Submitted to *Interfaces*
manuscript (Please, provide the manuscript number!)

Scheduling Multiple Sports Leagues with Travel Distance Fairness: An Application to Argentinean Youth Football

Guillermo A. Durán
Dto. de Matemtica e Inst. de Clculo FCEyN-UBA y CONICET, Argentina, Dto. de Ingeniera Industrial, FCFM-Univ. de Chile, Santiago, Chile, ,

Mario Guajardo
Dept. of Business and Management Science, NHH Norwegian School of Economics, Bergen, Noruega, ,

Agustina F. López
Dto. de Matemtica e Inst. de Clculo FCEyN-UBA, Argentina, ,

Javier Marenco
Inst. de Ciencias, UNGS y Dto. de Computacin, FCEyN-UBA, Argentina, ,

Gonzalo A. Zamorano
Dto. de Ingeniera Industrial, FCFM-Univ. de Chile, Santiago, Chile, ,

The first division clubs in Argentinean professional football maintain teams in each of six youth leagues, classed by age as major divisions (Under-20, Under-18, Under-17) and minor divisions (Under-16, Under-15, Under-14). Regular season play in these leagues typically follows a single round-robin format, the minor divisions playing the same schedule as the majors but with the home-away status of the matches reversed. This setup can give rise to very significant differences in travel distances between the major and minor division teams of a given club, a frustrating situation for club officials, coaches and players alike but almost impossible to avoid with manual season scheduling techniques. Nor are these methods able to take into account any number of other criteria that go into the design of a satisfactory match calendar. This paper reports on models developed using mathematical programming to simultaneously schedule multiple leagues while also meeting a series of other desirable conditions. The central criterion is a better balance in the travel distances of the various teams, pursued through the application of two alternative solution approaches, one based on regional team clusters and the other on explicit analysis of actual distances between the teams' home venues. The solutions generated by these approaches have been used by the organisers of the Argentinean youth leagues to draw up their season schedules since 2018, which has resulted in a series of benefits for all stakeholders.

Key words: integer linear programming; sports scheduling; travelling tournament problem
History:

Introduction

Football is without a doubt one of the most popular sports in the world, with some 200 million players registered in more than 200 countries around the globe. The top leagues are followed by hundreds of millions of fans and generate enormous sums of money. The last FIFA World Cup, held in Russia in 2018, generated profits of USD 5.3 billion, the most ever recorded in the history of the tournament, and attracted a television audience estimated at more than 3.5 billion viewers. But if most of the attention naturally focuses on the professional divisions, their future depends on the existence of a well-organised set of youth leagues where much of the upcoming talent is developed, often on teams run by professional-level sides. This is certainly the case with Argentina, which has two World Cups to its credit, one of only six national sides to win the quadrennial event more than once. Without a doubt, much of this success is due to the importance placed on the nurturing of players in the juvenile leagues, which is also reflected in Argentina's six Under-20 World Cup wins. A long list of the country's top footballers have come up from children's and youth teams belonging to the professional clubs. A good example is Lionel Messi, who has won four Ballon d'Or awards plus the Golden Ball award in the 2014 World Cup and has been World Cup runner-up with Argentina four times. The Argentine star began with the Newell's Old Boys children's teams, transferring directly from them to Barcelona while still very young. Another case is Diego Maradona, world champion with the national side and Golden Ball winner at the Mexico World Cup in 1986. Maradona played on the Argentinos Juniors children's and youth teams and did a brief stint with Boca Juniors before being transferred to Barcelona at the tender age of 21.

Thus, with the importance of juvenile player development virtually impossible to exaggerate, there are clear benefits to be had in introducing at youth league level the use of Operations Research and analytics, which has grown rapidly in recent years in the professional leagues with undeniable success. The present article reports on the application of OR techniques to season scheduling in Argentinean youth football. One of the main difficulties facing organisers of these leagues is that they must simultaneously schedule six different divisions classed by age level: Under-20, Under-18, Under-17, Under-16, Under-15 and Under-14. In addition to the various factors typically considered in the design of any good sports season calendar, for these leagues it was desired to reduce the often very significant differences in distance the teams in the various divisions belonging to any single club must travel over the course of a season due to certain aspects of the schedule formats. The resulting scheduling problem was virtually impossible to solve efficiently using traditional manual methods. Adopting an OR approach both simplified the task and improved the outcome significantly.

Sports Scheduling

Sports scheduling is a field of study that investigates methods for the design of sporting competition calendars satisfying a set of conditions decided by the competition organisers. Due to the wide variety of tournament and season formats used in different sports, a large number of interesting optimisation problems arise in the field, as was demonstrated in a recent classification of real-world sports scheduling problems and implementations by Van Bulck et al. (2019a). The broad range of applications is also clearly reflected in the number of articles published in (Fry and Ohlmann (2012)). To cite just a couple of the more recent studies, Alarcón et al. (2017) describes 12 years of experiences in scheduling Chilean professional football leagues, which culminated in a request by the South American Football Confederation (CONMEBOL) to define the calendar for the South American qualifiers of the 2018 World Cup as described in Durán et al. (2017). Cocchi et al. (2018) reports on the use of mathematical programming to schedule the Italian national volleyball tournament.

Although the sports scheduling literature has focussed mainly on adult professional leagues, a small number of papers have dealt with non-professional and youth leagues. These include the cases of German table tennis (Schönberger et al. (2004), Knust (2010)), Finnish ice hockey (Nurmi et al. (2014)), English cricket (Wright (2018)) and Belgian football (Toffolo et al. (2019), Van Bulck et al. (2019b)). And while most applications are designed to schedule a single league, a few have addressed the problem of scheduling multiple leagues. These include U.S. softball (Grabau (2012)), New Zealand rugby (Burrows and Tuffley (2015)), and German table tennis (Schönberger (2017)).

Among the various optimisation criteria applied in the literature, one of the most common is travel distance. In the *Travelling Tournament Problem* (Easton et al. (2001)), the problem most frequently studied, the objective is the minimisation of the sum of distances travelled by the teams over the course of a fictitious tournament. The difficulty of solving this problem has motivated a considerable number of methodological studies but its implementations in real-word cases are rather few (Bonomo et al. (2012), Durán et al. (2019)).

The present article aims to contribute a new and practical OR implementation for the simultaneous scheduling of multiple youth football league seasons. Central among the various optimisation criteria to be considered will be travel distance. The objective pursued by the organisers of these leagues is the reduction of the disparities in distances travelled by the teams from a given club in the various youth divisions and simultaneously by the different teams in a given division whose home grounds are in the same geographical region. Both of these objectives are key to attaining greater tournament schedule fairness and avoiding inequalities in player fatigue between teams from different clubs and teams from the same club in different divisions.

To address this scheduling problem, we develop two mathematical programming approaches. Whereas one of them explicitly incorporates the minimisation of travel distance differences between the teams from the same club in the various divisions, the other one takes a more innovative tack, defining clusters of teams by geographical region and then equalising, to the extent possible, the number of times each team from a given club plays away against teams in each of the clusters. The outcome of both approaches is the reduction of differences in the distances travelled by the various teams located within a particular region playing in a given division. As noted earlier, this is considered by league organisers to be a key component of schedule fairness, but it is not necessarily achieved by the criterion commonly used in the sports scheduling literature of minimising the sum of travel distances.

Description of the problem

The most important youth football leagues in Argentina are those whose teams belong to the same clubs that compete in the country's first division of professional football. Each of these clubs has a team in six such leagues, three of them for older youth denoted the major divisions (Under-20, Under-18 and Under-17) and the other three known as the minor divisions (Under-16, Under-15 and Under-14). The major divisions all follow a single schedule while the minor divisions follow the same schedule but with the matches' home-away status reversed.

In 2017, the last of the seasons (hereafter called "tournaments") played on a calendar-year basis, the youth divisions each had 30 teams. This number fell to 28 in the first half of 2018, a short transitional tournament as Argentinean football switched to a split-year schedule. In 2018-2019, the first tournament under the new setup, the number of teams fell again to 26. The models and their solutions reported in this study were developed for the 2018 transitional and 2018-2019 tournaments.

The 30 teams in the 2017 tournament played a single round-robin format, the champion in each division being the team with the most points. By contrast, the 28 teams in the 2018 transitional tournament were divided into two groups of 14 teams denoted Group A and Group B, respectively (see Table 1). The *Superliga Argentina de Ftbol* (SAF), the governing body responsible for scheduling the first professional division and their associated youth leagues, held a draw before the tournament started to determine the teams' group assignments. Also, fourteen pairs of classic rivalries were defined and once a team was assigned to a group, its classic rival was automatically assigned to the other one. The teams within each group played a single round-robin, meaning each team had 13 matches, one in each scheduled round. To this the SAF added a 14th "classics round" in which each team played its classic rival, which by the definition given above was necessarily a team from the other group. It was also decided that this extra round would be the seventh of the fourteen rounds, and would therefore be played more or less at the midway point of the tournament.

