ANALYSIS, DISCUSSION, AND FORMULATION OF SPORTS SCHEDULING CONSIDERATIONS, ALGORITHMS, AND PROGRAMMING METHODS

A Master's Project
Presented to
The Faculty of California Polytechnic State University,
San Luis Obispo

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Engineering with Integrated Technology Management
Specialization

by
Branden Shimamoto
June 2015

© 2015

Branden Shimamoto

ALL RIGHTS RESERVED

COMMITTEE MEMBERSHIP

TITLE:	Analysis, Discussion, and Formulation of Sports Scheduling Considerations, Algorithms, and Programming Methods
AUTHOR:	Branden Shimamoto
DATE SUBMITTED:	June 2015
COMMITTEE CHAIR:	Tali Freed, PhD Professor of Industrial Engineering
COMMITTEE MEMBER:	Jianbiao Pan, PhD Professor of Industrial Engineering
COMMITTEE MEMBER:	

Abstract

Analysis, Discussion, and Formulation of Sports Scheduling Considerations, Algorithms, and Programming Methods

Branden Shimamoto

Sports scheduling is a problem that often involves a very wide spanning solution set. There are many different combinations of home and away assignments, matchups, and opponents that can take place. This problem is commonly done manually, however, in recent years, the use of operations research techniques and methods have provided a more efficient process and more optimal solutions. Linear programming is often used to achieve desired optimization while remaining within the confines of the rules set by the league and its stakeholders. Large linear programming models can require excessive amounts of time so the use of algorithms and decomposition are used in order to generate schedules in a reasonable amount of time. This paper analyzes the strengths, weaknesses, and concerns of the use of linear programming and algorithms to schedule different professional sports leagues. It also discusses the considerations or constraints that are set along with the criteria for optimization regarded in each situation. Lastly, this paper provides a case study in which a proposed solution for the 2014 FIFA World Cup is presented along with the respective linear programming model, method of decomposition, and considerations.

Keywords: Operations Research, Linear Programming, Sports Scheduling

ACKNOWLEDGMENTS

Tali Freed	Project Advisor Providing guidance, assistance, and insight throughout the duration of the project
Fernando Calderon	Recommendations and research in case study portion of the paper
California Polytechnic State University, San Luis Obispo Industrial and Manufacturing Engineering Department	Providing the resources and facilities necessary for the completion this project

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. Introduction

The sports industry has grown to become one of the largest industries worldwide. In today's globalized world, sports have become increasingly prevalent and the industry generates over 100 billion dollars (USD) annually. Competitions range from recreational leagues on smaller scales to professional tournaments involving participants from all corners of the globe. While recreational leagues may only contain participants from a single city and location, professional competitions have many pertinent implications involving millions of fans, lucrative player salaries, stadium locations, television broadcasting, advertising, and other related logistical and financial components.

In recent years, the use of mathematics in the sports industry has become more and more widespread. A few examples of mathematics being used in sports are to predict the outcomes of games to set betting odds and payouts, to assemble high performing optimal teams, and to create effective metrics to gage performance. Another use of mathematics in sports is to create schedules. Sports scheduling most commonly uses methods regarding operations research, programming, and heuristic algorithms in order to generate feasible, optimal schedules in a practical amount of time.

Reaching a feasible schedule can be extremely challenging due to the large amount of interrelated requirements, considerations, and resource limitations that need to be addressed. These restrictions include but are not limited to location, availability, television broadcasting, holidays, home and away game assignments, stadium personnel, and tournament format. It is common for scheduling to be done manually by a group of individuals determined to represent the needs and wishes of all the participants. This task can be very time demanding and tedious if done manually or by using rudimentary methods. It is very rare that schedules made in this way meet all the requirements and needs set by teams.

Sports organizations, athletes, and related parties are highly invested in the industry, therefore, it is imperative that schedules seek to optimize profits by maximizing entertainment value while maintaining competitive fairness. There are many criterions that are considered to maximize entertainment value while fairness criterion are important to the schedule in order to maintain the integrity of the competition. Deciding on which criterion to use is specific to the sport, tournament format, and the projected audience. It is rare that schedules generated through remedial methods address entertainment value and fairness while meeting the previously discussed organizational and logistical requirements.

The goal of this academic paper is to present information regarding the use of operations research and heuristic algorithm methods. This includes which methods are suitable for different tournament formats, which considerations are applied for different situations, how to implement tools to meet certain goals, as well as the implications of each sport scheduling applications. This paper discusses and analyzes the methods used in several academic papers: Scheduling the Belgian Soccer League (Goossens and Spieksma, 2009), Scheduling the German Basketball League (Westphal, 2014), Scheduling the

Brazilian Soccer Tournament: Solution Approach and Practice (Ribeiro and Urrutia, 2012), A Decision Support System for Scheduling the Canadian Football League (Kostuk and Willoughby, 2012), Scheduling A Major College Basketball Conference (Nemhauser and Trick, 1997), Scheduling Soccer League of Honduras using Integer Programming (Fiallos, Pérez, Sabillón, and Licona, 2010), and The Chilean Soccer League paper (Durán, Guajardo, Miranda, Sauré, Souyris, Weintraub, and Wolf, 2007).

After reading this paper, a reader with basic previous knowledge of programming techniques should be able to produce a feasible, optimal sports schedule fit for a given situation. The reader should also see how the techniques in this paper could be applied to improve and optimize operations in other situations apart from sports scheduling.

This paper will also include a case study applying the information, techniques, and recommendations organized in this paper to generate a recommended schedule for the 2014 FIFA World Cup Tournament.

This paper is organized as follows. In the *Tournament Formats* section, basic elements of tournaments and schedules will be discussed. In the *Methodologies* section, the different approaches used in sports scheduling will be analyzed and the strengths and weaknesses of each will be discussed. In the *Considerations* section, the different constraints related to schedule feasibility and the imposed requirements related to the respective sports leagues. In the *Optimization* section, the criteria that determines optimal schedules is analyzed. In the *Notable Approaches* section, approaches that are unique, seldom witnessed in sports scheduling, and significant are explained and analyzed. The next half of the paper is a case study involving the scheduling of the 2014 FIFA World Cup. It will utilize methods discussed in the first half of the paper and apply them to the specifics of the World Cup. In the *Case Study – Background* section, the problem is introduced and the current methodology is explained. In the *Case Study – Literature Review* section, information justifying the reasoning for using certain approaches and use of assumptions is presented. In the *Case Study – Conditions Imposed on the Problem* section, the constraints and objective functions are described. In the *Cast Study – Formulation of the Mathematical Model* section, all of the constraints and objective functions, described in the previous section, are presented as mathematical formulas. In the *Case Study – Results* section, the results of the mathematical models are presented and measured against the results of the current scheduling method. In the *Conclusion* section, we briefly summarize the paper and discuss guidelines for future work in related areas.

II. Tournament Formats

There are various styles or formats of sport scheduling tournaments. In other words, there are different methods of pairing participants up to play each other to determine the “best” or winning team out of all the competitors. Over the years there have been many different tournament formats that have been used in sports. This paper will mostly focus on the following formats: single round robin and double (or multiple) round robin tournaments, mirrored and non-mirrored schedules, home and away location dependencies, and compact and loose tournament rounds. Each format has their different rules and organizational policies that they abide by. Each format also has their schedules generated in different ways to fit the organizational needs and specific wishes for each situation.

Round robin tournaments formats require that each team play every other team in the robin. The robin is the set of teams that will all play each other. Depending on the format of the tournament round robins can be all of the teams in the whole competition or can be just be all of the teams in a particular group or division of the competition. Single round robin tournaments are where teams play every other team in the robin exactly one time. Double (multiple) round robin tournaments are where teams play every other team in the robin twice (or more).

Sports schedule for a whole season of competitions may not appear to follow the round robin tournament format. Most seasons’ schedules actually do follow one round robin format or a combination between formats. Round robin tournaments can be a hybrid between playing all the teams in the competition and playing all of the teams in a group or division. For example, the National Football League (NFL), the professional American football league, has a hybrid round robin with a partial round robin between teams of different divisions (teams only play non-divisional teams once a season at the most and non-divisional matchups are rotated from year to year) and a double round robin between teams of the same division (teams play divisional teams twice a year).

Mirrored schedules are only applicable in double or multiple round robin tournaments formats. A mirrored schedule is one that follows the same sequence of opponents for the first section of the round robin as it does in the second section of the round robin, except with inverse home and away game assignments if applicable. A non-mirrored schedule does not require any specific sequence of opponents or home and away assignments. Mirrored schedules have the benefit of uniformity, distribution of opponents, and fairness. While non-mirrored schedules have less interrelated constraints and more flexibility to optimize entertainment and accomplish other goals.

Home and away dependencies refer to whether or not teams have a home location for their home games. In almost all professional sports leagues teams have their own home field or stadium. If teams have a home location, the schedule must assign for their home games to be played there. Some leagues will also put requirements on the number of home games in a row, no games at a location on a specific date, etc. In some cases teams of a city or region will even share a home stadium with each other. This complicates the problem a little bit more because both teams will have to be restricted from playing a home game at the same time. In the other case, some competitions do not specify a home or away team, or the home or away assignment can be arbitrary. This is common for recreational sports leagues where teams do not usually have a home location for their games. This is also seen in worldwide competitions where all of the games are in one country or region. Teams do not gain a clear home team advantage (in most sports) in a game where neither team is native to the country or region that the game is taking place in.

The last main element of a tournament format is whether the tournament is a compact or loose tournament. This simply refers to whether there are allowed to be byes, or rounds without a competition. Tournaments where no byes are allowed and each team must play in each round are called deemed compact while tournaments where byes are allowed are deemed loose. Compact tournaments are more restrictive while loose tournaments have flexibility to meet other goals and wishes. Compact tournaments have more structure but, due to being more restrictive, are usually only used for shorter spanning tournaments where it is easier to achieve a feasible schedule and where fan attention is on average higher for the short duration of the tournament. The reasoning for this is that short tournaments are composed of fewer games, which, in effect, are more important towards the outcome of the tournament. Loose tournaments are commonly seen in season long schedules where a lot of constraints and wishes make it hard to make a compact schedule. Also, fan attention is usually lower on average for the long duration of the tournament or season. Byes in loose schedules also have the benefit of allowing teams to rest for enhanced performance and energy. Robins with an odd number of teams automatically have loose schedules since not every team can play in each round.

It is important to understand the format of the tournament being scheduled. Each tournament format requires different sets and types of constraints. The tournament format will also determine suitable methodologies that are desirable to generate the schedule.

In the case where the tournament format is not already chosen, or is permitted to change, it is important to consider the competitive goals, logistics, and audience appeal when deciding on which format is most appropriate.

III. Methodologies

This paper will cover various operations research methods and algorithms used to create feasible and optimal professional sports schedules in a timely manner. Other methods exist outside of the ones presented in the paper. Methods in the paper are commonly encountered in sports scheduling research papers. This paper will cover linear programming, decomposition, and algorithms.

Linear Programming

Linear programming is the method that is most commonly witnessed in sports scheduling. It often serves as the basis for solving the problem while being accompanied by methods of decomposition and heuristics. Linear programming is a method of determining the most optimal solution using a mathematical model. A linear programming model is made up of unknown variables, an objective function, and constraints. The end solution will consist of values assigned to the variables that result in the maximum or minimum objective function value while adhering to the constraints provided.

This paper will include methods regarding integer linear programming and binary linear programming. Integer linear programming refers to models in which the variables can only take on integer values. Likewise, binary linear programming refers to models in which the variables can only take on the values of 1 or 0 (this can be thought of as true or false, yes or no, or the event occurs or does not occur). Therefore, binary linear programming can be considered a subset of integer linear programming.

Fiallos, et al. (2010) uses binary linear programming without any decomposition or algorithms to accompany it.

Decomposition and Algorithms

As previously mentioned, decomposition is a method that accompanies linear programming. Decomposition breaks the problem down instead of solving the entire problem as a whole. This helps to alleviate the long solve times. Large linear programming models with cross relationships between variables, the objective function, and constraints exponentially affect the solution set and therefore the solve time.

Along with decomposition, algorithms and heuristics can be paired with linear programming in order efficiently arrive at a solution.

The "*First-Break Then-Schedule*" method is a method of decomposition in sports scheduling in which HAPs are determined prior to actual matchups between opponents. Observed methods vary slightly from each other due to differences in problem description, problem size, and desired goals.

Goossens and Spieksma (2009):

1. Feasible HAPs are generated using a canonical schedule (explained in the section, "Considerations", in the subsection "Structural Constraints")
2. HAPs are assigned to each team using binary linear programming

    a. Considerations include requests such as teams avoiding home games on certain dates, top teams playing at home in televised rounds, etc.

3. Using the set HAPs, matchups between teams are assigned using binary linear programming
    a. Considerations include avoiding playing certain teams on specific dates, avoiding consecutive games against strong teams, etc.

The method is this paper includes the basics of the first-schedule then-break method. It reduces the solve time by eliminating many of the relationships between the HAPs and matchups between teams.

The previously used method by the Belgian Soccer League used a canonical schedule (referred to, in the paper, as the Basic Match Schedule or BMS) to produce a full schedule with numbers representing arbitrary teams. Then, actual teams would be manually assigned to the numbers by a "calendar committee". Utilizing the steps in the method listed above allows for more flexibility since teams are not restricted from being assigned to the full canonical schedule including set HAPs and matchups. Instead, matchups can be altered allowing for a wider solution set and the ability to reach the most optimal solution while still following the HAPs that are feasible and offer a minimal number of breaks.

One weakness of programming and objective functions is that they may not always produce what is actually the most optimal solution. In this case, the binary linear programming model in step 2 (HAP assignment) may not be capable of producing a high objective function value in step 3 (matchup assignment). This paper employs a *tabu search algorithm* in order to account for this weakness in order to find the most optimal solution.

The algorithm starts by following the previous steps which produces the highest objective function value in step 2 (HAP assignment) and then the highest objective function in step 3 (matchup assignment) using the inputs from step 2. The algorithm then swaps the HAPs of two teams and inputs the new HAP assignments to step 3 where a new schedule and objective function value is generated. If the sum of both new objective function values from both steps is greater than the previous sum of objective function values, then the new schedule is saved and the algorithm will continue checking the neighborhood resulting from this swap. If it is not, the HAP swap is reversed back to how it was and the swap is added to the tabu list. Swaps on this list are banned from being tested until five more swaps have been attempted. This process is stopped after 10 swaps have been done without any improvement in the sum of the objective function values.

See the flowchart in Figure 1 for a visual representation of the steps and algorithm:

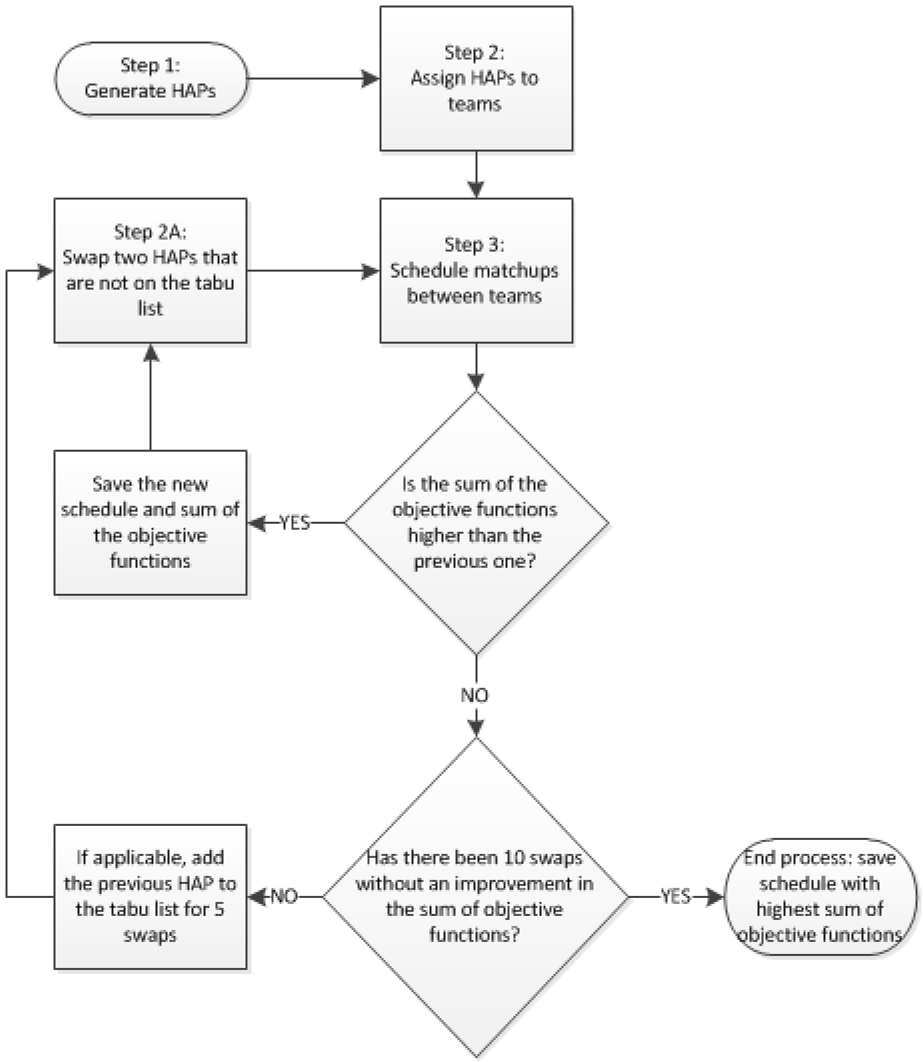


Figure 1: Goossens and Spieksma (2009) Flowchart

Ribeiro and Urrutia (2012):

1. All feasible HAPs are generated and infeasible patterns are discarded using enumeration
    a. HAPs must have the same number of home and away breaks in each half of the tournament, no breaks in certain rounds, as few breaks as possible, etc.
2. Randomly assign complementary HAPs to each pair of teams

   a. Pairs of teams are usually made up of teams that share the same home city and/or have a long standing rivalry

3. Matchups between teams are assigned using binary linear programming

The use of *pairs of teams* and *complementary HAPs* in the 2nd step of this method allows for a more extensive range of feasible schedules and therefore, likely, a more optimal schedule in the end. Pairs of teams that share the same city are prevented from both teams being assigned a home game or an away game in the same round. There will be exactly one team playing at home in the city each round. This is ideal for the logistical concerns of two matches occurring in the same round in the same city and the opportunity costs of having no matches occurring in a city.

