

Scheduling the Chilean Soccer League by Integer Programming

Guillermo Durán, Mario Guajardo, Jaime Miranda, Denis Sauré,
Sebastián Souyris, Andres Weintraub, Rodrigo Wolf

Department of Industrial Engineering, University of Chile, Santiago, Chile
{gduran@dii.uchile.cl, maguajar@dii.uchile.cl, jmiranda@dii.uchile.cl, dsauré@dii.uchile.cl,
ssouyris@dii.uchile.cl, aweintra@dii.uchile.cl, rwolf@dii.uchile.cl}

Since 2005, Chile's professional soccer league has used a game-scheduling system that is based on an integer linear programming model. The Chilean league managers considered several operational, economic, and sporting criteria for the final tournaments' scheduling. Thus, they created a highly constrained problem that had been, in practice, unsolvable using their previous methodology. This led to the adoption of a model that used some techniques that were new in soccer-league sports scheduling. The schedules they generated provided the teams with benefits such as lower costs, higher incomes, and fairer seasons. In addition, the tournaments were more attractive to sports fans. The success of the new scheduling system has completely fulfilled the expectations of the Asociación Nacional de Fútbol Profesional (ANFP), the organization for Chilean professional soccer.

Key words: Chilean soccer league; integer programming; sports scheduling; recreation/sports; OR/MS implementation; scheduling.

History: This paper was refereed.

Soccer is "the passion of multitudes" worldwide. The World Cup 2006, which was held in Germany, amply demonstrated this phenomenon. Beyond the purely sporting and emotional aspects of the game, managing soccer increasingly requires the application of scientific criteria. In Chile, soccer has faced competition from the international leagues, other televised sports, and different forms of entertainment such as shopping malls, cinema, video games, and the Internet. Organizers of professional sports in many countries are facing similar situations.

This competition has caused a reduction in Chileans' interest in soccer and a resulting decline in the revenue the sport generates. To reverse this decline and reduce costs, professional league officials are facing the challenge of having to increase the attractiveness of the league season. The planning of league game schedules is critical to achieving this objective. Considering the many aspects of developing a game calendar that is simultaneously fair to the teams, economically beneficial, and attractive to sports fans, the task of scheduling each regular-season matchup would be nearly impossible if attempted manually.

Beginning with the first tournament of 2005, the Asociación Nacional de Fútbol Profesional (ANFP), the organizing body for Chilean soccer, employed the services of the Centro de Gestión de Operaciones (CGO), a unit of the Industrial Engineering Department at the University of Chile, to assist in the planning of the main league's game schedule. We integrated sporting, operational, and economic criteria within an integer programming model to develop a schedule that meets the ANFP criteria and makes the season more interesting to soccer fans.

Our work falls within the area known as sports scheduling. In this paper, we present the criteria we used to define an efficient season schedule. We address sporting fairness or equity, the introduction of operational and economic considerations into the scheduling process, and how the model and its implementation provide a flexibility that was previously absent in Chilean soccer schedules. In addition to increasing the season's attractiveness, these factors combine to put the scheduling process on a more scientific basis, making it more transparent and more acceptable to team managers.

We organize the paper as follows. We begin with a description of the Chilean soccer tournaments and a review of the sports-scheduling literature. In the *Conditions Imposed on the Problem* section, we explain the conditions considered in the 2006 opening tournament. In *The Mathematical Model and Computational Solution* section, we describe the model and its computational solution. In the *Results* section, we discuss some recent statistics and qualitative factors of our model that have satisfied the ANFP and soccer teams and fans. In the *Conclusions* section, we summarize and provide guidelines for future work. Finally, in the appendix, we show the formulation of the mathematical model.

Background

The Chilean professional soccer league is composed of the First Division and the Second Division. The First Division has 20 teams and divides its annual playing calendar into two tournaments—the opening championship and the closing championship. Each championship comprises two phases: the regular season, which consists of 19 playing dates or rounds, and the playoffs. The teams are organized into four groups of five teams each. Each team must play once against each of the other 19 teams. The ANFP sets the date of each round and the composition of the groups in advance. Following the regular season, the two top teams in each group advance to the playoffs to determine the champion. The Mexican soccer-league system inspired this structure.

The Second Division has 12 teams. Each year the last two teams (measured by the sum of points in both tournaments) of the First Division are relegated to the Second Division; the best two teams of the Second Division are promoted to the First Division.

The country is geographically divided into 12 Regions and the Metropolitan Region (Santiago), which is located between Regions V and VI. We have classified the 20 teams of the First Division into three clusters by geographic location: North with five teams, Center with 10 teams, and South with five teams (Figure 1).

Prior to 2005, the method used to schedule the First Division was a random draw of teams and venues using a preset template; almost all soccer leagues in

South America and Europe use this system. While it facilitated manual scheduling, it did not take into account most of the criteria (e.g., fairness and economic considerations to provide greater revenue and lower costs for the teams) that are needed for efficiency. To ensure fairness, each team should play a balanced mix of home and away games against the strongest teams. Games against the strongest teams should not be scheduled consecutively; each team should play against two of its group opponents at home and against the other two opponents away. Scheduling two consecutive away games (e.g., Sunday–Wednesday or Wednesday–Sunday) for a given team in different opponents' venues, which are located relatively close to each other but far from the team's home venue, is an example of an economic consideration; it would spare the team a second long trip. Other examples include setting "attractive" games for appropriate dates, such as summer home games for teams located in popular beach towns against the most popular teams, and scheduling classic rivalries or matches between teams of the same group in the second half of the tournament when the stakes are higher. In addition, we can distribute weekday home games fairly; such dates are less attractive to the teams because attendance is lower than on weekends (revenue for any specific game goes entirely to the home team).

In previous years, scheduling based on these criteria was woefully deficient. Examples included classic matchups on inappropriate dates, weaker teams scheduled to play all their games against stronger ones away from home, and unbalanced distribution of weekday, home games.

