SCHEDULING THE BRAZILIAN FOOTBALL LEAGUE MINIMIZING EXTENDED CARRY-OVER EFFECTS ASSOCIATED TO STRENGTH GROUPS

Marcos Costa Roboredo[1]
Luiz Aizemberg[2]
Artur Alves Pessoa[3]
João Carlos Correia Baptista Soares de Mello[4]

Abstract: The Brazilian football league is a classical mirrored double round robin tournament with an even number of teams and place constraints. That league is composed of divisions (series) where the two strongest are A and B, in this order. Currently, the Serie A top four teams and the Brazilian Cup winner guarantee together the access to the important Libertadores Cup while the Serie B top four teams guarantee the access to Serie A. We consider a type of carry-over effect that occurs in the schedule when a team meets two teams from either a strong or a weak group in two consecutive rounds. A break occurs when a team plays at home (away) in two consecutive rounds. In this paper, we consider a scheduling problem that limits the number of breaks and minimizes the total number of the effects. We show that previously proposed techniques can be extended to solve this variation. In addition, we use a hypothesis test to provide an evidence that teams from Serie B in last season with access to Serie A at current season form a weak group while teams with access to Libertadores Cup and playing Serie A at the current season form a strong group.

Keywords: Carry-over effects, Mirrored double round robin tournament, Brazilian football league.

[1]Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, marcos.roboredo@ime.uerj.br
[2]Departamento de Engenharia de Produção, Universidade Federal Fluminense, luizaizemberg@gmail.com
[3] Departamento de Engenharia de Produção, Universidade Federal Fluminense, artur@producao.uff.br
[4] Departamento de Engenharia de Produção, Universidade Federal Fluminense, jcsmello@pq.cnpq.br

1. INTRODUCTION

In a single round robin tournament with an even number n of teams, each team meets each other exactly once. The tournament is divided in n-1 rounds, where each team plays once per round home or away. A mirrored double round robin tournament is composed of two single round robin tournaments where the second one is similar to the first one with just one difference: the home field of each match is swapped. Let us introduce the following classical notation on Round Robin tournaments. We define a pattern as a 0-1 sequence indicating if a team plays home (1) or away (0). A team has a break in the round r+1 if it plays home (away) in both rounds r and r+1. For example, the pattern 0 1 0 0 1 0 1 1 of a team A indicates that it plays home (away) in the rounds 2,5,7 and 8 (1,3,4 and 6) and it has two breaks, in the round 4 and 8. A pattern set is composedof n patterns.

In the literature, we can find many researches related to scheduling double round robin tournaments. Most of them solve the problem by using the procedure proposed by Nemhauserand Trick (1998) which consists of scheduling a single round robin tournament by decomposing the problem in three phases: the first one generates a large number of pattern sets, the second one tries to construct a feasible timetable with them and the last one assign teams to patterns ensuring all the place constraints. Next, the timetable is mirrored.

The schedules of football tournaments depends on global and place constraints. By global we mean constraints that are common in football tournaments. Some global constraints are presented by Henz et al. (2004) and Nurmi et al. (2010). The place constraints for the Brazilian football league are showed by Ribeiro and Urrutia (2011), while for the football leagues of Austria and Germany, Chile, Denmark and Netherlands, see, respectively, Bartsch et al. (2006), Dur´anet al. (2007),Rasmussen (2008) and Schreuder (1992).

To obtain good final schedules, an important issue is about the schedule fairness. To construct fair calendars, the main objective is to avoid breaks (Miyashiro et al.,2003). Besides, others important goals have been proposed. Della Croce and Oliveri (2006) regarded a group of seeded teams where games between teams belonging to this group do not occur in the first (last) three rounds. Ribeiro and Urrutia (2006) considers a group of elite teams where a team does not match teams from this group for five consecutive rounds. Russel(1980) defined carry-over effects and proposed to balance them.

A carry-over effect was defined by Russel (1980) in the following way. If a team A played team B in its previous match of a round robin tournament and is now playing team C, team C is said to receive a carry-over effect due team B. Briskorn (2009) and Briskorn and Knust (2010) considered partitions of strength groups of teams, showing whether there is a schedule without carry-over effects for several different numbers of teams and strength groups. Goossens and Spieksma (2012) studied the carry-over effects in whole Europe.
The Belgian football league was the first to balance carry-over effects according to Goossens and Spieksma(2009).

