# Baseball Game and Umpire Crew Scheduling Optimization

Christopher Arena, Coco Cai, Logan Donaldson, James Ingram, Eli Katz*, Benjamin Keever, Justin Nam, Sam Oberly, Devon Osgood, Chloe Warren*, Emily Zhai

Johns Hopkins University | Whiting School of Engineering | Baltimore, MD
Design Day 2021

* Team student leaders

## Introduction

The Johns Hopkins University Baseball Scheduling Optimization Research Group was launched in 2011 by Anton Dahbura and Donniell Fishkind. The group uses combinatorial optimization and combinatorial design, as well as state-of-the-art software and computing resources, to create schedules for professional baseball. We have created season schedules and umpire crew schedules for the majority of the leagues in Minor League Baseball (MiLB) at all levels. Before we came on the scene, leagues in MiLB utilized very suboptimal, by-hand schedules which took weeks to create. Our group has become well-known throughout professional baseball for pioneering mathematical optimization in scheduling for leagues in MiLB.

## Objectives

Create optimized schedules for MiLB using mathematical modeling, combinatorial optimization and design, advanced software, and supercomputing resources.

## Materials and Methods

**1)** We meet with league scheduling committees and leadership to determine the league's scheduling rules and priorities. Part of this task is to help the leagues themselves quantify a good schedule and resolve the competing interests and goals into a single objective function.

**2)** An appropriate skeleton ("template") is created, which includes gross features such as season midway point, off-day spread, holiday patterns, and division of the days into series units, etc.

**3)** We model the league constraints using appropriate variables and constraints. Quadratic constraints are converted into linear ones. The goal is the formation of a binary integer linear program whose objective function reflects the "badness" of a feasible schedule. Matrices and vectors with the specific problem parameters are created by building and running a computer program.

**4)** The matrices and vectors with the specific parameters of the binary integer linear program representing the schedule optimization are fed into a state-of-the-art solver, running on a 160-core computer until an optimal solution is produced.

**5)** The output is distilled into VBA spreadsheets that clearly summarize the descriptive statistics of the created schedule, and highlight important features.

**6)** We meet with the league scheduling committee and leadership to tune the model, and we adjust the priorities if this is useful.

*Our computers*:

**Ziggy** is an SGI UV2000 System with 8x Intel E5-4650V2 processors (10 cores each @ 2.4Ghz, 25M cache) and 256GB DDR3 1866MT/s Memory.

**Chillywilly** is a custom Penguin Computing system with 4x AMD Opteron 6378 processors (16 cores each @ 3.3Ghz, 16MB cache) and 128GB DDR3 1600MT/s Memory.

## Results

### Umpire Crew Scheduling –

Umpires are responsible for enforcing the rules of the game on the field and adjudicating all aspects of the game. There are half as many umpire crews as teams in a league. Umpire crew scheduling is built on the season schedule for the league, and the requirements and priorities of the crew schedule vary somewhat from league to league. Standard requirements include some of the following:

- Crew travel should be minimized;
- Differences between total travel of different crews should not be extreme;
- Crews are to avoid umpiring that involves the same team two series in a row;
- The number of times a Crew sees each team should be in a fixed range of values;
- The number of times a Crew visits each stadium should be in a fixed range of values;
- Crews cannot travel more than 500 miles without an off day;
- Crew travel patterns should be as sensible as possible to minimize ``ping-ponging" and other negative patterns specified by the league;

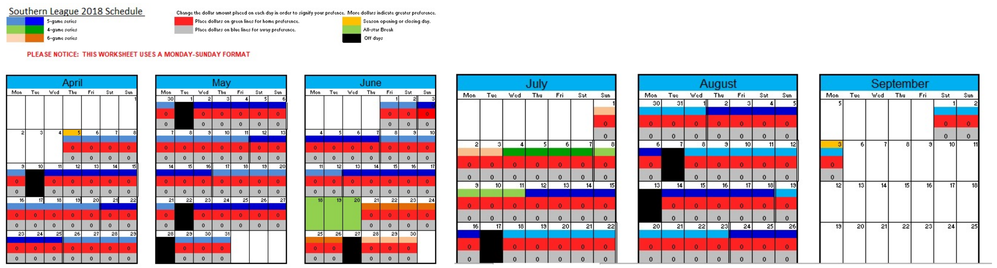


**Figure 2— Diamond Dollars for Season Scheduling**

In the Diamond Dollars concept, each team is given a fixed number of tokens, or chips, to assign to days of the season that they would prefer to be home or away. The number of chips a team places on a day reflects the importance of that request. A team could in theory place all of its chips on one day of the season, or they could spread the chips around the days of the season. Diamond Dollars supplement the other league and team constraints and are not intended to supersede or interfere with them. The group came up with a way of adjusting the priority of Diamond Dollars relative to the other constraints for the season. Surprisingly to many, the optimization was able to satisfy over 75% of the team Diamond Dollar preferences without disrupting the overall quality of the league schedule. The system was especially effective at identifying team requests that were complementary to the rest of the constraints. The Carolina League adopted the SAL Bucks/Diamond Dollars system for use in one of their recent schedules.