Because a season calendar that meets these standards of efficiency would be nearly impossible to develop manually, operations research can make a substantial contribution by allowing the application of technology to give more flexibility to the season-scheduling process. To help the reader appreciate the scale of the complexity involved, we note that for a tournament in which six teams play a simple round-robin, there are 720 different possible schedules (this does not even consider whether the games are at home or away); for a tournament with eight teams, there are more than 30 million possibilities. Clearly, the number of possible schedules for the 20 soccer

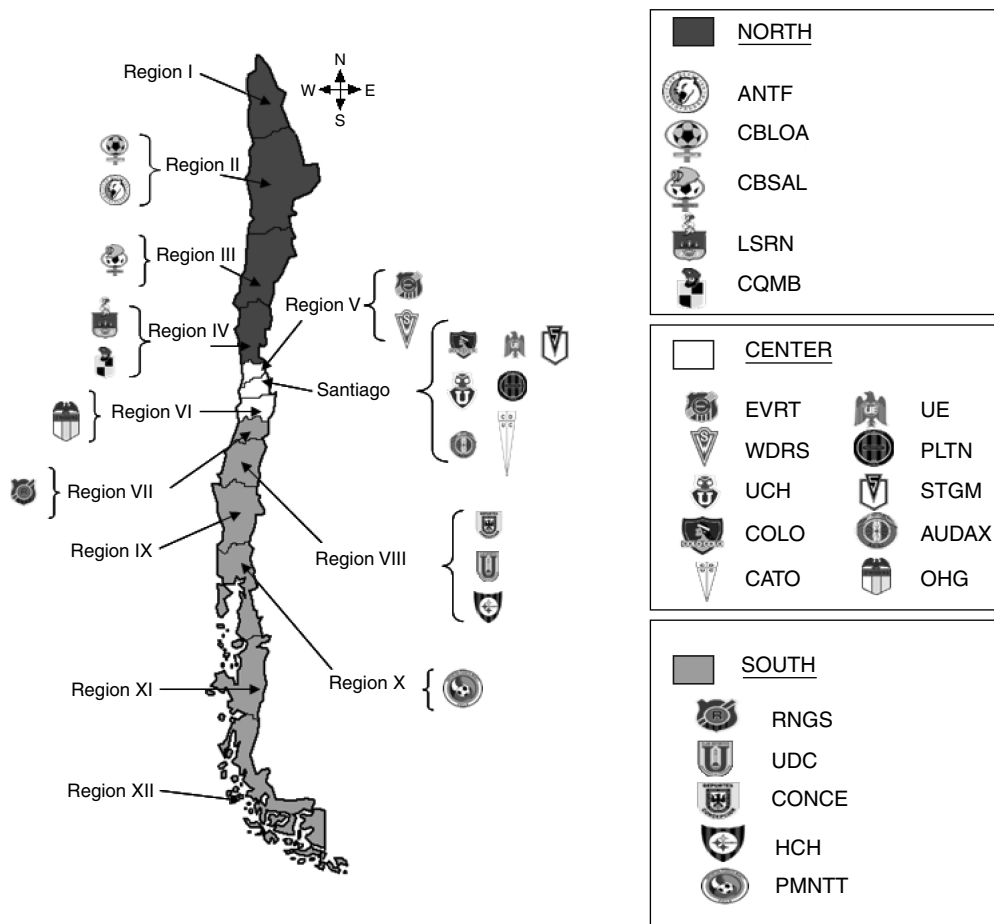


Figure 1: The map of Chile shows the locations of the 20 teams of the First Division in the opening tournament in 2006.

teams in Chile’s First Division would be too enormous to imagine.

The use of sports-scheduling techniques is still a novelty in South America. The only documented case was E. Dubuc’s development of such a method for the Argentinean soccer league in the 1995 season (Paenza 2006), but the practice was subsequently abandoned. Bartsch et al. (2006) have also reported isolated cases in European soccer. In the United States, on the other hand, sports scheduling is routinely employed by most major basketball, baseball, and football leagues that maintain their own teams of academics or hire third-party companies to design efficient regular-season schedules (Bean and Birge 1980, Nemhauser and Trick 1998, Henz 2001, Associated Press 2004).

At the academic level, the literature on sports scheduling has grown significantly in recent years. Various articles have been published proposing as yet untried scheduling applications for existing leagues (Costa 1995, Della Croce and Oliveri 2006, Schreuder 1992). Interest in the problem increased notably with the publication of the traveling tournament problem (TTP) (Easton et al. 2001), which addresses the design of a schedule to minimize the distances that teams in a sports league must travel. Although it is not set up as a real case, the TTP has provided a significant benchmark for using a range of methods and algorithms. Anagnostopoulos et al. (2006) propose a heuristic for solving the TTP based on simulated annealing, while Cardemil and Durán (2004)

present a tabu-search application with the same objective. Easton et al. (2003) offer a combination of integer programming and constraint programming as a method of finding the optimal solution for leagues of up to eight teams. Ribeiro and Urrutia (2006) develop heuristics for the mirrored version of the TTP (double round-robin tournament in which the game order in each single round-robin is the same).

As we will describe later, the goal of the Chilean tournament is not to minimize travel distances precisely, but rather to develop a schedule that satisfies a long and complex list of conditions, which together create a highly constrained problem.

Conditions Imposed on the Problem

The schedule must meet specific conditions to comply with the requirements that the ANFP established for the 2006 opening tournament. Most criteria were also among those applied to the two 2005 tournaments. The appendix shows the mathematical formulations for these conditions.

Basic Schedule Constraints

- (1) Each team must play each of the other teams once during the 19 tournament rounds.
- (2) Each team must play each round either at home or away.
- (3) Each team must play at least nine rounds, but not more than 10, at home.

Home and Away Game Sequence Constraints

- (4) Each team must play at most one sequence of two consecutive rounds at home (home stand). Note that this condition also implies that no team plays more than two consecutive rounds at home.
- (5) Each team must play at most one sequence of two consecutive rounds away (road trip). Note that this condition also implies that no team plays more than two consecutive rounds away.
- (6) Let A be a set of rounds such that if a team plays at home (away) in a round belonging to A , it must play away (at home) in the following round. In the 2006 opening game, the set was defined as $A = \{1, 16, 18\}$. The objective is to balance the teams' home and away games between the early and late stages of the tournament.

Home Game Balance Constraint for Matches Against Group Rivals

- (7) Each team must play against two of its group opponents at home and against the other two opponents away.

Geographic Constraints for Double Away Game Sequences

We have incorporated certain constraints to avoid consecutive long road trips.

- (8) When a North (South) team plays two consecutive away games, neither of these games will be played in the South (North).
- (9) When a North (South) team plays two consecutive away games, at least one of the games will be played in the North (South).
- (10) When a Center team plays two consecutive away games, at least one of the games will be played in the Center.