To define the division champions of the 2018 transitional tournament, the two top teams at the end of the tournament in Group A were cross-matched in a playoff semi-final with the two top teams in Group B. Thus, the first-place team in each group played against the second-place team in the other group. The winners of these two semi-finals then met in a playoff final.

For 2018-2019 the first half of the split-year schedule, known as the "Apertura" or opening tournament, was played from August through November 2018 while the second half, the "Clausura" or closing tournament, ran from March through June 2019. The format was similar to the previous tournament but in this case there were only 26 teams, of which 13 were again assigned by draw to Group A while their 13 classic rivals were automatically

	Group A	Group B
1	River Plate	Boca
2	San Lorenzo	Huracn
3	Vlez	Argentinos Jrs.
4	Tigre	Chacarita
5	Independiente	Racing
6	Banfield	Lans
7	Defensa y Justicia	Arsenal
8	Estudiantes	Gimnasia y Esgrima
9	Rosario Central	Newell's
10	Coln	Unin
11	Talleres	Belgrano
12	San Martn SJ	Godoy Cruz
13	Atltico Tucumn	Patronato
14	Temperley	Olimpo

Table 1 Teams in the 2018 transitional tournament.

placed in Group B. In the Apertura the teams in each group played a single round robin while the Clausura was just the mirror of this pattern, that is, the same single round-robin but with each match's home-away status reversed. Also, in each round the team that would otherwise have had a bye (because the number of teams was odd) played its classic rival (which necessarily also had a bye in that round).

In 2017, the last time the schedules were defined manually, there were glaring differences in the distances travelled over the course of the tournament between the major and minor division teams of any single club and also between the teams of a given division in the same geographic region. Such disparities naturally tend to arise from the very structure of a single round-robin in which each team plays half of the other teams at home and the other half away.

The main objective of the scheduling efforts reported in this paper was therefore to define schedules that narrowed the differences in tournament-long distances travelled by the major and minor division teams of each club and simultaneously reduce the travel distance differences for the teams in each division located in the same geographic region, while also satisfying a series of conditions imposed by SAF officials.

Solution approach

In what follows, we describe the modelling and solution of the problem as it arose for the 28 teams in the 2018 transitional tournament; with only minor adjustments to the analysis, the 26-team case of 2018-2019 can also be solved. The approach is based on integer linear programming, the formal details of which are presented in the Appendix.

Since the computational difficulties of scheduling 28 teams are likely to be formidable, we take advantage of the pre-defined classification of the teams into two groups to decompose the original problem into two separate ones. Furthermore, we only solve the model for the major youth divisions given that, as explained earlier, once their schedules are determined the same solution but with the matches' home-away status reversed is used for the minor divisions. The incorporation into the modelling of conditions peculiar to the minor divisions is trivial.

To schedule the major divisions, then, we first model Group A and then Group B. Note, however, that their respective schedules are not independent, for once we have the results for Group A, the home-away status is also set for the "classics round" in which all the games are intergroup matches. This information has then to be incorporated into the model for Group B.

Reducing the travel distance differences between the major and minor divisions of each club's teams can be approached in more than one way. Our models incorporate the following two alternatives:

- The first alternative is to divide all of the teams into regional clusters defined by geographical proximity. The overall idea is that each team plays half of its matches against the other teams in each cluster (including its own) at home and the other half away. Say, for example, that a Cluster Y has four teams. A major division team i in Cluster X would then play one home game each against two different Cluster Y teams and one away game each against the other two. Since the same club matchups are used for the minor division teams but with their home-away status reversed, team i from Cluster X in these leagues would also play twice at home and twice away in four games against the Cluster Y teams. Given that all the teams in any given cluster are relatively close to each other, the travel distances of the club i teams in the different divisions will then be reasonably similar. However, since there is no way to guarantee *a priori* that there will exist a solution in which the home-away match combinations as just described are so precisely balanced, the objective function attempts to find the result that, if not perfect, best approximates this goal.

- The second alternative does not use geographical clusters but rather incorporates explicitly the distances in kilometres between the various clubs' home grounds. A distance matrix is created in order to specify an objective function that minimises the travel distances between each club's major and minor division teams.

Intuitively, the second alternative should be more precise but we would also expect it to be computationally more costly. As will be seen, both approaches bring about a narrowing of the differences in travel distances of teams within a single geographical group in each division.

Group A

For either alternative, the Group A version of the complete scheduling model is built around a base submodel that incorporates a series of conditions defining the general structure of the tournament plus a number of further conditions required by the SAF to ensure the resulting schedules are as fair to the teams and attractive to fans as possible. These conditions are as follows:

- The maximum number of consecutive matches a team can play with the same home-away status is two. Such a pair of matches is called a break.
- No team may have more than one home break and one away break over the course of the tournament.
- A break may not include either the second or the last round of the tournament.
- Matches between any two of the “strong” teams (which in Group A are River Plate, San Lorenzo and Independiente) may not be scheduled for either the first two or the last two rounds.
- No team may play two consecutive matches against strong teams.

Each club was also allowed to submit its own conditions, such as not scheduling distant away matches for certain rounds. Similarly, the SAF requested that such matches not be scheduled on weekdays given that most of the players in these leagues are still attending school.

Another requested condition was not to schedule away breaks in which both matches involve long-distance travel In other words, if a team plays two away games in a row, at least one of them must be against an adversary located relatively close to home.

The complete Group A model also contains submodels for implementing the two alternatives. These are described in what follows.

Geographic clusters As was explained above, the first of the alternative approaches to the problem is to group the teams in regional clusters by geographical proximity. Ideally, each team will play half of its matches against the teams from the same group in each cluster (including its own) at home and the other half away.

IBA Cluster	GBA Cluster
River Plate - Boca	Independiente - Racing
San Lorenzo - Huracn	Banfield - Lans
Vlez - Argentinos Jrs.	Defensa y Justicia - Arsenal
Tigre - Chacarita	Estudiantes - Gimnasia y Esgrima
Santa Fe Cluster	**Crdoba/Cuyo Cluster**
Rosario Central - Newell's	Talleres - Belgrano
Coln - Unin	San Martn SJ - Godoy Cruz
Cluster Extra	
Atltico Tucumn - Patronato	
Temperley - Olimpo	

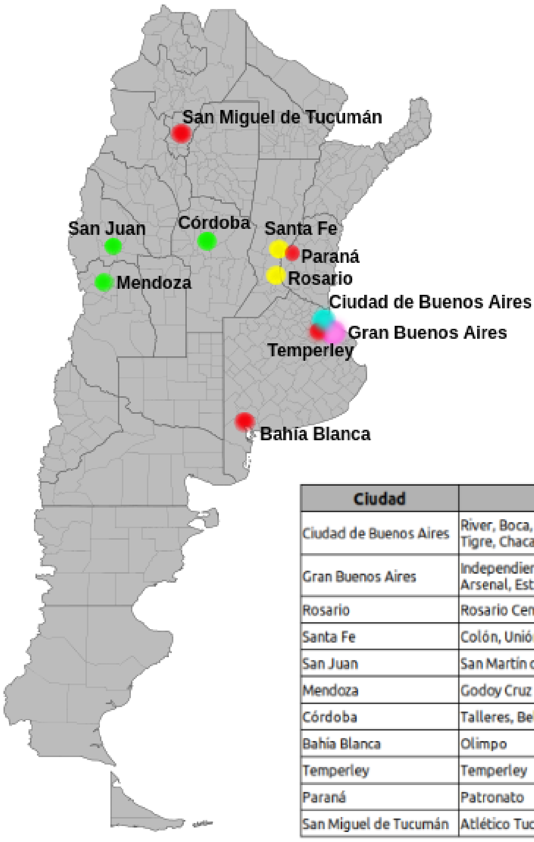
Table 2 **Geographic clusters for the 2018 transitional tournament.**

The geographical locations of the clubs in the 2018 transitional tournament resulted in the regional cluster definitions shown in Table 2. The location of each club is shown on the map in Figure 1.

Within each group, all of the clusters have an even number of teams. If a team in a given cluster is assigned to Group A, its classic rival, which is always in the same cluster, is assigned to Group B. This is illustrated in Table 2 where the teams within each cluster are arranged in pairs, the team on the left being in Group A and the team on the right its classic rival in Group B (see Table 1).

As can be seen on the map in Figure 1, teams within any single cluster are all located close to each other except those in the "Extra Cluster". In the latter case it was agreed that Atltico Tucumn and Temperley would be assigned to one group and Patronato and Olimpo to the other. This arrangement, though perhaps not ideal for all of the teams, is a workable solution for the Inner Buenos Aires (IBA) and Outer Buenos Aires (OBA) clusters where a clear majority of the clubs are located. For them, the teams in one of the groups (which turned out to be A) have one long trip (Atltico Tucumn) and one short one (Temperley) while for those in the other group, as compensation, there are two medium-length trips (Patronato, Olimpo).