In this paper, we focus on Brazilian football league which is a classical mirrored double round robin (MDRR) tournament with some place constraints. We propose to schedule such league avoiding a team to play against two strong or two weak teams consecutively. Some reasons to avoid these situations are the following:

i. When a team A loses (wins) a match and the next adversary is strong (weak), team A has more chance of losing (winning) the second match.

ii. When a team wins (loses) two consecutively games, it carries a psychological advantage (disadvantage)to the next matches.

iii. In Brazil, it is usual to fire the coach when a team loses a few consecutive matches. Besides, less supporters will attend the stadium, decreasing the income.

We call it Extended Carry-Over (EC) effects. Although we are not considering place constraints of the Brazilian football league, they can be easily incorporated in the model proposed in this paper. In the literature,several papers studied the schedule of that league (Ribeiro and Urrutia (2006); Ribeiro and Urrutia (2011); Kendall et al. (2010)). Besides, it is possible to find papers about results prediction (Alves et al., 2012) and analysis (Sant'Anna et al., 2010).

The Brazilian football league is composed of divisions (series) where the two strongest are A and B, in this order. Currently, the Serie A top four teams and the Brazilian Cup winner guarantee together the access to the Libertadores Cup (most important South American football tournament) while the Serie B top four teams guarantee the access to Serie A. Hence, two intuitive hypotheses were proposed. Hypothesis one (H1) says teams from Serie B in last season with access to Serie A at current season have a great chance of finishing the Brazilian Championship at the bottom half of the classification table. Hypothesis two (H2) says teams with access to Libertadores Cup and playing Serie A at the current season have a great chance of finishing the Brazilian Championship at the top half of the classification table.

For both cases, the null hypothesis (H0) is that the teams position in the final classification is independent of whether they played the Serie B last season or they have access to the Libertadores Cup. First, we provide evidences that H1 and H2 are true by rejecting H0 based on the number of teams that satisfy the conditions ofH1 and H2 in 19 past seasons. We believe that this can be used as an argument in favor of using constraints associated to the EC effects. Next, we give a schedule for the Serie A of the Brazilian football league by applying an integer programming approach similar to the proposed by Della Croce and Oliveri (2006) where, in order to obtain a fair final schedule, we ensure the following requirements: the final total number of effects related to strong (weak) teams group is minimal, each team has at most four breaks and there are no consecutive breaks.

In our model, we give preference to minimize the effects over breaks. An example of the EC effect occurred in 2010 when Vasco da Gama team matched consecutively Cruzeiro and Corinthians away and lost the games. Cruzeiro and Corinthians are traditional teams in Brazil and made heavy investments to play the Libertadores Cup in 2010.

Other issues about the Brazilian Soccer League have also been studied. Sant'anna et al (2010) analyzed alternative rules for the classification of the teams in the Brazilian Soccer Championship, ranking them by probabilistic criteria composition. Sant'anna (2008) did a similar analysis, but using Rough Sets Theory. Lever et al (1983) studied the history and structure of Brazilian soccer, and the place of soccer in the cultural lives of supporters.

Some papers studied soccer in other countries, by the similarity of the rules makes possible to use these papers to better understand the Brazilian Soccer League. Palacios (2004) and Bloyce and Murphy (2008) studied the effects of the "three points per win" rule. The importance of "home advantage" has been stressedin a large set of studies, Courneya and Carron (1992), Balmer et al (2001), Balmer et al (2004) and manyothers.

This paper is divided as follows. Section 2 talks about the history of soccer in Brazil. Section 3 shows the hypotheses test results. All phases of the used procedure are detailed in Section 4. Then, Section 5 shows our computational results considering the real Brazilian football league data from 2006 to 2010. Finally, the conclusions and future works are shown in Section 6.

