \qquad \forall e = 1,\dots,n/2, \quad k = 1,\dots,n-1. \tag{17}$$

Similarly, constraints (18) and (19) determine whether each even indexed team plays at home or away in each round and imply the satisfaction of requirement (A.2) for such teams:

$$\sum_{\substack{i\in T\\ i\neq 2e}} x_{2e,ik} = 1 - \sum_{\ell\in H} h(\ell,k)\cdot y_{e\ell}, \qquad \forall e = 1,\dots,n/2, \quad k = 1,\dots,n-1, \tag{18}$$

$$\sum_{\substack{i\in T\\ i\neq 2e}} x_{i,2e,k} = \sum_{\ell\in H} h(\ell,k)\cdot y_{e\ell}, \qquad \forall e = 1,\dots,n/2, \quad k = 1,\dots,n-1. \tag{19}$$

The complementarity of the home-away patterns of teams in the same pair is also ensured by constraints (16) to (19), corresponding to requirement (B.1).

Finally, constraints (20) handle the perfect matching of paired teams imposed by requirement (E.1), by ensuring that any team i will play at home against exactly one of the two teams indexed by $2e-1$ and $2e$, for $e = 1,\dots,n/2$:

$$\sum_{k=1}^{n-1} (x_{i,2e-1,k} + x_{i,2e,k}) = 1, \qquad e = 1,\dots,n/2, \quad \forall i \in T : i \neq 2e-1, i \neq 2e. \tag{20}$$

The integrality constraints (21) and (22) complete the problem formulation:

$$x_{ijk} \in \{0,1\}, \qquad \forall i \in T, \quad \forall j \in T : j \neq i, \quad k = 1,\dots,n-1, \tag{21}$$

$$y_{e\ell} \in \{0,1\}, \qquad e = 1,\dots,n/2, \quad \ell = 1,\dots,N. \tag{22}$$

Solution Approach

The problem has $O(n^3 + n\cdot \mid H \mid)$ variables and $O(n^2 + \mid H \mid)$ constraints. In the *Phase 1: HAP Generation* subsection, we show that the number $\mid H \mid$ of home-away patterns with exactly two breaks is bounded by $O(n^2)$. Therefore, the full model has $O(n^3)$ variables and $O(n^2)$ constraints. The complete integer programming formulation defined by the objective function (1) and constraints (2) to (22), enumerating all N candidate home-away patterns starting with a home game, could not be solved by a commercial solver such as CPLEX 10.2 within an entire day of computations.