Assigning complementary HAPs to teams that are rivals (teams of the same city are also considered rivals) means that the two rivals are able to matchup in any round of the tournament since one team will be playing at home and the other team will be playing away in each round. This allows for the matchup between the two rival teams to take place in the most optimal round (according to the objective function in the 3rd step of this method) instead of being restricted from rounds where both teams are playing at home or away (two teams cannot play each other if both have a home or away assignment).

The method in this paper also uses an algorithm paired with the previous steps in order to choose the best possible solution in a timely manner. The previous steps without the algorithm is less likely to produce an optimal solution due to the random assignment of HAPs to teams. The algorithm can best be described as a *randomized algorithm accompanied by bounds*.

The algorithm reaches an optimal solution with the use of upper and lower bounds. The first step of the algorithm computes the upper bound, or highest possible value, for the objective function. This is easy to compute since the objective function only seeks to maximize the number of games scheduled in certain rounds. A solution is generated using the steps listed above. If the objective function value of the solution is equal to the upper bound, then that solution is taken as the optimal solution and the process ends. If not, then the process continues.

The lower bound starts off with a value of 0. A solution is saved as long as its objective function value is higher than the lower bound. If it is higher, then its objective function value becomes the new lower bound. If it is not higher, the solution is discarded, HAPs are randomly assigned to teams again (step 2 above), and matchups are scheduled again (step 3). This process ends when the specified time limit is reached, or, as stated previously, a solution's objective function value is equal to the upper bound.

See the flowchart in Figure 2 below for a visual representation of the steps and algorithm:

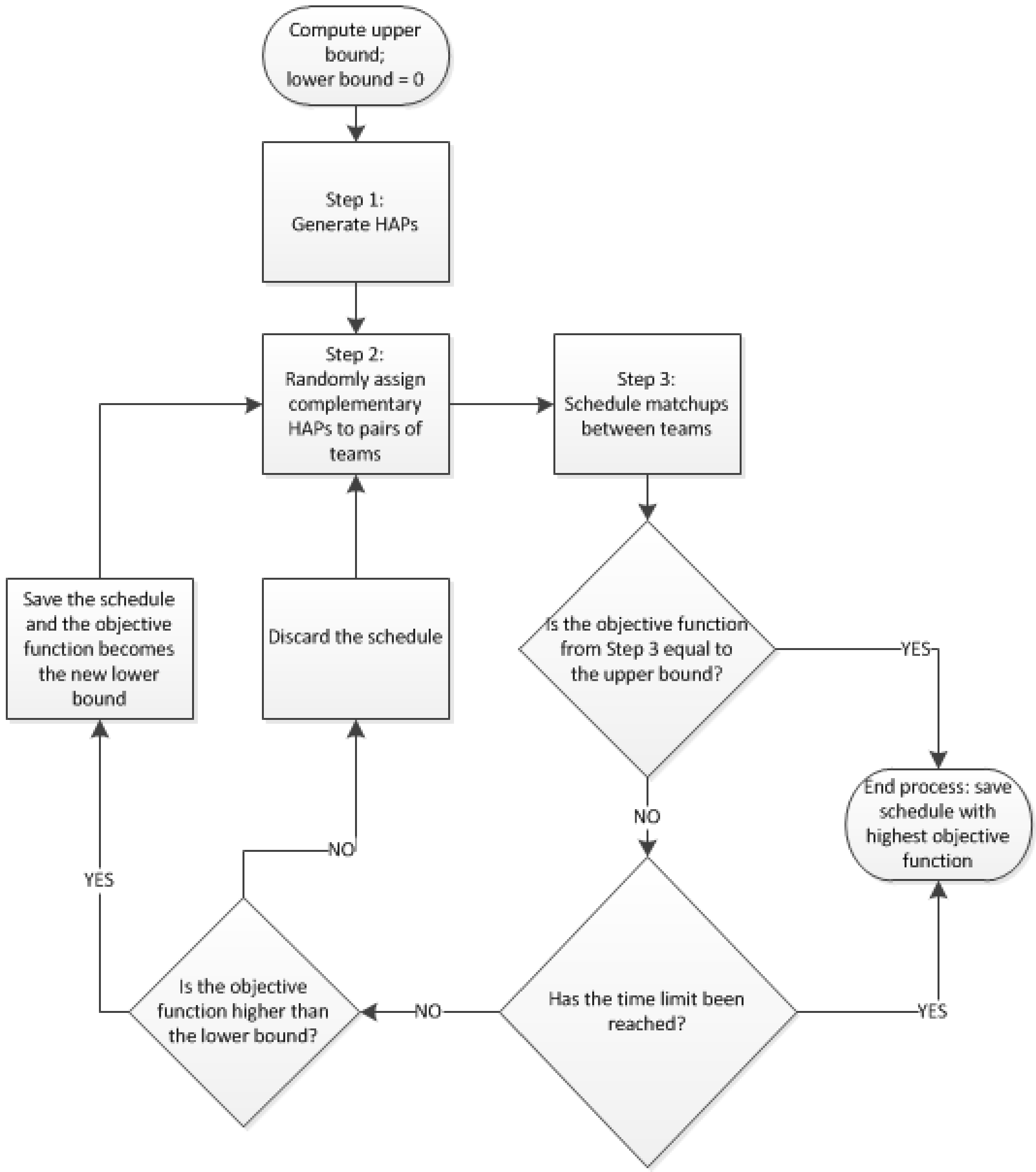


Figure 2: Ribeiro and Urrutia (2012) Flowchart

Nemhauser and Trick (1997):

1. All feasible HAPs are generated and infeasible patterns are discarded using enumeration

    a. HAPs must have appropriate length, number of home, away, and bye assignments, number of weekend assignments, limited number of breaks, etc.
2. All feasible pattern sets (sets of 9 patterns, one for each team) are generated using binary linear programming
    a. Pattern sets must have the correct number of home, away, and bye assignments in each round
3. All feasible matchups between patterns (referred to as timetables in the paper) within the different pattern sets are generated using binary linear programming
    a. Matchups between patterns must happen the correct number of times (total as well as at each location), patterns must play each other the correct number of times, the schedule must be mirrored, etc.
4. All combinations of teams assigned to patterns are generated and infeasible assignments are discarded using enumeration
    a. The specific request regarding consecutive matchups between strong teams, distribution of high entertainment games for televised rounds, etc. are taken into account
5. Using dominance aspects, optimal schedules were chosen and presented to the ACC

This method differs from the other method as it does not use an objective function to determine the most optimal schedule. There are objective functions in the binary linear programming models used in the 2nd and 3rd steps, however these models are used in a similar way as enumeration is used in the 1st and 4th steps (only to discard infeasible solutions). All feasible solutions are taken from each step and serve as inputs for the subsequent steps. Only in the 5th step are feasible solutions discarded. This is done by considering dominance aspects that are assumed to stem from wishes obtained from the ACC scheduling committee that cannot or may not be explicitly represented as constraints in the models.

This method produces and presents a very wide range of feasible solutions. This can be beneficial in the sense that objective functions typically are only used to produce one optimal solution with the highest objective function. Declaring this solution as the best requires the assumption that the objective function is the best measurement of the optimality of the schedule. Qualitative and manual analysis of more feasible solutions can be beneficial to encounter schedules that may end up being perceived as superior to the one with the highest objective function.

The next method uses a combination of the "First-Break Then-Schedule" method and the "*First-Schedule Then-Break*" method. In the first-schedule then-break method, matchups between opponents are determined before the HAPs.

Westphal (2014):

1. Feasible HAPs are generated using a canonical schedule

2. HAPs are assigned to each team using binary linear programming
    a. Considerations include availability restrictions, home and away wishes, etc.
3. Using the set HAPs, matchups between teams are assigned using binary linear programming
    a. Considerations include avoiding playing certain teams on specific dates, distributing entertaining games throughout televised rounds, etc.
4. HAPs are discarded and using the set matchups between teams, new HAPs are generated using binary linear programming
    a. Once more, considerations include availability restrictions, home and away wishes, etc.

The first three steps are the exact same as the entire method used in Goossens and Spieksma (2009). The 4th step utilizes the first-schedule then-break method. The rescheduling of home and away assignments in this last step allows for more flexibility to achieve a more optimal solution. If the benefit of meeting a home or away wish outweighs the cost of possibly creating a break (according to the objective function), then the change will be made.

Scheduling metrics were recorded for the different models. A method including the first 3 steps had 31 availability restrictions violated and 48 breaks. In the method including the 4th step, which allowed for the changing of HAPs to meet other goals, the number of violated availability restrictions decreased to 22, however, the number of breaks increased to 120. This shows the tradeoff between the two methods. Ultimately, neither method was chosen due to these undesirable metrics. It was determined that there was no feasible mirrored schedule. The mirroring constraint was discarded and a linear programming model without any decomposition or algorithms and with the same constraints and objective function was used to produce a feasible schedule.

Regarding the two decomposition methods, whichever entity is determined first, is the one in which should be seen as more important to the schedule. This is because the ensuing step will have more constraints and limitation from the already set entity from the first step.

Kostuk and Willoughby (2012):

The method in Kostuk and Willoughby (2012) does not use decomposition or any algorithms. It uses linear programming accompanied by an *iterative process* working with league officials in order to generate determine optimality. It combines the league officials' experience and knowledge of optimal schedules with the programmers' ability to meet these needs.

The original binary programming model sought to find a base model in which could be expanded upon through collaboration with the league officials. The objective function looked to maximize the number of games played where the number of games played was already defined by constraints. This method produced a feasible schedule with no real objective function but one that only adhered to constraints.

The next step was to rank constraint types by level of importance. In order from highest priority to lowest priority, the constraint types were structural constraints, stadium availabilities, pre-determined matchups on specific dates, and HAP assignments. One by one, constraint types were added to the model to ensure that a feasible model could be generated. The inclusion of the last constraint type, HAP assignments, made it impossible to achieve a feasible model. The league officials were asked to determine a identify tradeoffs between meeting the last constraint type and the highest three constraint types. The details of the tradeoffs and how they were accounted for were not included in the paper.

The next step was to produce, not only a feasible schedule, but also an optimal schedule. This was done through an iterative process closely working with league officials. League officials could not identify metrics of a good schedule, but could identify a good schedule when looking at one. By analyzing which schedules were deemed good schedules and further discussion with the league, six possible criteria were produced: number of back-to-back games, intradivisional games during the final four weeks of the season, number of Thursday games, number of Friday games, number of Saturday games, and number of games played on preferred dates.

To determine the optimal schedule, different objective functions with different mixes of weights for meeting each criterion were presented to league officials. The feedback from league officials regarding the schedule were used to determine tradeoffs and the amount of importance that should be associated with each criteria measure. The exact method for quantitatively determining this was left unmentioned. The schedule generated by the paper was presented, manually fine-tuned by the league, and used as the league's 2010 official schedule.

(Durán, et al., 2007):

(Durán, et al., 2007) also does not use decomposition or any algorithms. It uses one large binary linear programming model. Problems that can be solved in a reasonable amount of time (specific to each application) are best solved this way. All factors regarding optimality can be considered. In models where HAPs are assigned to teams in a model and subsequently matchups are assigned, differences in goals between the two models may not be able to be accurately represented in the solution.

Durán, et al. (2007) did, however, decide that the single large model was too time consuming. To resolve this problem, the paper generated feasible HAPs so that the numbers of HAPs equaled the number of teams. It then used the model to assign HAPs to teams and matchups between the teams in one calculation.

IV. Considerations

Scheduling considerations and respective constraints/formulations will be categorized and explained in this section. The categories are structural, logistical, entertainment value, and fairness.

Structural

There are four main structural constraints that are common in almost all round robin tournament formats.

1. Each matchup between opponents only occurs *X* times.

Westphal (2014), Ribeiro and Urrutia (2012), Goossens and Spieksma (2009), Fiallos, et al. (2010), Durán, et al. (2007):

$$\sum_k \left(X_{ijk} + X_{jik}\right) = 1 \ \forall i, \forall j, i \neq j$$

Where,

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

This constraint sets the number of matchups (regardless of location), for all rounds k, equal to 1. Altering the set of numbers for k can produce different results. Setting $k \in D_1$ in one constraint, and $k \in D_2$ in another constraint, where D_1 is the set of all rounds in the first half of the tournament and where D_2 is the set of all rounds in the second half of the tournament, forces the teams to play each other once in each half of the tournament.

Nemhauser and Trick (1997):

$$\sum_{k} X_{ijk} = 1 \ \forall i, \forall j, i \neq j$$

Where,

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

This constraint makes it so that each matchup happens twice, once at each team's home location.

Another method, as seen in another approach from Westphal (2014) and Goossens and Spieksma (2009), uses what is called a "canonical schedule".

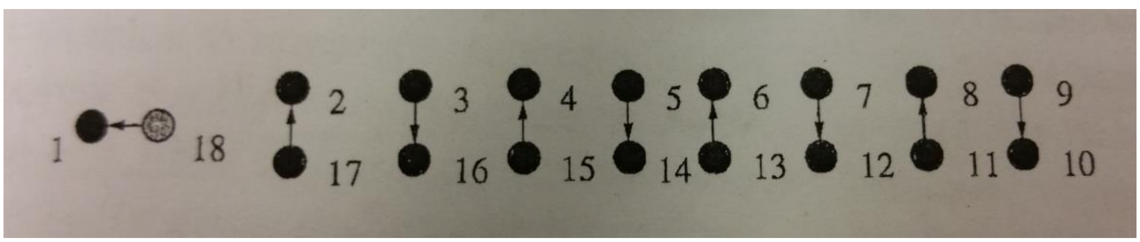


Figure 3: Canonical Schedule, Round 1

In Figure 3 above, each numbered node represents a team. In this diagram, node 1 plays at home versus node 18. The arrows designate matchups as well as game locations. This figure shows all the matchups and locations for the first round of the tournament.

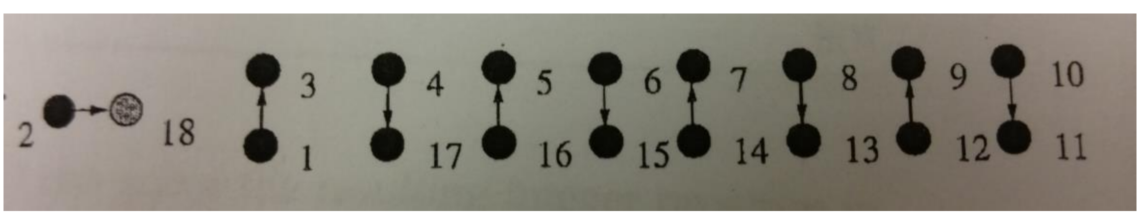


Figure 4: Canonical Schedule, Round 2

Figure 4 shows all of the matchups and locations for the next (second) round of the tournament. In this diagram all of the nodes have shifted counterclockwise with the exception of node 18, and the arrows have reversed directions.