2. SOCCERAND BRAZIL

Passion for soccer in Brazil developed along the twentieth century with limited influence of social factors like disputes between cities. There was a rivalry between the neighbor states of Rio de Janeiro, where was located the country capital during the largest part of the century and São Paulo, with the economic hegemony in the country long before it was accomplished the move of the capital to Brasilia, in the hinterland. During the decades of 50 and 60 the unique inter-states tournament involved only clubs of the states of Rio de Janeiro and São Paulo. In the other states there was frequently a double preference for a team of Rio de Janeiro or São Paulo and a local team, the national sympathy prevailing over the local preference. A National championship only started to happen by the end of the 60's. After that, in each state there remains some emulation between a small number of local clubs, usually two, and in matches between the team preferred in a national view and another antagonized in local grounds, the fan would strongly favor the team out of the state. This results in a culture where fidelity to the club plays an important role (Sant'anna et al, 2010).

3. HYPOTHESIS TEST

The hypothesis tests were performed using data from the Brazilian football league from 1990 to 2009. The first approach is to count the teams that fit in each hypothesis and to divide by the total number of weak (strong) teams for H1 (H2). As we expected, the results are far from 50% (H0): 65.51% for H1 and 68.57% for H2. Then, we calculate the p-value of H0 for the data of both H1 and H2: 0.0045 for H1 and 0.0004 for H2. With these values, we have a good evidence to reject H0 in both cases.

4. THE PROCEDURE

To generate the final schedule, we use a procedure composed of four phases:

Phase 1: Generate randomly a large number of complementary pattern sets subject to some constraints on breaks.

Phase 2: For each generated pattern set, test a condition for its feasibility through a linear programming problem proposed by Briskorn (2008).

Phase 3: For each feasible pattern set, construct a timetable scheduling with it.

Phase 4: For each timetable scheduling from phase 3, assign teams to patterns minimizing the effects related to strong (weak) teams.
Now, we detail each one of the four phases. Besides, in Subsection 4.5 we illustrate all these phases by a toy example.

4.1 PHASE 1: GENERATING PATTERN SETS

In this phase, we generate randomly a large number (about 200) of pattern sets ensuring three requirements: each pattern (mirrored pattern) is composed by at most two (four) breaks, there are no consecutive breaks, the generated pattern sets are only composed of complementary pattern pairs. The last requirement ensures one of the conditions for the pattern set feasibility: for all the rounds, half of the teams plays at home (away).

Note that the method used is not optimal because we do not use the all the possible pattern sets.

4.2 PHASE 2: TESTING A CONDITION FOR THE PATTERN SETS FEASIBILITY

In this phase we test a condition for the feasibility of each pattern set generated in phase 1 through the linear programming problem proposed by Briskorn (2008). That test is just a necessary condition for the feasibility of the pattern set. Here, two computational aspects are important: the computational time to test each pattern set is negligible since the model is a LP one and about 50% of total generated pattern sets are feasible.

4.3 PHASE 3: GENERATING TIMETABLES

In this phase, for each feasible pattern set, we schedule a timetable through the following IP formulation. Let the binary variable $\mathrm{x_{ijr}}$ indicate whether the pattern i matches the pattern j in the round $(\mathrm{i} > j)$. Let the constants Υ_{ir} indicate whether the pattern i plays at home $(\Upsilon_{\mathrm{ir}} = 1)$ or away $(\Upsilon_{\mathrm{ir}} = 0)$ at the round r.The model constraints are the following.

$$\sum_{j<i} \mathrm{x_{ijr}} = 1, \quad \forall i, r. \tag{1}$$

$$\sum_{r} \mathrm{x_{ijr}} = 1, \quad \forall i, j, j < i. \tag{2}$$

$$\mathrm{x_{ijr}} \leq \left|\Upsilon_{\mathrm{ir}} - \Upsilon_{\mathrm{jr}}\right|, \forall r, i, j, j < i. \tag{3}$$

$$\mathrm{x_{ijr}} \in \{0,1\}, \forall r, i, j, j < i. \tag{4}$$

Constraints (1) ensure for each patterni, for each round r, i plays against exactly one team. The group of constraints (2) requires for each pair of patterns i and j, that there is a round where i meets j while (3) ensures that a game between patterns i and j in the round r can only occur if i plays home (away) and j plays away (home) in this round. It is possible to schedule a timetable with a feasible pattern set in a few seconds.