It is important to note that in the 2006 opening tournament, all the rounds were played on weekends. (We note that it is very expensive for a team to stay away from its city for an entire week.)

Therefore, we could not schedule good trips in consecutive away games Sunday–Wednesday or Wednesday–Sunday.

Constraints on Highly Popular Teams

In our model, we placed constraints on the three most popular teams—Colo Colo, Universidad Católica, and Universidad de Chile. In the appendix, figures, and tables, we refer to these three teams as COLO, CATO, and UCH, respectively.

- (11) If a team plays at home (away) against Colo Colo, it plays away (at home) against Universidad de Chile. This contributes both to fairness and a better balance of revenue between the opening and closing tournaments because games against these teams usually generate higher receipts.
- (12) The three classic matchups between these popular teams must be played between the 10th and 16th rounds; however, parameters may change from tournament to tournament.
- (13) Each of these three teams plays exactly one classic matchup at home.

Operational Constraints on the Availability of Mobile Broadcasting Units for Televising Games

Four games from each round are televised. These include all of the games that involve the three popular teams plus a fourth match. The two opposing teams that have the highest combined point total play the fourth match; if there is a classic game in a given round, two games are selected in this way. Given Chile's large geographical area, it is desirable that these teams not play in venues that are located far apart; this limits travel distances and the associated costs that the television broadcasting company must incur for the transfer of mobile units to and from the games.

(14) When a popular team plays in the North (South), neither of the two other popular teams can play in the South (North).

(15) Given that the first five rounds are scheduled for the middle of the summer (when many events are televised and the availability of mobile units is lower), none of the three popular teams play away in those rounds in outlying areas of the country. We define outlying areas as the members of a set containing the teams whose home venues are located north of Chile's Region IV or south of Region VIII.

Constraints on Strong Teams, Defined as the Three Popular Teams Plus Cobreloa (CBLOA), Which Is a Fourth Strong Team

(16) No team may play two consecutive games against a strong team.

(17) Games between CBLOA and the popular teams are played between the 6th and 18th rounds.

Constraints on Home and Away Games for "Crossed" Teams

We refer to a pair of teams as *crossed* if their home venues are in the same region and if they alternate with each other playing home and away games in each round; they may alternate for operational reasons (e.g., they share the same home stadium), security reasons, or in order not to leave any region without a match for an entire weekend. In total, there were five crossed pairs: Wanderers (WDRS) and Everton (EVRT), both of which are Region V teams, and Colo Colo and Universidad de Chile, which are the most important teams in the country and are located in Santiago, are examples of crossed pairs.

(18) When one team of a crossed pair is playing at home, the other team plays away, and vice versa.

Constraints on Regional Classic Matchups

We define a number of pairs of teams from the same region and with a historic rivalry as a set of *regional classics*.

(19) Regional classic matchups are held between the 8th and 18th rounds.

Constraints on Santiago Games

(20) The number of games held in Santiago in each round cannot be less than two or more than four (there are seven Santiago teams). This enables the soccer activity in the capital to be regulated and ensures the availability of stadiums and municipal security personnel.

(21) It is desirable that the four Santiago teams with the lowest drawing power not play against each other in the first five rounds (all are in the summer) because the attendance would be relatively low. We define a set D that consists of these four teams.

Tourism-Related Constraints

We define a set T that contains teams located in tourist areas, where it is desirable that at least one attractive game (against a popular team) be scheduled for the first rounds of the tournament during the summer.

(22) Each team located in a tourist area plays at home against at least one of the popular teams in one of the first five rounds.

Special Constraints

(23) Some teams do not have their playing fields ready in time for the beginning of the tournament; therefore, they should be scheduled for away games in the first round. These teams form a set S .

(24) Not more than three games between teams of the same group are held in the last round. In 2006, the ANFP made this condition a requirement because in the past tournaments, the playoff qualifiers for most of the groups were defined before the very last round.

(25) Each North (South) team shall play at least once at home against a North (South) team. This constraint will ensure that, in the 2006 closing tournament, North and South teams play against at least one opponent in their respective clusters. This could

help to avoid “bad” two-game road trips. (Note that for Center teams, the same condition is guaranteed by constraints 11 and 13 because the three highly popular teams are from the Center cluster.)

We have also incorporated special constraints in the other tournaments we have scheduled. Examples include not scheduling home games for a team on dates when its stadium is booked for other events and avoiding road trips for teams on dates that are close to dates when they must travel to an international cup game. We also scheduled, in the first round of a tournament, a match between teams that are playing in the playoff finals of the immediately preceding championship. Therefore, this game could be rescheduled at a later date to allow these two teams an additional week of rest. Otherwise, the ANFP would have to reschedule two games.

Because the 2005 tournaments included weekday games, we took into account other factors that relate to fairness or economic considerations. In the 2005 opening, which included only one round played on Wednesday, the ANFP requested that a “good” road trip be scheduled for at least three teams between that round and an immediately adjacent one. The closing had three Wednesday rounds in the schedule; each team was required to play at home on at most two of these rounds (clearly, in these two tournaments, constraints 8, 9, and 10 were not considered for consecutive rounds including weekday games). Finally, for each of the championships, we formulated an objective function that maximized the concentration of “decisive” games among the final rounds of the schedule. In the opening tournaments of 2005 and 2006, decisive games were defined as those between teams in the same group. In the 2005 closing, this definition was broadened to include the games between teams that were expected, based on their performance in the opening tournament, to be fighting against relegation to the Second Division. The appendix shows the detailed formulation of the objective function.

We note the importance of the iterative process (by the group of academic experts and the ANFP) in arriving at a definitive schedule. Three or four different versions of the schedule may bounce back and forth between them as they fine-tune various details to incorporate new constraints into the model before putting a set of final proposals forward. This process

begins one month before the publication of the definitive schedule.

The Mathematical Model and the Computational Solution

We expressed the conditions described in the previous section in terms of an integer linear programming model (appendix). After we built the model, we used CPLEX 9.0 on a Pentium IV computer with a 2.4 GHz processor to solve it.

