REFEREE ASSIGNMENT IN THE CHILEAN FOOTBALL LEAGUE USING INTEGER PROGRAMMING AND PATTERNS

Fernando Alarcón
Departamento de Ingeniería Industrial, Universidad de Chile, Santiago 8370439, Chile
falarcon@ing.uchile.cl

Guillermo Durán
Departamento de Ingeniería Industrial, Universidad de Chile, Santiago 8370439, Chile
Departamento de Matemática e Instituto de Cálculo, FCEyN, Universidad de Buenos Aires; and CONICET, Buenos Aires C1428EGA, Argentina
gduran@dii.uchile.cl

Mario Guajardo
Department of Finance and Management Science, NHH Norwegian School of Economics
Helleveien 30, N-5045 Bergen, Norway
mario.guajardo@nhh.no

ABSTRACT

The referee assignment problem in sports scheduling is addressed for the First Division of the Chilean professional football league using integer linear programming. Various criteria that enhance the transparency and objectivity of the assignment process are considered. As well as better balances in the number of matches each referee must officiate, the frequency each is assigned to a given team, and the distances each must travel over the course of a season, these improvements include the generation of appropriate pairings of referee experience or skill category with the importance of certain matches. The model is solved using two approaches, one traditional and the other a pattern-based approach inspired by the well-known home-away patterns for scheduling season matches. The two approaches are implemented for real instances of the problem, reporting results that significantly improve the manual assignment. Also, the pattern-based approach achieves major reductions in solution times over the traditional formulation.

KEYWORDS. Referee assignment. Sports scheduling. Football. Patterns. Integer linear programming.

A Pesquisa Operacional em Grandes Eventos Esportivos

OA - Outras aplicações em PO

OC - Otimização Combinatória

1. Introduction

The use of sports scheduling techniques for match calendar planning has spread widely in recent decades, particularly among the world's football leagues. Examples can be found in the literature for The Netherlands (Schreuder 1992), Germany and Austria (Bartsch et al. 2006), Chile (Durán et al. 2007, 2012), Denmark (Rasmussen 2008), Belgium (Goossens and Spieksma 2009), Norway (Flatberg et al. 2009), Honduras (Fiallos et al. 2010) and Brazil (Ribeiro and Urrutia 2011).

Published applications for referee assignment are few in number, however. The first one was reported by Evans (1988), who specifies a multicriteria optimization problem for scheduling the assignment of baseball umpires in a North American League using a range of heuristic methods to obtain good solutions with reasonable execution times. Wright (1991) develops a computer system for assigning umpires to professional cricket matches in England, including hard and soft constraints and various optimization criteria. It is solved by finding an initial solution using only some of the constraints and then improving it by applying local perturbations consisting in swapping pairs of umpires. Farmer et al. (2007) formulate an integer programming model to assign umpires to professional tennis tournaments in the U.S. They propose a solution method consisting of a two-phase heuristic in which the first phase constructs an initial set of assignments and the second employs a simulated annealing heuristic to improve them. More recently, Trick et al. (2012) have reported an application of network optimization and simulated annealing to schedule umpires for Major League Baseball games in North America. Other published works have taken a more theoretical approach, such as Dinitz and Stinson (2005), Trick and Yildiz (2006), Duarte et al. (2007) and Yavuz et al. (2008).

The present article addresses the referee assignment problem for the First Division of the Chilean professional football league using an integer linear programming model. The model incorporates various user-defined criteria that enhance the transparency and objectivity of the assignment process. Two solution methods are developed: a traditional one that runs the model directly, and a novel two-stage approach in which a first model constructs referee patterns for the season and a second generates the actual assignments. This strategy is inspired by the successful use of home-away patterns in the match scheduling methods of various sports leagues around the world. To our knowledge, this is the first article developing a pattern-based approach to referee assignment.

2. Referee assignment in the Chilean context

The top league in Chile's professional football league system is the First Division, which is governed and managed by the *Asociación Nacional de Fútbol Profesional de Chile* (ANFP). One of the ANFP's responsibilities is the assignment of referees to the Division's scheduled matches. This task is handled by a group of experts normally made up of retired professional football referees known as the Referee Committee. Assignments are decided by the Committee based on relatively loose criteria using strictly manual methods, with results that are often disadvantageous. Some referees, for example, will typically be assigned significantly fewer games than others despite having similar experience and skill levels. It is also common for some referees to be assigned relatively many matches involving the same team while others are never assigned to certain teams. In addition, due to the long shape of Chile's physical territory, assigning referees using manual methods may result in some of them travelling considerably longer distances than others. Given that most referees live in Santiago, the nation's capital, they naturally prefer assignments within the city's greater urban area where they are close to their homes and workplaces, especially since the majority have full-time jobs during the week. But as long as there are First Division teams scattered along the 4,200 kilometres separating the country's northern and southern extremes, some referees must travel to the more outlying venues.

