Using Robust Optimization to Assign Events to MIT Women Swimmers

Daphne Kallinderi, Felicia Nappi, Kiera Patterson
15.C57 Fall 2024

Context of Problem

The MIT Women's Swim Team, consisting of 31 athletes, recently gained recognition after securing 5th place at the NCAA Division III Swimming and Diving Championship last year. However, in the 2024/2025 season, they lost to Harvard University during their second meet[1]. This project aims to develop an optimization-based method to assign MIT swimmers to specific events in a way that maximizes their points, with the goal of outperforming Harvard's team in future meets. The approach incorporates both deterministic and stochastic optimization techniques to account for uncertainty in swimmer performance. The project draws inspiration from a 2006 paper addressing a similar problem[2]. While the paper provided foundational ideas, the formulation used here differs and incorporates uncertainty to reflect real-world variability—since swimmers' performances can deviate from their average times. By optimizing event assignments under these uncertain conditions, this method offers a practical strategy that could help the team improve its rankings and secure future victories.

Relevance of Problem

In context, this is a relatively hard problem for a coach to solve, especially since it is dependent on the strengths and weaknesses of the opposing team. There is an inherent trade-off the coach (and in this case the model) must solve: should the best swimmer be allocated to a highly competitive event to challenge the opposing team's strongest swimmers, or should they be assigned to less competitive events where they have a higher likelihood of securing points? Trying to balance this to maximize overall points is difficult. There are also constraints with the amount a swimmer can participate, as the top swimmers can only swim a certain amount per meet. This dynamic swimmer assignment is challenging for a coach to reliably do, and the optimization model alleviates this challenge.

The optimization model presented in this project attempt to address these concerns without the risk of human error. The model avoids potential biases that could influence decisions—such as favoring certain swimmers or relying on heuristics—and ensures a data-driven, objective approach to event assignments. This in turn could help recruit faster

[1]https://swimswam.com/wp-content/uploads/2024/10/mit-harvard-northeastern-2024.pdf
[2]https://websites.umich.edu/ mepelman/research/swimmeet.pdf

swimmers and allow for long-term growth of the program in the coming years. While this problem focuses solely on the MIT Women's team, a similar formulation could scale and be used for other sports teams or other larger organizations such as the METS (Metropolitan Swimming Conference) or the Olympics.

Data

Data was taken from MIT's Athletic website which reports the athletes and their swim times in historical meets. For each swimmer-event combination, data points include recorded swim times across all meets during the previous season, along with times from the first meet of the current season. The challenge of lack of data for freshmen was addressed by the collection of times from the first few meets of the current season. This is crucial in helping the model make appropriate decisions when accounting for uncertainty. If there was not a recorded time for a certain swimmer-event combination their time for that event was set to big-M.

For Harvard's swimmers, performance data was collected from their most recent meet before facing MIT. These times were treated as point estimates, just to help make an appropriate comparison and validate the model. While point estimates for the opposing team may introduce some uncertainty, they reflect the most recent performance trends and help simulate real-world decision-making conditions, where opposing teams' future performances are unknown.

The data described above had to be taken from PDFs on the MIT Athletic Website. To extract the information, it had to be manually added to Excel. The data then went through preprocessing to standardize and enforce format consistency. It was then reconfigured using Python to make the parts of the problems (swimmer times, event placements) appropriate data frames for Gurobi formulations.

Problem Formulation

Three linear optimization formulations were developed to address the swimmer-assignment problem.

1 First Model

The first formulation minimizes the total swim time at a meet for MIT. This is the simplest model, as it does not take an opposing team into account. Due of this, there are no assigned points and no uncertainty in this problem. The constraints for this model were mainly taken from the NCAA rulebook which outlines how often a swimmer can participate. There are also a few constraints included for swimmer well-being, such as swimmers cannot swim back to back events. The goal of this model was to validate that the data and assignment formulation is compatible with Gurobi. The formulation is as follows:

Variables and Sets

$$
\begin{aligned}
r &= \{50 \text{ back}, 50 \text{ breast}, 50 \text{ fly}, 50 \text{ free}, 50 \text{ free}, 50 \text{ free}, 50 \text{ free}, 50 \text{ free}\} \\
i \in I &: \text{ set of individual events} \\
x_{si} &: \text{ binary, 1 if swimmer } s \text{ swims individual event } i \\
t_{si} &: \text{ time for swimmer } s \text{ to swim individual event } i \\
y_{srg} &: \text{ binary, 1 if swimmer } s \text{ swims relay event } r \text{ in group } g \\
t_{sr} &: \text{ time for swimmer } s \text{ to swim relay event } r
\end{aligned}
$$