The problem we face is essentially one of feasibility. Our main goal is to find a schedule that meets all of the conditions that we imposed on it. The objective function simply measures how much we can “push” the decisive games toward the end of the tournament. In our situation, it is not as important that we arrive at an optimal solution as it is in most other optimization problems. The ANFP selects the final schedule from among a series of options that we present to it (following the iterative process discussed above); the choice it makes is not necessarily the one that performs best according to the objective function.

The principal family of decision variables in the model, which we describe in the appendix, takes on a value of one when team i plays at home against team j in round k , and zero otherwise. Other variables express some of the more complex conditions. The model as formulated is extremely hard to solve. To simplify the solution, we incorporated an additional factor that establishes certain home-away patterns for the teams and adds some rigidity to the formulation.

A home-away pattern (Easton et al. 2003, Nemhauser and Trick 1998) is a sequence assigned to a given team that indicates the number of rounds it will play at home and away. A sequence consisting of the 19 elements (H, A, \dots, H) represents a pattern in which H signifies a home-game round and A an away-game round. The appendix shows the constraints that we applied to the assignment of home-away patterns. The advantage of this approach with home-away patterns is a solution-time reduction that is achieved without significant loss of solution quality. Furthermore, it can ensure a priori that certain constraints that normally account for the most serious prolongations of model run time are satisfied. Some of these constraints regard crossed pairs of teams and the avoidance of double home stands or road trips.

For the 2005 opening tournament, we found a feasible solution using constraint programming (CP) after approximately three hours of run time. The model we solved was similar to the one that we show in the appendix; however, it is simply a feasibility problem and is without the objective function. We used ILOG's Solver 5.2 software and CP because of its good performance history in solving similar sports-scheduling problems (Henz 2001).

Once we found this feasible solution, we assigned the set of patterns we obtained to each team. We executed the optimization process and obtained the optimum solution for this pattern assignment in minutes. The schedule approved for this tournament featured 24 of the 40 games between teams of the same group in the last three rounds, much to the ANFP's satisfaction. Note that this solution is the maximum possible number for three rounds because at most eight games between teams of the same group can be played in a given round.

For the 2005 closing tournament, we adopted a similar procedure; however, we partially incorporated the assignment of "feasible" patterns into the optimization process. To keep the solution time to a reasonable level, we allowed the process to assign some patterns and assigned others as fixed. We discovered through experimentation that with up to about 15 fixed patterns and five assigned by the process, we still arrived at the solution quickly; it required no more than 15 minutes, and was better than the solution obtained using the method employed for the opening. Thus, the calendar approved for the 2005 closing scheduled the great majority of games between teams from the same group and between teams that were expected to be fighting one other to avoid relegation to the Second Division in the second half of the tournament. A closing tournament is significantly more constrained because the home condition for a given pair of teams is already defined as the opposite of the one scheduled for the opening tournament (i.e., if team i played at home against team j in the opening, then team j must play at home against team i in the closing).

For the 2006 opening tournament, we also sought a feasible solution as a first step. This involved imposing a set of patterns that we derived from the contents of the definitive solution for the 2005 closing, but

with modifications that would eliminate all double home stands and road trips and impose several different sequences for rounds in which a concentration of intergroup games was desired. Once we imposed these patterns, we found a feasible solution in minutes. We then took the set of patterns in this solution and, rather than assigning them to the teams as is, we completely incorporated them into the optimization process based on the solution of the relaxed problem. We solved the LP problem (i.e., without the constraint of integrality of the variables) and then set the "pattern" variable z_{ip} with the highest value in the LP solution to 1. We repeated this procedure sequentially to arrive at a feasible solution. When the model did not generate a feasible result, we backtracked through the iterations and changed the pattern assignment in a logical way. This heuristic guided us quickly to solutions that were better than our original solution because the LP problem can be solved in seconds. Once the 20 patterns were fixed, we proceeded as we had with the 2005 closing, applying this solution to the model as the initial solution while allowing a few patterns (between two and five) to be reassigned in an attempt to further improve the objective function value. Figure 2 shows the solution that ANFP selected for the 2006 opening tournament.

In this schedule, 100 percent of the games between teams in a given group were set for the 10th round or later rounds, and slightly more than 80 percent were concentrated between the 14th and the 19th rounds—the rounds most highly weighted in the objective function. The value of the feasible solution in the objective function was 607, with a gap of 5.6 percent separating it from the relaxed problem solution (taking into account the last cut that the appendix shows).

Results

It is not easy to measure the impact of using the system we describe because there are many factors other than scheduling that influence variables, e.g., attendance at stadiums. Nevertheless, we can make certain observations. In both 2004 tournaments, the last under the old system, the classic matchup between Universidad de Chile and Colo Colo was held in the first round and drew crowds of 26,000 people in the opening and 22,000 in the closing. By contrast, in

Teams\Match	1	2	3	4	5	6	7	8	9
UCH	UE	@RNGS	UDC	@WDRS	AUDAX	@CBSAL	PMNTT	@ANTF	PLTN
COLO	@OHG	HCH	@LSRN	EVRT	@PLTN	WDRS	@CONCE	CBSAL	@STGM
CATO	@EVRT	PMNTT	UE	@CQMB	ANTF	@CBLOA	@PLTN	LSRN	@OHG
ANTF	@HCH	PLTN	@EVRT	STGM	@CATO	OHG	@LSRN	UCH	@CONCE
CBLOA	@CBSAL	CQMB	@STGM	CONCE	@OHG	CATO	@HCH	RNGS	@WDRS
CBSAL	CBLOA	@LSRN	@OHG	PMNTT	@STGM	UCH	EVRT	@COLO	CQMB
LSRN	@CONCE	CBSAL	COLO	@RNGS	UE	@AUDAX	ANTF	@CATO	HCH
CQMB	UDC	@CBLOA	@PLTN	CATO	@PMNTT	HCH	@WDRS	OHG	@CBSAL
EVRT	CATO	@AUDAX	ANTF	@COLO	@HCH	CONCE	@CBSAL	PLTN	@PMNTT
WDRS	@PLTN	OHG	@CONCE	UCH	RNGS	@COLO	CQMB	@UE	CBLOA
AUDAX	@PMNTT	EVRT	@HCH	OHG	@UCH	LSRN	@STGM	CONCE	@RNGS
UE	@UCH	CONCE	@CATO	HCH	@LSRN	PLTN	@RNGS	WDRS	UDC
PLTN	WDRS	@ANTF	CQMB	@UDC	COLO	@UE	CATO	@EVRT	@UCH
STGM	RNGS	@UDC	CBLOA	@ANTF	CBSAL	@PMNTT	AUDAX	@HCH	COLO
OHG	COLO	@WDRS	CBSAL	@AUDAX	CBLOA	@ANTF	UDC	@CQMB	CATO
RNGS	@STGM	UCH	@PMNTT	LSRN	@WDRS	@UDC	UE	@CBLOA	AUDAX
UDC	@CQMB	STGM	@UCH	PLTN	@CONCE	RNGS	@OHG	PMNTT	@UE
CONCE	LSRN	@UE	WDRS	@CBLOA	UDC	@EVRT	COLO	@AUDAX	ANTF
HCH	ANTF	@COLO	AUDAX	@UE	EVRT	@CQMB	CBLOA	STGM	@LSRN
PMNTT	AUDAX	@CATO	RNGS	@CBSAL	CQMB	STGM	@UCH	@UDC	EVRT