4.4 PHASE 4: ASSIGN TEAMS TO PATTERNS

In this phase, we present an integer programming model where for each timetable from phase 3, each pattern is assigned to exactly one team minimizing the EC effects. We give priority to those effects over breaks and do not distinguish EC effects over breaks at home or away because, for example, if a team matches two strong teams at home consecutively then this team matches the same strong teams away in the mirrored callendar. The model is the following. Let the binary variable $\mathrm{x_{ij}}$ indicate whether the pattern i is assigned to team j. Let the binary variable $\mathrm{y_{ijr}}$ $(\mathrm{z_{ijr}})$ indicate whether a EC effect occurs when the team which is assignedto pattern i matches the pair of strong (weak) teams j in the rounds $r$and $r + 1$. We define the following notation. R is the rounds set, B is the set of rounds with breaks, T is the teams set, $P_1$is the scheduled pattern set in phase 3, $O_{|P_1| \times |R|}$is a matrix where each cell O_{ir} indicates the opponent of the pattern i in the round $r$according to timetable from phase 3, E is a set of team pairs which use the same stadium for home games, $C$is the set of complementary pattern pairs, $S(W)$is a set composed of pairs of strong (weak) teams.

$$\min \sum_{i\in P}\sum_{j\in S}\left(\sum_{r\in(R-B)} \mathrm{y_{ijr}} + 2\sum_{r\in B} \mathrm{y_{ijr}}\right) + \sum_{i\in P}\sum_{j\in W}\left(\sum_{r\in(R-B)} \mathrm{z_{ijr}} + 2\sum_{r\in B} z_{\mathrm{ijr}}\right) \quad (5)$$

$$s.t. \sum_{j\in T} \mathrm{x_{ij}} = 1, \quad \forall i \in P_1. \quad (6)$$

$$\sum_{i\in P_1} \mathrm{x_{ij}} = 1, \quad \forall j \in T. \quad (7)$$

$$\mathrm{x}_{c_1 j_1} - \mathrm{x}_{c_2 j_2} = 0, \forall (c_1, c_2) \in C, (j_1, j_2) \in E. \quad (8)$$

$$\mathrm{x}_{O_{ir} j_1} + \mathrm{x}_{O_{ir+1} j_2} - \mathrm{y_{ijr}} \le 1, \forall j = (j_1, j_2) \in S, \forall r \in \{1, \dots, |R| - 1\}. \quad (9)$$

$$\mathrm{x}_{O_{i|R|} j_1} + \mathrm{x}_{O_{i1+1} j_2} - \mathrm{y_{ij|R|}} \le 1, \forall j = (j_1, j_2) \in S. \quad (10)$$

$$\mathrm{x}_{O_{ir} j_1} + \mathrm{x}_{O_{ir+1} j_2} - \mathrm{z} \le 1, \forall j = (j_1, j_2) \in W, \forall r \in \{1, \dots, |R| - 1\}. \quad (11)$$

$$\mathrm{x}_{O_{i|R|} j_1} + \mathrm{x}_{O_{i1+1} j_2} - \mathrm{z_{ij|R|}} \le 1, \forall j = (j_1, j_2) \in W. \quad (12)$$

The objective function (5) aims to minimize the total number of effects on the calendar. Note that we prioritizeto avoid EC effects over breaks giving weight two to the variables y and z associated to rounds with breaks. The group of constraints (6) indicates that each pattern i is assigned to exactly one team. Constraints (7) ensure that each team j is assigned to exactly one pattern. The group of constraints (8) ensures that teams which use the same stadium for home games are assigned to complementary patterns. Constraints (9), (10), (11) and (12) compute the effects occurrence. The computational time at Phase 4 depends on n; $|S|$and$|W|$. The worst tested case is considering $n = 20$, $|S| = 6$ and $|W| = 4$. In this case, the average time for the first ten solutions is five minutes (PC Core i5 2,27GHz, 4GB, Windows 7, Visual Basic).