This is done for each ensuing round in order to create a feasible schedule for each node that includes alternating home and away assignments as well as matchups. Each node plays all the other nodes one time per revolution. This is accompanied by assigning one team per node,

$$\sum_{i} X_{iv} = 1 \ \forall v$$

and one node per team,

$$\sum_{v} X_{iv} = 1 \ \forall i$$

Where,

$$X_{iv} = \begin{cases} 1, if\ team\ i\ is\ assigned\ to\ node\ v, \\ 0, otherwise. \end{cases}$$

2. Teams only play one game per timeslot.

Westphal (2014), Goossens and Spieksma (2009), Nemhauser and Trick (1997), Fiallos, et al. (2010), Durán, et al. (2007):

$$\sum_{j} \left(X_{ijk} + X_{jik}\right) = 1 \ \forall i, \forall k, i \neq j$$

Where,

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

This constraint forces each team to have exactly 1 game per timeslot. In Nemhauser and Trick (1997), however, the equal sign is replaced with a less than or equal to sign (≤). This is because the tournament format is not compact; teams are allowed to have a byes.

In the canonical schedule approach in Westphal (2014), one game per timeslot is defined by the canonical schedule, and limiting one team per node and one node per team.

In Ribeiro and Urrutia (2012), feasible home and away patterns (HAPs) are generated. Each one only has one game per timeslot. All feasible HAPs are open to being assigned to a team. Each team has exactly one HAP, and each HAP has one or less teams.

3. Teams cannot play themselves.

In Ribeiro and Urrutia (2012), Goossens and Spieksma (2009), Nemhauser and Trick (1997), Durán, et al. (2007) and Fiallos, et al. (2010), teams are prevented from being scheduled against themselves by setting $i \neq j$ in the “Each matchup between opponents only occurs *X* times” and “Teams only play one game per timeslot” constraints.

Westphal (2014) differs as in the canonical schedule approach a team is prevented from playing itself, as there is only 1 of each node. In the other two approaches in Westphal (2014) this constraint is used,

$$\sum_{d} X_{iid} = 0 \; \forall i$$

Where,

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

4. The second half of the tournament is mirrored.

In the canonical approach of Westphal (2014) and Fiallos, et al. (2010), only the first half was scheduled using programming. Constraints relevant to the second half of the tournament were given to the first half round. For example, in a league featuring 10 teams, a constraint relevant to round 15 would simply be applied to round 5. The second half was simply mirrored manually after a feasible first half solution was created.

In the other two approaches in Westphal (2014), this constraint was used,

$$X_{ijk} - X_{ji(k+n-1)} = 0\ \forall i, \forall j, k \in D_1$$

Where,

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

$$n = the\ number\ of\ teams$$

$$D_1 = the\ set\ of\ first\ half\ rounds$$

In Nemhauser and Trick (1997), this constraint was used,

$$X_{ijk} = X_{jik'}$$

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

$$k' = \ the\ second\ half\ round\ mirroring\ the\ first\ half\ round\ k$$

Logistical

Logistical constraints take into consideration concerns regarding the physical impossibilities and the ability to meet league requirements.

Security is one main concern. Sporting events attract a lot of fans and cities may not be able to provide enough police resources in order to safely facilitate the events. This is most common when there are other large events taking place within the vicinity.

In Goossens and Spieksma (2009), the mayor of the city can forbid a "*risk game*" from taking place on a certain date. A risk game is a game between teams of high hooliganism levels. These games have a higher possibility of riots and law breaking. These games are usually forbidden on dates where another event in the city requires a large portion of the police force. Another logistical constraint for the Belgian Soccer League (Goossens and Spieksma, 2009) is preventing games from occurring in the same city or local area as another game. This is for security reason as well.

Similar to Goossens and Spieksma (2009), Ribeiro and Urrutia (2012) prohibits more than one "*Classic games*" or games between teams of the same city from occurring in the same round. These games are considered rivalry games, attract more fans than the average game, and will require more security than is regularly available.

Fiallos, et al. (2010) does not allow two teams that share the same home city to both play at home or away in the same round. This is to account for extra security required as well as to avoid the opportunity costs of having not having a single game in the city in any particular round.

In Kostuk and Willoughby (2012), a set of constraints is employed in order to account for “*stadium blocks*”. These are dates in which a team’s home stadium is already being occupied, the team wishes to play at home, or the team wishes to play away.

In Nemhauser and Trick (1997), the first five weekends of the season are considered recruiting weekends. In these weekends, prospective student athletes are brought onto campus and the basketball facilities in hopes that they will choose that school to attend. Therefore, it is a league-mandated requirement for each team to be able to be present for the visits. Teams must play at home for at least two of the first five weekends of the season, or have one home game and one bye.

Entertainment Value

Leagues often set requirements to the schedule that they believe will enhance the entertainment value of the games and schedule, therefore, benefitting the leagues and their stakeholders economically.

In Nemhauser and Trick (1997), the ACC sets certain requirements for *rivalry matchups*. The ACC requires that all rivalry games take place in the last weekend round (17). There is also a requirement from the television network that two of the rival teams, Duke and UNC, play each other in round 10. Since round 10 and 17 are not naturally mirrored rounds in a double round robin tournament schedule with 9 teams, an adjustment must be made when creating the HAPs. Rounds 1 and 10 mirror each other, as well as rounds 8 and 17. These pairs of mirrored rounds were switched with each other so that rounds 1 and 8 now mirror each other and rounds 10 and 17 mirror each other.

Aside from predefining matchups with the use of constraints, the ACC also considers the entertainment of value of certain matchups between teams.

Table 1: ACC Matchup Ratings

	Away								
Home	Clem	Duke	FSU	GT	UMD	UNC	NCSt	UVA	Wake
Clem	—					B1/A2			
Duke		—		B	A	A2		B	B
FSU			—						
GT		B1		—	B	A1			B
UMD		A1		B1	—	A1		B1	
UNC	B1	A1		B	B	—			
NCSt		B				B	—		B
UVA		B						—	B1
Wake		B		B		B	B		—

As seen in the table above, each matchup is assigned a rating. Ratings differ depending on which team is at home and which team is away. An A rating denotes a high entertainment game, a B rating denotes a normal entertainment game, an A1 rating denotes a high entertainment game if it scheduled for a weekday round, an A2 rating denotes a high entertainment game if it scheduled for a weekend round, and B games are similarly noted with regards to weekday and weekend rounds. A round receives an A rating if there is an A rated matchup occurs or if 2 B rated matchups occur. A round receives a B rating if one B rated matchup occurs. A round receives no rating if neither an A nor a B rated matchup occurs. The number of A and B rated matchups that occur in a schedule are considered when all feasible schedules are evaluated. The paper does not explicitly state how much weight these matchups have on how strong a schedule is considered.

Westphal (2014) uses a similar rating system for the entertainment level of certain matchups. The difference in the method in Westphal (2014) is that it attempts to distribute these entertaining games throughout the televised rounds. See the subsection, “Distribution of High Entertainment Games”, in the section of the paper labeled “Notable Approaches”.

Durán, et al. (2007) addresses concerns regarding the availability of broadcasting units for televised games. The country of Chile spans a very large geographical area and the televising company cannot always adequately cover games all over the country without incurring unreasonable traveling costs. Therefore, the league prohibits popular teams’ games (which are always televised) from occurring across the country from each other. It also considers times when broadcasting units are less available and forces popular teams to play in the center of the country. This allows for the televising company to travel between games and other events in a shorter amount of time.

In Kostuk and Willoughby (2012), rivalry games can be pre-assigned to certain dates. These pre-assignments are mostly made up of holidays where there is a long-standing tradition of matchups. This is to maximize fan attendance and to enhance league ratings. Kostuk and Willoughby (2012) also considers televised games. The Canadian Football League only has one television broadcaster and, therefore, can only broadcast one game at a time. The paper limits the number of games that are scheduled on specific day to one since it cannot broadcast more games. The paper does not mention how these specific days are chosen or if there are goals for scheduling a certain type of game on broadcasting days.

The approach in Ribeiro and Urrutia (2012) treats its high entertainment matchups slightly differently than the previously mentioned papers. The paper groups teams into several sets: Group of Twelve (G12) teams are teams that are the founding teams of the league, Regional games are between teams sharing the same state, and Classic games are between teams sharing the same home city (Classic games ⊆ Regional games). G12 teams are treated as strong teams, while regional and classic games are treated as rivalry games.

To maximize entertainment, the objective function of this paper seeks to maximize the number of Classic games that are played in double weekend rounds. Since the schedule is mirrored, matchups scheduled in a certain round in the first half of the tournament are required to take place in the mirrored second half equivalent round. There exists several rounds in which both mirroring rounds take place on the weekend. These rounds are double weekend rounds. These rounds are believed to be optimal for high entertainment matchups as weekend games have larger attendances and television audiences. Ribeiro and Urrutia (2012) also puts restrictions on regional games occurring during the first three rounds as fan interest is relatively low.

There are restrictions on Classic games occurring in the same round as a matchup between G12 teams. The league also ensures that there are two to four games between two G12 teams in each round. These two types of restrictions are to achieve a wider distribution and avoid overlapping of high entertainment games.

Durán, et al. (2007) considers whether teams are home or away when facing group rivals and popular teams. It seeks to prevent teams from having an imbalance of these games. It is a requirement for teams to play half their group rivalry games at home. The league also requires to play at least one of their games against popular teams at home. These measures are to ensure that each team has the opportunity to host important matches at home. This distributes entertainment throughout the cities as well as ensures that each team benefits from the increased revenue.

Similar to the other papers, Fiallos, et al. (2010) prevents games between the popular teams of the league from occurring in the first or last two rounds of each half of the tournament. The schedule is mirrored so this is necessary in preventing these games from occurring in the first or last two rounds of the whole tournament. The paper also prevents these games from occurring in consecutive rounds. It does not list the reason but it is possible that it does not want to bunch all the games up so that entertainment levels are high throughout the tournament instead of just in one section.

Unlike Fiallos, et al. (2010), Kostuk and Willoughby (2012) seeks to push its high entertainment games towards the end of the season in the final four weeks. These games are between intradivisional teams. The league does this because these games have the possibility of having a high impact on the team’s placement in the standings and whether or not the team makes the playoffs.

Durán, et al. (2007) employs a similar approach. It pushes games between intradivisional teams or teams of the same group to the very end of the tournament. It does this using the objective function. Games between teams of the same group are weighted higher (maximization objective function) the later on in the tournament they take place. This is with the exception of the very last round where the league speculates that the fate of each team might be determined already. The last round still as a high weight to it, but not as much as the previous rounds.

While Goossens and Spieksma (2009) does not list any formulated constraints, it does state several considerations and rules that it abides by. The first rules is that there is one top game (games between two teams that have the best record over the past 10 years) in each round with the exception of the first four rounds. This aims to distribute top games throughout the season while avoiding the first few rounds where fan interest is perceived to be low. The other rules are that at least one of the four top teams play away in each round, and at least three of the six teams that had the best record in the previous season play away in each round. Achieving this distribution of strong teams playing away in each round is to create the possibility for upsets throughout the season. This stems from the theory that strong teams playing at their home stadium lopsidedly result in easy wins.

Goossens and Spieksma (2009) also seeks to maximize entertainment by allowing teams with cash flow problems to request home games versus top teams in the beginning of the season. Games versus top teams often bring in more revenue thus allowing teams with cash flow problems to have monetary resources throughout the rest of the season.

Increasing the entertainment value of schedules is important to the economic success of the league. There are often high opportunity costs associated with creating schedules that do not adequately address the relationship between matchups of different entertainment values and the dates in which they are scheduled on. On the contrary, the entertainment value of a schedule also includes the fairness of the schedule as to give each team an equal chance at success. Schedule fairness will be discussed in the following section.

Fairness

As mentioned in the previous section, fairness is extremely important to the schedule. A fair schedule is important to create the most competitive competition possible which affects the entertainment value and economic success of the season and league. Fairness is important in maintaining the integrity of the competition. There are many rules set in place in order to ensure that this happens.

The strength of schedule is one thing that is considered and rated in almost every sport. This takes into consideration the strength of opponents. However, in round robin tournaments where teams play every team the same amount of times this method of determining the strength of schedule becomes obsolete. Instead, varying schedule difficulties can be controlled by limiting the sequence of opponents that a team will face. This is usually done by creating a set of teams that are considered strong and preventing any single team from playing a certain amount of consecutive games against those teams.

In Goossens and Spieksma (2009), a set of “top teams” is created and teams are prevented from playing two top teams in a span of four consecutive games. The paper also requires that teams play at home versus a top team at least once in each half of the season. This gives teams a fair chance to defeat a top team, as well as equal economic benefits of hosting a top, and most likely popular, team. The paper does not determine how the top teams are determined or how many of them there are. In Nemhauser and Trick (1997), teams are prevented from facing three teams in three consecutive games. These three teams (Duke, UNC, and Wake Forest) have traditionally strong basketball programs. And of those three teams, the two strongest (Duke and UNC) are prevented from being faced in back to back games. Ribeiro and Urrutia (2012) prevents teams from facing “Group of 12” (G12) more than 5 consecutive games in a row. This group is made up of the founding teams that are traditionally the strongest, and as the name suggests, the group is made up of 12 teams. Ribeiro and Urrutia (2012) also prevents teams from playing back to back “Classic games”. Classic games are between two teams that share the same city. While this may not mean the teams are strong, it is believed that close proximities between teams create competitive rivalries and that rivalry games are more unpredictable in their outcomes. Durán, et al. (2007) prevents two consecutive games against strong teams while Westphal (2014) does not consider the strengths of opponents at all.

On the contrary, in Ribeiro and Urrutia (2012), teams are prevented from having an unfair advantage by playing consecutive games in their same state. Teams from the state of Sao Paulo are prohibited from playing five or more consecutive games in the state of Sao Paulo. Teams from states other than Sao Paulo are prohibited from playing four or more consecutive games in their home state. Sao Paulo has a higher threshold because there are significantly more teams from that state. These limits are imposed in order to prevent a team from gaining an advantage of minimal traveling distances while other teams are traveling a lot.

Goossens and Spieksma (2009) also considers an element concerning opponent matchup sequences. This is called the carry over effect. It is described in the subsection, “Carry Over Effect”, in the section of the paper labeled “Notable Approaches”.

Another major factor that comes into play when trying to generate fair schedules is whether a team plays home or away, how many times, and the home and away sequence. In all of the papers, teams are given the same amount of home, away, and byes throughout the season. HAPs are also generated with minimal breaks and teams are given the same amount of home and away breaks. This is with the exception of the first-schedule then-break method in Westphal (2014). HAPs are set early in the process however have the ability to be altered at the very end satisfy other wishes if the associated benefit outweighs the cost of changing the balanced and fair HAPs. In this last step, teams can end up with more or less breaks than others, and more or less home or away assignments than others. The paper does not mention any limits to these changes. Also, in regards to breaks, Nemhauser and Trick (1997) and Durán, et al. (2007) limit the number of consecutive breaks to two (teams cannot play three consecutive home or away games). Ribeiro and Urrutia (2012) does not mention consecutive breaks and consecutive breaks are prevented in Goossens and Spieksma (2009) via the method of the canonical schedule. Durán, et al. (2007) considers traveling between regions during away breaks. Teams playing locations during away breaks are prevented from taking place in faraway regions. This is to prevent teams from being at a disadvantage of playing in an away break and having to travel long distances.

There is also a lot of consideration that goes into how a team begins and ends a season. It is believed that the beginning and end of a season has increased economic and competitive impact. In Nemhauser and Trick (1997), teams alternate home and away assignments for the first and last games from season to season. This means that a team that started with an away game and ended with a home game will, in the following season, start with a home game and end with an away game. This is similar to Ribeiro and Urrutia (2012); teams are assigned home or away in the first round depending on “historical data”. We can assume that this means teams will alternate home and away assignments in the first round. Also, in Ribeiro and Urrutia (2012), teams that start the season with a home game, end the season with an away game, and vice versa.

Several papers take into consideration breaks at the beginning and end of the season. In Nemhauser and Trick (1997), a team cannot start or end the season with an away break, Kostuk and Willoughby (2012) prevents teams from ending the season with a break, and in Ribeiro and Urrutia (2012), no breaks are allowed in the first four rounds or the last round.

In the Canadian Football League, games can take place on Thursday, Friday, Saturday, Sunday, and Monday. Unlike in the other leagues where rounds are used, it cannot be assumed that there is an adequate amount of time in between games in the Canadian Football league. Days of rest between games must be taken into consideration, as teams scheduled to play without enough rest will be at a disadvantage (the health and safety of the athletes also become at risk). Teams that play on Sunday cannot play on the following Thursday or Friday, and teams playing on Monday cannot play on the follow Thursday, Friday, or Saturday.

V. Optimization

After finding a solution that is feasible, the next step is creating one that is optimal. It should be noted that these steps do not take place one after another, but concurrently (with the exception of Nemhauser and Trick, 1997). This section will discuss the criteria in which schedules are chosen instead of other schedules and how the criteria are represented in the objective function.