These various factors point up the importance of improving the efficiency of referee assignment in the Chilean league. The dissatisfaction expressed by players, fans and team managers with the current quality of officiating is hardly surprising, and only serves to increase the pressures on the referees. Though the problems just described are not entirely attributable to poor assignment, the establishment of objectively defined criteria implemented by mathematical programming models would do much to ensure the process was both fair and transparent in the eyes of all relevant actors and would raise the league's general level of professionalism.

A previous application of mathematical programming tools to top-tier Chilean football was reported in Durán et al. (2007), which developed an optimization model for defining the First Division's annual match calendar. This formulation has had a considerable impact since it was first applied in 2005 and has been used by the league ever since. A modified version was adopted two years later by the Second Division (Durán et al. 2012). Further details on the organization of Chilean professional football and its competition formats may be found in the two just-cited papers. The involvement of the authors in scheduling referee assignments as described in the present article is one of various projects that have grown out of these earlier experiences.

3. Integer linear programming model

The various conditions that should be satisfied by the First Division referee assignment were defined in the light of conversations with officials of the ANFP and its Referee Committee that focussed on the weaknesses of the manual assignment methods detailed above and various suggestions by the participants. Since the season match calendar is assumed to be already known, the referee assignment determines which referee will officiate at each scheduled match. In practice, changes may be made as the season progresses if, for example, unforeseen circumstances affect the availability of certain referees. The model we propose is sufficiently flexible to be re-executed before each round (i.e., match date) using updated information incorporating these eventualities as well as the assignment experience to that moment.

In Appendix A, we set out an integer linear programming model for addressing the referee assignment problem. In the remaining sections of this article, this model will be referred to as Model 1, or simply the "original" or "traditional" model.

4. Pattern-based solution approach

Solving the traditional formulation is likely to be difficult due to the combinatorial structure, the nature of the data and the size of the instances. To illustrate this point, a football season organized as a double round-robin with 6 teams and 4 referees available for each match (thus requiring 3 of the 4 referees for each round) would have 63 billion possible referee assignments. For the current season format of Chile's First Division, with 18 teams, 34 rounds and about 15 referees, the possibilities would be almost literally endless. However, as we will see in Section 5, real instances of Model 1 can be solved in 14 to 72 minutes using a commercial solver. Though such execution times are reasonably acceptable, reducing them still further would be desirable so that solutions could be readily generated at meetings of the Referee Committee or when conducting multiple tests with different parameter values.

An alternative approach to producing good solutions in relatively little time that has been widely and successfully used for sports scheduling problems involves using an additional formulation that generates structures for defining each team's home-away match sequences known as patterns (see, for example, Bartsch et al. 2006, Durán et al. 2012, Goossens and Spieksma 2009, deWerra 1988 and Nemhauser and Trick 1998). Once these patterns have been constructed and assigned to the various teams, the complete season schedule is then determined. This two-stage approach usually cuts computation times significantly while delivering solutions that, though not necessarily optimal, perform well in terms of the objective function value. In what follows we briefly describe the home-away pattern approach in sports season scheduling and then set out our adaptation of it to referee assignment.

4.1 The pattern-based approach in season scheduling

A pattern as used in a sports season scheduling model refers to an ordered array of characters *H*, *A* and *B* denoting "home", "away" and "bye", respectively. In a pattern assigned to a particular team, the j^{th} element indicates whether, in round *j*, the team plays at home or away or has a bye. For example, in an illustrative season with five rounds, the pattern *P*(*Team 1*) = (*H,A,H,B,A,A*) indicates that *Team 1* plays at home in the first and third rounds, away in the second, fifth and sixth rounds, and has a bye in the fourth round.

Various strategies have been suggested in the literature for generating these home-away patterns, including logical rules and integer programming models. However they are constructed, they are generally used in a first stage of the solution approach to decide which teams play at home and which teams play away in each round, assuring that the teams will satisfy the various applicable home-away sequence constraints. In a second stage, it is decided which team plays against each other in each round, respecting the home-away decisions imposed by the patterns in the previous stage. The use of the patterns in the second stage allows to eliminate the constraints assured by the first stage, thus the aim becomes to obtain feasibility on the remaining constraints and, if there is an objective function, to look for the optimal solution.