4.5 TOY ILLUSTRATIVE EXAMPLE

To illustrate each phase of the procedure, we propose a toy example. It has just $n = 4$ teams: $A, B, C$and D. Phase 1 generates a number of pattern sets. $P_1 : 101, P_1 : 010, P_1 : 100andP_4 :$ 011 is one pattern set for the toy example. Phase 2 tests whether each pattern set from phase 1 is feasible. Our chosen pattern set is feasible. The next phase constructs timetables with the feasible pattern sets from phase 2. A possible timetable for the pattern set example is in the Table 1 . The phase 4 assigns patterns to teams. For example $P_1 \to B, P_2 \to D, P_3 \to A$and $P_4 \to C$. The final schedule for the illustrative instance is in the Table 2.

Tabela1: A possible timetable for the pattern set of the toy example.

round 1	round 2	round 3
$P_1 - P_2$	$P_1 - P_4$	$P_1 - P_3$
$P_3 - P_4$	$P_2 - P_3$	$P_2 - P_4$

Tabela2: The final schedule for the toy example.

round 1	round 2	round 3
$B - D$	$B - C$	$B - A$
$A - C$	$D - A$	$D - C$

5. COMPUTATIONAL RESULTS

We tested our algorithm on the real Brazilian football leagues data from 2006 to 2010. For all these years, only two teams use the same stadium for home games: Flamengo and Fluminense.

Summary results are shown in Table 3. Column 1 indicates the year considered. Columns 2, 3 and 4 show for the current year, the number of teams, the number of strong teams and the number of weak teams, respectively. Statistics about effects on the official calendar are shown in columns 5,6, 7 and 8, while columns 9,10,11 and 12 show the average effects of the ten first calendars found by the proposed algorithm.

Tabela 3: Computational Results

Year	n	$\|S\|$	$\|W\|$	Oficial Calendar (Effects)					ThisPaper (Effects)				
				Without breaks		With breaks		Total	Without breaks		With breaks		Total
				Strong	Weak	Strong	Weak		Strong	Weak	Strong	Weak	
2010	20	5	4	8	13	4	1	26	5.4	2.1	2.2	0	*9.7*
2009	20	5	4	27	15	0	3	45	5.4	2.1	2.2	0	9.7
2008	20	5	4	17	11	2	2	32	4.4	2.1	2.8	0	9.3
2007	20	6	4	27	6	4	0	37	10.9	2.1	3.8	0	16.8
2006	20	5	2	11	0	6	0	17	5.4	0	2.2	0	7.6

These results show that, for all the years, the number of effects in the proposed solution is much smaller than the official one. Although we are not considering place constraints of the Brazilian football league, the large reduction obtained on the average number of the EC effects suggests that it can be substantially reduced in acomplete model.

6. CONCLUSIONS AND FUTURE WORKS

This paper proposes a new way to schedule the Brazilian football league by extending a known decomposition method. The proposed algorithm obtains fair schedules using a small number of breaks and minimizing the total number of extended carry-over (EC) effects regarding two strength groups and we justify the use of such groups by statistical analysis of previous tournaments. The proposed solution find a small number of EC effects compared to the official one for all the considered years. Moreover, it is easy to incorporate the objective function and the constraints proposed in phase 4 to any Integer Programming model to schedule sport timetables also based on decomposition method. As a future work, we plan to include Brazilian football league requirements to obtain more practical schedules.

REFERENCES

ALVES A.M., SOARES DE MELLO J.C.C.B., RAMOS T.G. & SANT'ANNA, A.P. 2012. Logit models for the probability of winning football games. *PesquisaOperacional*, 31(3), 459-465.

BALMER, N.J., NEVILL, A.M. & WILLIAMS, A.M. 2001. Home advantage in the Winter Olympics (1908-1998). *Journal of Sports Sciences*, 19(2), 129-139.

BARTSCH T., DREXL, A.& KROGER, S. 2006. Scheduling the professional soccer leagues of Austria andGermany. *Computers & Operations Research*, 33(7), 1907-1937.

BLOYCE, D. & MURPHY, P. 2008. Sports administration on the hoof: the three points for a win experiment in

English soccer. *Soccer & Society*, 9(1), 14-27.