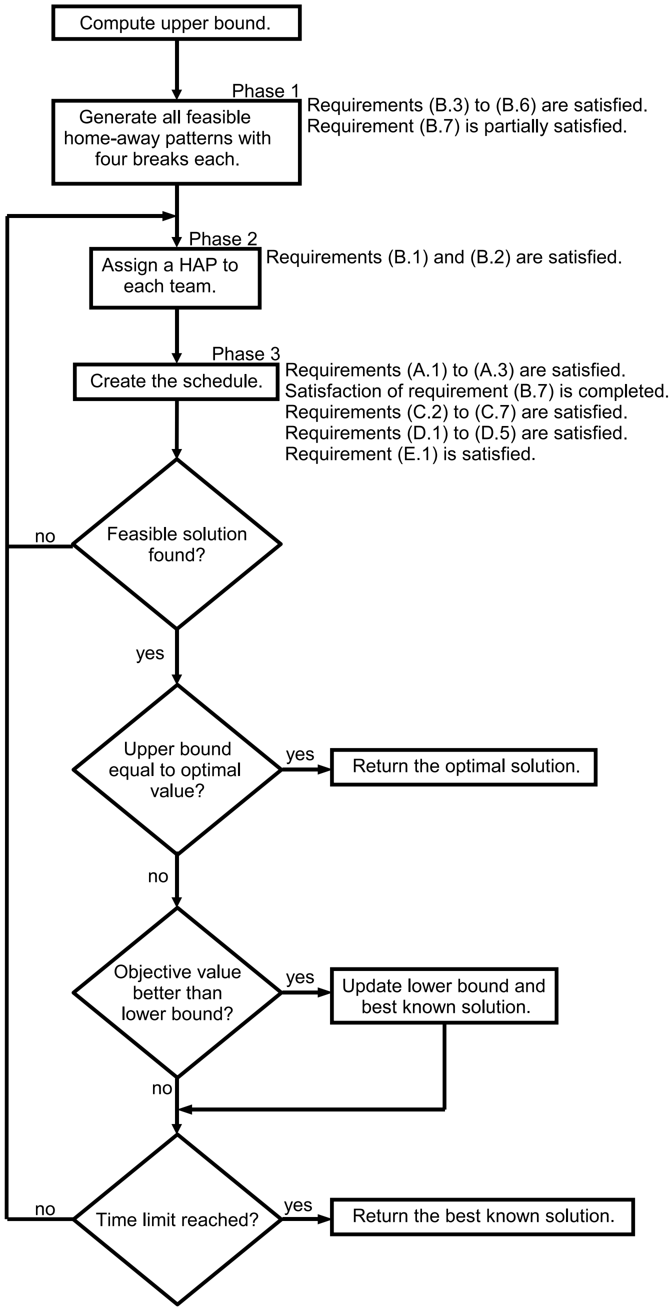
We developed a three-phase solution approach based on a "first-break, then-schedule" decomposition scheme similar to the one that Nemhauser and Trick (1998) proposed to solve the problem of scheduling a basketball league. In the first phase, we create feasible home-away patterns. In the second phase, we assign a different feasible home-away pattern to each team, because distinct teams must have different home-away patterns in every feasible round-robin schedule. Finally, in the third and final phase, we attempt to find an optimal schedule by solving a simpler version of the integer programming model we presented in the *Integer Programming Model* section, derived from the latter by variable fixations. In this simpler model, the binary variables $y_{e\ell}$ no longer exist, because each team has already been assigned to a previously determined home-away pattern. Figure 3 illustrates the approach, whose three phases are described in the subsections below. Similar approaches have been used for other leagues and sports; see Della Croce and Oliveri (2006), Bartsch et al. (2006), and Goossens and Spieksma (2009).

Before performing the first phase of the decomposition solution approach, the algorithm computes an upper bound to the maximum number of classic games that can be played on double weekend rounds. There are three classic games that can be easily played in any double weekend round: Atlético Mineiro vs. Cruzeiro, Grêmio vs. Internacional, and Goiás vs. Atlético Goianiense. Because four teams have São Paulo as their home city, they play six games between them in each phase of the tournament. If six or more double weekend rounds are available for game playing, all such six games may be played in double weekend rounds. Otherwise, the number of games that can be played in double weekend rounds is bounded by the number of double weekend rounds. The same reasoning applies to Rio de Janeiro, which is also the home city of four teams. As an example, we consider the 2010 calendar, in which only four rounds are assigned to weekends in both tournament phases: 11, 12, 13, and 15. Therefore, no more than four of the six classic games to be played in São Paulo can take place in double weekend rounds. However, if three classic games are played in the three consecutive rounds, 11, 12, and 13, in any of these two cities, at least one team will play two consecutive classic games, which is forbidden by requirement (C.6). Therefore, only three of a total of six classic games can be played in São Paulo in double weekend rounds (and, similarly, in Rio de Janeiro). Consequently, at most 9 of 15 classic games can be played in double weekend rounds; this number establishes an upper bound to the value of the optimal solution: three classic games in the three cities, which host two teams,. three in São Paulo and three in Rio de Janeiro.

Phase 1: HAP Generation

Home-away patterns of mirrored schedules may be seen as divided into two symmetric halves. The second half is completely determined by the first. Therefore, we may determine which properties the first half of a HAP must satisfy, so that the entire home-away pattern is feasible. If the number of breaks in the first half of a HAP is even, then the total number of breaks is also even and equal to twice the number of breaks in the first half. However, if the number of breaks in the first half is odd, then the total number of breaks is also odd and equal to twice the number of breaks in the first half plus one break that occurs in the first round of the second half of the tournament. We notice that if there are no breaks, then a team playing at home in the first round will also play at home in every odd round and away in every even round. After a break, this sequence is reversed: this same team will play at home in the following odd rounds and away in the following even rounds.