Teams\Match	10	11	12	13	14	15	16	17	18	19
UCH	@HCH	COLO	OHG	@CQMB	STGM	@EVRT	@CATO	LSRN	@CBLOA	CONCE
COLO	RNGS	@UCH	@AUDAX	CATO	@UE	CBLOA	CQMB	@PMNTT	ANTF	@UDC
CATO	CONCE	@STGM	RNGS	@COLO	WDRS	@AUDAX	UCH	@UDC	CBSAL	@HCH
ANTF	WDRS	@RNGS	CBSAL	@UDC	AUDAX	@PMNTT	CBLOA	@CQMB	@COLO	UE
CBLOA	LSRN	@UDC	UE	@PLTN	PMNTT	@COLO	@ANTF	AUDAX	UCH	@EVRT
CBSAL	@UDC	PLTN	@ANTF	HCH	@RNGS	WDRS	@UE	CONCE	@CATO	AUDAX
LSRN	@CBLOA	@OHG	WDRS	@PMNTT	UDC	@STGM	EVRT	@UCH	CQMB	@PLTN
CQMB	STGM	AUDAX	@CONCE	UCH	@EVRT	UE	@COLO	ANTF	@LSRN	RNGS
EVRT	OHG	@UE	UDC	@STGM	CQMB	UCH	@LSRN	WDRS	@RNGS	CBLOA
WDRS	@ANTF	HCH	@LSRN	AUDAX	@CATO	@CBSAL	UDC	@EVRT	STGM	@PMNTT
AUDAX	UE	@CQMB	COLO	@WDRS	@ANTF	CATO	PLTN	@CBLOA	UDC	@CBSAL
UE	@AUDAX	EVRT	@CBLOA	@OHG	COLO	@CQMB	CBSAL	@STGM	PMNTT	@ANTF
PLTN	PMNTT	@CBSAL	STGM	CBLOA	@HCH	CONCE	@AUDAX	RNGS	@OHG	LSRN
STGM	@CQMB	CATO	@PLTN	EVRT	@UCH	LSRN	@CONCE	UE	@WDRS	OHG
OHG	@EVRT	LSRN	@UCH	UE	@CONCE	RNGS	PMNTT	@HCH	PLTN	@STGM
RNGS	@COLO	ANTF	@CATO	CONCE	CBSAL	@OHG	HCH	@PLTN	EVRT	@CQMB
UDC	CBSAL	CBLOA	@EVRT	ANTF	@LSRN	HCH	@WDRS	CATO	@AUDAX	COLO
CONCE	@CATO	@PMNTT	CQMB	@RNGS	OHG	@PLTN	STGM	@CBSAL	HCH	@UCH
HCH	UCH	@WDRS	PMNTT	@CBSAL	PLTN	@UDC	@RNGS	OHG	@CONCE	CATO
PMNTT	@PLTN	CONCE	@HCH	LSRN	@CBLOA	ANTF	@OHG	COLO	@UE	WDRS

Figure 2: The schedule approved for the 2006 opening tournament shows games between teams of the same group in grey. Note that all of them are scheduled for the second half of the championship.

INFORMS holds copyright to this article and distributed this copy as a courtesy to the author(s). Additional information, including rights and permission policies, is available at <http://journals.informs.org/>.

Average attendance and ticket revenue per game			
	2004 opening	2004 closing	2006 opening
Attendance	3,756	3,557	4,953
Ticket revenue	5,852	6,095	11,803

Table 1: Comparisons of average game attendance and ticket revenue (in thousands of Chilean pesos) in the 2004 opening and closing tournaments and the 2006 opening tournament. Figures refer only to regular season games. (Data supplied by ANFP.)

the two 2005 tournaments and the 2006 opening, this matchup was scheduled further into the season and the attendance jumped to 45,000, 37,000, and 49,000, respectively.

Two additional indicators of interest are the attendance and revenue averages, both measured per game, during the regular season (the new scheduling criteria do not affect the playoffs). Table 1 summarizes the data for the last two tournaments using the old scheduling system (i.e., 2004 opening and closing) and the latest tournament played (2006 opening). They reveal that the average rise in attendance was 32 percent when compared to the 2004 opening and 39 percent over that year’s closing; the corresponding average increases for ticket sales were 102 and 94 percent, respectively.

Table 2 compares attendance and ticket revenue for the classic games of the 2004 tournaments (still using the old scheduling method) with those of the 2005 and 2006 tournaments scheduled using the new system.

When the same team was at home, all of the 2005 matchups drew better crowds and generated more revenue than in 2004; total attendance increased 74 percent and receipts increased 142 percent. The 2006 opening classics (once again, with the same team at home) drew 124 percent more than their 2004 equivalents. This translated into a revenue increase of 347 percent.

Table 3 displays comparisons for home games played in summer rounds by teams in Region IV—Coquimbo (CQMB) and La Serena (LSRN)—and Region V—Everton (EVRT) and Wanderers (WDRS)—against the popular teams.