BRISKORN, D. 2008. Feasibility of home-away-pattern sets for round robin tournaments. *OperationsResearch Letters*, 36(3), 283-284.

BRISKORN, D. 2009. Combinatorial properties of strength groups in round robin tournaments. *EuropeanJournal of Operational Research*, 192(3), 744-754.

BRISKORN, D.& KNUST, S. 2010. Constructing fair sports league schedules with regard to strength groups.*Discrete Applied Mathematics*, 158(2), 123-135.

COURNEYA, K.S. & CARRON, A.V. 1992. The home advantage in sport competitions: A literature review.*Journal of Sport and Exercise Psychology*, 14(1).

DELLA CROCE, F. & OLIVERI, D. 2010. Scheduling the Italian football league: An ILP-based approach.*Computers & Operations Research*, 33(7), 1963-1974.

DURÁN,G., GUAJARDO, M., MIRANDA, J., SAURÉ , D., SOUYRIS, S., WEINTRAUB, A. & WOLF, R. 2007. Scheduling theChilean soccer league by integer programming. *Interfaces*, 37(6), 539-552.

GOOSSENS, D.R.& SPIEKSMA, F.C.R. 2009. Scheduling the Belgian soccer league. *Interfaces*, 39(2), 109-118.

GOOSSENS, D.R. & SPIEKSMA, F.C.R.2012. Soccer schedules in Europe: an overview. *Journal of Scheduling*,*15*(5), 641-651.

HENZ, M., MULLER, T.& THIEL, S. 2004. Global constraintsfor round robin tournamentscheduling. *EuropeanJournal of Operational Research*, 153(1), 92-101.

KENDALL, G., KNUST, S.,RIBEIRO, C.C.& URRUTIA, S. 2010. Scheduling in sports: An annotated bibliography.*Computers & Operations Research*, 37(1), 1-19.

LEVER, J. 1983. Soccer madness. *Soccer madness*.

MIYASHIRO, R., IWASAKI, H.& MATSUI, T. 2003. Characterizing feasible pattern sets with a minimumnumber of breaks. In: *Practice and Theory of AutomatedTimetablingIV*, Springer Berlin / Heidelberg, 78-99.

NEMHAUSER, G.L. & TRICK, M.A. 1998. Scheduling a major college basketball conference. *Operations Research*, 46(1), 1-8.

NURMI, K., GOOSSENS, D., BARTSCH, T., BONOMO, F., BRISKORN, D., DURAN, G., KYNGAS, J., MARENCO,J., RIBEIRO, C.C., SPIEKSMA, F.& OTHERS. 2010. A framework for a highly constrained sports scheduling problem.*Proceedings of the International MultiConference of Engineers and Computer Scientists*, 3, 1991-1997.

PALACIOS-HUERTA, I. 2004. Structural changes during a century of the world's most popular sport.*Statistical methods & applications*, 13(2), 241-258.

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.

RIBEIRO, C.C.& URRUTIA, S. 2006. Scheduling the Brazilian soccer tournament with fairness and broadcastobjectives. *Proceedings of the 6th international conference on Practice and theory of automatedtimetabling VI*, 147–157.

RIBEIRO, C.C. & URRUTIA, S.2011. Scheduling the Brazilian soccer tournament: Solution approach andpractice. *Interfaces*. To appear.

RUSSELL, K.G. 1980. Balancing carry-over effects in round robin tournaments. *Biometrika*, 67(1), 127.

SANT'ANNA, A. P. 2008. Rough sets analysis with antisymmetric and intransitive attributes: classificationof Brazilian soccer clubs. *Pesquisa Operacional*, 28, 217-230.

SANT'ANNA, A.P., BARBOZA, E.U. & MELLO, J.C.C.B.S. 2010. Classification of the teams in theBrazilian soccer championship by probabilistic criteria composition. *Soccer & Society*, 11(3): 261-276.

SCHREUDER, J.A.M. 1992. Combinatorial aspects of construction of competition Dutch professional footballleagues. *Discrete Applied Mathematics*, 35(3), 301-312.