Objective Function

$$\min \sum_s \sum_i t_{si} x_{si} + tg$$

$$\sum_i x_{si} \leq 3 \quad \forall s \quad \text{(each swimmer swims at most 3 events)}$$

$$\sum_{s=1}^{\text{S}} \sum_{k=1}^{\text{K}} x_{s,i,k} = 3, \quad \forall i \quad \text{(3 swimmers per individual event)}$$

$$\sum_{r=1}^{4} \sum_s y_{srg} = 4 \quad \forall g \quad \text{(4 people swim in the first relay)}$$

$$\sum_{r=5}^{8} \sum_s y_{srg} = 4 \quad \forall g \quad \text{(4 people swim in the second relay)}$$

$$x_{si} + x_{s,i+1} \leq 1 \quad \forall s, \forall i \quad \text{(swimmers can't swim back-to-back events)}$$

$$\sum_g \sum_{r=1}^{4} y_{srg} \leq 1 \quad \forall s \quad \text{(swimmers can only swim once in the first relay)}$$

$$\sum_g \sum_{r=5}^{8} y_{srg} \leq 1 \quad \forall s \quad \text{(swimmers can only swim once in the second relay)}$$

$$\sum_g \sum_{r=1}^{4} y_{srg} + x_{s1} \leq 1 \quad \forall s \quad \text{(swimmers that swim in the first relay can't swim in the first individual event)}$$

$$\sum_g \sum_{r=5}^{8} y_{srg} + x_{s12} \leq 1 \quad \forall s \quad \text{(swimmers that swim in the second relay can't swim in the last individual event)}$$

Group Time Calculation

$$t_g = \sum_s \sum_{r=1}^{4} t_{sr} y_{srg} + \sum_s \sum_{r=5}^{8} t_{sr} y_{srg} \quad \forall g$$

Relay Event Constraints

$$\sum_s y_{s1g} = 1 \quad \forall g \in \{1,2,3\} \quad \text{(only 1 swimmer can swim first position in a relay group)}$$

$$\sum_s y_{s2g} = 1 \quad \forall g \in \{1,2,3\} \quad \text{(only 1 swimmer can swim second position in a relay group)}$$

$$\sum_s y_{s3g} = 1 \quad \forall g \in \{1,2,3\} \quad \text{(only 1 swimmer can swim third position in a relay group)}$$

$$\sum_s y_{s4g} = 1 \quad \forall g \in \{1,2,3\} \quad \text{(only 1 swimmer can swim fourth position in a relay group)}$$

2 Second Model

The second linear optimization formulation aims to maximize the points scored for MIT against the swim times of Harvard. This formulation has more real-world application, and as a result, is more complicated. Certain teams are better at different events, so this model also has to take into account the opponents' strengths and weaknesses. This formulation captures that dynamic element the first lacks, and the approach is more similar to the strategic decisions a coach must make. To maintain a realistic scope of the project within the time frame allotted, only individual points were focused on for this formulation. The problem was formulated as such:

Points and Placing Parameters

$$\begin{aligned} P_i &: \text{ points vector for individual events } [9,5,3,2,1] \\ X_{sik} &: 1 \text{ if swimmer } s \text{ finishes event } i \text{ in place } k \\ t_{si} &: \text{ mean time for swimmer } s \text{ to swim individual event } i \end{aligned}$$

Opponent Variables

$$\begin{aligned} t_{oi} &: \text{ time for opponent } o \text{ to finish event } i \\ z_{iok} &: 1 \text{ if opponent } o \text{ finishes event } i \text{ in place } k \end{aligned}$$

$$\max \sum_s \sum_i \sum_k P_i X_{isk}$$

Placing Constraints The constraint ensuring that the time of the swimmer or opponent in place k is less than or equal to the time in place $k+1$ is:

$$\sum_{s=1}^{\mathrm{S}} X_{sik_1} \cdot t_{si} + \sum_{o=1}^{\mathrm{O}} z_{iok_1} \cdot \min(t_{oi}, M) \leq \sum_{s=1}^{\mathrm{S}} X_{sik_2} \cdot t_{si} + \sum_{o=1}^{\mathrm{O}} z_{iok_2} \cdot \min(t_{oi}, M)$$

$$\forall i, \quad k_1 \in \{1, \ldots, 6-1\}, \quad k_2 \in \{k_1+1, \ldots, 6\}.$$

Additional Constraints

$$\sum_s x_{isk} + \sum_o z_{iok} \leq 1 \quad \forall i, \forall k \quad \text{(only one MIT swimmer or opponent per place)}$$

$$\sum_i x_{si} \leq 3 \quad \forall s \quad \text{(each swimmer swims at most 3 events)}$$

$$x_{si} + x_{s,i+1} \leq 1 \quad \forall s, \forall i \quad \text{(swimmers can't swim back-to-back events)}$$

$$\sum_{s=1}^{\mathrm{S}} \sum_{k=1}^{\mathrm{K}} x_{s,i,k} = 3, \quad \forall i \quad \text{(3 swimmers per individual event)}$$

3 Third Model

The final formulation aims to build on the second formulation to incorporate uncertainty. This step allows for a more applicable model given swimmers' performance and therefore their times fluctuate due to external factors. To do this, it was assumed that swimmer's times approximately follow a normal distribution. The following constraints were added to the previous formulation:

μ_{si} : mean time for swimmer s to swim individual event i

σ_{si} : standard deviation of the time for swimmer s to swim individual event i

$$t_{si} \leq \mu_{si} + \sigma_{si}$$
$$t_{si} \geq \mu_{si} - \sigma_{si}$$

Findings

The first formulation produced a total time of 2412.8 seconds (approximately 40 minutes), including individual events and relays. A total of seventeen out of the thirty-one swimmers participated, with none swimming more than three events, except for Sydney Smith. This makes logical sense, as Smith is consistently mentioned in MIT Athletic Articles and is known for being a strong member of the team.