With the teams thus grouped, the model attempts to have each team play half of the teams in each cluster in its group with one home-away status and the other half with



Ciudad	Equipos	Cluster
Ciudad de Buenos Aires	River, Boca, San Lorenzo, Huracán, Vélez, Argentinos Jrs., Tigre, Chacarita	CABA
Gran Buenos Aires	Independiente, Racing, Banfield, Lanús, Defensa y Justicia, Arsenal, Estudiantes, Gimnasia y Esgrima	Prov. Buenos Aires
Rosario	Rosario Central, Newell's	Santa Fe
Santa Fe	Colón, Unión	Santa Fe
San Juan	San Martín de San Juan	Córdoba/Cuyo
Mendoza	Godoy Cruz	Córdoba/Cuyo
Córdoba	Talleres, Belgrano	Córdoba/Cuyo
Bahía Blanca	Olimpo	Extra
Temperley	Temperley	Extra
Paraná	Patronato	Extra
San Miguel de Tucumán	Atlético Tucumán	Extra

Figure 1 **Map showing the geographic clusters for the 2018 transitional tournament.**

the opposite status. However, since the clusters all have an even number of teams, the matches each team plays against its own-group opponents in the same cluster are uneven in number. The imbalance that would otherwise result in home-away assignments is corrected using the classic rivalry matches. For example, within its own cluster River Plate plays San Lorenzo, Vlez and Tigre, the other three sides in its group, plus Boca, its classic rival in Group B. Ideally, it would play two of the above-named three teams away (at home), one of them at home (away) and then Boca also at home (away).

Another important point regarding Group A should be mentioned. Some of the teams in, for example, the major divisions might have to play away against Atltico Tucumn and San Martn de San Juan, the two clubs located furthest from most of their opponents (and which belong to different clusters despite being relatively close to each other). This by itself could result in a considerable travel distance difference between the major and minor divisions, especially for the Inner and Outer Buenos Aires teams, given that the (in this example) minor divisions, with the reverse home-away status, would play against Atltico

Tucumn and San Martn de San Juan at home. We therefore impose an additional constraint requiring that the home-away status of the pair of matches between the aforementioned two sides and any centrally located team cannot be the same. But note that the very fact this situation arises suggests that the clusters could have been constituted differently, a possibility we will return to in our conclusions.

Distance matrix The second alternative for addressing the problem is to minimise the actual differences in travel distance between the clubs' major and minor division teams. For this approach we define two possible objective functions. The first one minimises the sum, over all of the clubs, of the travel distance differences between each club's major and minor division teams, and the second one minimises the maximum difference between the major and minor division teams of the various clubs.

For either function, a matrix D is created containing the distances between the teams. The matrix is necessarily symmetric with all the elements along the diagonal equal to 0. The odd rows/columns represent the distances from the Group A teams while the even rows/columns represent the distances from the Group B teams. Also, the team in the row/column immediately following each Group A team is the latter's classic rival.

Group B

The base submodel for the complete Group B model is for the most part the same as the one we have just described for Group A, apart from a few modifications due to the sequential nature of the solution procedure (the solution to the first group solved being input to the second).

Another difference, however, is that in Group B the set of strong teams contains only two (Boca and Racing) rather than three as in Group A. The clusters are still used because they are needed by the constraints and no changes are made to the objective functions.

In the case of the constraints, those requested by the Group B clubs will obviously be different since by definition they relate to situations peculiar to each club. The specific conditions imposed by the SAF will also be formulated to reflect the teams' particular needs. The only new restrictions of a structural nature are those indicating the home-away status for each team in the classic rival matches (round 7), which as we saw earlier, was already determined by the Group A schedules.

The complete Group B model also requires submodels for implementing the two alternative approaches, that is, the geographical cluster and distance matrix options. These are constructed in the same fashion as was done for the Group A version.

Results

The models for both groups were solved using CPLEX 12.4 on a computer with an Intel Core i7 processor and 8 GB of RAM. Every instance of the clusters models was solved to optimality and run times ranged from 30 minutes to 10 hours. The objective function was in every case equal to 0, meaning that the models were indeed able to perfectly balance each team's home-away combinations for its matches with every cluster. As for the distance matrix models (both the sum and maximum versions), they were run in each case for 24 hours and the best solution then selected. In most cases these models were not solved to optimality but the SAF considered the resulting schedules to be highly satisfactory.

Tables 3 and 4 set out the results obtained for Group A and B, respectively, in the 2018 transitional tournament (with 28 teams) while Tables 5 and 6 do the same for the 2018-2019 Apertura tournament (with 26 teams). For each of the three models (clusters, sum OF, maximum OF) the tables report the travel distance (round trip) in kilometres for each club's major and minor divisions, the sum of the differences between the divisions over all of the clubs, and the maximum of these differences.

Club	**Clusters**			**Matriz (Sum)**			**Matriz (Maximum)**		
	Min.	**Maj.**	**Diff.**	**Min.**	**Maj.**	**Diff.**	**Min.**	**Maj.**	**Diff.**
River Plate	5014.2	3242.6	54.64%	4125.8	4131.0	0.13%	4080.4	4176.4	2.35%
San Lorenzo	3063.6	5179.4	**69.06%**	4094.2	4148.8	1.33%	4244.6	3998.4	6.16%
Vlez	3225.6	5179.4	55.08%	4072.2	4155.8	2.05%	4592.2	3635.8	**26.31%**
Tigre	5155.8	3255.8	58.36%	4225.0	4186.6	0.92%	4233.6	4178.0	1.33%
Independiente	5073.8	3208.4	58.14%	4147.8	4134.4	0.32%	4234	4048.2	4.59%
Banfield	3272.8	5100.4	55.84%	4186.2	4187.0	0.02%	4298.8	4074.4	5.51%
Defensa y Justicia	3303.4	5360.2	62.26%	4345.6	4318.0	0.64%	4132.8	4530.8	9.63%
Estudiantes	5652.8	3759.8	50.35%	4704.8	4707.8	0.06%	4729.4	4683.2	0.99%
Rosario Central	5450.0	5336.4	2.13%	5076.4	5710	**12.48%**	5320.4	5466.0	2.74%
Coln	6129.2	7118	16.13%	6597.2	6650.0	0.80%	6436.0	6811	5.83%
Talleres	9202.0	7596.2	21.14%	8308.2	8490.0	2.19%	8406.2	8392	0.17%
San Martn SJ	13400	14704	9.73%	13330	14774.0	10.83%	13944	14160	1.55%
Atltico Tucumn	15722.0	14918.0	5.39%	15268.0	15372.0	0.68%	14896.0	15744.0	5.69%
Temperley	4899.8	4776.4	2.58%	4826.6	4849.6	0.48%	5034.0	4642.2	8.44 %
Sum of diffs.	20146.6 km.			2665.8 km.			4199.4 km.		
Max. absolute diff.	2115.8 km. (San Lorenzo)			1444 km. (S. M. San Juan)			956.4 km. (Vlez)		

Table 3 Results for the 2018 transitional tournament, Group A.

To appreciate these results, we compare them to the differences in travel distance between the major and minor divisions in 2017, the tournament previous to the ones studied here when the SAF drew up the schedules using manual methods.

Club	Clusters			Matriz (suma)			Matriz (mximo)		
	Min.	Maj.	Diff.	Min.	Maj.	Diff.	Min.	Maj.	Diff.
Boca	4277.4	3496.0	23.30%	3900.6	3845.8	1.42%	4009.8	3736.6	7.31%
Huracn	3949.2	3717.6	6.23%	3604.4	4062.4	12.71%	3351	4315.8	**28.79%**
Argentinos Jrs.	3715.4	3893.2	4.79%	3922.4	3686.2	6.41%	3371.0	4237.6	25.71%
Chacarita	3506.8	4271.4	21.80%	3883.4	3894.8	0.29%	4312.4	3465.8	24.43%
Racing	3908.8	3811.2	2.56%	3724.0	3996.0	7.30%	3832.8	3887.2	1.42%
Lans	4201.4	3538.6	18.73%	3857.8	3882.2	0.63%	3929.9	3811.0	3.10%
Arsenal	3627.0	4250.8	17.20 %	4304.8	3573.0	**20.48%**	3803.8	4074.0	7.10%
Gimnasia y Esgrima	4248.6	4487.2	5.62%	4326.8	4409.0	1.90%	4186.8	4549.0	8.65%
Newell's	4732.4	5182	9.50%	5172.0	4742.4	9.06%	5118.0	4796.4	6.71%
Unin	6686	5780.2	15.67%	6680.0	5786.2	15.45%	6389.2	6077.0	5.14%
Belgrano	8392.2	8584	2.29%	8626.0	83500.2	3.30%	8646.0	8330.2	3.79%
Godoy Cruz	14008	13140	6.61%	13602.0	13546.0	0.41%	13984.0	13164.0	6.23%
Patronato	6213	8022	**29.12%**	7445	6790	9.65%	7459.0	6776.0	10.08%
Olimpo	10866	10192	6.61%	10540	10518	0.21%	9922.0	11136.0	12.42%
Sum of diffs.	8503.4 km.			4203.0 km.			7422.6 km.		
Max. absolute diff.	1809 km. (Patronato)			893.8 km. (Unin)			1214 km. (Olimpo)		

Table 4 Results for the 2018 transitional tournament, Group B.