Durán, et al. (2007)

$$Maximize \sum_{10\leq k\leq 18} \sum_{e} \sum_{i\in t(e)} \sum_{j\in t(e)} k \cdot x_{ijk} + \sum_{e} \sum_{i\in t(e)} \sum_{j\in t(e)} 15 \cdot x_{ij19}$$

Where,

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

$$t(e) = the\ set\ of\ teams\ in\ group\ e$$

$$e \in E = \{1, 2, 3, 4\}$$

This objective function aims to push all games between teams of the same group towards the end of the tournament. This is done by formulating the objective function as a maximization problem and giving a weight equal to the round number to each of the group matches if they occur in the later rounds (10 through 18). The last round (19) did not receive as high of a weight and received a weight of 15. This objective function attempted to maximize entertainment and fan interest.

Goossens and Spieksma (2009)

$$Minimize \sum_i \sum_j \sum_k d_{ijk} x_{ijk} + \sum_c q_c y_c$$

Where,

$$d_{ijk} = penalty\ associated\ with\ team\ i\ playing\ at\ home\ in\ round\ k.$$

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

$$q_c = penalty\ (or\ weight)\ associated\ with\ violating\ constraint\ c.$$

$$y_c = \begin{cases} 1, if\ constaint\ c\ is\ violated, \\ 0, otherwise. \end{cases}$$

This objective function aims to meet the home, away, and encounter wishes of the teams as well as avoiding violating constraints. Both penalty variables are determined using the approach discussed in the “Notable Approaches” section in the subsection “Determining Weighted Coefficients”. The paper does not explicitly state the constraints imposed.

Westphal (2014):

$$Minimize \sum_{(i,d)\in F, j\in T, d\in D} 100 \cdot X_{ijd}$$

$$+ \sum_{(i,d)\in AW, j\in T} \alpha_{(i,d)} \cdot X_{ijd}$$

$$+ \sum_{(i,d)\in HW, j\in T} \beta_{(i,d)} \cdot X_{jid}$$

$$+ \sum_{(i,j,d)\in EW, j\in T} \gamma_{(i,j,d)} \cdot (1 - X_{ijd}) + \sum_{i\in T, d\in D} 10 \cdot w_{i,d} + \sum_{d\in D} 0.5 \cdot (p_d^o + p_d^a$$

$$+ p_d^b) + \sum_{i\in T} 0.05 \cdot (dist(i,j)x_{i,j,\hat{d}} + z_i)$$

Where,

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

$$F = the\ set\ of\ all\ availability\ wishes$$

$$T = the\ set\ of\ all\ teams$$

$$D = the\ set\ of\ all\ rounds$$

$$AW = the\ set\ of\ all\ away\ wishes$$

$$HW = the\ set\ of\ all\ home\ wishes$$

$$EW = the\ set\ of\ all\ encounter\ wishes$$

$$\alpha_{(i,d)} = weight\ of\ violating\ team\ i's\ away\ wish\ in\ round\ d$$

$$\beta_{(i,d)} = weight\ of\ violating\ team\ i's\ home\ wish\ in\ round\ d$$

$$\gamma_{(i,j,d)} = weight\ of\ violating\ the\ encounter\ wish\ of\ team\ i\ and\ j\ playing\ each\ other\ in\ round\ d$$

$$w_{i,d} = \begin{cases} 1, if\ team\ i\ has\ a\ break\ in\ round\ d \\ 0, otherwise. \end{cases}$$

$$p_d^o = the\ number\ of\ A - games\ for\ every\ round\ d\epsilon D/D^{TV}$$

$$p_d^a = \begin{cases} 1, if\ there\ is\ no\ A - game\ in\ round\ d\epsilon D^{TV}, \\ 0, otherwise. \end{cases}$$

$$p_d^b = \begin{cases} 1, if\ there\ is\ neither\ an\ A - game\ or\ a\ B - game\ in\ round\ d\epsilon D^{TV}, \\ 0, otherwise. \end{cases}$$

$$\hat{d} = derby\ day\ round$$

$$dist(i,j) = the\ distance\ (km)\ between\ the\ locations\ of\ team\ i\ and\ j$$

$$z_i = the\ derby\ day\ traveling\ penalty$$

This objective function is the most complex and encompassing out of the papers. It seeks to avoid availability restrictions, meet away wishes, meet home wishes, meet encounter wishes, avoid breaks, distribute high entertainment games amongst televised rounds, and control how far teams travel on the league's "derby day". Each portion of the objective function is weighted, however, the method of how weights are determined was not stated.

Ribeiro and Urrutia (2012)

$$Maxmize \sum_{k\in D} \sum_{c\in C} \sum_{i\in T(c)} \sum_{j\in T(c):j\neq i} (x_{ijk} + x_{jik})$$

Where,

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

$$D = the\ set\ of\ all\ double\ weekend\ rounds$$

$$C = the\ set\ of\ all\ cities\ that\ host\ two\ or\ more\ teams$$

$$T(c) = the\ set\ of\ teams\ whose\ home\ city\ is\ c$$

This objective function, similar to Durán, et al. (2007), seeks to schedule high entertainment games, in this case, rivalry games between teams that share the same city, in certain rounds. The schedule has weekday and weekend games. The schedule is mirrored and is only one half of the tournament is actually scheduled using the binary linear programming model. There are certain weekend rounds in the first half of the tournament which also correspond to weekend rounds in the second, mirrored half of the tournament. These are the double weekend rounds.

Kostuk and Willoughby (2012)

$$Maxmize \sum_{i}\sum_{j}\sum_{k} x(i,j,k)$$

Where,

$$x(i,j,k) = \begin{cases} 1, if\ team\ i\ plays\ at\ home\ versus\ team\ j\ in\ round\ k, \\ 0, otherwise. \end{cases}$$

This objective function was not directly used to achieve an optimal solution. It seems to maximize the number of games, which, combined with constraints to limit the number of games, simply produces feasible schedules. As stated in the “Methodology” section, the paper worked very closely with league officials in order to determine what an optimal schedule was and how to represent it in the objective function. The exact objective function that was arrived on was not included in the paper.

Nemhauser and Trick (1997)

Nemhauser and Trick (1997), like Kostuk and Willoughby (2012), did not employ an objective function to reach an optimal solution. Instead it produced as many feasible schedules as possible and eliminated them according to “dominance aspects” which were not explicitly mentioned.

VI. Notable Approaches

This section will lay out notable approaches witnessed in the chosen sports scheduling papers. These approaches are ones that are unique to the paper and/or are seldom used. These approaches are considered significant as the formulations and ideas can be used in applications differing from their use in these papers.

Determining Weighted Coefficients

One challenge, when using quantitative methods for decision making, is factoring in qualitative aspects as well as weighing quantitative factors against each other.

Goossens and Spieksma (2009) openly describes the method in which it determined weighted coefficients for its objective function. The authors of the paper obtained a list of wishes from the Belgian Soccer League calendar committee in which they asked the committee to categorize and prioritize their wishes. Five priority levels were created with physical impossibilities making up the highest priority level, police and local government wishes making up the second highest priority level, and wishes pertaining to television companies, fairness, sporting motives, etc. being assigned to the lowest three priority levels. Then, the committee was asked how many lower level priority wishes they would sacrifice in order to satisfy a higher level priority wish.

[Figure (of Table 2 in Belgian Soccer League paper with 5 levels and weights)]

The table above shows the weights given to each of the priority levels. In order to satisfy one level 3 priority level wish, the committee was willing to sacrifice satisfying 20 level 5 priority level wishes or 5 level 4 priority level wishes. Using this method essentially assigns a quantitative amount of utility to the wishes. The tradeoff between wishes of different priority levels is quantified in terms of relative importance.

Penalties and their respective weights were accounted for using the following relevant portion of the objective function, constraint, and variable definitions:

$$Minimize \sum_{c} q_c y_c$$

$$\sum a_{ijk} x_{ijk} \leq b_c + y_c$$

Where,

$$q_c = penalty\ (or\ weight)\ associated\ with\ violating\ constraint\ c$$

$$y_c = \begin{cases} 1, if\ constaint\ c\ is\ violated, \\ 0, otherwise. \end{cases}$$

$$a_{ijk}, b_c = parameters\ to\ model\ wishes\ that\ we\ cannot\ express\ directly\ in\ the$$

$$the\ goal\ function$$

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

The above formulations is a generic way of applying penalties for violations of constraints and wishes. Variables must be modified in order to fit the specific application. An example of this can be seen in the "Carry Over Approach" subsection of this section.

The case study in the second half of this paper also includes a method in which weights are determined. It takes into consideration multiple factors regarding games and uses the Analytic Hierarchy Process (AHP) in order to assign weights. Please see the case study for more information.

Distribution of High Entertainment Games

In professional sports leagues it is often sought after for high entertainment games, games taking place between high profile and/or strong teams, to be televised.

This is the case in Westphal (2014), the most interesting games are denoted as A-games and the second most interesting games are denoted as B-games. A portion of the objective function and several constraints are dedicated to not only scheduling these interesting games in televised rounds, but to distribute them as evenly as possible throughout all the broadcasted rounds (as to not overload a round with too many). This is the relevant portion of the objective function:

$$Minimize \sum_{d \in D} 0.5 \cdot (p_d^o + p_d^a + p_d^b)$$

Where,

$$d = round$$

$$D = the\ set\ of\ all\ rounds$$

$$D^{TV} = the\ set\ of\ rounds\ where\ broadcasting\ slots\ are\ available$$

$$p_d^o = the\ number\ of\ A - games\ for\ every\ round\ d \epsilon D/D^{TV}$$

$$p_d^a = \begin{cases} 1, if\ there\ is\ no\ A - game\ in\ round\ d \epsilon D^{TV}, \\ 0, otherwise. \end{cases}$$

$$p_d^b = \begin{cases} 1, if\ there\ is\ neither\ an\ A - game\ or\ a\ B - game\ in\ round\ d \epsilon D^{TV}, \\ 0, otherwise. \end{cases}$$

This portion of the objective function penalizes A-games that are not scheduled in broadcasted rounds, broadcasted rounds without A-games, and broadcasted rounds without an A-game or a B-game. Since p_d^a and p_d^b are binary variables, interesting games are distributed as evenly as possible throughout all the broadcasting rounds. This is due to the fact that there is no benefit to having multiple interesting games in one round. In addition, scheduling multiple interesting games in the same round will leave less interesting games available for distribution throughout the other broadcasted rounds. Broadcasted rounds with neither an A-game nor a B-game are penalized in order to encourage the scheduling of B-games in broadcasted rounds without an A-game scheduled.

The aforementioned variables are defined by these constraints:

$$\sum_{(i,j) \in A} X_{i,j,d} \leq p_d^o \ \ \forall d \in D \backslash D^{TV}$$

$$\sum_{(i,j) \in A} X_{i,j,d} \geq 1 - p_d^a \ \ \forall d \in D^{TV}$$

$$\sum_{(i,j) \in A \cup B} X_{i,j,d} \geq 1 - p_d^b \ \ \forall d \in D^{TV}$$

Where,

$$A = the\ set\ of\ all\ A - games$$

$$B = the\ set\ of\ all\ B - games$$

$$X_{i,j,d} = \begin{cases} 1, if\ team\ i\ plays\ at\ home\ against\ team\ j\ in\ round\ d, \\ 0, otherwise. \end{cases}$$

Derby Day Traveling Penalty

In Westphal (2014), the league determined that it was undesirable to have games the day after New Year's Day in which fans would have to travel long distances to see their team play. This day was deemed "Derby Day" and the goal was to minimize the total traveling distance between teams playing each other on this day. It was discouraged to allow teams to travel any amount over 200 kilometers. This was represented in the objective function as well as in the constraints. This is the relevant portion of the objective function:

$$Minimize \sum_{i \in T} 0.05 \cdot (dist(i,j)x_{i,j,\hat{d}} + z_i)$$

Where,

$$X_{i,j,\hat{d}} = \begin{cases} 1, if\ team\ i\ plays\ at\ home\ against\ team\ j\ in\ round\ \hat{d}, \\ 0, otherwise \end{cases}$$

$$\hat{d} = round\ of\ derby\ day$$

$$T = the\ set\ of\ all\ teams$$

$$dist(i,j) = the\ distance\ between\ the\ home\ location\ of\ team\ i\ and\ team\ j$$

$$z_i = the\ derby\ day\ traveling\ penalty\ in\ the\ game\ involving\ team\ i$$

The variable z_i is defined in the following constraint:

$$\sum_{j \in T} dist(i,j) \cdot x_{i,j,\hat{d}} \leq 200 + z_i \ \ \forall i \in T$$

In the constraint above, the variable $\boldsymbol{z_i}$ takes on the amount a team's distance traveled exceeding 200 kilometers. Since the objective function penalizes for both the distance traveled and the variable, $\boldsymbol{z_i}$, every kilometer traveled over 200 kilometers receives twice the penalty as kilometers traveled under 200 kilometers. The effect of kilometers traveled on objective function penalty points can be seen in Figure 5 below:

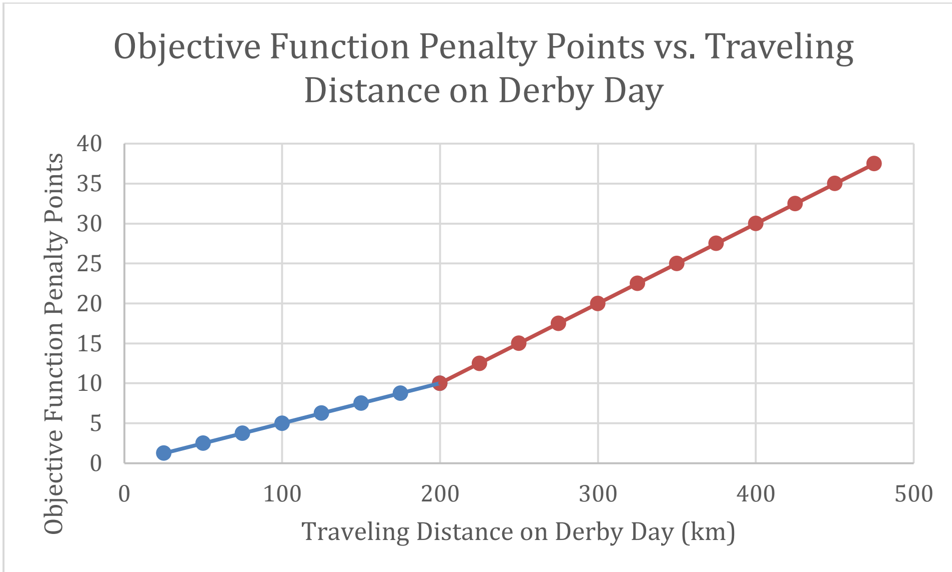


Figure 5: Derby Day Traveling Penalty

The number 200 in the derby day constraint can be altered to achieve a different affect. Likewise, more weight can be added to kilometers traveled. For example, a coefficient of 2 can be multiplied by the variable z_i in the objective function to increase the penalty for kilometers traveled over 200 kilometers to three times the penalty as kilometers traveled under 200 kilometers.

Carry Over Effect

The carry over effect takes into consideration the effect of the sequence in which teams face other teams. Consider a situation in which there are three teams: Team X, Team Y, and Team Z. Team X is considered a strong team, while the strength of Team Y and Team Z are irrelevant and can be considered arbitrary. Team X and Team Y play each other in the first game and Team Y and Team Z play each other in the game immediately following it. The carry over effect says that since Team Y played a strong team in the game before, due to fatigue, minor injuries, etc. that are more likely to occur when a team faces a strong team, that Team Y will be weaker in the next game when it faces Team Z. Team Z benefits off of the strength of Team X. This effect can occur in the case where Team X is a low strength team as well. In this situation, Team Z would be at a disadvantage.

Goossens and Spieksma (2009) is one of the few sports scheduling papers to consider this effect. While the paper does not show the constraint, the league prevents this carry over effect between the two teams from taking place more than twice in a season. The following constraints will limit this carry over effect:

$$\sum_{z}\left(X_{izk} + X_{zik} + X_{zj(k+1)} + x_{jz(k+1)}\right) \geq 1 + C_{ij} \quad \forall i, \forall j : j \neq i, \forall k$$

$$C_{ij} \leq 1 \quad \forall i, \forall j$$

Where,

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

$$C_{ij} = \begin{cases} 1, if\ there\ is\ a\ carry\ over\ effect\ with\ team\ i\ carrying\ over\ to\ team\ j \\ 0, otherwise. \end{cases}$$

The method in Goossens and Spieksma (2009) schedules the first half of the season and then mirrors the second half accordingly. Therefore, the carry over effect is limited to one in the previous constraint.