Figure 3 Our solution approach is based on a three-phase decomposition scheme.



Every additional break reverses this sequence in the same manner. Therefore, if a team playing at home in the first round has an odd number of breaks in the first half of the schedule, then it will play away in round 19 (i.e., the last round of the first half). This team must play away in the first round of the second half because the schedule is mirrored; therefore, it will have a break in the first round of the second half.

HAPs satisfying constraints (B.4) are those with an even number of breaks, because the first round is odd and round 38 (i.e., the last) is even. Therefore, we consider only HAPs with an even number of breaks in the first half. With constraints (B.5) and (B.6), this requirement implies that we must consider only HAPs with the same even number of breaks and with as few breaks as possible. Because every team must have a different HAP in every feasible schedule and there are only two home-away patterns without breaks (i.e., those that perfectly alternate between home and away games), we must consider HAPs with exactly two breaks in each phase of the tournament.

HAPs satisfying constraints (B.3) are those without breaks in the second, third, fourth, and last rounds. Because the schedule is mirrored, they cannot have a break in the last round of the first phase (round $n-1$). Therefore, there are $n-6$ rounds (those numbered from 5 to $n-2$) in which the breaks may occur. To ensure that every team has one home break and one away break, as implied by requirement (B.5), the two rounds in which each team has a break must be both even or both odd. There are $(n-6)/2$ even rounds and $(n-6)/2$ odd rounds in which the teams may have their two breaks; therefore, there are $2 \times \binom{(n-6)/2}{2} = ((n-6)/2-1) \cdot (n-6)/2$ possible break configurations. Moreover, there is exactly one HAP starting with a home game for every possible break configuration, making the number of feasible home-away patterns with exactly two breaks equal to $2 \cdot ((n-6)/2-1) \cdot (n-6)/2 = 84$ for $n = 20$.

The HAPs generated in this phase satisfy requirements (B.3) to (B.6). All home-away patterns generated in this phase have either 9 home games and 10 away games or 10 home games and 9 away games; therefore, requirement (B.7) is partially fulfilled for such assignments. The complete satisfaction of this requirement is enforced by setting the game between teams of the same pair at the venue of the team playing 10 home games in the first phase. The small number of feasible home-away patterns with exactly two breaks each makes it possible to completely enumerate all of them in this phase.

Phase 2: HAP assignments

In this phase, we randomly choose and assign a pair of complementary HAPs to each pair of teams. Therefore, each pair of teams takes a different pair of HAPs. Assigning pairs of complementary home-away patterns to pairs of teams (instead of assigning one home-away pattern to each individual team) makes it possible to automatically satisfy requirements (B.1) and (B.2): the first HAP in the pair is assigned to the odd indexed team starting at home, while its complement is assigned to the even indexed team in the same pair.

Phase 3: Schedule creation

At this point, a home-away pattern has been assigned to each team. In this phase, we build and solve a simplified integer program derived from the formulation presented in the *Integer Programming Model* section by considering the HAPs currently assigned to each team.

Because each team has been already assigned to a home-away pattern in Phase 2, the variables $y_{e\ell}$ are no longer necessary and constraints (14) and (15) no longer exist. Furthermore, the right side of constraints (16) to (19) is set to either 0 or 1, depending on the HAP assigned to each team. This characteristic greatly simplifies the model by trivially setting to zero one half of all x_{ijk} variables. Consequently, this problem can be solved by a commercial solver in reasonable computation times.

This integer program enforces that requirements (A.1) to (A.3), (B.7), (C.1) to (C.7), (D.1) to (D.5), and (E.1) are satisfied. We first assume that it is feasible. If its optimal value is equal to the upper bound on the number of classic games that can be played in double weekend rounds, then the algorithm terminates with an optimal solution. If its optimal value is smaller than this upper bound but improves the best lower bound, then the best-known solution is updated. If the maximum time limit is reached or another stopping criterion is met, then the best-known solution is returned. Otherwise, if the integer program is infeasible or if its optimal value is smaller than the upper bound but the time limit was not reached, the algorithm returns to the second phase to perform new HAP assignments.