According to the data, attendance grew by 46 percent for Region IV matches between 2004, when they were not played in summer, and 2006, when they

Attendance and ticket revenue for classic games					
	UCH @ COLO		COLO @ UCH		
	2004 op.	2005 op.	2004 cl.	2005 cl.	2006 op.
Date	Feb-08-04	Apr-10-05	Aug-01-04	Aug-28-05	Apr-09-06
Attendance	25,743	45,236	21,750	37,420	48,996
Ticket revenue	55,900	114,879	59,967	137,394	240,557
	CATO @ UCH		UCH @ CATO		
	2004 cl.	2005 op.	2004 op.	2005 cl.	2006 op.
Date	Nov-04-04	Apr-30-05	May-04-04	Nov-13-05	May-14-06
Attendance	18,093	24,450	7,881	18,292	14,409
Ticket revenue	55,173	71,499	12,241	69,099	67,466
	COLO @ CATO		CATO @ COLO		
	2004 op.	2005 op.	2004 cl.	2005 cl.	2006 op.
Date	Mar-06-04	Mar-20-05	Sep-12-04	Sep-25-05	Apr-23-06
Attendance	9,887	24,352	13,333	18,138	32,654
Ticket revenue	16,575	100,408	29,595	61,074	147,442

Table 2: Comparisons of attendance and ticket revenue (in thousands of Chilean pesos) for classic games in 2004, 2005, and 2006. Figures refer to matches in which the home team was the same. (Data supplied by ANFP.)

were. The improvement in Region V was particularly impressive at 156 percent. In both cases, revenue also increased significantly—by 84 and 313 percent, respectively.

Attendance and ticket revenue for games in tourist areas				
Region IV	COLO @ LSRN		CATO @ CQMB	
	2004	2006 op.	2004	2006 op.
Date	May-09-04	Feb-11-06	Mar-13-04	Feb-18-06
Attendance	5,373	7,533	4,673	7,178
Ticket revenue	10,175	22,101	10,681	16,175
Region V	CATO @ EVRT		UCH @ WDRS	
	2004	2006 op.	2004	2006 op.
Date	Apr-21-04	Jan-29-06	Oct-29-04	Feb-19-06
Attendance	3,314	6,638	3,494	10,787
Ticket revenue	3,840	13,857	6,563	29,091

Table 3: Comparisons of attendance and ticket revenue (in thousands of Chilean pesos) for games held in tourist areas in 2004, 2005, and 2006. Figures refer to matches in which the home team was the same in both years. (Data supplied by ANFP.)

Furthermore, the adoption of criteria that reduce the costs of broadcasting games has been highly celebrated by the television-station management and the ANFP, which has the 80 percent share of the soccer television company. For example, two televised games played in the north may result in savings for the television company of approximately US\$20,000 over the cost of broadcasting one game held in the north and the other in the south. In the future, it may be possible to pass along some of these savings to the teams. The income derived from television rights is an important source of revenue for the Chilean teams. In the last negotiation, the league sold its television rights for approximately US\$3,700,000. Forty percent of this amount was assigned to the three most popular teams.

Although we may partly explain the quantitative information in the preceding paragraphs by the new schedules, it is difficult to control for some exogenous factors that are present. These factors, which include the performance of the national team in World Cup qualifying rounds, weather patterns, acts of violence committed by team supporters, and the quality of the teams (fundamentally those that draw larger crowds) may distort the measurements.

The positive impact of the sporting-fairness criteria that we adopted in our model deserves special mention. Both the ANFP and the league teams have expressed their satisfaction with how these criteria have been put into practice. Senior ANFP officials also noted that, for the first time in many years, they have received almost no complaints from the teams regarding the schedules.

Conclusions

To improve the solution process, the procedures we described above for finding good solutions raise certain issues that remain to be explored in future work. These include:

- Formalization of the solution to the integer problem starting from the relaxed version via a heuristic that sequentially structures the fixing of the z_{ip} pattern variables at one, thus detecting and repairing infeasibilities.
- Incorporation of the creation of patterns in the optimization process or use of a set of patterns of

higher cardinality than the number of teams. This could lead to better solutions and greater flexibility in the search for feasible solutions; however, the price will likely be increased solution times.

- Experimentation to determine how much the cuts shown in the appendix contribute in terms of reducing solution time (recall that as with every integer programming problem, the use of cuts to adjust the feasible polyhedron of the linear relaxation may be very useful).

The incorporation of these modern techniques into the Chilean soccer-league scheduling process since 2005 has provided an excellent opportunity to demonstrate that the use of operations research can be effective in making soccer-season schedules more attractive to the public as well as fairer and more profitable for the teams and organizing bodies. It is also important to highlight the transparency that the model we discussed has brought to the system. Once the constraints to be applied have been defined and made known to all concerned, they are incorporated as part of the mathematical model. Then, we can generate some possible solutions and allow the ANFP to select the definitive solution. This procedure also requires that the ANFP submit its objectives for the schedule to the teams; this facilitates consensus building and the creation of new mechanisms for improving the league's scheduling process.

Thus far, we have scheduled the last five tournaments of the First Division (2005 opening, 2005 closing, 2006 opening, 2006 closing, and 2007 opening), the annual 2007 tournament of the Second Division, and the Chilean proposal for the scheduling of South American qualifiers to the South Africa 2010 World Cup.

There are several anecdotes that clearly illustrate the impact this scheduling application has had in Chilean soccer circles; one is particularly revealing. The day that the 2005 closing-tournament schedule was publicized, the ANFP's operational manager noted that he had received a complaint from the president of the Coquimbo team regarding its home game against Colo Colo; the game had been scheduled on the "Fiesta de la Pampilla," a local holiday. This meant that the necessary police presence for the match would not be available and that the stadium attendance would be significantly reduced. Within

a few hours, the game was switched to a different round. In addition, because the overall schedule had already been published, the number of other changes was minimized so that the modified version was as similar as possible to the original schedule. The ability to rely on a rapid and agile tool that would enable us to solve the problem in minimal time was crucial to the task.

Participants have also expressed other significant reactions to us. A player for Palestino, a Santiago team, commented that he was surprised by the “chance occurrence” that his team was scheduled to play against its four group rivals in the last five rounds of the 2005 opening, thus leading to more significant games. During the 2006 opening, a television company executive remarked on the “luck” of televising games involving Universidad de Chile on a Saturday and Universidad Católica the next day while the two teams were on road trips in the north. This saved his company a significant amount of money because of the proximity of the two venues and the correspondingly low mobile-unit transfer costs.