4.2 A pattern-based approach for referee assignment

Obviously, the "home" and "away" concepts have no meaning in the referee assignment context. The patterns we use for the referee assignment rather define the zone in which each match assigned to a referee is played. For this purpose, the country is segmented into *North* (*N*), *Centre* (*C*) and *South* (*S*) zones, and each league team is classed into one of them as determined by the geographical location where its home venue is located. Since the number of referees is usually greater than the number of matches, we also incorporate a value denoted *Unassigned* (*U*) that indicates the rounds in which a referee has no game assigned. A pattern is thus defined as an ordered array of characters in {*N, C, S, U*} whose dimension is equal to the number of rounds in the season.

For example, in an illustrative season with nine rounds, the pattern *Q*(*Referee 1*) = (*C,N,S,N,U,C,S,S,N*), indicates that *Referee 1* officiates somewhere in the North zone of the country in rounds 2, 4 and 9; in the Centre zone in rounds 1 and 6; and in the South zone in rounds 3, 7 and 8. In round 5, he is unassigned. Note also that once the pattern of a referee *r* is defined, so is the total number of matches that referee will officiate over the course of the season. In *Q*, for example, with only one round unassigned *Referee 1* will officiate in 8 of the 9 rounds.

The segmentation of the teams by geographical zone was motivated by the peculiarities of Chile's physical territory referred to earlier, but any other relevant criterion could of course be used. One might be to group teams by level of popularity; another might be simply to classify them randomly.

With the foregoing definitions and explanations we can now develop our proposed solution methodology. The first stage consists in formulating a model that generates the patterns for each referee, incorporating some of the constraints defined in the original ILP problem (the ones that seem *a priori* particularly relevant to the pattern sequences). Hereafter the pattern-generation model will also be referred to as Model 2a. Its formulation is presented in Apendix B.

Once the patterns have been generated by solving Model 2a, another integer linear model incorporates the remaining constraints and generates the definitive assignments of referees to matches. We present this pattern-based assignment model in Appendix C and it will be referred to as Model 2b. Since the number of matches each referee will officiate is already defined by the patterns, so are the values for the variables Δ. This being the case, the ILP model will make feasible assignments based solely on these patterns without reference to the objective function. Note that the optimal values of variables y in the solution to Model 2a will be parameters in Model 2b.

If Model 2b cannot find a feasible solution, we generate a new set of patterns and try again to solve Model 2b with this new set. Several alternatives could be used to generate the set of patterns. For example, we have empirically realized that the mere change in the order of the constraints in the computational code of the model makes the solver to find a different pattern set. One could also try swapping patterns from one to another referee, or at least some rounds of their patterns. Alternatively, we could use an heuristic that relaxes the pattern specifications iteratively. Thus, if a set of patterns generated by Model 2a and used in Model 2b does not yield a solution, the patterns for two chosen referees are eliminated so that certain rounds can be interchanged between them. Model 2b is then run again, but this time all of the original problem constraints are imposed on the two referees now without patterns and the objective function expresses only the difference between their target and actual number of match assignments. If again no solution is found, the pattern of a third referee is eliminated and the process is iterated, each time stripping the pattern of another referee. In the worst possible case, this heuristic will terminate with all referee patterns eliminated and all the original problem constraints restored, in effect returning the original model.

However, as will be reported below, our experience solving the four actual instances of the problem for the 2007 through 2010 seasons was that Model 2b always found a feasible assignment using the set of patterns generated by Model 2a such that all referees officiate their target number of matches.

5. Results

In this section we evaluate a range of characteristics of the solutions obtained by our model as well as the solution times. The solution characteristics were derived from data for the regular 2007 season of Chile's First Division. That year there were 21 teams, 15 referees and 420 matches scheduled in 42 rounds across two half-seasons. In each round, 10 matches were played and one team had a bye. Model 1 for this instance contained about 6,300 binary variables, 15 integer variables and 15,503 constraints. The parameter values were defined in consultation with the Referee Committee.

To assess the solution times, all four instances of the problem covering the First Division's 2007-2010 seasons were solved. The models were implemented in AMPL and solved using the CPLEX 10.0 solver on an Intel Core 2 Duo 2.26GHz processor with 2 GB of RAM.

5.1 Characteristics of the solution

The problem was solved to optimality under both approaches (with and without patterns). The objective function value was zero, meaning that the solution satisfied the targets regarding the number of match assignments for every referee.

Our assignment results are compared in Table 1 with the actual manually produced results for the 2007 season. As can be seen, the model solution is superior on every point of comparison.

A useful indicator for measuring the balance of an assignment is its standard deviation. For the number of assignments to a referee, the actual 2007 standard deviation was 3.01 whereas in our solution it was only 0.63. The lowest and highest absolute numbers of actual matches assigned in 2007 were 24 and 36 respectively whereas in our solution they were 27 and 29. These minimum and maximum values of 27 and 29 are indeed well-balanced, considering that the ratio of the number of matches to the number of referees ($|M|/|R|$) was 420/15=28.