Club	Clusters			Matriz (suma)			Matriz (mximo)		
	Min.	Maj.	Diff.	Min.	Maj.	Diff.	Min.	Maj.	Diff.
Boca	4602.8	4230.4	8.80%	4747.2	4086	16.18%	4634.8	4198.4	10.39%
Argentinos Jrs.	4232.4	4572.4	8.03%	4165.6	4639.2	11.37%	4464	4340.8	2.84%
Huracn	4212	4606.8	9.37%	4372.8	4446	1.67%	4410.4	4408.4	0.05%
Racing	4715.6	4104.2	14.90%	4965.6	3854.2	**28.84%**	4134.6	4685.2	13.32%
Lans	4102.6	4756.2	15.93%	4167	4691.8	12.59%	4221.2	4637.6	9.86%
Defensa y Justicia	5016.4	4176.2	20.12%	4374	4818.6	10.16%	4037.8	5154.8	**27.66%**
Gimnasia y Esgrima	4867.8	4787	1.69%	5074.6	4580.2	10.79%	5021.6	4633.2	8.38%
Aldosivi	9140	9662	5.71%	8352	10450	25.12%	9432	9370	0.66%
Rosario Central	4860	5948.4	22.40%	5212	5596.4	7.38%	5842.4	4966	17.65%
Coln	7018	6029.2	16.40%	6580	6467.2	1.74%	6379.2	6668	4.53%
Belgrano	8414.2	7912	6.35%	8716	7610.2	14.53%	7724	8602.2	11.37%
Godoy Cruz	11948	13568	13.56%	12760	12756	0.03%	12856	12660	1.55%
S. M. de Tucumn	15392	12003.2	**28.23%**	13290	14105.2	6.13%	13648	13747.2	0.73%
Sum of diffs.	11403.4 km.			8303.4 km.			5434.6 km.		
Max. absolute diff.	3388.8 km. (S. M. de Tucumn)			2098 km. (Aldosivi)			1117 km. (Defensa y Justicia)		

Table 5 Results for the 2018-2019 Apertura tournament, Group A.

Club	Clusters			Matriz (suma)			Matriz (mximo)		
	Min.	Maj.	Diff.	Min.	Maj.	Diff.	Min.	Maj.	Diff.
River Plate	4084.2	4890.2	19.73%	3525.2	5449.2	54.58%	4528.2	4446.2	1.84%
Vlez	4923.2	4042	21.80%	4546.8	4418.4	2.91%	4428.6	4536.6	2.44%
San Lorenzo	4885.8	4148.8	17.76%	5533.6	3501	58.06%	4407.2	4627.4	5.00%
Independiente	4094.6	5013.2	**22.43%**	4553.2	4554.6	0.03%	4895.4	4212.4	16.21%
Banfield	4279.6	4969.6	16.12%	4600	4649.2	1.07%	4670.6	4578.6	2.01%
Tigre	4108.4	5015.6	22.08%	5948.2	3175.8	**87.30%**	4981.6	4142.4	**20.26%**
Estudiantes	5471.6	4738.4	15.47%	5061.2	5148.8	1.73%	5169	5041	2.54%
Patronato	6202	6929	11.72%	6732	6399	5.20%	6570	6561	0.14%
Newell's	4992.4	4690	6.45%	4586.4	5096	11.11%	4988	4694.4	6.25%
Unin	5112.2	6162	20.54%	5236.2	6038	15.31%	5759	5515.2	4.42%
Talleres	7564	7092.2	6.65%	7612.2	7044	8.07%	7064.2	7592	7.47%
San Martn SJ	11948	13460	12.65%	12580	12828	1.97%	12608	12800	1.52%
Atltico Tucumn	13423.2	12104	10.90%	12319.2	13208	7.21%	12735.2	12792	0.45%
Sum of diffs.	11055.4 km			10345 km.			3475.4 km.		
Max. absolute diff.	1512 km. (S.M. de San Juan)			2772.4 km. (Tigre)			839.2 (Tigre)		

Table 6 Results for the 2018-2019 Apertura tournament, Group B.

Recall, first of all, that in 2017 the format was different. Although the 30 teams that season played a single round robin, they were not divided into two groups. The home-away status of the minor division matches was, however, the reverse of the scheduled for the major division as in the later tournaments, so the reduction of differences in travel distance between the major and minor divisions was a genuine issue that year as well.

The schedule for 2017 was drawn up using the system traditionally employed by many football leagues around the world, based on a match template in which numbers representing teams are allocated to the various game dates or rounds. The SAF simply held a draw to determine which team would be assigned to each number. No consideration was given either to reducing the travel distance disparities or to any of the other conditions set by the SAF for the models described here. The differences between the distances travelled by the major and minor division teams of each club in 2017 expressed in percentage terms are set out in Table 7.

River Plate	126.07%	Belgrano	44.26%
Boca	119.19%	Estudiantes	167.43%
Vlez	135.81%	Gimnasia y Esgrima	128.55%
San Martn SJ	10.95%	Independiente	113.81%
Godoy Cruz	1.92%	Racing	158.34%
Banfield	81.97%	San Lorenzo	**201.16%**
Lans	27.25%	Huracn	45.66%
Tigre	157.81%	Coln	44.49%
Defensa y Justicia	175.39%	Unin	30.92%
Arsenal	34.22%	Rosario Central	2.47%
Atltico Tucumn	39.81%	Newell's	5.98%
Patronato	28.83%	Aldosivi	**1.90%**
Temperley	4.36%	Atltico Rafaela	36.66%
Olimpo	26.55%	Quilmes	147.61%
Talleres	29.97%	Sarmiento	15.67%

Table 7 Percentage difference between distances travelled by major and minor division teams of each club, 2017.

As these figures show, the differences that year were very significant, ranging from 1.90% for Aldosivi to 201.16% for San Lorenzo. In simple terms, whereas for some clubs the travel distances of their major and minor division teams were practically the same, for others they differed by as much as a factor of three. By contrast, the schedules generated by our models for the great majority of clubs resulted in differences that were significantly smaller.

As well as reducing travel distance disparities between the major and minor divisions of each club, our models also achieved a better balance of distances between the different teams in each of the divisions within any one geographical region. Thus, teams close to each other travel a similar number of kilometres over the course of a tournament, a significant improvement in terms of schedule fairness. The reason for this result is that for a given pair of teams whose home venues are geographically in close proximity, the total distance each one would have travelled if both played all of their rivals away would be very similar. Since the models tend to generate travel distances for a club's teams in the various divisions of approximately 50% of that total, the distance the two teams travel are similar.

Tournament	Majors		Minors	
	Mean	Std Dev	Mean	Std Dev
2017: Manual	9903.3	3578.5	9022.7	3557.4
2018 transitional: Clusters	4152.1	697.4	4107.9	590.4
2018 transitional: Sum	4127.6	316.2	4132.5	308.7
2018 transitional: Maximum	4120.3	341.7	4139.8	414.7
2018-2019 Apertura: Clusters	4575.0	349.7	4542.6	427.4
2018-2019 Apertura: Sum	4429.5	585.0	4688.2	586.3
2018-2019 Apertura: Maximum	4545.9	283.5	4571.8	330.8

Table 8 Standard deviation of teams' travel distances by tournament: IBA and OBA regions

Tournament	Majors		Minors	
	Mean	Std Dev	Mean	Std Dev
2017: Manual	14744.1	2796.9	13255.1	1788.2
2018 transitional: Clusters	6100.1	1078.8	6029.7	736.4
2018 transitional: Sum	5935.7	739.9	6194.1	922.7
2018 transitional: Maximum	5985.3	774.5	6144.5	849.2
2018-2019 Apertura: Clusters	5707.4	592.3	5495.7	883.4
2018-2019 Apertura: Sum	5799.4	509.6	5403.7	727.4
2018-2019 Apertura: Maximum	5460.9	757.0	5742.2	496.2

Table 9 Standard deviation of teams' travel distances by tournament: Santa Fe region

To measure this effect, note first that the mean number of kilometres travelled on round trips by the major divisions in 2017 was more than double that for any of the later tournaments, regardless of which model would have been used to schedule them. This was to be expected since the 2017 season was an entire year and so had twice as many matches as 2018 or the 2018-2019 Apertura. We therefore use as our indicator the standard deviations of the distances travelled by the major and minor division teams by region for the last

Tournament	Majors		Minors	
	Mean	Std Dev	Mean	Std Dev
2017: Manual	27705.1	5450.2	23863.1	7048.2
2018 transitional: Clusters	11081.6	3004.4	11175.0	2476.5
2018 transitional: Sum	11290.1	2903.0	10966.6	2503.8
2018 transitional: Maximum	11011.6	2673.8	11245.1	2720.3
2018-2019 Apertura: Clusters	10508.1	3020.1	9968.6	2002.1
2018-2019 Apertura: Sum	10059.6	2739.9	10417.1	2287.4
2018-2019 Apertura: Maximum	10413.6	2344.3	10063.1	2680.6

Table 10 Standard deviation of teams' travel distances by tournament: Crdoba/Cuyo region

manually scheduled tournament (2017) with the following tournaments scheduled by our models.