In the second formulation that did not account for uncertainty, 74 points were scored by MIT in individual events. For context, in reality MIT scored 77 points in this match against Harvard on October 25th, 2024 while Harvard won with 151 points. This model preformed worse than the coach's assignment. However, in the real meet many swimmer performed better than their average which this model does not account for. In addition, five swimmers are assigned to events without historical times, so it is assumed by the model that they will place last. Twenty-two of the swimmers participated. It is important to note this model does not take uncertainty into account.

Fortunately, the third and final formulation shows extreme promise. In this scenario, MIT scored 117 points while Harvard scored 111, leading to a victory for MIT in individual events. In this case, only one swimmer is assigned to an event they typically do not swim. Only nineteen swimmers of the thirty-one swam in this meet. MIT is able to increase the points acquired in this meet by 40 points by incorporating uncertainty. This is a 51.9 percent increase in points won in individual events, as opposed to the 23.7 percent increase in the original paper (their individual points had increased from 59 to 73 using their optimization formulation). This model also helps to show which swimmers are more important in order to win. Specifically, the optimization model outputted a quicker than average time for some of the swimmers and an average or worse than average time for others. In a real meet, it would be more important for those swimmers who have a quicker than average output to perform their best.

Event	**Swimmers Assigned by the Coach**	**Swimmers Assigned by the Robust Optimization**
1000 Yard Freestyle	Jessie Crane, Christina Beggs, Kim Jolie	Christina Beggs, Lauren Adler, Jessie Crane
200 Yard Freestyle	Ella Robertson, Sydney Smith, Mary Feliz	Ella Robertson, Sydney Smith, Belise Swartwood
100 Yard Backstroke	Iris Yang, Kate Augustyn, Olivia Chen	Christina Beggs, Kate Augustyn, Olivia Chen
100 Yard Breaststroke	Sarah Bernard, Kailey Simons, Anna Li	Sarah Bernard, Kailey Simons, Katherine Yao
200 Yard Butterfly	Belise Swartwood, Natalie Tang, Sydney Chun	Belise Swartwood, Mary Felix, Christina Beggs
50 Yard Freestyle	Alexandra Turvey, Annika Naveen, Rachel Yang	Kailey Simons, Sydney Smith, Natalie Tang
100 Yard Freestyle	Ella Robertson, Alexandra Turvey, Annika Naveen	Ella Robertson, Alexandra Turvey, Iris Yang
200 Yard Backstroke	Kate Augustyn, Iris Yang, Mary Feliz	Kate Augustyn, Sydney Smith, Natalie Tang
200 Yard Breaststroke	Sarah Bernard, Anna Li, Katherine Yao	Katie Kudela, Olivia Chen, Katherine Yao
500 Yard Freestyle	Belise Swartwood, Jessie Crane, Lauren Adler	Belise Swartwood, Jessie Crane, Sarah Bernard
100 Yard Butterfly	Sydney Smith, Alexandra Turvey, Lauren Levy	Natalie Tang, Mary Feliz, Lauren Levy
200 Yard IM	Kate Augustyn, Belise Swartwood, Lauren Adler	Kate Augustyn, Ella Robertson, Katherine Yao

Table 1: Swimmer Assignments by Event

As seen above in Table 1, every event had at least one different assigned swimmer when comparing the assignments made by the coach and the assignments made by the robust optimization model. However, in every event besides 50 Freestyle, there is at least one overlapping swimmer.

Impact of Findings

The results of this paper, particularly in the final formulation, show that the use of optimization can have a big impact on optimizing event assignments in a swim meet. While the deterministic nature of the first two models served as a good baseline, it was truly incorporating the realistic uncertainty element that allowed MIT to win. Notably, the model was much more willing to sacrifice results in one event in order to achieve better results in another event, than the assignments from the coach. This could be a useful tool for swim coaches, as it removes the human element of bias. Instead, they can focus on the historical data of their swimmers and leverage uncertainty to make better decisions.

Extension of Problem

As stated before, this is a smaller-scale project; however, there are many opportunities for extension to a higher scale. With more time, this can scale to include relay events as well as the MIT men's team to maximize the benefit for the entire swimming program. Different distributions for the uncertainty could be explored, such as a log-normal or Bernoulli distribution. Another valuable development could be to build a live model: that is, have the model automatically adjust as new data becomes available. The results of this study demonstrate the model's promise in optimizing event assignments.

Model Implementation

https://github.com/felicianappi8/Optimization-Final-Projects