Variation Reduction

There are situations in which reducing variation can aid in producing a fair schedule. In the scheduling of the FIFA World Cup (case study at the end of this paper), teams are delegated into groups of four. This process was normally done randomly, however, this often led to unbalanced groups. This process often led to the popular term, “Group of Death”, a moniker given to a group each year that composed of exceptionally stronger teams than the rest.

In order to create as fair groups as possible, a new method sought to decrease the variation the strengths of each group with regards to FIFA rankings. The main working variable and parameter in the model are:

$$X_{ijk} = \begin{cases} 1, if\ team\ i\ is\ placed\ in\ group\ j\ with\ a\ group\ ranking\ of\ k, \\ 0, otherwise. \end{cases}$$

$$R_i = FIFA\ Ranking\ of\ Team\ i$$

The group ranking is the ranking of the team within the group (with regards to the teams FIFA ranking). A team can be group ranked as the 1st, 2nd, 3rd, or 4th since the groups are made up of four teams. The method chosen in the model seeks to reduce the variation between FIFA rankings of the 2nd, 3rd, and 4th ranked teams of each group (1st ranked teams of each group were excluded since they were automatically assigned to each group). Another way of saying that is, all the 2nd best teams in each group should have as close of FIFA rankings as possible, and same with the 3rd and 4th best teams in each group. The following objective function and constraint accomplishes this effect:

$$Minimize \sum_{i} \sum_{j} \sum_{k \in 2,3,4} d_{ijk}^{+} + d_{ijk}^{-}$$

$$\left(X_{ijk}R_i - \frac{\sum_i \sum_j X_{ijk}R_i}{8}\right) + d_{ijk}^{-} - d_{ijk}^{+} = 0\ \forall i, \forall j, \forall k = 2,3,4$$

Where, $\boldsymbol{d_{ijk}^{-}}$ and $\boldsymbol{d_{ijk}^{-}}$ are deviation variables that take on the value of a team's deviation from the mean of all teams that have the same group ranking. These deviation variables are minimized in the objective function.

This method can be altered and applied accordingly to encourage equality in schedules. Many schedules will place limits/constraints on certain attributes. This prevents attributes from taking on values above or below the limit, however, those attributes can still take on values that vary significantly from the rest. See the scheduling of the FIFA World Cup (case study at the end of this paper) to see the effects of this method.

VII. Case Study – Problem Definition

The FIFA World Cup is the biggest soccer tournament and sporting event in the world. It occurs every 4 years and takes place over a span of 30 days. The first World Cup was hosted in Uruguay in 1930 and has occurred ever since, with the exception of 2 consecutive times, which were cancelled due to World War II. The tournament, which is hosted in a new country each time, starts with 32 teams and ends with one champion. There are currently 208 Member Associations in FIFA who compete in friendlies, continental tournaments, and World Cup qualifiers. The World Cup Qualifiers, which takes place in each of the 6 different regions, gives each nation an opportunity to qualify to compete in the actual World Cup and starts two years before the tournament. At the end of the 2-year process the best teams from each geographic region qualify for the World Cup. Throughout this paper two different kinds of rankings will be used in discussion: World Cup ranking and FIFA ranking. The World Cup ranking is the rank of the team compared to only the 32 teams in the tournament. World Cup rankings span from 1 (best) to 32. The FIFA ranking is the ranking given to the team by FIFA in the October before the tournament. The FIFA ranking is the rank of the team compared to every team in the world. FIFA Rankings span from 1 (best) to 208. This year, the team with the lowest FIFA Ranking (worst, but highest number) is 59 who has a World Cup ranking of 32. The number of teams from each region varies and depends on the perceived strength of the teams in the region:

Africa (CAF) – 4 or 5 teams

Asia (AFC) – 4 or 5 teams

Europe (UEFA) – 13 teams

North & Central America (CONCACAF) – 4 or 5 teams

Oceania (OFC) – 0 or 1 teams

South America (CONMEBOL) – 5 or 6 teams

After the qualification process ends, 32 nations attend the World Cup in the chosen host nation. The 2014 World Cup will be hosted by Brazil.

The World Cup consists of two stages:

1) Group (Round Robin) stage:
   - The 32 qualified teams are placed into 8 groups of 4 teams each
   - Each team plays 3 matches against opponents from their own group
   - Points are awarded by match results: 3 for a win, 1 for a tie, and 0 for a loss
   - The 2 top teams from each group advance to the next stage

2) Knockout (Single Elimination) stage:
   - There are 16 teams compete in a “bracket” style tournament
   - 4 total rounds: the first round with 16 teams, the quarterfinals with 8 teams, the semifinals with 4 teams, and the championship game with 2 teams

- Each team plays one match in each round, losing teams are eliminated while the winning teams play the other winning teams

Due to the immensity and importance of the World Cup to people around the globe it is critical to manage the event as best as possible. This starts with the group allocation and scheduling of World Cup matches. In this paper we will use the information of the teams and the 2014 FIFA World Cup to a) allocate teams into groups and b) to schedule the actual matches for the Group (Round Robin) stage. Both steps will be solved with binary programming methodologies.

Currently the 8 groups for the round robin stage are created 7 months prior to the tournament. FIFA uses a Random Drawing procedure to allocate the 32 teams into 8 groups:

Table 2: 2014 FIFA World Cup Pots

Pot 1 (Host & Seeds)	Pot 2 (Africa & South America)	Pot 3 (Asia & North America)	Pot 4 (Europe)
1. Brazil (Host) 2. Argentina 3. Colombia 4. Uruguay 5. Belgium 6. Germany 7. Spain 8. Switzerland	1. Algeria 2. Cameroon 3. Ivory Coast 4. Ghana 5. Nigeria 6. Chile 7. Ecuador	1. Australia 2. Iran 3. Japan 4. South Korea 5. Costa Rica 6. Honduras 7. Mexico 8. United States	1. Bosnia and Herzegovina 2. Croatia 3. England 4. France 5. Greece 6. Italy (drawn into Pot 2) 7. Netherlands 8. Portugal 9. Russia

Table 2 shows the 4 pots of 8 teams each that were used to perform the group allocation. The first team in Pot 1 is always the host nation; hence Brazil got the

#1 seed. The remaining 7 teams in Pot 1 are made up of the top 7 FIFA Ranked teams from the Month of October. Pot 2 contains to the unseeded African and South American teams, Pot 3 contains the unseeded Asian and North American teams, and Pot 4 contains to the unseeded European teams. Manual changes are made to the pots if necessary. Since there were 9 unseeded European teams this year, Italy was moved to Pot 2 in order to have 8 teams in each pot.

The goal of the Pot procedure is to ensure there is an equal geographical representation among the 8 groups formed. The problem with focusing mainly on geographic distribution is that FIFA forgets to consider creating evenly talented groups. Even though they consider and prevent the top 8 seeded teams from being in the same group, they forget to identify the strength of the other 24 teams. This creates an uneven balance of talent between each group. For example, one group every World Cup is marked as the “Group of Death” where each team is relatively evenly matched and there are no clear favorites. The “Group of Death” has an uneven balance of strong teams leaving other groups with excess weak teams. Our new method of allocating teams into groups, which will be described furthermore in the paper, attempts to minimize the difference in FIFA rankings across the different groups. The goal of this objective function is to create equally talented made groups, making the tournament “fair”.

The draw procedure is as follows:

1. One European team was first randomly drawn from Pot 4 and placed into Pot 2, in order to create four even pots of eight teams *(in the draw Italy were drawn out)*.

2. The draw then proceeded with the drawing of the other seven seeded teams from Pot 1 into Groups B–H, with Brazil having been predetermined to be in Group A.
3. To maximize geographic separation, an ancillary pot ("Pot X") was created during the draw into which the four seeded South American teams (from Pot 1) were placed. One of these four teams was then drawn out *(in the draw Uruguay were drawn out)*.
4. The sole European team from Pot 2 was then automatically placed into the group of the South American team that was drawn from "Pot X" *(Italy were therefore placed into Uruguay's group)*; this process prevented three European teams being grouped together.
5. All remaining teams were then drawn sequentially from the pots (i.e. Pots 2, 3, then 4) into the groups in alphabetical order (i.e. Group A, then Group B etc.). During the drawing of Pot 2, groups could be skipped over as the two South American teams in Pot 2 were not permitted to be drawn into the (remaining three) groups headed by South American seeds.
6. The positions within the eight groups were then drawn for the non-seeded teams, in order to determine the order of the fixtures within each group. The eight seeded teams were automatically designated the position of Team 1 within their group (i.e. Brazil would be Group A Seed 1).

Table 3: 2014 FIFA World Cup groups created by the drawing procedure

Group A	Group B	Group C	Group D
Brazil Croatia	Spain Netherlands	Colombia Greece	Uruguay Costa Rica

Mexico Cameroon	Chile Australia	Ivory Coast Japan	England Italy
Group E	Group F	Group G	Group H
Switzerland Ecuador France Honduras	Argentina Bosnia and Herzegovina Iran Nigeria	Germany Portugal Ghana United States	Belgium Algeria Russia South Korea

This year's World Cups groups are displayed in *Table 3*. This year, Group G is the "Group of Death" with Germany, Portugal, the United States, and Ghana with FIFA rankings of 2, 13, 14, and 23, respectively, and World Cup rankings of 3, 14, 15, and 23, respectively. The lowest ranked group was Group H containing Belgium, Russia, Algeria, and South Korea with FIFA rankings of 5, 19, 32, and 56, respectively, and World Cup rankings of 6, 20, 26, and 31. As seen, the distribution of rankings between groups can become very unbalanced with the random selection method. The "unfairness" of the generated groups was the motivation behind our own methodology for the first part of the solution method.

After the drawing procedures ends and the groups are created, FIFA schedules the Round Robin matches. Unfortunately, FIFA could not be reached for their exact scheduling method. However we did find that the solution is done mainly by random, and that the only constraint they consider is so that the last round of group matches to be scheduled takes place at the same time. This is done to preserve fairness among the teams in each group by insuring one team does not have additional knowledge of a favorable result, caused by an earlier outcome from the same group.

Furthermore, FIFA schedules the next phase - the single elimination phase - with a simple format. Although it's unknown which teams will make it to the elimination phase, FIFA lays out the group leader matchups. For example, 1A vs. 2B, 1C vs. 2D, 1E vs. 2F, 1G vs. 2H; 1A vs. 2B stands for the first place team of group A against the second place team in group B. These teams are all put on the left side of the single elimination tree branch as shown in *Figure 3* on the next page. Similarly the right side of the tree branch contains the following matches 1B vs. 2A, 1D vs. 2C, 1F vs. 2E, 1H vs. 2F. Each of these 16 potential spots is filled by the 2 winners out of each group from results of the Round Robin Stage. After the Round of 16 the two-sided tree branch shortens to 8 teams to create a quarterfinal round. This occurs until all teams are eliminated and there are only two teams left standing which is the final. The single elimination format is shown below.

Figure 6: World Cup Knock/Single Elimination Round Format

The team who wins the final match becomes World Cup champions. They are awarded the World Cup trophy made of gold, receive financial awards, earn a prestigious social and athletic status, and much more.

VIII. Case Study - Literature review

Information regarding linear programming methods, algorithms, considerations, optimization, and notable approaches encountered in academic papers can be found in the first half of the paper. The paper, The Impact of Seeding, Home Continent, and Hosting on FIFA World Cup Results (Monks and Husch 2009), gave us guidelines to the economics of the World Cup based on team performances. The points taken away from this paper were: delayed confrontations between rival/higher seeded teams create greater excitement which leads to higher revenue (Goosens and Spieksma 2009), sincerity to be awarded for higher seeded teams, and minimizing favoritism. The last two guidelines, which considered tournament “fairness”, were heavily considered for our solution method. Additionally, another paper (Russel and Urban 2006) talked about how the host team receives an advantage since they are playing at home. As a result we developed a new rating method for all the stadiums in order to distribute the “desired” stadiums evenly. No team shall have an advantage over the other because of an environmental factor such as, stadium size, tourist location, the city’s economy, etc.

IX. Cast Study - Conditions Imposed on the Problem

Group Assignment Problem

Currently, the groups are assigned by randomly selecting one team from each of the four “pots”. One pot contains the highest FIFA seeded teams and the other three pots compose of the unseeded teams from 5 different geographic regions. The 4 pots are arranged this way to ensure an even geographical representation across the 8 groups. Because there are 13 European teams that qualify for the World Cup, we were forced to create a constraint that limited the quantity of European teams per group to a maximum of 2 but a minimum of 1. The remaining constraints only allowed for one team from the other 4 geographic regions to be in a group. For example since Chile and Argentina are both from South America, they can not be allowed to be in the same group.

A binary programming model, including newly developed constraints and others required by FIFA, will be used to optimize the process with regards to group fairness, higher seed advantage, and entertainment value.

1. The objective function will be to minimize the difference between all the 2nd, 3rd, and 4th best teams of each group.
    a. This is to create an even distribution of rankings between all the groups and to eliminate the “Group of Death”.
2. There will be a maximum of two European league teams in a group and a maximum of one team from each of the other leagues in a group.

3. Each group has one seeded team and three other teams. The average of the three non-seeded teams in the group of a higher/better seeded team will be greater than or equal to the average of the three non-seeded teams in the group of a lower/worse seeded team.
    a. For example, if the FIFA rankings of the three non-seeded teams of the group with the #1 FIFA ranked team has an average of 15 (i.e. the three non-seeded teams are ranked 10, 15, and 20) then the FIFA rankings of the three non-seeded teams of the group with the #2 FIFA ranked team has to have an average of 15 or higher.
    b. This is to reward the higher seeded teams and to provide incentives so that teams aim to place in highest possible FIFA ranking before the tournament starts. As a result, teams that are not in the seeded pot (seeds 1-7) but are still within the top ranks (seeds 8-20) will benefit from being less likely to be placed in a group with the best FIFA ranked teams.
    c. Also the host nation will receive the 1^{st} seed with the top 7 ranked teams receiving the next 7 seeds in order (1^{st} FIFA ranked team with 2^{nd} seed, 2^{nd} FIFA ranked team with 3^{rd} seed, etc.)
4. Teams that are identified as rivals (Germany and England, Brazil and Portugal) and that rank in the top 16 of the total 32 teams in the World Cup will not be placed in the same group. If teams are identified as rivals, but one or more of the teams rank in the bottom 16 of the total 32 teams, then the two rival teams will be placed in the same group.

a. Teams that are ranked in the top 16 are expected to move on to the single-elimination tournament portion of the tournament therefore we want to set up the schedule so a match between rivals occurs in the next stage of the tournament. The entertainment value of a rival match in the single elimination stage will be higher, than the value of the same match in the group stage. Secondly, teams that are not ranked in the top 16 but are rivals, are not expected to make it past the round-robin portion of the tournament therefore allocating the two low ranked rivals to be in the same group phase will ensure a match between the two rival teams.

5. Each team will only be placed in one group.
6. There will be 4 teams per group.
7. The will be one 1^{st}, 2^{nd}, 3^{rd}, and 4^{th} best team per group.

The “group of death”, as previously mentioned, is a group that has drawn unusually competitive teams as opposed to teams from other groups. This unnaturally competitive bundle represents an uneven distribution of skill level among the groups. In order to avoid such an event from occurring our objective function will minimize the regression from the mean between the first, second, third, and fourth ranked teams within the groups (If a group consists of teams with teams with FIFA rankings of 1, 8, 15, and 25, the first ranked team of the group is the #1 FIFA ranked team, the second ranked team of the group is the #8 FIFA ranked team, the third ranked team of the group is the #15 FIFA ranked team, and the fourth ranked team of the group is the #25 FIFA ranked team). In other words, it will attempt to have the lowest possible difference between the second ranked teams of all the groups combined with the lowest possible difference of the third and the fourth ranked teams of all the groups. This is meant to distribute the FIFA rankings for each group as fair and evenly as possible.

Round Robin Scheduling Problem

Recognizing FIFA's current method was important for our paper because we wanted to identify the current issues and shortcomings of their method as well as considering other constraints FIFA might have in place such as days of rest between games, stadium usage, etc. Our schedule of the Round Robin stage considers many more constraints such as desirability of stadiums, matches, and rounds.

Each group is composed of four teams and each of the teams play three games against the other three teams in the groups. The times of the games will take place at the same time that FIFA currently schedules the World Cup games at; these will be called timeslots. Constraints will be modeled after several current FIFA World Cup necessities.

1. Every team will play only 1 match in each of the 3 rounds
2. There will be at least 12 timeslots/games (equivalent to 3 days in the first two rounds of the stage) of rest between the matches for each team.

3. All teams in a group will play their third round game at the same time.
4. Each team in the group plays the other team once.
5. There are 4 games at each of the 12 stadiums.
6. Each group only has no more than 1 game at each stadium.
7. There are at least 8 timeslots/games between consecutive games at the same stadium.
8. Brazil (the Host nation) will play in the first game of the tournament in Sao Paulo.
    a. The tournament opener is always the Host nation team and this year it has been designated to take place in Sao Paulo.