To conclude this section, we summarize the input data used by the decomposition solution approach:

• Participating teams: the list of 20 participating teams changes from year to year; the last four teams in Series A in the preceding year are replaced by the best four teams in Series B.

• Paired teams: the pairs of teams are usually formed by those with the same home city and with a long history of rivalry. Cities hosting three or more teams offer different possibilities for grouping them. The pairings defined by CBF may vary from year to year, depending not only on the participating teams in each year, but also on criteria defined by the sponsors and security authorities.

• Teams playing their first games at home: such decisions are made by CBF. Teams upgraded from Series B usually play their first game at home to give their fans a chance to celebrate the ascension.

• Games with predefined dates: such decisions are also made by CBF to advantage some dates. For example, one may be interested in scheduling the game between two teams with the same home city on a major local holiday.

• Dates available for game playing: this information establishes the midweek rounds and the weekend rounds. Double weekend rounds are of major importance, because they may impose a bound on the maximum number of classic games that can be played in weekend rounds. Such dates change from year to year. Only four double weekend rounds were in the tight calendar that CBF imposed for 2010, because of the interruption of the competition during the World Cup in June and July. The restricted dates available for game playing made it hard to find a good schedule and the optimization software was instrumental in creating such a schedule.

Development and Practical Experience

The complete optimization model and the software system coded in C++ have been developed, tuned, and validated over the last three years. CBF and TV Globo staff participated actively in formulating the problem and validating its results. The optimization software, which runs on a standard processor with any efficient integer programming solver, generates solutions in a few minutes of processing time. A database with historical tournament data supports the user to avoid repetitions of schedules and situations observed in previous years, such as playing at home or away in the first round.

The resulting system was first validated with data from the 2005 and 2006 Series A editions of the tournament, which 22 and 20 teams, respectively, played. The schedules obtained by integer programming have been proved to be much better than those manually built and used in practice. The official schedules used in 2005 and 2006 violated some requirements, while the optimized schedules met all constraints. There were 156 breaks in the 2005 schedule and 172 breaks in the 2006 schedule; both numbers were far greater than the minimum number of breaks established by the requirements imposed by CBF. At that time, the main objective was to find schedules in which a maximum number of games between teams from Rio de Janeiro and São Paulo could be broadcast by open TV channels. The new approach led to schedules in which all 56 of the most attractive games could be broadcast by open TV channels without conflicts; the ad hoc rules used for scheduling the 2005 and 2006 tournament editions made it possible to broadcast only 43 and 47 games, respectively.

In 2009, the system was first used as the official scheduler for both the first and second divisions of the Brazilian soccer tournament; each division had 20 participating teams. We provided multiple schedules to the users, who selected their preferred choice. New criteria have been proposed and introduced in the system along the decision process based on successive refinements of the solution, as the decision makers evaluated and filtered the different solutions. The organizers (e.g., CBF, sponsors, and security authorities) checked each proposed schedule and imposed additional constraints (or removed existing constraints) to handle specific situations that might be desired to fine-tune the schedule. The final, complete schedule was announced in January 2009.

The 2009 tournament was the most attractive in recent times, with four teams (Flamengo, Internacional, São Paulo, and Palmeiras) still in contention for the title when the last round started. All games in the last round must start simultaneously. The title changed hands several times during almost two hours of play, as the scores of the 10 games underway changed. The goal that decided the tournament for Flamengo was scored only 20 minutes before the end of the tournament. The champion was not known until the last game ended, contrary to what had happened in previous years when the winners were known up to four rounds before the end of the tournament, making the games of the last rounds very uninteresting. Because of the fight for the title, many attractive games have been played in the last rounds, maintaining high public interest and achieving large gate attendance. This scenario was partly the result of a fair and balanced schedule of games, in which no team had specific advantages or disadvantages. .