The values for the various tournaments and modelling approaches are compared for the combined Inner and Outer Buenos Aires regions in Table 8. They show that the standard deviation in 2017 ranged from 5 times to more than 12 times greater than the values for the later tournaments in the major divisions and from 6 times to more than 11 times greater in the minor divisions. Clearly, the model-generated schedules succeeded in significantly reducing the differences in travel distance for the teams in the Buenos Aires regions.

For the region of Santa Fe, a province located northwest of the Buenos Aires urban area, the comparisons are given in Table 9. In this cluster, the standard deviation in 2017 ranged from 3 to more than 5 times greater than the values for the later, model-scheduled tournaments in the major decisions and was more than double in every case but one for the minor divisions. Again, these results reveal how little balance there was in the teams' travel distances under the manual schedules, and how much the situation improved when the schedules were defined by our models.

For a third region, that of Cuyo (San Juan and Mendoza) and Cordoba in the west central area of the country, the comparisons again reveal marked improvements when the models were used. The standard deviation in manually scheduled 2017 was in most cases roughly double the value for the later tournaments in the case of the major divisions and more or less triple in the minor divisions.

In addition to the better balance in travel distances, a number of other improvements were achieved by the models due to the incorporation of certain conditions. For example, in 2017 Defensa y Justicia played at home in both the first and second rounds, an undesirable situation that was prevented by model constraints from recurring in the later tournaments.

Another desirable constraint incorporated in the models was that no matches between the strong teams should be scheduled for the first two or last two rounds. The models satisfied this condition in every case whereas in the manual 2017 schedule, Racing played River Plate in the second round and Boca and Independiente met in the second-to-last one.

The models also included constraints ensuring that when away breaks were scheduled the travel distances for the two matches were relatively short, something that could not be achieved using manual methods. Just such a situation arose in 2017 in the major divisions when Coln (450 km northwest of the Buenos Aires urban area) played away against Aldosivi (400 km south of the Buenos Aires urban area) in the third round and Boca (Inner Buenos Aires) in the fourth, a total travel distance of 2,730 km for the two consecutive weekend matches. Another case that year was the minor divisions teams of Patronato (located in Entre Ros province), which were scheduled to play away against San Martn de San Juan (in San Juan province) in the fifth round and Aldosivi in the sixth, which meant a total travel distance of 3,748 km.

Conclusions and future research

An account was given of the mathematical models used to schedule the 2018 and 2019 football tournaments of the Argentinean youth leagues whose teams belong to the Argentina's first division professional clubs. The main objectives of the scheduling were to narrow the differences in distance travelled between the various teams belonging to a given club in the major (i.e., older) and minor (i.e., younger) divisions and between teams within a single division in the same region, while also satisfying a series of conditions imposed by the leagues' governing body.

Although the format of these youth tournaments is somewhat unconventional, the ideas applied in the proposed solution approaches can be extrapolated to the double round-robin season format commonly used by most football leagues around the world. Under this setup, the second tournament of the season is the same as the first but with the home-away status of the matches reversed, thus potentially producing the same type of travel distance disparities between the first and second tournaments that have been studied here.

As regards future research, models could be constructed that impose an upper bound on the sum over all the teams of the differences in travel distance, on the maximum such difference, or on both. Identifying appropriate upper bounds may well prove arduous, but suitable values could be derived based on the results already obtained. Thus, the reduction in travel distance differences would be pursued through the application of constraints, not through the objective function as in our models, and making explicit use of the distance matrix as in one of our two alternative methods. This would free up the objective function to be employed for some other purpose such as minimising the number of breaks over the course of the tournament.

Another promising idea *a priori* is to construct the clusters on some basis other than the pairing up of classic rivals. Thus, the clusters in the two groups could have different numbers of teams. This would require some modifications to the design of the model but could potentially generate very good results.

Yet another possibility would be to use a different approach for each group. For example, Group A of a tournament could be scheduled using the cluster approach while Group B would be scheduled using one of the two versions of the distance matrix approach. Better still, all possible combinations could be tried to find the best solution, which could include reversing the sequence in which the two groups are solved.

The results of our models for the 2018 transitional tournament and the 2018-2019 Apertura and Clausura tournaments were submitted to the Superliga Argentina de Ftbol, which then made the final decision on which ones to use in each case. The authors have also presented a series of proposed schedules for the 2019-2020 season, in which 24 teams will participate.

The satisfaction of the Superliga with the results of the mathematical scheduling project has led to the decision to broaden its application. According to Mariano Elizondo, president of the organisation, "in light of the success of the project with the youth leagues, we have extended our involvement with the group of mathematicians at the University of Buenos Aires to the scheduling of the professional league's first division, including the assignment of the exact days and kickoff times for the various matches in each round. We're very pleased with this collaboration with the University, which began almost two years ago, and hope it will continue in the future, including further possible extensions to other areas of the Superliga 's activities."

Acknowledgments

The work presented here was carried out under the terms of an agreement between the *Instituto de Clculo* of the Faculty of Exact and Natural Sciences at the University of Buenos Aires and the Superliga Argentina de Ftbol for provision to the latter of consulting services pertaining to the scheduling of sports tournaments. We are grateful to the Superliga for its constant collaboration in bringing this work to fruition, and wish particularly to thank Enrique Sacco and Juan Pablo Paterniti, the officials responsible at Superliga for tournament scheduling. The first author of this paper was partly financed by the Instituto Sistemas Complejos de Ingeniería (ISCI), Chile (CONICYT PIA/BASAL AFB180003) and grant nos. UBACyT 20020170100495BA (Argentina) and ANPCyT PICT 2015-2218 (Argentina).

Appendix. Integer programming models

In what follows we present in more formal terms the model used to schedule the 2018 transitional tournament, which featured 28 teams. The fictitious extra team added to represent the matches played by the Group A real teams against their respective classic rivals in Group B will be referred to as "team number 15".

We begin with the *base submodel*, used for scheduling both groups together with the submodels for the alternative cluster and distance matrix approaches. This model requires the following sets:

- Rounds: $\mathrm{R} = \{1, \ldots, 14\}$.
- Teams: $\mathrm{T} = \{1, \ldots, 15\}$.
- Real teams: $\mathrm{T}^* = \{1, \ldots, 14\}$.
- Strong teams: $\mathrm{BT} = \{1, 2, 5\}$.
- Inner Buenos Aires: $\mathrm{IBA} = \{1, \ldots, 4\}$.
- Outer Buenos Aires: $\mathrm{OBA} = \{5, \ldots, 8\}$.
- Santa Fe: $\mathrm{SF} = \{9, 10\}$.
- Cordoba/Cuyo: $\mathrm{CC} = \{11, 12\}$.
- Extra: $\mathrm{E} = \{13, 14\}$.

The model contains four variables as set out below. For each team $i \in \mathrm{T}$, $j \in \mathrm{T}$ and every round $k \in \mathrm{R}$, these variables take the following values:

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

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

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

$$\boldsymbol{z_{ij}} = \begin{cases} 1 \text{ if team } i \text{ plays at home against } j, \\ 0 \;\text{ otherwise.} \end{cases}$$

The model also includes a number of constraints. The first group of constraints are those that govern the relationships between the variables:

$$z_{ij} = \sum_{k \in \mathrm{R}} x_{ijk} \quad \forall\, i,j \in \mathrm{T} \tag{1}$$

$$\sum_{j \in \mathrm{T}} x_{ijk} + \sum_{l \in \mathrm{T}} x_{ilk+1} \leq 2(1 - w_{ik}) \quad \forall\, i \in \mathrm{T}^*, k \in \mathrm{R}, k \neq 14 \tag{2}$$

$$\sum_{j \in \mathrm{T}} x_{ijk} + \sum_{l \in \mathrm{T}} x_{ilk+1} \geq 1 - w_{ik} \quad \forall\, i \in \mathrm{T}^*, k \in \mathrm{R}, k \neq 14 \tag{3}$$

$$\sum_{j \in \mathrm{T}} x_{jik} + \sum_{l \in \mathrm{T}} x_{lik+1} \leq 2.(1 - y_{ik}) \quad \forall i \in \mathrm{T}^*, k \in \mathrm{R}, k \neq 14 \tag{4}$$

$$\sum_{j \in \mathrm{T}} x_{jik} + \sum_{l \in \mathrm{T}} x_{lik+1} \geq 1 - y_{ik} \quad \forall i \in \mathrm{T}^*, k \in \mathrm{R}, k \neq 14 \tag{5}$$

The variable $\boldsymbol{z_{ij}}$, defined in (1), is introduced solely for the purpose of simplifying the notation. If team i plays at home against team j in round k, then obviously team i plays at home in that round against team j.