In conclusion, we observe that management techniques can make other contributions to South America’s most popular sport. Issues such as resource management, management of the creation of lower divisions, new tournament formats, optimal ticket prices, strategic alliances with other countries in the region, policies for encouraging the return of top players currently playing abroad, and the efficient administration of the economic and operational aspects of teams and related organizations, are some of the areas that could benefit from a more scientific and quantitative approach.

Appendix: Formulation of the Mathematical Model

The integer linear programming model used to generate the 2006 Opening Tournament is described below.

Variables

To define the games to be held in each round, we define $\forall i \neq j \in I$ (the set of teams) and $\forall k \in K$ (the set

of rounds), a family of binary variables, as follows:

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

To represent simply certain home and away game sequence constraints, we define $\forall i \in I$ and $\forall k = 1, \dots, 18$, the following auxiliary variables, also binary:

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

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

A total of more than 7,900 variables are included.

Objective Function

To conform with ANFP requirements, our objective function maximized the concentration of games between teams in the same group toward the final rounds of the tournament.

The experience of previous tournaments suggests it is not advisable to overload the last round because by that time the playoff qualifying teams tend to already be determined. More specifically, it was requested that the majority of games involving teams of the same group are concentrated between the 14th and 19th rounds, that not more than three of these games be scheduled for the last round, and that, to the extent possible, none of them be played before the 10th round.

With the foregoing in mind, a game between teams of the same group played on the final round was assigned a weight of 15, while all other games held after the 9th round were assigned a weight equal to the round number. Thus, the objective function used in the model was the following:

$$\max \left\{ \sum_{10 \leq k \leq 18} \sum_e \sum_{i \in \mathcal{I}(e)} \sum_{j \in \mathcal{I}(e)} k \cdot x_{ijk} + \sum_e \sum_{i \in \mathcal{I}(e)} \sum_{j \in \mathcal{I}(e)} 15 \cdot x_{ij19} \right\},$$

where

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

In the 2005 opening tournament, the objective function incorporated a weighted sum of the round numbers for the games between teams of the same group. In the 2005 closing tournament, games between relegation rivals were also weighted (a set of six candidates was deduced by the performance of the teams in the opening tournament).

Constraints

The formulation of the constraints is given below. They are numbered to match the numbering of the constraints described in §3.

$$\sum_k [x_{ijk} + x_{jik}] = 1 \quad \forall i, j \in I, \quad (1)$$

$$\sum_j [x_{ijk} + x_{jik}] = 1 \quad \forall i \in I, k \in K, \quad (2)$$

$$9 \leq \sum_j \sum_k x_{ijk} \leq 10 \quad \forall i \in I, \quad (3)$$

$$\sum_{k < 19} y_{ik} \leq 1 \quad \forall i \in I. \quad (4)$$

To this must be added the constraint relating the y variables to the x variables:

$$\sum_j [x_{ijk} + x_{ij(k+1)}] \leq 1 + y_{ik} \quad \forall i \in I, k < 19, \quad (5)$$

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

To this must be added the constraint relating the w variables to the x variables:

$$\sum_j [x_{jik} + x_{ji(k+1)}] \leq 1 + w_{ik} \quad \forall i \in I, k < 19, \quad (6)$$

$$\sum_j [x_{ijk} + x_{ij(k+1)}] = 1 \quad \forall i \in I, k \in A, \quad (7)$$

$$\sum_k \sum_{j \in t(e)} x_{ijk} = 2 \quad \forall e \in E, i \in t(e), \quad (8)$$

$$\sum_{i \in \text{South}} [x_{ijk} + x_{ij(k+1)}] \leq 1 - w_{jk} \quad \forall j \in \text{North}, k < 19, \quad (9)$$

$$\sum_{i \in \text{North}} [x_{ijk} + x_{ij(k+1)}] \leq 1 - w_{jk} \quad \forall j \in \text{South}, k < 19, \quad (10)$$

$$w_{ik} \leq \sum_{j \in \text{North}} [x_{jik} + x_{ji(k+1)}] \quad \forall i \in \text{North}, k < 19, \quad (9)$$

$$w_{ik} \leq \sum_{j \in \text{South}} [x_{jik} + x_{ji(k+1)}] \quad \forall i \in \text{South}, k < 19, \quad (10)$$

$$w_{ik} \leq \sum_{j \in \text{Center}} [x_{jik} + x_{ji(k+1)}] \quad \forall i \in \text{Center}, k < 19, \quad (10)$$

$$\sum_k [x_{hik} + x_{hjk}] = 1 \quad \forall h \neq i, h \neq j, i = \text{COLO}, j = \text{UCH}, \quad (11)$$

$$\sum_{i, j \in \text{PopularTeams}} \sum_{(10 > k \vee k > 16)} x_{ijk} = 0, \quad (12)$$

$$\sum_k [x_{hik} + x_{jik}] = \sum_k [x_{hjk} + x_{ijk}], \quad h = \text{CATO}, i = \text{COLO}, j = \text{UCH}, \quad (13)$$

$$\sum_{i \in \text{North}} x_{ijk} \leq 1 - \sum_{i \in \text{South}} x_{ihk} \quad \forall j, h \in \text{PopularTeams}, j \neq h, k \in K, \quad (14)$$

$$\sum_{k \leq 5} \sum_{j \in \text{OutlyingTeams}} x_{jik} = 0 \quad \forall i \in \text{PopularTeams}, \quad (15)$$

$$\sum_{j \in \text{StrongTeams}} [x_{ijk} + x_{jik} + x_{ij(k+1)} + x_{ji(k+1)}] \leq 1 \quad \forall i \in I, k < 19, \quad (16)$$

$$\sum_{j \in \text{PopularTeams}} \sum_{(6 > k \vee k > 18)} [x_{ijk} + x_{jik}] = 0, \quad i = \text{CBLOA}, \quad (17)$$

$$\sum_h [x_{ihk} + x_{jhk}] = 1 \quad \forall (i, j) \in \text{CrossedTeams}, k \in K, \quad (18)$$

$$\sum_{(8 > k \vee k > 18)} [x_{ijk} + x_{jik}] = 0 \quad \forall (i, j) \in \text{RegionalClassics}, \quad (19)$$

$$2 \leq \sum_{i \in \text{Santiago}} \sum_j x_{ijk} \leq 4 \quad \forall k \in K, \quad (20)$$

$$\sum_{6 > k} x_{ijk} + x_{jik} = 0 \quad \forall i, j \in D, \quad (21)$$

$$\sum_{j \in \text{PopularTeams}} \sum_{k \leq 5} x_{ijk} \geq 1 \quad \forall i \in \text{TouristTeams}, \quad (22)$$

$$\sum_i x_{ijk} = 1 \quad \forall j \in S, k = 1, \quad (23)$$

$$\sum_e \sum_{i \in t(e)} \sum_{j \in t(e)} x_{ijk} \leq 3, \quad k = 19, \quad (24)$$

$$\sum_k \sum_{j \in \text{North}} x_{ijk} \geq 1 \quad \forall i \in \text{North}, \quad (25)$$

$$\sum_k \sum_{j \in \text{South}} x_{ijk} \geq 1 \quad \forall i \in \text{South}.$$

In total, the model considers around 3,000 constraints.