The optimization system was used for the second time in 2010. Once again, the decision makers were happy with the alternative schedules the system computed and with their choices. The official schedule obtained by integer programming is available at `http://www.cbf.com.br/php/tabela.php?ct=1&cc=40&aa=2010`. This was a particularly difficult tournament to schedule. It had to be interrupted in June and July during the 2010 World Cup; thus, few dates were available for game playing. As a result, there were more midweek rounds and fewer weekend rounds, making it impossible to schedule all classic games in double weekend rounds. The system sought a schedule with a maximum number of classic games played at double weekend rounds. Furthermore, for the first time in many years, all G12 teams participated in the tournament, augmenting the number of constraints. The increase in the number of teams from the state of São Paulo also made it harder to find balanced schedules, because these teams could benefit from short trips within the state. As in 2009, the title was decided in the last round, with three teams (Fluminense, Corinthians, and Cruzeiro) still in contention for the title when their matches started. The goal that decided the tournament for Fluminense was scored 25 minutes before the end of the tournament.

Soccer is a sport for which it is hard to predict results in advance, especially for long tournaments. Previous attempts to develop schedules that pair the title-contending teams at the end of the tournament have largely failed because of the difficulty in accurately identifying these teams beforehand when the schedules are being developed. Therefore, the proposed approach, which leads to balanced schedules in which all teams face the same difficulties, proved to be the best strategy to maximize the attractiveness of the matches throughout the tournament.

Operations research has proved its usefulness in sports management. Besides the quality of the schedules generated, the main advantages of the integer programming optimization system reported in this paper are its ease of use and its ability to construct multiple schedules, making it possible for the organizers to compare and select the most attractive schedule from among different alternatives, while contemplating other secondary goals: avoiding two rival visiting teams in the same city on the same day to avoid clashes between their fans; ensuring that at least one game is played at every round in large cities; and offering alternative attractive games to other pay-per-view or open channels.

Acknowledgments

The authors are grateful for the support of Virgílio Elísio da Costa Neto, Director of Competitions of CBF, and to Telmo Zanini, from Rede Globo, for his encouragement and assistance in the initial stage of this work. Research of the first author was sponsored by CNPq research grants 301694/2007-9 and 485328/2007-0 and by FAPERJ grant E-26/102.805/2008. We are also grateful to two anonymous referees whose remarks contributed to improving this paper.

References

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

Della Croce, F., D. Oliveri. 2006. Scheduling the Italian football league: An ILP-based approach. *Comp. Oper. Res.* **33**(7) 1963–1974.

Durán, G., M. Guajardo, J. Miranda, D. Sauré, S. Souyris, A. Weintraub, R. Wolf. 2007. Scheduling the Chilean soccer league by integer programming. *Interfaces* **37**(2) 539–552.

Durán, G., M. Guajardo, A. Weintraub, R. Wolf. 2009. O.R. & Soccer: Scheduling the Chilean league using mathematical programming. *OR/MS Today* **36**(2) 42–47.

Fiallos, J., J. Pérez, F. Sabillón, M. Licona. 2010. Scheduling soccer league of Honduras using integer programming. A. Johnson, J. Miller, eds., *Proc. 2010 Indust. Engr. Res. Conf.*. Cancún, Mexico.

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

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

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

Noronha, T. F., C. C. Ribeiro, G. Duran, S. Souyris, A. Weintraub. 2007. A branch-and-cut algorithm for scheduling the highly-constrained Chilean soccer tournament. *Practice and Theory of Automated Timetabling VI*, *Lecture Notes in Computer Science*, vol. 3867. Springer, Berlin, 174–186.

Nurmi, K., D. Goossens, T. Bartsch, F. Bonomo, D. Briskorn, G. Duran, J. Kyngäs, J. Marenco, C. C. Ribeiro, F. Spieksma, S. Urrutia, R. Wolf. 2010. A framework for a highly constrained sports scheduling problem. *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. III. Hong-Kong, 1991–1997.

Rasmussen, R. V. 2008. Scheduling a triple round robin tournament for the best Danish soccer league. *European Journal of Operational Research* **185**(2) 795–810.

Rasmussen, R. V., M. A. Trick. 2008. Round robin scheduling – A survey. *Eur. J. Oper. Res.* **188**(3) 617–636.

Ribeiro, C. C., S. Urrutia. 2007. Scheduling the Brazilian soccer tournament with fairness and broadcast objectives. *Practice and Theory of Automated Timetabling VI*, *Lecture Notes in Computer Science*, vol. 3867. Springer, Berlin, 147–157.