Recall that variable $\boldsymbol{w_{ik}}$ is equal to 1 if team i has an away break beginning round k and 0 otherwise. Then if $\boldsymbol{w_{ik}} = 1$, constraint (2) ensures that team i cannot play at home in either round k or $k+1$. Observe, however, that if $\boldsymbol{w_{ik}} = 0$, the equation does not indicate anything that is not already known, so in this case another constraint is needed to clarify the role of $\boldsymbol{x_{ijk}}$. Also, if $\boldsymbol{w_{ik}} = 0$, team i does not have an away break beginning round k, meaning that one of the two matches in rounds k and $k+1$ must be at home. This is ensured by constraint (3). Constraints (4) and (5) are analogous to (2) and (3) for the case of home breaks.

Constraints (2)-(5) express the conditions required for the teams $i \in \mathrm{T}^*$ for rounds $k \in \mathrm{R}$ where $k \neq 14$. The set of real teams is specified here because the fictitious team only plays in one round so no reference to its breaks is necessary, while the inequality is needed because no break can start in $k = 14$, it being the last round of the tournament.

Below is the second group of constraints, those that define the global structure of the tournament.

$$x_{iik} = 0 \quad \forall\, i \in \mathrm{T}, k \in \mathrm{R} \tag{6}$$

$$z_{ij} + z_{ji} = 1 \quad \forall\, i,j \in \mathrm{T},\ \ i \neq j \tag{7}$$

$$\sum_{\substack{j \in \mathrm{T} \\ i \neq j}} x_{ijk} + x_{jik} = 1 \quad \forall\, i \in \mathrm{T}^*, k \in \mathrm{R} \tag{8}$$

$$\sum_{i \in \mathrm{T}^*} \sum_{\substack{j \in \mathrm{T}^* \\ j \neq i}} x_{ijk} = 7 \quad \forall\, k \in \mathrm{R}, k \neq 7 \tag{9}$$

$$\sum_{\substack{j \in \mathrm{T} \\ j \neq i}} z_{ij} = 7 \quad \forall\, i \in \mathrm{T} \tag{10}$$

$$\sum_{j \in \mathrm{T}} x_{15,j,7} = 7 \tag{11}$$

$$\sum_{j \in \mathrm{T}} x_{j,15,7} = 7 \tag{12}$$

$$\sum_{j\in \mathrm{T}}\sum_{\substack{k\in \mathrm{R}\\ k\neq 7}} x_{j,15,k} + x_{15,j,k} = 0 \tag{13}$$

Constraint (6) specifies that no team can play against itself. Constraint (7) imposes that every team play every other team once (recall that the model is for just one of two groups). Given a team i (other than the fictitious one) and a round k, constraint (8) requires that i must have a single rival to play against, whether at home or away. Constraint (9) ensures that in each round, 7 matches are played (recall that there are 14 teams), and in each match one team must obviously play at home. For the classic rivals round and the fictitious team this restriction is handled differently, as will be seen below.

Constraint (10) requires that each team $i \in \mathrm{T}$ play 7 matches at home. Since we have already imposed (by (8)) that team i must play in every round and that the total number of rounds is 14, it would be redundant to have another constraint imposing that each team also play 7 matches away.

As just noted above, the fictitious team, denoted number 15, plays only in round 7 (chosen by the SAF to be the classic rivals round) and is therefore subject to special treatment. So that the real teams play 7 matches at home and 7 away over the whole tournament, number 15 must do the same in round 7, which is ensured by constraints (7), (11) and (12).

Having presented all of the constraints needed for the basic format of the tournament, we now turn to those relating to breaks and the special conditions applied to the strong teams, as required by the SAF.

$$y_{ik} + y_{ik+1} \leq 1 \qquad \forall\, i \in \mathrm{T}^*, k \in \mathrm{R}, k \neq 14 \tag{14}$$

$$w_{ik} + w_{ik+1} \leq 1 \qquad \forall\, i \in \mathrm{T}^*, k \in \mathrm{R}, k \neq 14 \tag{15}$$

$$\sum_{k\in \mathrm{R}} y_{ik} \leq 1 \qquad \forall\, i \in \mathrm{T}^* \tag{16}$$

$$\sum_{k\in \mathrm{R}} w_{ik} \leq 1 \qquad \forall\, i \in \mathrm{T}^* \tag{17}$$

$$y_{i1} + w_{i1} + y_{i13} + w_{i13} = 0 \qquad \forall\, i \in \mathrm{T}^* \tag{18}$$

$$x_{ij1} + x_{ji1} + x_{ij2} + x_{ji2} + x_{ij13} + x_{ji13} + x_{ij14} + x_{ji14} = 0 \qquad \forall\, i,j \in \mathrm{BT} \tag{19}$$

$$1 - x_{ijk+1} - x_{jik+1} \geq x_{ilk} + x_{lik} \qquad \forall\, i \in \mathrm{T}^*, j,l \in \mathrm{BT}, k \in \mathrm{R} \tag{20}$$

$$\sum_{j\in \mathrm{CABA}\cup\mathrm{BA}} x_{1j3} + x_{j1,3} = 1 \tag{21}$$

$$\sum_{j\in \mathrm{CABA}\cup\mathrm{BA}} x_{3j3} + x_{j3,3} = 1 \tag{22}$$

$$\sum_{j\in \mathrm{CABA}\cup\mathrm{BA}} x_{2j1} + x_{j2,1} = 1 \tag{23}$$

$$\sum_{j\in \mathrm{CABA}\cup\mathrm{BA}} x_{2j6} + x_{j2,6} = 1 \tag{24}$$

$$\sum_{j\in \mathrm{CABA}\cup\mathrm{BA}} x_{14j1} + x_{j14,1} = 1 \tag{25}$$

$$\sum_{j\in \mathrm{CABA}\cup\mathrm{BA}} x_{14j4} + x_{j14,4} = 1 \tag{26}$$

$$\sum_{j\in \mathrm{CABA}\cup\mathrm{BA}} x_{14j6}+x_{j14,6}=1 \tag{27}$$

$$x_{13,12,11}+x_{12,13,11}=1 \tag{28}$$

$$x_{9,4,11}+x_{4,9,11}=1 \tag{29}$$

$$x_{11,10,11}+x_{10,11,11}=1 \tag{30}$$

Note once again that the above constraints on breaks apply only to the real teams, that is, those belonging to the set $\mathrm{T}^*=\{1,\dots,14\}$, given that the fictitious team has no breaks since it plays in only one round.

The SAF-imposed condition that no team may play more than two consecutive matches with the same home-away status is captured by constraints (16) and (17). They also ensure that no team plays three times in a row at home or away, which would be the equivalent of two consecutive breaks. Constraint (18) specifies that no breaks may be scheduled to start in either the first or the last round of the tournament. Put another way, no team may play twice in a row with the same home-away status in rounds 1 and 2 or rounds 13 and 14.

Constraint (19) prohibits the scheduling of matches between strong teams (recall that in Group A, these are River Plate, San Lorenzo and Independiente) for either the first two or the last two rounds of the tournament. The idea here is to save these games, which are especially "attractive" to fans, for other stages of the tournament. Constraint (20) ensures that no real team may play two matches in a row against a strong team, regardless of their home-away status.

Next, we have the following restrictions requested by the teams:

• Constraint (21): River Plate requested that their third round match not involve long-distance travel, meaning in effect that the game must be against an IBA or an OBA team.

• Constraints (22) – (27): Analogous to (21), but requested by Vlez for the third round, San Lorenzo for the first and sixth rounds, and Temperley for the first, fourth and sixth rounds.

• Constraint (28): Atltico Tucumn requested that it play against San Martn de San Juan in round 11.