Now let us examine certain additional aspects of formulating the constraints on patterns. We shall call P a set of 20 patterns and $\text{Home}(k)$ the subset of the set of patterns that assigns home games in round k . Consider the following variables:

$$z_{ip} = \begin{cases} 1 & \text{if pattern } p \text{ is assigned to team } i, \\ 0 & \text{otherwise.} \end{cases}$$

The constraints on the patterns are as follows:
 Exactly one pattern is assigned to each team:

$$\sum_p z_{ip} = 1 \quad \forall i \in I. \quad (26)$$

Exactly one team is assigned to each pattern. This constraint is useful to strengthen the relaxation of the integer problem and holds only if the number of patterns in P is equal to 20. Alternatively, one might consider more patterns in this set. This might lead to better solutions but at the expense of higher running times:

$$\sum_i z_{ip} = 1 \quad \forall p \in P. \quad (27)$$

A team plays at home in a given round if the assigned pattern so indicates; otherwise, it plays away (Associated Press 2004):

$$\sum_j x_{ijk} = \sum_{p \in \text{Home}(k)} z_{ip} \quad \forall i \in I, k \in K. \quad (28)$$

Given the complexity of the solution, we assume that it may be useful to incorporate cuts that help reduce the size of the feasible domain of the relaxed problem.

Note that because each group contains five teams, there can be at most two games between teams of the same group for each group and round. We therefore add the following cut:

$$\sum_{i \in t(e)} \sum_{j \in t(e)} x_{jik} \leq 2 \quad \forall e \in E, k \geq 10. \quad (29)$$

Finally, we add the following upper bound for the objective function. Note that this condition is implied by constraints (1), (24), and (29), but its inclusion showed to be efficient in our experiments.

Given that the total number of games between teams of the same group is 40, the value attainable by the objective function is upper bounded by 643:

$$\begin{aligned} & \sum_{10 \leq k \leq 18} \sum_e \sum_{i \in t(e)} \sum_{j \in t(e)} k \cdot x_{ijk} + \sum_e \sum_{i \in t(e)} \sum_{j \in t(e)} 15 \cdot x_{ij19} \\ & \leq 3 \cdot 15 + 8 \cdot 18 + 8 \cdot 17 + 8 \cdot 16 + 8 \cdot 15 + 5 \cdot 14 = 643. \end{aligned} \quad (30)$$

Acknowledgments

We thank the anonymous referees for the valuable suggestions that improved this work. We also thank the Asociación Nacional de Fútbol Profesional de Chile, and in particular Alejandro Carmash and Felipe Chaigneau, the ANFP's operational manager and executive secretary, respectively, during the period this project was carried out. This project would not have been possible without their help. Finally, we are grateful to the Complex Engineering Systems unit of the Millennium Sciences Institute for financial support in the completion of this work. FONDECYT grant number 1050747, Conicyt, Chile, and UBACyT X184, Universidad de Buenos Aires, Argentina partially financed the first author. The first and sixth authors received partial funding from PROSUL 490333/2004—4, CNPq, Brazil.

References

- Anagnostopoulos, A., L. Michel, P. Van Hentenryck, Y. Vergados. 2006. A simulated annealing approach to the traveling tournament problem. *J. Scheduling* 9 177–193.
- Associated Press. 2004. Husband-wife team out bid after 24 years. Retrieved July 23, 2007 <http://sports.espn.go.com/mlb/news/story?id=1936328&CMP=OTC-DT9705204233>.
- Bartsch, T., A. Drexler, S. Kröger. 2006. Scheduling the professional soccer leagues of Austria and Germany. *Comput. Oper. Res.* 33(7) 1907–1937.
- Bean, J., J. Birge. 1980. Reducing traveling costs and player fatigue in the National Basketball Association. *Data Mining Knowledge Discovery* 10 98–102.
- Cardemil, A., G. Durán. 2004. Un algoritmo tabú search para el traveling tournament problem (in Spanish). *Revista Ingeniería de Sistemas* 18(1) 95–115.
- Costa, D. 1995. An evolutionary tabu search algorithm and the NHL scheduling problem. *INFOR* 33 161–178.
- Della Croce, F., D. Oliveri. 2006. Scheduling the Italian football league: An ILP-based approach. *Comput. Oper. Res.* 33(7) 1963–1974.
- Easton, K., G. Nemhauser, M. Trick. 2001. The traveling tournament problem: Description and benchmarks. *Proc. 7th Internat. Conf. Principles and Practice of Constraint Programming*. Paphos, Cyprus, 580–584.
- Easton, K., G. Nemhauser, M. Trick. 2003. Solving the traveling tournament problem: A combined integer programming

- and constraint programming approach. E. Burke, P. De Causmaecker, eds. *Lecture Notes in Computer Science*, Vol. 2740. Springer, Berlin/Heidelberg, Germany, 100–109.
- Henz, M. 2001. Scheduling a major college basketball conference—Revisited. *Oper. Res.* **49** 163–168.
- Nemhauser, G., M. Trick. 1998. Scheduling a major college basketball conference. *Oper. Res.* **46**(1) 1–8.
- Paenza, A. 2006. *Matemática... Estas Ahí? Episodio 2* (in Spanish). Siglo XXI, Buenos Aires, Argentina, 194–206.
- Ribeiro, C. C., S. Urrutia. 2006. Heuristics for the mirrored traveling tournament problem. *Eur. J. Oper. Res.* **179**(3) 775–787.
- Schreuder, J. 1992. Combinatorial aspects of construction of competition Dutch professional football leagues. *Discrete Appl. Math.* **35** 301–312.