Measure	Actual	Model
Min. nr. of match assignments for a referee	24	27
Max. nr. of match assignments for a referee	36	29
Min. nr. of assignments of a referee to the same team	0	1
Max. nr. of assignments of a referee to the same team	7	4
Min. average travel distance per match by a referee	308	650
Max. average travel distance per match by a referee	1192	1146
Max. nr. of consecutive rounds in which a referee was unassigned	4	2

Table 1: Comparison of actual 2007 season assignment and optimization model assignment.

The actual 2007 standard deviation in the number of referee assignments to the same team for all of the referees and teams was 1.55 while the model value was only 1.11. Since this is a highly sensitive issue for fans and the media, a well-balanced result on this indicator is particularly important. A good idea of what the average number of referee-team assignments should be in a perfectly balanced scenario is given by the ratio of the number of matches played by each team to the number of referees, which in 2007 was 40/15 = 2.67. The actual minimum and maximum results were 0 and 7 respectively, contrasting sharply with the model results of 1 and 4.

As for average travel distance, the maximum difference between two referees was 884 km in the actual assignment, while in our solution it was 497 km. In addition, the actual 2007 standard deviation was 268 whereas the model result was 181.

5.2 Solution times

A comparison of the solution times recorded by the original model with those of the pattern-based approach (the latter being the sum of the individual Model 2a and Model 2b times) is given in Table 2. Note that the number of teams in the 2008 season was 20, and in the 2009 and 2010 seasons was 18, thus fewer than the 21 teams in 2007. In every case, both the traditional and the pattern-based models reached their optimal values with the OF equal to 0, even though for the pattern-based formulation this was not assured *a priori*. Note also that in all cases the pattern-based approach found the optimum in a single run of Model 2a and Model 2b, with no need for generating another set of patterns.

Instance	T_1	T_{2a}	T_{2b}	$T_{2a} + T_{2b}$
2007	4314.9	1.3	5.7	7.0
2008	3359.4	1.0	0.7	1.7
2009	827.9	0.7	0.3	1.0
2010	2764.7	0.9	0.8	1.7

Table 2: Solution times (in seconds) of the traditional formulation model (T_1), the pattern-generation model (T_{2a}) and the pattern-based assignment model (T_{2b}).

The second column of Table 2 shows the solution times for Model 1 while the fifth column (the sum of columns 3 and 4) displays those for the pattern-based approach. As can be appreciated, the latter reduces the solution times dramatically. The traditional formulation takes from 827.9 to 4314 seconds or, equivalently, from 14 to 72 minutes. The pattern-generation model Model 2a arrived at its solutions in less than 2 seconds, while the referee assignment Model 2b, using the patterns generated, delivered its solutions in times ranging from 0.3 to 5.7 seconds. The enormous reduction in solution times of the pattern-based approach expressed in percentage terms accounts for more than 99%.

6. Discussion and conclusions

We have proposed the use of integer linear programming for improving the assignment of referees to scheduled matches in the First Division of the Chilean professional football league. We developed two solution approaches. The first is a traditional approach that runs the model directly; the second is a novel two-stage approach in which a first model constructs referee patterns for the season and a second incorporates this information to generate the actual assignments.

The models were tested on the real-world cases of the referee assignments in the First Division of the Chilean league for the years 2007 through 2010, delivering significant improvements over the actual manual assignments. Also, the pattern-based version achieved major reductions in solution times over the traditional formulation. Our approach also simplifies the assignment process and renders it more transparent by establishing clearly defined decision criteria.

The model was used for First Division referee assignment on a trial basis in the 2010 season. For this purpose, a friendly interface was developed in Microsoft Excel so that the model can be applied easily and directly as a tool by league officials. They also requested that the model be extended to handle referee assignment for the Under-17 and Under-18 youth leagues, where it was used successfully for much of 2010. However, due to changes in the governing body's referee committee over the last couple of years the application of the referee assignment model was dropped. Efforts are continuing to have the league incorporate our Operations Research approach on a permanent basis, as it has occured with the match scheduling application the authors developed in conjunction with other academics which has been used every season since 2005 (Durán et al. 2007, 2012).

The model is also amenable to a series of useful extensions that would address a range of significant issues. If the season calendar at a given point schedules a mid-week round followed immediately by a weekend round (e.g., Wednesday and then Saturday) or vice versa, better advantage could be taken of the extensive travel involved by assigning a referee to both rounds within one of the outlying zones, thus obviating the need to return to Santiago between matches. Another useful extension would be to incorporate the assignment of these assistant referees as well. Yet another extension would be to formulate the model so that it integrates match scheduling and referee assignment in a single problem. Existing developments, including the present one, generate the referee assignment on the basis of a previously defined match calendar. This puts conditions on the setting of the assignment problem in that some of its constraints will be determined by the calendar scheduling. The simultaneous generation of match schedules and referee assignments could also be pursued at a theoretical level by combining the Travelling Umpire Problem (Trick and Yildiz 2006) with the Travelling Tournament Problem (Easton et al. 2001). This would provide a conceptual benchmark for integrated formulations of the two problems that till now have always been addressed separately.