• Constraints (29) and (30): Analogous to (28), but for Rosario Central vs. Tigre and Coln vs. Talleres

The following additional constraints are needed to ensure that in the case of an away break, not more than one of the matches involves long-distance travel:

$$w_{ik}\le\left(\sum_{j\in \mathrm{CABA}\cup\mathrm{BA}} x_{jik}+x_{j,i,k+1}\right)+x_{15,i,k}+x_{15,i,k+1}\quad \forall i\in \mathrm{CABA}\cup\mathrm{BA},\ \forall k\in R, k\neq 14 \tag{31}$$

$$w_{ik}\le\left(\sum_{j\in \mathrm{SF}} x_{jik}+x_{j,i,k+1}\right)+x_{15,i,k}+x_{15,i,k+1}\quad \forall i\in \mathrm{SF},\ \forall k\in R, k\neq 14 \tag{32}$$

$$w_{11k}\le\left(\sum_{j\in \mathrm{CC}\cup\mathrm{SF}} x_{j,11,k}+x_{j,11,k+1}\right)+x_{15,11,k}+x_{15,11,k+1}\quad \forall k\in R, k\neq 14 \tag{33}$$

$$w_{14k}\le\left(\sum_{j\in \mathrm{CABA}\cup\mathrm{BA}} x_{j,14,k}+x_{j,14,k+1}\right)\quad \forall k\in R, k\neq 14 \tag{34}$$

Constraint (34) applies to the classic rival match between Temperley and Olimpo (in Bahia Blanca), preventing it from being part of a break given that although the two teams are in the same cluster, they are located relatively far from each other.

Note also that although these constraints were said above to apply to away breaks, they must also be imposed for breaks at home. The reason is that the model as presented here is written for the major divisions whereas the minor divisions, although they follow the same schedule, do so with the home-away status reversed so that home breaks scheduled for the majors imply away breaks for the minors. Thus,

$$y_{ik} \leq \left(\sum_{j \in \mathrm{CABA} \cup \mathrm{BA}} x_{ijk} + x_{i,j,k+1} \right) + x_{i,15,k} + x_{i,15,k+1} \quad \forall i \in \mathrm{CABA} \cup \mathrm{BA},\ \forall k \in R, k \neq 14 \tag{35}$$

$$y_{ik} \leq \left(\sum_{j \in \mathrm{SF}} x_{ijk} + x_{i,j,k+1} \right) + x_{i,15,k} + x_{i,15,k+1} \quad \forall i \in \mathrm{SF},\ \forall k \in R, k \neq 14 \tag{36}$$

$$y_{11k} \leq \left(\sum_{j \in \mathrm{CC} \cup \mathrm{SF}} x_{11,j,k} + x_{11,j,k+1} \right) + x_{11,15,k} + x_{11,15,k+1} \quad \forall k \in R, k \neq 14 \tag{37}$$

$$y_{14k} \leq \left(\sum_{j \in \mathrm{CABA} \cup \mathrm{BA}} x_{14,j,k} + x_{14,j,k+1} \right) \quad \forall k \in R, k \neq 14 \tag{38}$$

Clusters submodel

The objective of the clusters submodel as explained earlier is that each team plays half of its matches against the other teams in each cluster at home and the other half away, using the classic rivals intergroup match to maintain this balance for each team's matches within its own cluster. The objective function attempts to reduce the differences between the number of matches of each home-away status that each team plays against its opponents both in its own and the other clusters. Thus, in the ideal case the OF value will be 0.

This model contains the following variables:

$$Caba_i^+,\, Caba_i^- \;\; \forall i \in T^*.$$
$$intraCaba_i^+,\, intraCaba_i^- \;\; \forall i \in CABA.$$
$$BA_i^+,\, BA_i^- \;\; \forall i \in T^*.$$
$$intraBA_i^+,\, intraBA_i^- \;\; \forall i \in BA.$$
$$SF_i^+,\, SF_i^+ \;\; \forall i \in T^*.$$
$$intraSF_i^+,\, intraSF_i^- \;\; \forall i \in SF.$$
$$CC_i^+,\, CC_i^- \;\; \forall i \in T^*.$$
$$intraCC_i^+,\, intraCC_i^- \;\; \forall i \in CC.$$
$$E_i^+,\, E_i^- \;\; \forall i \in T^*.$$
$$intraE_i^+,\, intraE_i^- \;\; \forall i \in E.$$

Note that in this set of variables there are two for each cluster, identifiable by their abbreviations. In the case of the IBA cluster, for example, $Caba_i^+ + Caba_i^-$ is the difference between the number of matches team i not in IBA plays against IBA teams away and the number it plays against them at home. $intraCaba_i^+ + intraCaba_i^-$, on the other hand, is the difference between the number of matches team i *in* IBA plays against other IBA teams away and the number it plays against them at home (counting in the latter the classic rivals intergroup match). The variables for the other clusters are defined analogously.

These new variables are related to the ones in the base submodel by the following constraints:

$$\begin{aligned}
Caba_i^+ - Caba_i^- &= \sum_{j\in \mathrm{CABA}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{T}^* \\
intraCaba_i^+ - intraCaba_i^- &= \sum_{\substack{j\in \mathrm{CABA}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{CABA} \\
BA_i^+ - BA_i^- &= \sum_{j\in \mathrm{BA}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{T}^* \\
intraBA_i^+ - intraBA_i^- &= \sum_{\substack{j\in \mathrm{BA}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{BA} \\
SF_i^+ - SF_i^- &= \sum_{j\in \mathrm{SF}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{T}^* \\
intraSF_i^+ - intraSF_i^- &= \sum_{\substack{j\in \mathrm{SF}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{SF} \\
CC_i^+ - CC_i^- &= \sum_{j\in \mathrm{CC}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{T}^* \\
intraCC_i^+ - intraCC_i^- &= \sum_{\substack{j\in \mathrm{CC}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{CC} \\
E_i^+ - E_i^- &= \sum_{j\in \mathrm{E}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{T}^* \\
intraE_i^+ - intraE_i^- &= \sum_{\substack{j\in \mathrm{E}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji} \quad \forall\, i \in \mathrm{E}
\end{aligned}$$

As can be observed, the right-hand side of these constraints is a value that may be negative, positive or zero, and represents the difference between the number of home and away matches played by team i against the corresponding set of j teams. The use of superindexes "+" and "-" reflects the fact that the objective function in the program below includes the modulus of these values. In its basic form, the OF to be minimised can be written as follows:

$$\begin{aligned}
\min \quad & \sum_{i\in \mathrm{T}^*\setminus \mathrm{CABA}} \Big|\sum_{j\in \mathrm{CABA}} z_{ij} - z_{ji}\Big| + \sum_{i\in \mathrm{T}^*\setminus \mathrm{BA}} \Big|\sum_{j\in \mathrm{BA}} z_{ij} - z_{ji}\Big| \\
& + \sum_{i\in \mathrm{T}^*\setminus \mathrm{SF}} \Big|\sum_{j\in \mathrm{SF}} z_{ij} - z_{ji}\Big| + \sum_{i\in \mathrm{T}^*\setminus \mathrm{CC}} \Big|\sum_{j\in \mathrm{CC}} z_{ij} - z_{ji}\Big| \\
& + \sum_{i\in \mathrm{T}^*\setminus \mathrm{E}} \Big|\sum_{j\in \mathrm{E}} z_{ij} - z_{ji}\Big| + \sum_{i\in \mathrm{CABA}} \Big|\sum_{\substack{j\in \mathrm{CABA}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji}\Big| \\
& + \sum_{i\in \mathrm{BA}} \Big|\sum_{\substack{j\in \mathrm{BA}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji}\Big| + \sum_{i\in \mathrm{SF}} \Big|\sum_{\substack{j\in \mathrm{SF}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji}\Big| \\
& + \sum_{i\in \mathrm{CC}} \Big|\sum_{\substack{j\in \mathrm{CC}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji}\Big| + \sum_{i\in \mathrm{E}} \Big|\sum_{\substack{j\in \mathrm{E}\cup\{15\}\\ j\neq i}} z_{ij} - z_{ji}\Big|.
\end{aligned}$$

Using the definitions given above, this can be rewritten as follows:

$$
\begin{aligned}
\min \quad & \sum_{i\in \mathrm{T}^*\setminus \mathrm{CABA}} |Caba_i^+ - Caba_i^+| \; + \sum_{i\in \mathrm{T}^*\setminus \mathrm{BA}} |BA_i^+ - BA_i^+| \\
& + \sum_{i\in \mathrm{T}^*\setminus \mathrm{SF}} |SF_i^+ - SF_i^+| \; + \sum_{i\in \mathrm{T}^*\setminus \mathrm{CC}} |CC_i^+ - CC_i^+| \\
& + \sum_{i\in \mathrm{T}^*\setminus \mathrm{E}} |E_i^+ - E_i^+| \; + \sum_{i\in \mathrm{CABA}} |intraCABA_i^+ - intraCABA_i^-| \\
& + \sum_{i\in \mathrm{BA}} |intraBA_i^+ - intraBA_i^-| \; + \sum_{i\in \mathrm{SF}} |intraSF_i^+ - intraSF_i^-| \\
& + \sum_{i\in \mathrm{CC}} |intraCC_i^+ - intraCC_i^-| \; + \sum_{i\in \mathrm{E}} |intraE_i^+ - intraE_i^-|.
\end{aligned}
$$