The objective function or goal of this binary programming model will be to maximize the entertainment value of games and economic gains for organizations, teams, and TV broadcasters. This will be done by giving each coefficient a certain weight (either a high or low value) determined by the desirability of that matchup or stadium. The stadium factor will encourage the "better" games to be played at more ideal stadiums for fan attendance and fan/viewer appeal, while the matches between closer seeded teams are the most competitive and exciting. For example in the 2014 FIFA World Cup, there are two pivotal games between closely ranked teams that took place in the very beginning of the tournament, Spain (#1 FIFA Ranking and defending champion) vs. Netherlands (#8 FIFA Ranking) in the 3rd game of the tournament, and England (#10 FIFA Ranking) vs. Italy (#9 FIFA Ranking) in the 7th game of the tournament. While there is no quantitative way of assessing this undesirable scheduling, many dissatisfied and perplexed statements have been made by announcers and journalists. Because of this, we want those desired matchups and desired stadiums to be “pushed” towards the last round of the round-robin stage.

X. Case Study - Formulation of the Mathematical Model

Group Assignment Problem

The binary linear programming model suggested to generate the FIFA World Cup groups is described below.

Variables:

$$X_{ijk} = \begin{cases} 1 \text{ if team } i \text{ is placed in group } j \text{ with a group ranking of } k, \\ 0 \text{ otherwise.} \end{cases}$$

Parameters:

$R_i = \text{FIFA Ranking of Team } i$

Sets :

$$UEFA = The\ set\ of\ teams\ in\ the\ Union\ of\ European\ Football\ Associations$$

$$AFC = The\ set\ of\ teams\ in\ the\ Asian\ Football\ Confederation$$

$$CAF = The\ set\ of\ teams\ in\ the\ Confederation\ of\ African\ Football$$

$$CONCACAF = The\ set\ of\ teams\ in\ the\ Confederation\ of\ North, Central$$
$$American\ and\ Carribean\ Association\ Football$$

$$OFC = The\ set\ of\ teams\ in\ the\ Oceania\ Football\ Confederation$$

$$CONMEBOL = The\ set\ of\ teams\ in\ the\ South\ American\ Football\ Confederation$$

$$TopPastWinners = The\ set\ of\ Past\ Winning\ teams\ ranked\ in\ the\ upper\ half$$

$$BottomPastWinner = The\ set\ of\ Past\ Winning\ teams\ ranked\ in\ the\ lower\ half$$

Objective Function:

In order to create the fairest round robin schedule as possible it is necessary that the distribution of talent, gauged by FIFA rankings, is as similar as possible within the groups. This is to avoid such occurrences as the “group of death”, as described previously.

$$Minimize\ Z\colon \sum_{i}\sum_{j}\sum_{k\in 1,2,3} d_{ijk}^{+} + d_{ijk}^{-}$$

Constraints:

1. $\left(X_{ijk}R_i - \frac{\sum_i \sum_j X_{ijk}R_i}{8}\right) + d_{ijk}^{-} - d_{ijk}^{+} = 0\ \forall i, \forall j, \forall k = 2,3,4$
2. $\sum_{i\epsilon EUFA} \sum_k X_{ijk} \leq 2\ \ \forall j$

   $\sum_{i\epsilon AFC} \sum_k X_{ijk} \leq 1\ \ \forall j$

   $\sum_{i\epsilon CAF} \sum_k X_{ijk} \leq 1\ \ \forall j$

   $\sum_{i\epsilon CONCACAF} \sum_k X_{ijk} \leq 1\ \ \forall j$

   $\sum_{i\epsilon OFC} \sum_k X_{ijk} \leq 1\ \ \forall j$

   $\sum_{i\epsilon CONMEBOL} \sum_k X_{ijk} \leq 1\ \ \forall j$
3. $\sum_i \sum_k F_i X_{ijk} - \sum_i \sum_k F_i X_{i(j+1)k} \geq 0\ \ \forall j$
4. $\sum_{i\epsilon TopPastWinners} \sum_k X_{ijk} \leq 1\ \forall j$

   $\sum_{i\in BottomPastWinners} \sum_k X_{ijk} - \sum_{i\in TopPastWinners} \sum_k X_{ijk} < 1\ \forall j$
5. $X_{ii1} = 1\ \forall i, \forall j \leq 8$
6. $\sum_j \sum_k X_{ijk} = 1\ \ \forall i$
7. $\sum_i \sum_k X_{ijk} = 4\ \ \forall j$

8. $\sum_i X_{ijk} = 1 \; \forall j, k$

9. $d_{ijk}^{-} \geq 0$

$d_{ijk}^{+} \geq 0$

Formulation of the Round Robin Scheduling Problem

The binary linear programming model suggested to generate the FIFA World Cup Round Robin Scheduling problem is described below.

Variables:

$$X_{ijk} = \begin{cases} 1 \text{ if matchup } i \text{ is placed in timeslot } j \text{ at stadium } k, \\ \quad 0 \text{ otherwise.} \end{cases}$$

$$t(y) = the\ set\ of\ all\ matchups\ involving\ team\ y$$

$$g(z) = the\ set\ of\ all\ matchups\ within\ group\ z$$

Parameters:

$$R_i = \text{Weighted Coefficient of matchup } i$$

$$S_k = \text{Weighted Coefficient of stadium } k$$

$$W_j = \text{Weighted Coefficient of timeslot } j$$

$$D_j = \text{Date of timeslot } j$$

Sets:

$$Teams = the\ set\ of\ all\ teams$$

$$Groups = the\ set\ of\ all\ groups$$

Objective Function:

In order to create the fairest and most optimal round robin schedule as possible with regards to entertainment value and economic gains, the “better” games, defined by proximity of FIFA rankings and rivalries will take place towards the end of the round robin portion of the tournament. They will also take place in more ideal stadiums, defined by the geographic location, prosperity, and popularity of the city it’s located in, the size of the stadium, and the climate of the city.

$$Maximize\ Z\colon \sum_{i}\sum_{j}\sum_{k} X_{ijk} R_i W_j S_k$$

Constraints:

1. $\sum_{i\in t(y)} \sum_{1\le j\le 16} \sum_k D_j X_{ijk} + 4 \le \sum_{i\in t(y)} \sum_{17\le j\le 32} \sum_k D_j X_{ijk}\ \ \forall y \in Teams$

   $\sum_{i\in t(y)} \sum_{17\le j\le 32} \sum_k D_j X_{ijk} + 4 \le \sum_{i\in t(y)} \sum_{32\le j\le 48} \sum_k D_j X_{ijk}\ \forall y \in Teams$

2. $\sum_{j\le 16} \sum_k X_{ijk} = 1\ \forall i$

   $\sum_{17\le j\le 32} \sum_k X_{ijk} = 1\ \forall i$

   $\sum_{33\le j\le 48} \sum_k X_{ijk} = 1\ \forall i$

3. $\sum_{i\in g(z)} \sum_k (X_{ijk} + X_{i(j+1)k}) \ne 1\ \ \forall z \epsilon Groups\ j = 33,35,37,39,41,43,45, \&47$

4. $\sum_{i\in g(z)} \sum_j X_{ijk} \le 1\ \ \forall k, i \epsilon Groups$

5. $\sum_i \sum_j X_{ijk} = 4\ \forall k$

6. $\sum_i \sum_{x\le j\le x+8} X_{ijk} \le 1\ \forall k$

7. $\sum_{i=1,2,3} \sum_{j=1} X_{ij3} = 1$

The weighted coefficients for matchups, stadiums, and timeslots were based on several difference factors. The coefficients for matchups used the difference in FIFA Rankings between the two teams in the matchup. The coefficients for the stadiums were determined using Analytic Hierarchy Process (AHP) regarding the city that the stadium is located in, the size of the stadium, and the estimated climate during the World Cup dates. The coefficients for timeslots were based on how far into the round robin stage the timeslot was.

Interim Coefficients for Matchups:

Two different solutions were generated using different models to determine the coefficients for matchups: a linear model and a non linear model. First, all of the differences in FIFA Rankings for all the matchups within each group (determined in the solution of the Group Assignment problem) were compiled. Since the objective function is set to maximize, higher coefficients needed to be assigned to "better" matchups or matchups with a smaller difference in FIFA rankings. Therefore, in order to make the smaller differences in FIFA Rankings have higher coefficients, all of the differences were simply subtracted from 100, in the linear model. In the non linear model, 1 was simply divided by the differences. The purpose of the non linear model was to assign significantly higher coefficients to smaller differences in rankings compared to the matchups with larger differences. This will be further explained in the *Result* section of this paper. The coefficients generated up to this point were interim coefficients and will be altered later on.

Example:

$$Difference\ in\ FIFA\ Rankings = 15$$

$$Interim\ Coefficient\ for\ the\ Linear\ Model = 100 - 15 = 85$$

$$Interim\ Coefficient\ for\ the\ Non\ Linear\ Model = 1/15 = 0.0667$$

Interim Coefficients for Stadiums:

Analytic Hierarchy Process was used to determine these coefficients. AHP uses ratings from 1-9 to compare each element to each other, in this case each stadium was compared to all of the other stadiums. This is done separately for each of the three criteria we deemed relevant to the World Cup games: city the stadium was located in, size of the stadium, and climate of the stadium during the World Cup games. Then each of the criteria was compared to each other in order to determine the importance of each one. Lastly, the ratings from the first step were combined with the importance determined in the second step and an effective value was given to stadium. Higher values were given to more desirable stadiums so no further adjustments were necessary for the interim coefficient.

Interim Coefficients for Timeslots:

Interim timeslot coefficients were determined using the following formulas:

$$R = Round\ that\ the\ timeslot\ is\ in\ (1,2,3)$$

$$T = \begin{cases} 1, if\ the\ timeslot\ is\ the\ 2nd\ game\ of\ the\ day \\ 0, otherwise. \end{cases}$$

$$Interim\ Timeslot\ Coefficient_{Rounds\ 1\ \&\ 2} = Timeslot^{(R+T)}$$

$$Interim\ Timeslot\ Coefficient_{Round\ 3} = Timeslot^{R}$$

The goal of the *R* exponent was to give exponentially higher weights for the later round games compared to the earlier round games. The goal of the *T* exponent was to push the “better” games to the second games each day because it lines up with the European and African time zones making it the more ideal time to have games. These formulas assign higher values towards later, more desirable timeslots so no further adjustment is necessary for the interim coefficients.

Weighted Coefficients:

In order to appropriately weight each of the interim coefficients, AHP was used in order to determine the importance of the matchups, stadiums, and timeslots (like we did to determine the importance of criteria in the stadium portion). The importance/weights of the matchups, stadiums, and timeslots were .67, .09, and .24, respectively. In order to weight the interim coefficients appropriately we set the average of the interim coefficients equal to weights determined in for their respective category (matchups, stadiums, timeslots).

Example:

$$Matchup\ \#1\ Interim\ Coefficient = 85$$

$$Mean\ of\ Matchup\ Interim\ Coefficients = 77.5$$

$$Importance/Weight\ of\ Matchups = .67$$

$$Matchup\ \#1\ Weighted\ Coefficient = \frac{85 * .67}{77.5} = .735\ , done\ for\ all\ matchups$$

$$Mean\ of\ Weighted\ Matchup\ Coefficients = .67$$

Tables of all the interim coefficients and weighted coefficients can be found in *Appendices D, E, F, and G*.

XI. Case Study - Results

The first portion of the problem was the Group Assignment problem in which we altered the method of placing the teams in their groups. *Tables 4 and 5* show the groups and the FIFA Rankings of the teams in the groups for both the current method and our proposed solution method.

Table 4: FIFA World Cup groups created by the current procedure

Group A		Group B		Group C		Group D	
Team	FIFA Ranking	Team	FIFA Ranking	Team	FIFA Ranking	Team	FIFA Ranking
Brazil (Host)	11	Spain	1	Colombia	4	Uruguay	6
Bosnia and Herzegovina	16	Chile	12	England	10	Portugal	14
Croatia	18	Ghana	23	Russia	19	France	21
Cameroon	59	Australia	57	Iran	49	Costa Rica	31
Group E		Group F		Group G		Group H	
Team	FIFA Ranking	Team	FIFA Ranking	Team	FIFA Ranking	Team	FIFA Ranking
Switzerland	7	Argentina	3	Germany	2	Belgium	5
Italy	9	Netherlands	8	United States	13	Greece	15
Ivory Coast	17	Mexico	24	Algeria	32	Ecuador	22
Honduras	34	South Korea	56	Japan	44	Nigeria	33

Table 5: FIFA World Cup groups created by the our proposed procedure

Group A		Group B		Group C		Group D	
Team	FIFA Ranking	Team	FIFA Ranking	Team	FIFA Ranking	Team	FIFA Ranking
Brazil (Host)	11	Spain	1	Colombia	4	Uruguay	6
Croatia	18	Netherlands	8	Greece	15	Italy	9
Mexico	24	Chile	12	Ivory Coast	17	England	10
Cameroon	59	Australia	57	Japan	44	Costa Rica	31
Group E		Group F		Group G		Group H	
Team	FIFA Ranking	Team	FIFA Ranking	Team	FIFA Ranking	Team	FIFA Ranking
Switzerland	7	Argentina	3	Germany	2	Belgium	5
France	21	Bosnia and Herzegovina	16	United States	13	Russia	19
Ecuador	22	Nigeria	33	Portugal	14	Algeria	32
Honduras	34	Iran	49	Ghana	23	South Korea	56

Table 6 shows the sum of the FIFA rankings of the non-seeded/host teams in each group. It shows that with the current method FIFA uses, that some groups have much more difficult competition than others. For example, the group of death in this years tournament is Group G who has the lowest sum of the non-seeded/host teams out of all the groups, 52, while

Group H has the highest sum, 112. Also in the current method you can see that the sums are relatively random with Group B, the group with the 1st seeded team, having a much more difficult group than group H, the group with the 5th seeded team. Therefore, the team that performed better prior to the tournament has not been rewarded adequately. In our proposed solution method, you can see that there is a lesser difference between the groups (lowest sum of 60 and highest sum of 93) and that the groups with the higher seeds are placed into easier groups to reward them for better performance. One metric used to measure the difference between the groups and our attempt to create even distributions of FIFA rankings between the groups we looked at the standard deviation of all the sums. The standard deviation for the current method is 21.43 while our proposed method had a standard deviation of 12.85.

Table 6: Current vs. Proposed Method Analysis of Groups

Group	Seed of Group	Sum of FIFA Rankings of non-seeded/host teams (Current Method)	Sum of FIFA Rankings of non-seeded/host teams (Proposed Method)
A	Host	101	93
B	1st	78	92
G	2nd	52	89
F	3rd	101	88
C	4th	80	78
H	5th	112	70
D	6th	56	66
E	7th	84	60
Standard Deviation		21.43	12.85

The second portion of the problem was scheduling the round robin stage of the tournament. In this portion of the problem there were two solutions using both linear and nonlinear weighted coefficients for the matchups. They will be referred to as the Linear Model and the Non Linear Model. Complete schedules of the round robin stage for the Non Linear Model, the Linear Model, and the Current Model can be found in *Appendices A, B, and C*.

Table 7: Average Difference in FIFA Rankings by Stadium

	Average Difference in FIFA Rankings per Stadium		
Stadium	Current	Linear Model	Non Linear
Rio de Janeiro	5.75	4	9.75
Sao Paulo	8	11.75	19
Salvador	11.75	10	16.5
Recife	15.5	17	16.5
Brasilia	27.25	22.25	23.75
Curitiba	15	25.75	24.25
Natal	22.5	30.25	19.25
Porto Alegre	38.75	36	29
Fortaleza	35.25	27	17.75
Belo Horizonte	16.75	23	26.25
Manaus	40.5	27.75	17.5
Cuiaba	33.25	35.5	34.75

Table 7 shows the average difference in FIFA Rankings by stadium. The stadiums are in order from most desired to least desired determined by AHP, as explained earlier in the paper. In both the Linear and Non Linear Models it shows a much higher emphasis on placing matches between closer FIFA Ranked teams in stadiums that are more desirable. Graphs are shown in *Appendices H and I* for a more visual representation of this outcome.

Table 8: Difference in FIFA Rankings as the Stage Progresses

	Difference in FIFA Ranking per Round	Difference in FIFA Ranking per Day	Difference in FIFA Ranking per Timeslot
Current	2.11	0.14	0.04
Linear Model	-4.94	-1.18	-0.39
Non Linear Model	-4.81	-1.07	-0.35

Table 8 shows the differences in FIFA Rankings as the stage progresses. For example, in the current model, an increase in the difference of FIFA Rankings per match of 2.11 can be expected with an increase in the round number for and, in the Linear Model, a decrease of -4.94 FIFA Ranking difference can be expected with an increase in the round number. In the current model, it is seen that as the round robin stage progresses, that the matches are actually between teams that are further apart in rankings. This is undesirable because closer matches are desired further on into the tournament. In both the Linear and Non Linear Models, there is clearly a push for closer matches to be placed towards the end of the round robin stage.