As regards national team competitions, an interesting topic would be to analyze how often each country's squad is officiated by referees of a given nationality. An analysis by the present authors of the South American zone qualifying stages for the 2010 World Cup revealed that some teams were officiated relatively frequently by referees from certain countries while others were officiated by referees with a greater variety of national origins. Just as our model attempted to balance the frequency of assignments of a given referee to a specific team, referee assignments by nationality in international tournaments could be similarly balanced.

Finally, greater use of sports scheduling techniques for referee assignment could reduce much of the controversy and criticism among referees, players, team officials, fans and the media that often surrounds the choice of referees for sporting events. Although the use of OR techniques for match calendar scheduling has spread widely in recent years, their implementation for referee assignment is not yet firmly established.

References

Bartsch T, Drexl A, Kröger S. (2006). Scheduling the professional soccer leagues of Austria and Germany. *Computers & Operations Research* 33(7):1907-1937.

deWerra D. (1988). Some models of graphs for scheduling sports competitions. *Discrete Applied Mathematics* 21(1):47-65.

Dinitz JH, Stinson DR. (2005). On assigning referees to tournament schedules. *Bulletin of the Institute of Combinatorics and its Applications* 44:22-28.

Duarte A, Ribeiro C, Urrutia S. (2007). Referee assignment in sport tournaments. *Lecture Notes in Computer Science* 3867:158-173.

Durán G, Guajardo M, Miranda J, Sauré D, Souyris S, Weintraub A, Wolf R. (2007) Scheduling the Chilean Soccer League by Integer Programming. *Interfaces* 37(6):539-552.

Durán G, Guajardo M, Wolf R. (2012). Operations Research Techniques for Scheduling Chile's Second Division Soccer League. *Interfaces* 42(3):273-285.

Easton K, Nemhauser G, Trick M. (2001). The traveling tournament problem: description and benchmarks. In: *Proceedings of the 7th. International Conference on Principles and Practice of Constraint Programming*, Paphos, Cyprus, p.580-584.

Evans JR. (1988) A Microcomputer-Based Decision Support System for Scheduling Umpires in the American Baseball League. *Interfaces* 18(6):42-51.

Farmer A, Smith JS, Miller LT. (2007). Scheduling Umpire Crews for Professional Tennis Tournaments. *Interfaces* 37(2):187-196.

Fiallos J, Pérez J, Sabillón F, Licona M. (2010). Scheduling soccer league of Honduras using integer programming. In: Johnson A, Miller J, editors. *Proceedings of the 2010 Industrial Engineering Research Conference*, Cancún, Mexico.

Flatberg T, Nilssen EJ, Stølevik M. (2009). Scheduling the topmost football leagues of Norway. In: *23rd European Conference on Operational Research Book of Abstracts*, Bonn, Germany, p.240.

Goossens D, Spieksma F. (2009). Scheduling the Belgian Soccer League. *Interfaces* 39(2):109-118.

Nemhauser GL, Trick MA. (1998). Scheduling a major college basketball conference. *Operations Research* 46(1):1-8.

Rasmussen RV. (2008). Scheduling a triple round robin tournament for the best Danish soccer league. *European Journal of Operational Research* 185(2):795-810.

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

Schreuder J. (1992). Combinatorial aspects of construction of competition Dutch professional football leagues. *Discrete Applied Mathematics* 35(3):301-312.

Trick MA, Yildiz H. (2006). The Traveling Umpire Problem. *Informs Annual Conference*, Pittsburgh, USA.

Trick MA, Yildiz H, Yunes T. (2012). Scheduling Major League Baseball Umpires and the Traveling Umpire Problem. *Interfaces* 42(3):232-244.

Wright MB. (1991). Scheduling English cricket umpires. *Journal of the Operational Research Society* 42(6):447-452.

Yavuz M, Inan UH, Figlali A. (2008). Fair referee assignments for professional football leagues. *Computers & Operations Research* 35(9):2937-2951.

APPENDIX A: TRADITIONAL ILP FORMULATION (Model 1)

Sets

M: The set of matches.
R: The set of referees.
T: The set of teams.
K: The set of rounds.
FIX: The set of pairs (r, m) predetermining that referee r must officiate at match m.
$NOFIX$: The set of pairs (r, m) predetermining that referee r must not officiate at match m.