Since these functions contain absolute values, they are not linear. To linearise them we use the fact that

$$\min \left\{ \sum_{i=1}^{n} |x_i| : Ax \leq b,\; x \in \mathbb{R}^n \right\}$$

is equivalent to

$$\min \left\{ \sum_{i=1}^{n} x_i^+ + x_i^- : A(x^+ - x^-) \leq b,\; x_i^+, x_i^- \geq 0 \text{ para } i = 1, \ldots, n \right\}.$$

We can therefore rewrite the OF again as

$$
\begin{aligned}
\min \quad & \sum_{i\in \mathrm{T}^*\setminus \mathrm{CABA}} Caba_i^+ + Caba_i^+ \; + \sum_{i\in \mathrm{T}^*\setminus \mathrm{BA}} BA_i^+ + BA_i^+ \\
& + \sum_{i\in \mathrm{T}^*\setminus \mathrm{SF}} SF_i^+ + SF_i^+ \; + \sum_{i\in \mathrm{T}^*\setminus \mathrm{CC}} CC_i^+ + CC_i^+ \\
& + \sum_{i\in \mathrm{T}^*\setminus \mathrm{E}} E_i^+ + E_i^+ \; + \sum_{i\in \mathrm{CABA}} intraCABA_i^+ + intraCABA_i^- \\
& + \sum_{i\in \mathrm{BA}} intraBA_i^+ + intraBA_i^- \; + \sum_{i\in \mathrm{SF}} intraSF_i^+ + intraSF_i^- \\
& + \sum_{i\in \mathrm{CC}} intraCC_i^+ + intraCC_i^- \; + \sum_{i\in \mathrm{E}} intraE_i^+ + intraE_i^-.
\end{aligned}
$$

A final constraint required by the clusters approach in addition to those in the base submodel is that the home-away status of matches pitting IBA and OBA teams against San Martn de San Juan and Atltico Tucumn be different. This is necessary because the latter two are located far from the former but are in different clusters. Without this condition, large disparities in travel distance could result between the major and minor divisions for the Inner and Outer Buenos Aires teams.

$$z_{12i} = z_{i13} \;\; \forall i \in \mathrm{CABA} \cup \mathrm{BA} \tag{39}$$

Distance matrix submodel

The main component of the distance matrix submodel is a variable d_i for each team i that represents the difference in travel distance (outbound trip only) between the major and minor divisions, as follows:

$$d_i = \sum_{j \in \mathrm{T}^*} D_{2i-1,2j-1}(2z_{ji} - 1) + D_{2i-1,2i}(2z_{15,i} - 1)$$

This definition of d_i includes only the outbound trip distance given that including the return trip would simply mean multiplying the objective function by a constant, which would not affect the optimal solution. If, now, team i plays against team j away, the travel distances of the two are summed. This is so because $\boldsymbol{z_{ji}} = 1$ since j is at home, meaning the expression $2z_{ji} - 1$ in the above equation is also equal to 1 so that the matrix value $D_{2i-1,2j-1}$ is added. And since i and j belong to Group A and set $\{1, \ldots, 14\}$, its value is equal to the distance between the two teams' home grounds). By contrast, if i plays against j at home, then $2z_{ji} - 1$ and the aforementioned distance is subtracted. This is equivalent to summing the trip distances travelled by the major (minor) divisions and subtracting those travelled by the minor (major) divisions given the opposite home-away status of their respective matches. In the case of the special match in round 7, if i plays away then $2z_{15,i} - 1 = 1$ and the distance between the team and its classic rival (equal to $(D_{2i-1,2i})$ is added, and if i plays at home, that distance is subtracted. Therefore, d_i is effectively the difference between the (one way) distance travelled by the major and minor divisions of team $i \in \mathrm{T}^*$. Note finally that d_i is not necessarily positive.

Also included for each team i are the variables $\boldsymbol{d_i^+}$ and $\boldsymbol{d_i^-}$, which are defined as follows:

$$d_i^+ - d_i^- = \sum_{j \in \mathrm{T}^*} D_{2i-1\,2j-1}.(2z_{ji} - 1) + D_{2i-1\,2i}.(2z_{15i} - 1) \;\; \forall\, i \in \mathrm{T}^*$$

$$d_i^+, d_i^- \geq 0 \;\; \forall\, i \in \mathrm{T}^*$$

With these definitions we can now specify the two objective functions. For the first model (sum of differences), the OF is

$$\min \sum_{i \in \mathrm{T}^*} |d_i|,$$

whereas for the second model (minimisation of the maximum difference), the OF is

$$\min \max_{i \in \mathrm{T}^*} |d_i|.$$

Neither of these functions are linear due to the presence of absolute values in both of them and the maximum in the second one. We must therefore reformulate them to satisfy the requirements of this type of modelling. To linearise the first OF we need only rewrite the modulus as we did for the clusters model. To linearise the second OF we make use of the fact that the model

$$\min \; \left\{ \max\{|x_1|, \ldots, |x_n| : Ax \geq b, \; x \in \mathbb{R}^n \; \right\}$$

is equivalent to the model

$$\min \left\{ M : Ax \le b,\ x_i \le M,\ -M \le x_i \text{ para } i = 1, \ldots, n \right\},$$

which in turn is equivalent to

$$\min \left\{ M : A(x^+ - x^-) \le b,\ x_i^+ + x_i^- \le M,\ x_i^+, x_i^- \ge 0 \text{ para } i = 1, \ldots, n \right\}.$$

These linearised forms of the two objective functions can then be used for the models based on minimising actual travel distances.

References

Alarcón F, Durán G, Guajardo M, Miranda J, Muñoz H, Ramírez L, Ramírez M, Sauré D, Siebert M, Souyris S, et al. (2017) Operations research transforms the scheduling of chilean soccer leagues and south american world cup qualifiers. *Interfaces* 47(1):52–69.

Bonomo F, Cardemil A, Durán G, Marenco J, Sabán D (2012) An application of the traveling tournament problem: The argentine volleyball league. *Interfaces* 42(3):245–259.

Burrows W, Tuffley C (2015) Maximising common fixtures in a round robin tournament with two divisions. *arXiv preprint arXiv:1502.06664* .

Cocchi G, Galligari A, Nicolino FP, Piccialli V, Schoen F, Sciandrone M (2018) Scheduling the italian national volleyball tournament. *Interfaces* 48(3):271–284.

Durán G, Durán S, Marenco J, Mascialino F, Rey PA (2019) Scheduling argentina's professional basketball leagues: A variation on the travelling tournament problem. *European Journal of Operational Research* 275(3):1126–1138.

Durán G, Guajardo M, Sauré D (2017) Scheduling the south american qualifiers to the 2018 fifa world cup by integer programming. *European Journal of Operational Research* 262(3):1109–1115.

Easton K, Nemhauser G, Trick M (2001) The traveling tournament problem: Description and benchmarks. *Principles and Practice of Constraint Programming*, 580–585 (Springer).

Fry MJ, Ohlmann JW (2012) Introduction to the special issue on analytics in sports, part ii: Sports scheduling applications.

Grabau M (2012) Softball scheduling as easy as 1-2-3 (strikes you're out). *Interfaces* 42(3):310–319.

Knust S (2010) Scheduling non-professional table-tennis leagues. *European Journal of Operational Research* 200(2):358–367.

Nurmi K, Goossens D, Kyngäs J (2014) Scheduling a triple round robin tournament with minitournaments for the finnish national youth ice hockey league. *Journal of the Operational Research Society* 65(11):1770–1779.

Schönberger J (2017) The championship timetabling problem-construction and justification of test cases. *Proceedings of MathSport International 2017 Conference*, 330.

Schönberger J, Mattfeld DC, Kopfer H (2004) Memetic algorithm timetabling for non-commercial sport leagues. *European Journal of Operational Research* 153(1):102–116.

Toffolo TA, Christiaens J, Spieksma FC, Berghe GV (2019) The sport teams grouping problem. *Annals of Operations Research* 275(1):223–243.

Van Bulck D, Goossens D, Schönberger J, Guajardo M (2019a) Robinx: A three-field classification and unified data format for round-robin sports timetabling. *European Journal of Operational Research* .

Van Bulck D, Goossens DR, Spieksma FC (2019b) Scheduling a non-professional indoor football league: a tabu search based approach. *Annals of Operations Research* 275(2):715–730.

Wright M (2018) Scheduling an amateur cricket league over a nine-year period. *Journal of the Operational Research Society* 69(11):1854–1862.