Finally, this brings us to the difference between the Linear and Non Linear Models. In the Linear Model each difference in the FIFA Ranking is treated the same. In each group there are 6 matches that take place with 2 matches in each round. In the Linear Model, the total combined difference of FIFA Rankings of the two matches in the first round than the second round, and the second round than the third round. However, in the Non Linear Model, each difference in FIFA Ranking is not treated the same. There is a much higher weighted coefficient given to matches with low differences in FIFA Rankings. Let's take a look at Tables 9 and 10 for an example:

Table 9: Difference of FIFA Rankings, Linear Model

Linear Model, Group A	Round 1	Round 2	Round 3
Difference of FIFA Rankings in Game 1	43	58	41
Difference of FIFA Rankings in Game 2	17	2	15
Total	60	60	56

Table 10: Difference of FIFA Rankings, Nonlinear Model

Non Linear Model, Group A	Round 1	Round 2	Round 3
Difference of FIFA Rankings in Game 1	43	41	58
Difference of FIFA Rankings in Game 2	17	15	2
Total	60	56	60

In the Linear Model it is seen that the 2 matches with the lowest total difference in FIFA Rankings is placed in the third round. However, in the Non Linear Model, those 2 matches are placed in round 2. This is because game 2 in the third round has a difference of only 2 and in the Non Linear Model this was given a significantly higher coefficient. And since $\frac{1}{2}+\frac{1}{58}>\frac{1}{41}+\frac{1}{15}$, those 2 matches were placed later on in the stage. The Non Linear Model should be used if it desired for very close matches to be placed at the end even though it may schedule a non-close match at the end also.

XII. Conclusion

Our solution methods used 1,024 variables for the Group Assignment Model and 27,648 variables for the Round Robin Scheduling Model. Variables, constraints, and the objective functions were generated using Microsoft Access 2010 and 2013 on various computers. The problems were solved using Gurobi Optimizer version 5.6.3 build v5.6.3rc2 (mac64) on a MacBook Pro with a 2.4 GHz Intel Core i5 processor on a Mac OS X Lion 10.7.5 operating system. Solving the problem using Gurobi took less than one minute for both models. However, generating the constraints for the Round Robin Scheduling Model took approximately 3 hours total.

To improve our solution method, the procedures we described above for finding good solutions raise certain issues that should be further explored in future work. The most prominent area to look into is the process of converting qualitative factors into quantitative factors to be used in the solution method. The coefficients generated were based off of our own intuition about the World Cup and the competitors. For the matchup coefficients, the only factor used to determine the desirability of a matchup was the difference in FIFA Rankings. Other factors that could be included are rivalries between teams (both from historical soccer happenings as well as current world events), proximities of countries to each other and to the World Cup, and expected attendance of fans from each country. Also different methods of modeling the coefficients should be explored. In this model we used a linear and nonlinear method to spread matchup coefficient values, but other models should be explored to see which one accurately pushes to achieve FIFA's objectives (the same goes for timeslot coefficients). For generating stadium coefficients, someone more knowledgeable about the host nation, its stadiums, and FIFA World Cup economic tradeoffs and objectives should be used to complete the AHP.

For the Group Assignment problem, there are several things that should be investigated. Since the format of the Knockout Stage is known, it might be beneficial to explore assigning teams to groups in a way that will create desirable matchups in the Knockout stage. Also, taking into consideration rivalries between teams and the entertainment/economic value of having certain teams in the same group should be explored. In our model treated past winners of the FIFA World Cup as rivals and accounted for that, however, a more knowledgeable FIFA follower should be able to come up with more reasons for teams to either be placed in the same group or to not be placed in the same group.

In Conclusion, we observe that management techniques can be used to lead to higher indices when it comes to performance metrics. Issues such as delayed confrontation, fairness, operational constraints, etc. can be improved with the use of binary programming. There are several areas in which our models can be improved on but we believe that many sports scheduling circumstances can benefit from the ideas and concepts presented in this paper.

XIII. References

Duran, G., Guajardo, M., Miranda, J., Saure, D., Souyris, S., Weintraub, A., & Wolf, R. (n.d.). Scheduling the Chilean Soccer League by Integer Programming. Interfaces, 539-552.

Fiallos, J., Perez, J., Sabillon, F., Licona, M., & Miller, J. (2010). Scheduling Soccer League of Honduras using Integer Programming (A. Johnson, Ed.). Proceedings of the 2010 Industrial Engineering Research Conference.

Final Draw Procedures. (2014). In 2014 FIFA World Cup (8th ed.). Brazil: FIFA.

Goossens, D., & Spieksma, F. (n.d.). Scheduling the Belgian Soccer League. Interfaces, 109-118.

Gurobi Optimization, Inc. (2014) Gurobi Optimizer Reference Manual. <http://www.gurobi.com>

Kendall, G., Knust, S., Ribeiro, C., & Urrutiad, S. (2009). Scheduling in sports: An annotated bibliography. Computers & Operations Research, 37(1), 1-19. doi:10.1016

Kostuk, K., & Willoughby, K. (n.d.). A Decision Support System for Scheduling the Canadian Football League. Interfaces, 286-295.

Monks, J., & Husch, J. (n.d.). The Impact Of Seeding, Home Continent, And Hosting On FIFA World Cup Results. Journal of Sports Economics, 391-408.

Nemhauser, G., & Trick, M. (n.d.). Scheduling A Major College Basketball Conference. Operations Research, 1-8.

Ribeiro, C., & Urrutia, S. (n.d.). Scheduling the Brazilian Soccer Tournament: Solution Approach and Practice. Interfaces, 260-272.

Russell, R., & Urban, T. (n.d.). A constraint programming approach to the multiple-venue, sport-scheduling problem. Computers & Operations Research, 1895-1906.

Vickery, T. (2013, December 2). 2014 Fifa World Cup: Luck of the draw 'an essential component' Retrieved June 9, 2015, from http://www.bbc.com/sport/0/football/25187878

Westphal, S. (n.d.). Scheduling the German Basketball League. Interfaces, 498-508.

XIV. Appendices

Appendix A: Proposed Round Robin Stage Schedule, Non Linear Model

Team 1	Team 2	Difference in FIFA Rankings	Day	Game of Day	Stadium
Brazil	Croatia	17	1	1	Sao Paulo
Germany	Algeria	30	1	2	Cuiaba
Netherlands	South Korea	48	2	1	Manaus
United States	Japan	31	2	2	Recife
Colombia	Russia	15	2	3	Belo Horizonte
England	Iran	39	3	1	Fortaleza
Argentina	Mexico	21	3	2	Natal
Italy	Honduras	25	3	3	Porto Alegre
Chile	Australia	45	4	1	Brasilia
Belgium	Ecuador	17	4	2	Salvador
Spain	Ghana	22	4	3	Cuiaba
Portugal	Costa Rica	17	5	1	Manaus
Uruguay	France	15	5	2	Curitiba
Greece	Nigeria	18	5	3	Belo Horizonte
Bosnia and Herzegovina	Cameroon	43	6	1	Fortaleza
Switzerland	Ivory Coast	10	6	2	Recife
Germany	Japan	42	6	3	Porto Alegre
Colombia	Iran	45	7	1	Brasilia
England	Russia	9	7	2	Rio de Janeiro
Argentina	South Korea	53	7	3	Cuiaba
Spain	Australia	56	8	1	Manaus
Chile	Ghana	11	8	2	Salvador
United States	Algeria	19	8	3	Belo Horizonte
Uruguay	Costa Rica	25	9	1	Natal

Portugal	France	7	9	2	Sao Paulo
Netherlands	Mexico	16	9	3	Curitiba
Switzerland	Honduras	27	10	1	Fortaleza
Greece	Ecuador	7	10	2	Rio de Janeiro
Belgium	Nigeria	28	10	3	Cuiaba
Croatia	Cameroon	41	11	1	Manaus
Italy	Ivory Coast	8	11	2	Brasilia
Brazil	Bosnia and Herzegovina	15	11	3	Belo Horizonte
Spain	Chile	11	12	1	Recife
Ghana	Australia	34	12	1	Natal
Germany	United States	11	12	2	Salvador
Algeria	Japan	12	12	2	Curitiba
Colombia	England	6	13	1	Sao Paulo
Russia	Iran	30	13	1	Porto Alegre
Argentina	Netherlands	5	13	2	Rio de Janeiro
Mexico	South Korea	32	13	2	Fortaleza
Belgium	Greece	10	14	1	Natal
Ecuador	Nigeria	11	14	1	Brasilia
Uruguay	Portugal	8	14	2	Salvador
France	Costa Rica	10	14	2	Recife
Switzerland	Italy	2	15	1	Sao Paulo
Ivory Coast	Honduras	17	15	1	Curitiba
Brazil	Cameroon	58	15	2	Porto Alegre
Bosnia and Herzegovina	Croatia	2	15	2	Rio de Janeiro

Appendix B: Proposed Round Robin Stage Schedule, Linear Model

Team 1	Team 2	Difference in FIFA Rankings	Day	Game of Day	Stadium
Brazil	Croatia	17	1	1	Sao Paulo
Belgium	Ecuador	17	1	2	Cuiaba
Argentina	Mexico	21	2	1	Manaus
Netherlands	South Korea	48	2	2	Brasilia
Colombia	Russia	15	2	3	Belo Horizonte
Portugal	Costa Rica	17	3	1	Fortaleza
Bosnia and Herzegovina	Cameroon	43	3	2	Curitiba
England	Iran	39	3	3	Porto Alegre
Spain	Australia	56	4	1	Natal
Chile	Ghana	11	4	2	Recife
Italy	Honduras	25	4	3	Cuiaba
Greece	Nigeria	18	5	1	Manaus
Switzerland	Ivory Coast	10	5	2	Salvador
Germany	Algeria	30	5	3	Belo Horizonte
United States	Japan	31	6	1	Fortaleza
Uruguay	France	15	6	2	Brasilia
Argentina	South Korea	53	6	3	Porto Alegre
Colombia	Iran	45	7	1	Natal
Bosnia and Herzegovina	Croatia	2	7	2	Rio de Janeiro
Brazil	Cameroon	58	7	3	Cuiaba
Chile	Australia	45	8	1	Manaus
England	Russia	9	8	2	Salvador
Spain	Ghana	22	8	3	Belo Horizonte
Belgium	Nigeria	28	9	1	Fortaleza
Greece	Ecuador	7	9	2	Sao Paulo

Netherlands	Mexico	16	9	3	Curitiba
United States	Algeria	19	10	1	Recife
Portugal	France	7	10	2	Rio de Janeiro
Germany	Japan	42	10	3	Cuiaba
Switzerland	Honduras	27	11	1	Manaus
Italy	Ivory Coast	8	11	2	Natal
Uruguay	Costa Rica	25	11	3	Belo Horizonte
Croatia	Cameroon	41	12	1	Porto Alegre
Brazil	Bosnia and Herzegovina	15	12	1	Brasilia
Spain	Chile	11	12	2	Salvador
Ghana	Australia	34	12	2	Curitiba
Colombia	England	6	13	1	Sao Paulo
Russia	Iran	30	13	1	Recife
Argentina	Netherlands	5	13	2	Rio de Janeiro
Mexico	South Korea	32	13	2	Fortaleza
Germany	United States	11	14	1	Porto Alegre
Algeria	Japan	12	14	1	Natal
Belgium	Greece	10	14	2	Curitiba
Ecuador	Nigeria	11	14	2	Brasilia
Uruguay	Portugal	8	15	1	Recife
France	Costa Rica	10	15	1	Salvador
Switzerland	Italy	2	15	2	Rio de Janeiro
Ivory Coast	Honduras	17	15	2	Sao Paulo

Appendix C: Current Round Robin Stage Schedule

Team 1	Team 2	Difference in FIFA	Day	Game of	Stadium

		Rankings		Day	
Brazil	Croatia	17	1	1	Sao Paulo
Mexico	Cameroon	35	2	1	Natal
Spain	Netherlands	7	2	2	Salvador
Chile	Australia	45	2	3	Cuiaba
Colombia	Greece	11	3	1	Belo Horizonte
Uruguay	Costa Rica	25	3	2	Fortaleza
England	Italy	1	3	3	Manaus
Ivory Coast	Japan	27	3	4	Recife
Switzerland	Ecuador	15	4	1	Brasilia
France	Honduras	13	4	2	Porto Alegre
Argentina	Bosnia and Herzegovina	13	4	3	Rio de Janeiro
Germany	Portugal	12	5	1	Salvador
Iran	Nigeria	16	5	2	Curitiba
Ghana	United States	10	5	3	Natal
Belgium	Algeria	27	6	1	Belo Horizonte
Brazil	Mexico	23	6	2	Fortaleza
Russia	South Korea	37	6	3	Cuiaba
Australia	Netherlands	49	7	1	Porto Alegre
Spain	Chile	11	7	2	Rio de Janeiro
Cameroon	Croatia	41	7	3	Manaus
Colombia	Ivory Coast	13	8	1	Brasilia
Uruguay	England	4	8	2	Sao Paulo
Japan	Greece	29	8	3	Natal
Italy	Costa Rica	22	9	1	Recife
Switzerland	France	14	9	2	Salvador
Honduras	Ecuador	12	9	3	Curitiba
Argentina	Iran	46	10	1	Belo Horizonte

Germany	Ghana	21	10	2	Fortaleza
Nigeria	Bosnia and Herzegovina	17	10	3	Cuiaba
Belgium	Russia	14	11	1	Rio de Janeiro
South Korea	Algeria	24	11	2	Porto Alegre
United States	Portugal	1	11	3	Manaus
Netherlands	Chile	4	12	1	Sao Paulo
Australia	Spain	56	12	2	Curitiba
Cameroon	Brazil	58	12	3	Brasilia
Croatia	Mexico	6	12	4	Recife
Italy	Uruguay	3	13	1	Natal
Costa Rica	England	21	13	2	Belo Horizonte
Japan	Colombia	40	13	3	Cuiaba
Greece	Ivory Coast	2	13	4	Fortaleza
Nigeria	Argentina	30	14	1	Porto Alegre
Bosnia and Herzegovina	Iran	33	14	2	Salvador
Honduras	Switzerland	27	14	3	Manaus
Ecuador	France	1	14	4	Rio de Janeiro
Portugal	Ghana	9	15	1	Brasilia
United States	Germany	11	15	2	Recife
South Korea	Belgium	51	15	3	Sao Paulo
Algeria	Russia	13	15	4	Curitiba

Appendix D: Coefficients for Matchups, Non Linear Model

Linear Model					
Match Number	Team 1	Team 2	FIFA Ranking Difference	Interim Coefficient	Weighted Coefficient
1	Brazil	Bosnia and Herzegovina	15	0.066666667	0.5380000
2	Brazil	Croatia	17	0.058823529	0.4750000
3	Brazil	Cameroon	58	0.017241379	0.1390000
4	Bosnia and Herzegovina	Croatia	2	0.5	4.0380000
5	Bosnia and Herzegovina	Cameroon	43	0.023255814	0.1880000
6	Croatia	Cameroon	41	0.024390244	0.1970000
7	Spain	Chile	11	0.090909091	0.7340000
8	Spain	Ghana	22	0.045454545	0.3670000
9	Spain	Australia	56	0.017857143	0.1440000
10	Chile	Ghana	11	0.090909091	0.7340000
11	Chile	Australia	45	0.022222222	0.1790000
12	Ghana	Australia	34	0.029411765	0.2380000
13	Germany	United States	11	0.090909091	0.7340000
14	Germany	Algeria	30	0.033333333	0.2690000
15	Germany	Japan	42	0.023809524	0.1920000
16	United States	Algeria	19	0.052631579	0.4250000
17	United States	Japan	31	0.032258065	0.2610000
18	Algeria	Japan	12	0.083333333	0.6730000
19	Argentina	Netherlands	5	0.2	1.6150000
20	Argentina	Mexico	21	0.047619048	0.3850000
21	Argentina	South Korea	53	0.018867925	0.1520000
22	Netherlands	Mexico	16	0.0625	0.5050000
23	Netherlands	South Korea	48	0.020833333	0.1680000
24	Mexico	South Korea	32	0.03125	0.2520000
25	Colombia	England	6	0.166666667	1.3460000
26	Colombia	Russia	15	0.066666667	0.5380000
27	Colombia	Iran	45	0.022222222	0.1790000