Parameters

$\alpha_{m,k}$: 1 if match m is played in round k, 0 otherwise.
$\beta_{m,t}$: 1 if team t plays in match m, 0 otherwise.

a_r: Minimum total number of match assignments for referee r.
$\bar{a}_r$: Maximum total number of match assignments for referee r.
$n_{r,t}$: Minimum number of referee-team assignments involving referee r and team t.
$\bar{n}_{r,t}$: Maximum number of referee-team assignments involving referee r and team t.
c: Minimum number of consecutive rounds between assignments of a referee r to the same team ($c \geq 1$).
$\delta_{r,m}$: Distance (round trip) between home town of referee r and city of venue of match m, in kilometres.
$\bar{\delta}$: Maximum difference allowed between any two referees' average match travel distances.
u_r: Maximum number of consecutive rounds for which a referee r is unassigned, that is, has no match assigned.
τ_r: Target number of match assignments for referee r.

Variables

$x_{r,m}$= 1 if referee r is assigned to match m, 0 otherwise.
Δ_r = Absolute value of the difference between target and actual number of match assignments for referee r.

Objective function

The objective function (OF) is the one suggested by Duarte et al. (2007), which consists of minimizing the sum over all referees of the absolute value of the difference between the target and the actual number of games assigned to each referee.

$$\min f = \sum_{r \in R} \Delta_r \tag{1}$$

Constraints

Basic constraints. Each match must be assigned one and only one referee.

$$\sum_{r \in R} x_{r,m} = 1 \qquad \forall\, m \in M. \tag{2}$$

Referee-round constraints. Each referee can be assigned to a maximum of one match per round.

$$\sum_{m \in M} \alpha_{m,k} \cdot x_{r,m} \leq 1 \qquad \forall\, r \in R,\ k \in K. \tag{3}$$

Season match assignment balance constraints. Minimum and maximum numbers of total season match assignments for each referee, limited by lower and upper bounds (the target number τ_r is thus a value between these two bounds).

$$\sum_{m \in M} x_{r,m} \geq a_r \qquad \forall\, r \in R. \tag{4}$$

$$\sum_{m \in M} x_{r,m} \leq \bar{a}_r \qquad \forall\, r \in R. \tag{5}$$

Referee-team balance constraints. Minimum and maximum referee-team assignments are limited by lower and upper bounds.

$$\sum_{m \in M} \beta_{m,t} \cdot x_{r,m} \geq n_{r,t} \qquad \forall\, r \in R,\ t \in T. \tag{6}$$

$$\sum_{m \in M} \beta_{m,t} \cdot x_{r,m} \leq \bar{n}_{r,t} \qquad \forall\, r \in R,\ t \in T. \tag{7}$$

Also, in c consecutive rounds a given referee cannot be assigned to the same team more than once.

$$\sum_{d=0}^{c-1} \sum_{m \in M} \beta_{m,t} \cdot \alpha_{m,k+d} \cdot x_{r,m} \leq 1 \qquad \forall\, r \in R,\ t \in T,\ k \leq |K| - c + 1. \tag{8}$$

Average travel distance balance constraints. The differences between the referees' average match travel distances are subject to an upper bound. The distances are calculated assuming the number of referee assignments are equal to the target values. This assumption is not necessarily satisfied *a priori*, but since the objective function attempts to satisfy it, we would expect that calculating the distances this way should yield good estimates while allowing us to conserve the problem's linearity.

$$\frac{1}{\tau_r}\sum_{m\in M}\delta_{r,m}\cdot x_{r,m} - \frac{1}{\tau_{\hat{r}}}\sum_{m\in M}\delta_{\hat{r},m}\cdot x_{\hat{r},m} \leq \bar{\delta} \quad \forall\, r,\hat{r}\in R. \tag{9}$$

No assignment constraint. Sets the maximum number of consecutive rounds for which a referee may have no assignment.

$$\sum_{i=0}^{u_r}\sum_{m\in M}\alpha_{m,k+i}\cdot x_{r,m} \geq 1 \qquad \forall\, r\in R,\ k\leq |K|-u_r. \tag{10}$$