$$\forall j, \sum_{i=1}^{\#\,series} (\#games\ in\ series\ i)x_{i,j,j} - s_{j1} \leq n$$

$$\forall j, \sum_{i=1}^{\#\,series} -(\#games\ in\ series\ i)x_{i,j,j} - s_{j1} \leq -n$$

```
for i=1:nslots
    for k=1:nteams
        req=req+1;
        for j=1:nteams
            if j==k
                Aeq(req,(i-1)*nteams*nteams+(j-1)*nteams+(k-1)+1)=1;
            else
                Aeq(req,(i-1)*nteams*nteams+(j-1)*nteams+(k-1)+1)=-1;
            end
        end
        beq(req,1)=0;
    end
end
```

**Figure 3a,b— Constraint modeling, very simplified examples**

Modeling is done through Binary Integer Linear Programming. BIP is used to solve a system of binary linear inequalities. The decision variables are constructed as follows:

$X_{i,j,k}$ where the subscript i is series index, j is team index, and k is stadium index.

$X_{i,j,k} = 1$ would indicate that team j is playing team k at team k's stadium during series slot i.

The constraints are designed to enforce the league requirements and priorities.

Artificial variables are created as needed to allow penalized violations. This is important, as there is no perfect schedule, and there are always competing goals that need to be optimized.

**3a) Every team must have a specified number of home games:** For each team, the number of games played at home is n, with an allowance of deviation by one game at a penalty, enforced through an artificial variable.

**3b) Every game is home versus visitor:** We iterate through teams within stadiums within slots. We add to the iterative count of req because we are adding an equality constraint. Within our iteration over teams, we are seeing that if the team is at its stadium, then there should be another team there as well and that if it is not, there should be no teams there.

| | | Total Mileage: | | 2,527 | Total Mileage: | | 3,122 | Total Mileage: | | 2,780 | Total Mileage: | | 3,006 | Total Mileage: | | 2,568 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Rank: | | 5 | Rank: | | 1 | Rank: | | 3 | Rank: | | 2 | Rank: | | 4 |
| | | Average Trip: | | 109.9 | Average Trip: | | 135.7 | Average Trip: | | 103.0 | Average Trip: | | 111.3 | Average Trip: | | 102.7 |
| Day | Date | | Crew 1 | | | Crew 2 | | | Crew 3 | | | Crew 4 | | | Crew 5 | |
| Thursday | 3-Jun | BLU | KNG | | JCY | BRS | | GRN | ELZ | | BUR | PRN | | PUL | DAN | |
| Friday | 4-Jun | BLU | KNG | | JCY | BRS | | GRN | ELZ | 123 | BUR | PRN | 12 | PUL | DAN | 40 |
| Saturday | 5-Jun | JCY | KNG | | ELZ | BRS | | PRN | PUL | | GRN | BLU | | DAN | BUR | |
| Sunday | 6-Jun | JCY | KNG | 132 | ELZ | BRS | 189 | PRN | PUL | 123 | GRN | BLU | 0 | DAN | BUR | 181 |
| Monday | 7-Jun | OFF | OFF | | OFF | OFF | | OFF | OFF | | OFF | OFF | | OFF | OFF | |
| Tuesday | 8-Jun | BRS | PRN | | PUL | DAN | | JCY | GRN | | BUR | BLU | | KNG | ELZ | |
| Wednesday | 9-Jun | BRS | PRN | 61 | PUL | DAN | 40 | JCY | GRN | 143 | BUR | BLU | 159 | KNG | ELZ | 8 |
| Thursday | 10-Jun | PRN | PUL | | BRS | BUR | | KNG | GRN | | DAN | BLU | | ELZ | JCY | |
| Friday | 11-Jun | PRN | PUL | 143 | BRS | BUR | 181 | KNG | GRN | 159 | DAN | BLU | 103 | ELZ | JCY | 134 |
| Saturday | 12-Jun | BUR | PUL | | BLU | ELZ | | DAN | PRN | | KNG | BRS | | GRN | JCY | |
| Sunday | 13-Jun | BUR | PUL | 120 | BLU | ELZ | 38 | DAN | PRN | 0 | KNG | BRS | 41 | GRN | JCY | 23 |
| Monday | 14-Jun | OFF | OFF | | OFF | OFF | | OFF | OFF | | OFF | OFF | | OFF | OFF | |
| Tuesday | 15-Jun | BUR | DAN | | BRS | GRN | | BLU | PRN | | PUL | JCY | | ELZ | KNG | |
| Wednesday | 16-Jun | BUR | DAN | 189 | BRS | GRN | 257 | BLU | PRN | 140 | PUL | JCY | 141 | ELZ | KNG | |
| Thursday | 17-Jun | JCY | BRS | | PRN | BUR | | GRN | ELZ | | DAN | BLU | | PUL | KNG | |
| Friday | 18-Jun | JCY | BRS | 56 | PRN | BUR | 38 | GRN | ELZ | 26 | DAN | BLU | 127 | PUL | KNG | 214 |
| Saturday | 19-Jun | PUL | GRN | | BLU | BUR | | BRS | ELZ | | JCY | KNG | | PRN | DAN | |
| Sunday | 20-Jun | PUL | GRN | 257 | BLU | BUR | 143 | BRS | ELZ | 26 | JCY | KNG | 132 | PRN | DAN | 154 |
| Monday | 21-Jun | OFF | OFF | | OFF | OFF | | OFF | OFF | | OFF | OFF | | OFF | OFF | |