28	England	Russia	9	0.111111111	0.8970000
29	England	Iran	39	0.025641026	0.2070000
30	Russia	Iran	30	0.033333333	0.2690000
31	Belgium	Greece	10	0.1	0.8080000
32	Belgium	Ecuador	17	0.058823529	0.4750000
33	Belgium	Nigeria	28	0.035714286	0.2880000
34	Greece	Ecuador	7	0.142857143	1.1540000
35	Greece	Nigeria	18	0.055555556	0.4490000
36	Ecuador	Nigeria	11	0.090909091	0.7340000
37	Uruguay	Portugal	8	0.125	1.0100000
38	Uruguay	France	15	0.066666667	0.5380000
39	Uruguay	Costa Rica	25	0.04	0.3230000
40	Portugal	France	7	0.142857143	1.1540000
41	Portugal	Costa Rica	17	0.058823529	0.4750000
42	France	Costa Rica	10	0.1	0.8080000
43	Switzerland	Italy	2	0.5	4.0380000
44	Switzerland	Ivory Coast	10	0.1	0.8080000
45	Switzerland	Honduras	27	0.037037037	0.2990000
46	Italy	Ivory Coast	8	0.125	1.0100000
47	Italy	Honduras	25	0.04	0.3230000
48	Ivory Coast	Honduras	17	0.058823529	0.4750000

Appendix E: Coefficients for Matchups, Linear Model

Linear Model					
Match Number	Team 1	Team 2	FIFA Ranking Difference	Interim Coefficient	Weighted Coefficient
1	Brazil	Bosnia and Herzegovina	15	85	0.7360000
2	Brazil	Croatia	17	83	0.7190000
3	Brazil	Cameroon	58	42	0.3640000
4	Bosnia and Herzegovina	Croatia	2	98	0.8490000
5	Bosnia and Herzegovina	Cameroon	43	57	0.4940000
6	Croatia	Cameroon	41	59	0.5110000
7	Spain	Chile	11	89	0.7710000
8	Spain	Ghana	22	78	0.6760000
9	Spain	Australia	56	44	0.3810000
10	Chile	Ghana	11	89	0.7710000
11	Chile	Australia	45	55	0.4760000
12	Ghana	Australia	34	66	0.5720000
13	Germany	United States	11	89	0.7710000
14	Germany	Algeria	30	70	0.6060000
15	Germany	Japan	42	58	0.5020000
16	United States	Algeria	19	81	0.7020000
17	United States	Japan	31	69	0.5980000
18	Algeria	Japan	12	88	0.7620000
19	Argentina	Netherlands	5	95	0.8230000
20	Argentina	Mexico	21	79	0.6840000
21	Argentina	South Korea	53	47	0.4070000
22	Netherlands	Mexico	16	84	0.7280000
23	Netherlands	South Korea	48	52	0.4500000
24	Mexico	South Korea	32	68	0.5890000
25	Colombia	England	6	94	0.8140000
26	Colombia	Russia	15	85	0.7360000

27	Colombia	Iran	45	55	0.4760000
28	England	Russia	9	91	0.7880000
29	England	Iran	39	61	0.5280000
30	Russia	Iran	30	70	0.6060000
31	Belgium	Greece	10	90	0.7800000
32	Belgium	Ecuador	17	83	0.7190000
33	Belgium	Nigeria	28	72	0.6240000
34	Greece	Ecuador	7	93	0.8060000
35	Greece	Nigeria	18	82	0.7100000
36	Ecuador	Nigeria	11	89	0.7710000
37	Uruguay	Portugal	8	92	0.7970000
38	Uruguay	France	15	85	0.7360000
39	Uruguay	Costa Rica	25	75	0.6500000
40	Portugal	France	7	93	0.8060000
41	Portugal	Costa Rica	17	83	0.7190000
42	France	Costa Rica	10	90	0.7800000
43	Switzerland	Italy	2	98	0.8490000
44	Switzerland	Ivory Coast	10	90	0.7800000
45	Switzerland	Honduras	27	73	0.6320000
46	Italy	Ivory Coast	8	92	0.7970000
47	Italy	Honduras	25	75	0.6500000
48	Ivory Coast	Honduras	17	83	0.7190000

Appendix F: Coefficients for Stadiums

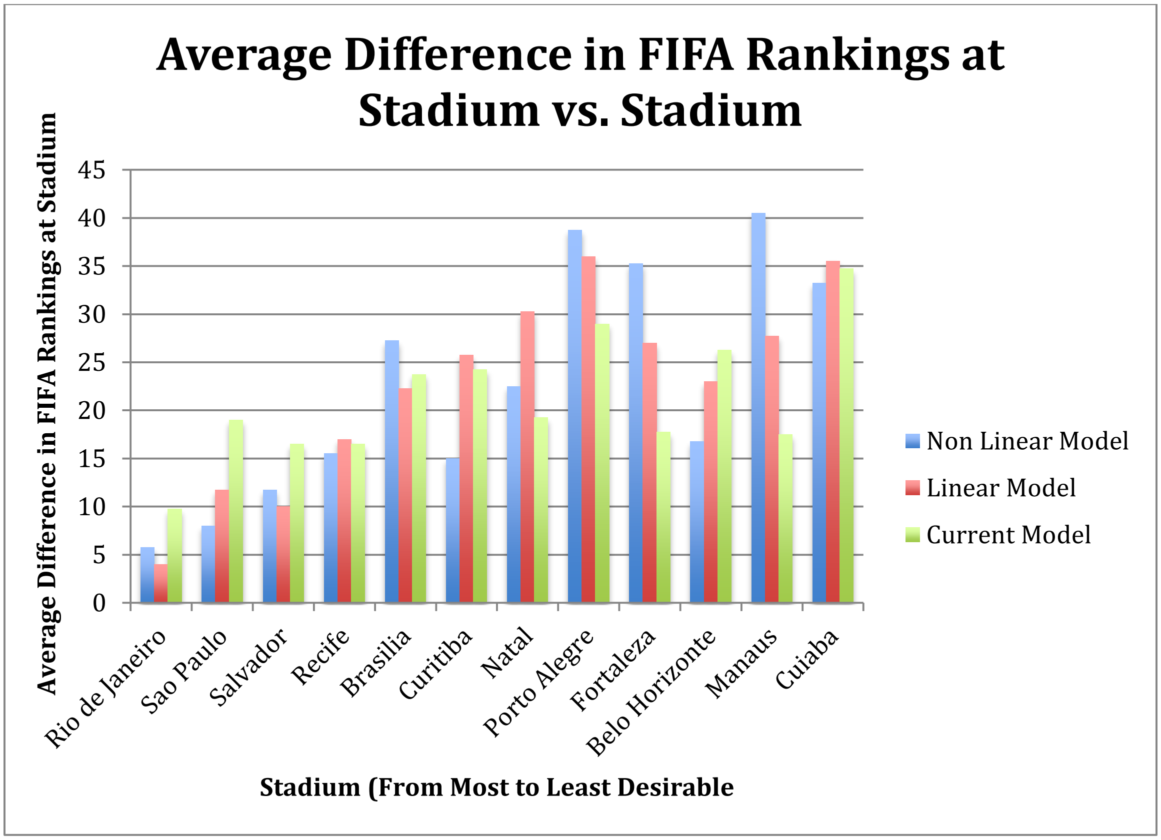
Stadium	Interim Coefficient	Weighted Coefficient
Rio de Janeiro	0.169474276	0.174
Brasilia	0.085589436	0.088
Sao Paulo	0.130717063	0.134
Fortaleza	0.065805068	0.067
Belo Horizonte	0.053205753	0.055
Porto Alegre	0.080834708	0.083
Salvador	0.095234842	0.098
Recife	0.092827509	0.095
Cuiaba	0.028212018	0.029
Manaus	0.028416922	0.029
Natal	0.084196258	0.086
Curitiba	0.085486147	0.088
Average	0.083333333	0.085

Appendix G: Coefficients for Timeslots

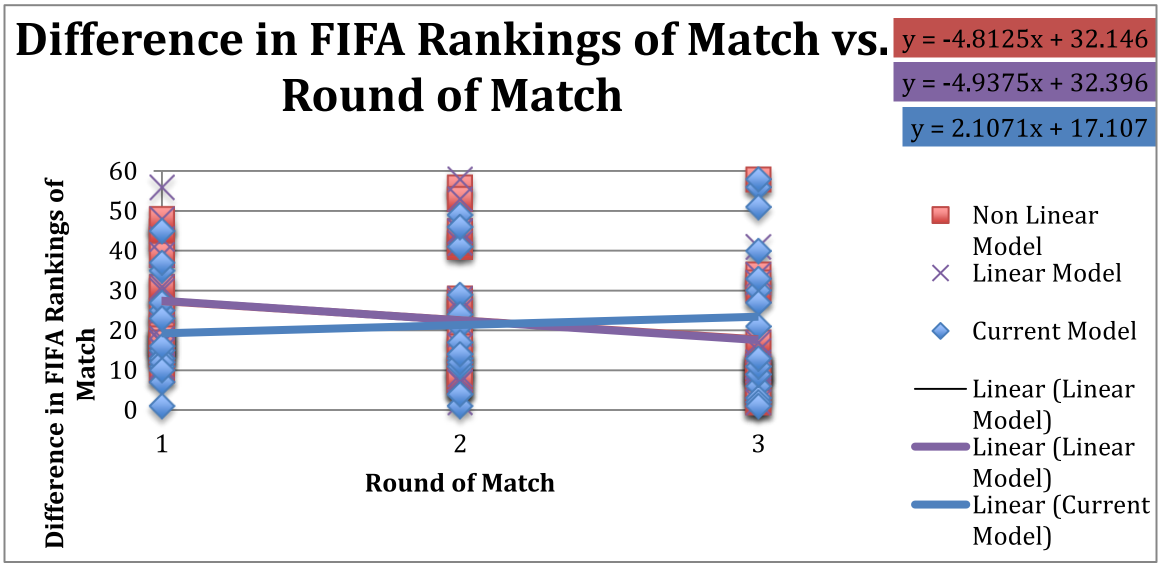
Day of Game	Round	Timeslot	1 If Game of Day is 2 in First 2 Rounds	Interim Coefficient	Weighted Coefficient
1	1	1		1	0
1	1	2		2	0
2	1	3		3	0
2	1	4	1	16	0
2	1	5		5	0
3	1	6		6	0
3	1	7	1	49	0
3	1	8		8	0
4	1	9		9	0
4	1	10	1	100	0.001
4	1	11		11	0
5	1	12		12	0
5	1	13	1	169	0.002
5	1	14		14	0
6	1	15		15	0
6	1	16	1	256	0.003
6	2	17		289	0.003
7	2	18		324	0.003
7	2	19	1	6859	0.067
7	2	20		400	0.004
8	2	21		441	0.004
8	2	22	1	10648	0.104
8	2	23		529	0.005
9	2	24		576	0.006
9	2	25	1	15625	0.153
9	2	26		676	0.007
10	2	27		729	0.007
10	2	28	1	21952	0.214

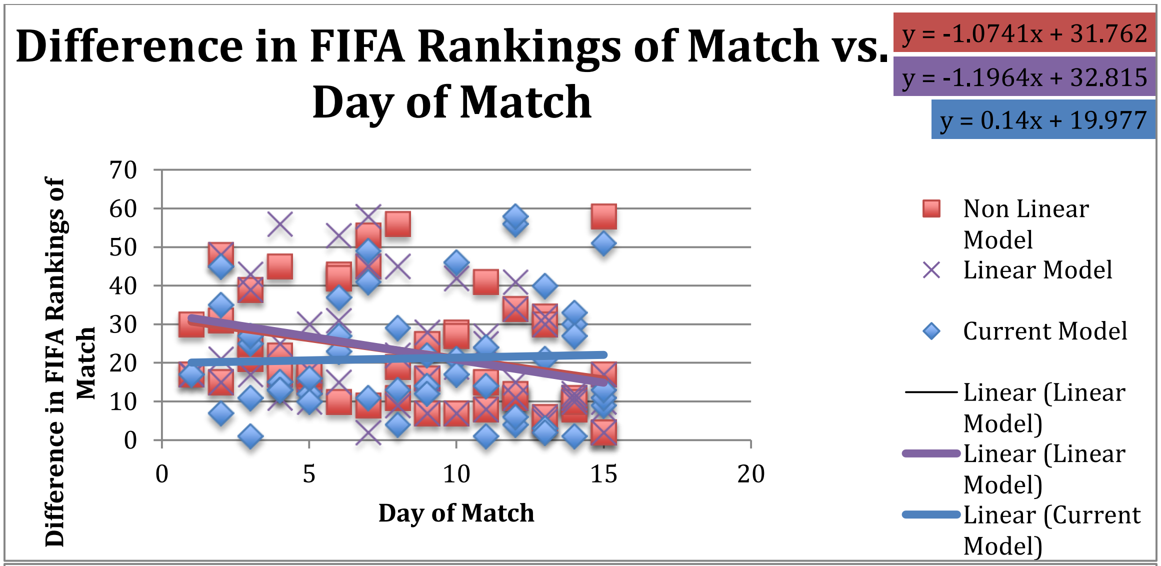
10	2	29		841	0.008
11	2	30		900	0.009
11	2	31	1	29791	0.291
11	2	32		1024	0.01
12	3	33		35937	0.351
12	3	34		39304	0.384
12	3	35		42875	0.419
12	3	36		46656	0.456
13	3	37		50653	0.495
13	3	38		54872	0.536
13	3	39		59319	0.58
13	3	40		64000	0.625
14	3	41		68921	0.673
14	3	42		74088	0.724
14	3	43		79507	0.777
14	3	44		85184	0.832
15	3	45		91125	0.89
15	3	46		97336	0.951
15	3	47		103823	1.014
15	3	48		110592	1.08

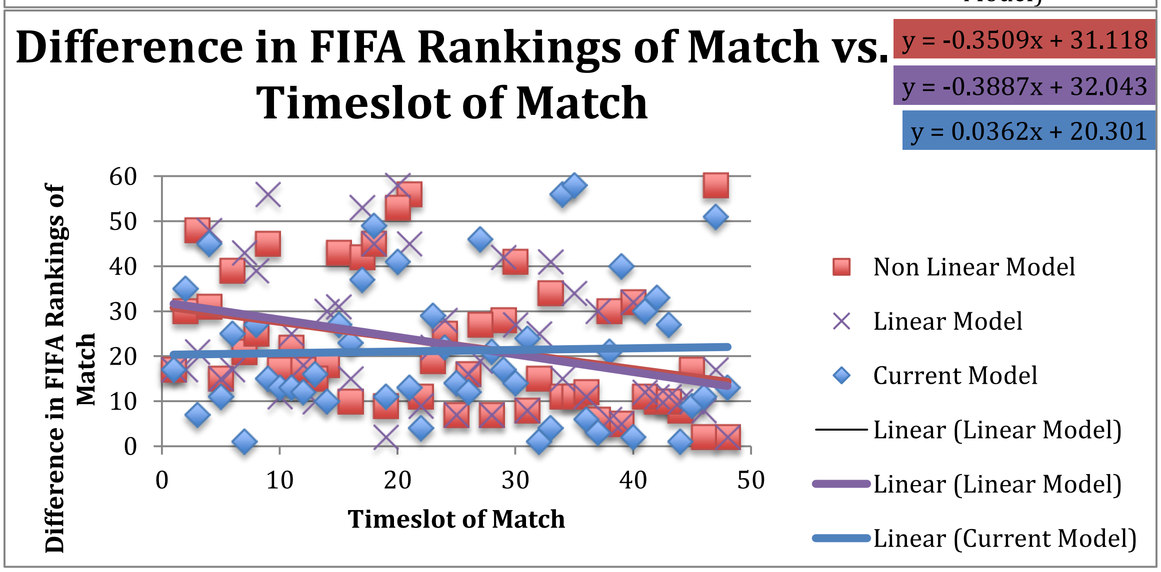
Appendix H: Bar Graph comparing Differences in Stadiums



Appendix I: Graphs showing Difference in FIFA Rankings as Tournament Progresses




Difference in FIFA Rankings of Match vs. Day of Match
y = -1.0741x + 31.762
y = -1.1964x + 32.815
y = 0.14x + 19.977
Difference in FIFA Rankings of Match
Day of Match
0
10
20
30
40
50
60
70
0
5
10
15
20
Non Linear Model
Linear Model
Current Model
Linear (Linear Model)
Linear (Linear Model)
Linear (Current Model)


Difference in FIFA Rankings of Match vs. Timeslot of Match
y = -0.3509x + 31.118
y = -0.3887x + 32.043
y = 0.0362x + 20.301
Difference in FIFA Rankings of Match
Timeslot of Match
0
10
20
30
40
50
60
0
10
20
30
40
50
Non Linear Model
Linear Model
Current Model
Linear (Linear Model)
Linear (Linear Model)
Linear (Current Model)