Referee category and match importance level. Certain matches during the season must be officiated by more experienced or higher "category" referees. To formulate this restriction, the set M of matches is partitioned into three subsets denoted M_V, M_H and M_N containing matches of *very high*, *high* and *normal* expectation level, respectively ($M = M_V \cup M_H \cup M_N$). Similarly, the referees are partitioned into three subsets denoted R_A, R_B and R_C corresponding to skill level categories A, B and C (in decreasing skill level order) as determined by the Referee Committee ($R = R_A \cup R_B \cup R_C$). An R_A referee can officiate any match, an R_B referee can officiate M_H or M_N matches and an R_C referee can only be assigned to M_N matches. These constraints are expressed as follows:

$$\sum_{r\in R_A} x_{r,m} = 1 \qquad \forall\, m\in M_V. \tag{11}$$

$$\sum_{r\in R_A\cup R_B} x_{r,m} = 1 \qquad \forall\, m\in M_H. \tag{12}$$

Special assignments and non-assignments. A referee may be unable to officiate a certain match, for example, due to a suspension or an injury. To accommodate such cases the following constraint is included:

$$x_{r,m} = 0 \qquad \forall\, (r,m)\in NOFIX. \tag{13}$$

The Referee Committee may wish to impose the assignment of a given referee to a certain match. This can be done through the following constraint:

$$x_{r,m} = 1 \qquad \forall\, (r,m)\in FIX. \tag{14}$$

Logical constraints for Δ_r. The final two constraints ensure that Δ_r is the absolute difference, for each referee, between the target number of assignments defined *a priori* and the actual number assigned.

$$\sum_{m\in M} x_{r,m} + \Delta_r \geq \tau_r \qquad \forall\, r\in R. \tag{15}$$

$$\sum_{m\in M} x_{r,m} - \Delta_r \leq \tau_r \qquad \forall\, r\in R. \tag{16}$$

Nature of the variables.

$$\Delta_r \in \mathbb{Z} \text{ and } x_{r,m}\in\{0,1\} \qquad \forall\, r\in R,\ m\in M. \tag{17}$$

APPENDIX B: PATTERN-GENERATION MODEL (Model 2a)

In the first model of our two-stage approach we introduce a family of variables for constructing the patterns and select certain constraints from the original model that are then partially or wholly captured in the pattern-generation specification. The sets and parameters include some from the original model which retain their definitions plus some additional ones that are set out below.

Additional sets

$Z = \{N, C, S, U\}$: A set of characters indicating that the referee is either assigned to officiate in the specified geographic zone (North, Centre or South) or is unassigned.

RN: The set of triples (r, z, k) such that referee r cannot officiate in zone z in round k.

RY: The set of triples (r, z, k) such that referee r must officiate in zone z in round k.

Additional parameters

$\gamma_{m,z,k}$: 1 if match m is played in zone z in round k, 0 otherwise.

$\rho_{t,z,k}$: 1 if team t plays in zone z in round k, 0 otherwise.

$\tilde{\delta}_{r,z}$: Average distance between home town of referee r and cities where home venues of teams in zone z are located.

Variables

$y_{r,z,k} = 1$ if pattern of referee r indicates that he officiates in zone z or is unassigned in round k.

Δ_r = Difference between target and actual number of match assignments for referee r.

Objective function

The same as that for Model 1.

$$\min f = \sum_{r \in R} \Delta_r \tag{18}$$

Constraints

Basic constraints on patterns and season schedule. The number of patterns indicating a match to be officiated in zone z in round k must equal the number of matches specified by the match calendar for that zone in that round.

$$\sum_{r \in R} y_{r,z,k} = \sum_{m \in M} \gamma_{m,z,k} \qquad \forall\, z \in \{N, C, S\},\ k \in K. \tag{19}$$

This family of constraints is similar to constraints (2) of Model 1, except that here it is applied to the variables y.

Referee-round constraints. For every round, each referee must either be assigned to officiate in some zone or be unassigned.

$$\sum_{z \in Z} y_{r,z,k} = 1 \qquad \forall\, r \in R,\ k \in K. \tag{20}$$

This family of constraints is analogous to constraints (3) of Model 1.

Season match assignment balance constraints for each referee. Echoing the restrictions (4) and (5) in Model 1, lower and upper bounds impose minimum and maximum values for the total number of matches each referee can officiate in a season.

$$\sum_{z \in \{N,C,S\}} \sum_{k \in K} y_{r,z,k} \geq a_r \qquad \forall\, r \in R. \tag{21}$$

$$\sum_{z \in \{N,C,S\}} \sum_{k \in K} y_{r,z,k} \leq \bar{a}_r \qquad \forall\, r \in R. \tag{22}$$

Referee-team balance constraints. To partially capture the restrictions on referee assignments to particular teams expressed in constraints (6) of Model 1, we impose for each team and referee a lower bound on the number of times any pattern may be assigned to the zone in which that team plays.

$$\sum_{z \in \{N,C,S\}} \sum_{k \in K} \rho_{t,z,k} \cdot y_{r,z,k} \geq n_{r,t} \qquad \forall\, r \in R,\ t \in T. \tag{23}$$

However, the upper bound $\bar{n}_{r,t}$ in Model 1 is unnecessary because various matches are normally played in a given zone in a given round so that assigning referee r to a zone does not necessarily mean he officiates team t, and therefore undesirable because its mere application would reduce the range of assignment options.