| Club Balance | | Crew 1 | | | Crew 2 | | | Crew 3 | | | Crew 4 | | | Crew 5 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | v | @ | TOTAL | v | @ | TOTAL | v | @ | TOTAL | v | @ | TOTAL | v | @ | TOTAL |
| | KNG | 4 | 7 | 11 | 6 | 4 | 10 | 7 | 4 | 11 | 6 | 4 | 10 | 4 | 7 | 11 |
| | BRS | 6 | 5 | 11 | 8 | 3 | 11 | 4 | 6 | 10 | 5 | 6 | 11 | 4 | 6 | 10 |
| | JCY | 5 | 4 | 9 | 7 | 4 | 11 | 4 | 8 | 12 | 8 | 4 | 12 | 2 | 7 | 9 |
| | ELZ | 7 | 4 | 11 | 6 | 6 | 12 | 4 | 5 | 9 | 2 | 7 | 9 | 8 | 4 | 12 |
| | GRN | 8 | 3 | 11 | 2 | 8 | 10 | 7 | 4 | 11 | 4 | 8 | 12 | 5 | 4 | 9 |
| | BLU | 5 | 4 | 9 | 4 | 6 | 10 | 6 | 5 | 11 | 4 | 8 | 12 | 7 | 4 | 11 |
| | PRN | 4 | 6 | 10 | 8 | 3 | 11 | 6 | 6 | 12 | 5 | 3 | 8 | 4 | 8 | 12 |
| | PUL | 6 | 6 | 12 | 3 | 6 | 9 | 2 | 8 | 10 | 8 | 3 | 11 | 7 | 4 | 11 |
| | DAN | 2 | 8 | 10 | 4 | 7 | 11 | 7 | 4 | 11 | 6 | 4 | 10 | 8 | 3 | 11 |
| | BUR | 6 | 6 | 12 | 5 | 6 | 11 | 6 | 3 | 9 | 5 | 6 | 11 | 4 | 6 | 10 |
| | OFF | 14 | 14 | | 14 | 14 | | 14 | 14 | | 14 | 14 | | 14 | 14 | |

**Figure 4 – Appalachian Umpire Crew Schedule**

Due to its compact size, travel distance constraints are less concerning when developing Appalachian League umpire schedules. Apart from penalizing consecutive 180-mile trips, the optimization is free to prioritize each umpire crew's distribution amongst the teams resulting in some particularly pretty schedules.

For instance, in the above example, the total number of games each crew sees each club is near uniform. The largest discrepancy, a mere four games, occurs in Crew 4, which sees Princeton 8 times and Elizabethton 12 times.

Also, note that this uniformity was achieved without the implementation of an explicit constraint, which penalized large discrepancies between the number of times a crew saw each team. Rather, this distribution arose naturally as a byproduct of setting upper and lower bounds on the number of times each crew should see each team; and then requiring crews to move around semi-frequently.

This schedule was adopted immediately without tweaks. In general, leagues have the option of adjusting their requirements and priorities and tradeoffs in response to a schedule presented to them.

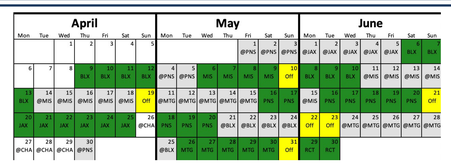


**Figure 5 – Future Directions**

Currently under development is a new scheduling-making paradigm utilizing machine learning rather than binary integer linear programming. The machine learning framework involves incorporating analogous concepts to that of the program AlphaGo -- the artificial intelligent version that was developed by DeepMind and acquired by Google. The scheduling Python framework is like that of games, which determines which scheduling moves are optimal in the immediate and long term. More specifically, within the game, each team is represented by a machine learning agent. These agents then take sequential turns. A turn consists of selecting a series and choosing whether to be home or away. If a team chooses to be away, they may also choose their opponent. Turns are evaluated based on how well they adhere to the scheduling constraints. The game ends when every series is filled, and a schedule is created. The goal is to leverage machine learning to perform robust scheduling as well as determine which of these moves are most beneficial. In doing so, statistical analysis and tree-based methodologies are explored to determine which avenues and lines of play are worth evaluating and pursuing. The figure above shows part of an example team's schedule resulting from a recent simulation.

## Conclusion

Our methods have proven to be effective and highly adaptable. Future directions include advancing the state of the art in season scheduling and umpire crew scheduling for professional baseball leagues.