Average travel distance balance constraints for each referee. This constraint, similar to constraints (9) in Model 1, aims at achieving a balance between the referees' average travel distances. We partially capture this, by considering the average distance $\tilde{\delta}_{r,z}$ between home town of referee r and cities where home venues of teams in zone z are located.

$$\frac{1}{\tau_r} \sum_{z \in \{N,C,S\}} \sum_{k \in K} \tilde{\delta}_{r,z} \cdot y_{r,z,k} - \frac{1}{\tau_{\hat{r}}} \sum_{z \in \{N,C,S\}} \sum_{k \in K} \tilde{\delta}_{\hat{r},z} \cdot y_{\hat{r},z,k} \leq \bar{\delta} \qquad \forall\, r, \hat{r} \in R. \tag{24}$$

No assignment constraint. As with constraints (10) in Model 1, this constraint bounds the number of consecutive unassigned rounds for each referee pattern.

$$\sum_{i=0}^{u_r} y_{r,z,k+i} \leq u_r \qquad \forall\, r \in R,\ k \leq |K| - u_r,\ z \in \{U\}. \tag{25}$$

Referee category and match level. This captures constraints (11) and (12) of Model 1 by imposing that the number of patterns assigned to A category referees officiating in zone z in round k be equal to the number of matches in zone z in round k that require this referee category.

$$\sum_{r\in R_A} y_{r,z,k} \geq \sum_{m\in M_V} \gamma_{m,z,k} \qquad \forall\, z \in \{N, C, S\},\ k \in K. \tag{26}$$

The same condition is imposed for matches requiring referees of at least B category (obviously, A category referees can also officiate such matches).

$$\sum_{r\in R_A\cup R_B} y_{r,z,k} \geq \sum_{m\in M_V\cup M_H} \gamma_{m,z,k} \qquad \forall\, z \in \{N, C, S\},\ k \in K. \tag{27}$$

Special assignments and non-assignments. To capture constraints (13) of Model 1, we impose the constraint that the referee cannot officiate in the zone where the match in question is to be played.

$$y_{r,z,k} = 0 \qquad \forall\, (r, z, k) \in RN. \tag{28}$$

Analogously, for constraints (14) we impose the constraint that the referee must officiate in the zone where the home team's ground is located:

$$y_{r,z,k} = 1 \qquad \forall\, (r, z, k) \in RY. \tag{29}$$

Logical constraints to calculate Δ_r. The variable Δ_r is calculated in analogous fashion to constraints (15) and (16), except that here it is the variables y in the geographic zones that are summed instead of x.

$$\sum_{z\in\{N,C,S\}} \sum_{k\in K} y_{r,z,k} + \Delta_r \geq \tau_r \qquad \forall\, r \in R. \tag{30}$$

$$\sum_{z\in\{N,C,S\}} \sum_{k\in K} y_{r,z,k} - \Delta_r \leq \tau_r \qquad \forall\, r \in R. \tag{31}$$

Nature of the variables.

$$\Delta_r \in \mathbb{Z} \text{ and } y_{r,z,k} \in \{0, 1\} \qquad \forall\, r \in R,\ z \in Z,\ k \in K. \tag{32}$$

APPENDIX C: PATTERN-BASED ASSIGNMENT MODEL (Model 2b)

As with Model 2a, in Model 2b the parameters and sets defined for Model 1 retain their previous definitions except that now the optimal values of variables y in the solution to Model 1 will be parameters in Model 2a, here denoted $\tilde{y}$ to avoid confusion. This predefines the pattern that will be assigned to each referee.

Parameters

$\tilde{y}_{r,z,k}$: 1 if in round k the pattern of referee r indicates that he is assigned to officiate in zone z or is unassigned, 0 otherwise.

Variables

$x_{r,m}$= 1 if referee r is assigned to match m, 0 otherwise.

Constraints

Constraints on patterns and their logical relationship with variable x. A condition is imposed on the relationship between variable x and parameter $\tilde{y}$ ensuring that a match m is assigned to referee r only if the corresponding pattern assigns r to the zone z in which the venue of m is located. This restriction is modelled as follows:

$$\sum_{m\in M} \gamma_{m,z,k} \cdot x_{r,m} = \tilde{y}_{r,z,k} \qquad \forall\, r \in R,\ z \in \{N, C, S\},\ k \in K. \tag{33}$$

Model 2b, also includes explicitly the Model 1 constraints (2), (6), (7), (8), (9), (11), (12) and (14), which are not necessarily guaranteed by the patterns generated by Model 2a. On the other hand, constraints (3), (4), (5), (10), (13), (15) and (16) are guaranteed by the patterns generated by Model 2a, thus we do not include them in Model 2b.