Assignment of Swimmers to Dual Meet Events

Maciek Nowak
Georgia Institute of Technology, Atlanta, GA

Stephen M. Pollock
Marina Epelman
University of Michigan, Ann Arbor, Michigan

February 7, 2003

Abstract

Every fall, thousands of high school swimming coaches across the country begin the arduous process of preparing their athletes for competition. With a grueling practice schedule and a dedicated group of athletes, a coach can hone the squad into a cohesive unit, poised for any competition. However, oftentimes all preparation is in vain, as coaches assign swimmers to events with a lineup that is far from optimal. This paper provides a model that may help a high school (or other level) swim team coach make these assignments. Following state and national guidelines for swim meets, we describe a binary integer model that allows an examination of the times for swimmers on the squad, compares these to those of the expected opponent, and determines an overall assignment that maximizes the total number of points scored by the squad.

Every fall, thousands of high school swimming coaches across the country begin the arduous process of preparing their athletes for competition. With a grueling practice schedule and a dedicated group of athletes, a coach can hone the squad into a cohesive unit, poised for any competition. However, oftentimes all preparation is in vain, as coaches assign swimmers to events with a lineup that is far from optimal. Prior to each meet, a coach must decide which athletes will compete in which events. In a sport where every point counts, these decisions are extremely important. Making one poor assignment may cost the team a victory. Coach Denny Hill, winner of twenty state championships at Ann Arbor (Michigan) Pioneer High School and former national swim coach of the year, says, "a lot of times we wonder why a coach used the swimmers that they did."

Each meet consists of multiple events in various disciplines, such as freestyle and backstroke, with points awarded in each event based on placement. Swimmers can compete in individual or relay events. However, each swimmer is restricted in the number of events he/she can race, due to both meet rules and physical limitations.

Determining which athletes to assign to which events is a difficult task, often taking years to master. Analyzing the individual performances of a squad of sixty swimmers in order to determine which two or three should compete in one event while keeping in mind the ten other events to be assigned can be next to impossible. Coach Hill believes that the main cause of poor assignments is that many young coaches have neither the time nor the experience to create a competitive lineup.

This paper provides a model that may help a high school (or other level) swim team coach to assign swimmers to events at a meet. Although there may be many different objectives (maximizing the number of points won, maximizing the probability that the team wins, providing an opportunity for swimmers to qualify for later meets, etc.), this model only addresses the goal of maximizing the number of points won by a team.

1 Guidelines

Because determining the optimal placement of swimmers in events is a constrained assignment problem, a binary integer program is used for the model formulation. The size of a *squad* (the group of athletes eligible to compete) may vary widely among teams. We assume (without loss of generality) that the squad has enough swimmers to fill all the events. Although in practice swimmers may be assigned to an event just prior to that event, the *roster* (the assignment of swimmers to events) is generally fixed at the beginning of the meet.

Therefore, we assume that the allocation of swimmers is completed prior to the first event and no changes are made during the meet. Although swim teams can compete in a variety of meets, we focus on a dual meet, in which two teams compete.

The assignment of swimmers is limited by constraints currently in effect in many high school competitions:

- each team can enter at most three swimmers in any one individual (non relay) event. Thus each individual event usually has six swimmers (three from each team), unless a team fails to (or chooses not to) enter three;
- each team can provide at most three "entries" in each relay event, where an entry is a group of four swimmers;
- each swimmer can enter at most four events;
- each swimmer can enter at most two individual events;
- in a relay event, a team cannot be awarded points for more than two finishing places.

Coaches make decisions about which events their swimmers should enter using information such as their event times, their capabilities of performing a particular sequence of events, their performance in competition compared to training, etc. The coaches also have some limited information about the opposing team's swimmers. We assume that all assignments are made based on an estimated time for each swimmer on "our" squad (hereafter referred to as the "squad") for each event. These estimates can be made using information from earlier meets, training sessions or previous seasons. They may also be adjusted depending on how a coach believes a swimmer will compete under certain conditions, such as swimming two consecutive events.

Opponent times are also estimated using previously observed times from past meets, which are generally available to a coach. An opponent's roster is then estimated using these times, and is assumed to be set prior to determining the squad assignments.

2 Model Formulation

Our model compares the times for swimmers on the squad to those of the opponent, and determines an overall assignment that maximizes the total number of points.

Definitions

- $A = \{1, 2, ..., E\}$ is the set of all events. Typically $E = 11$ in a high school dual meet (diving is generally a 12th event, but is not included in this model).

- I is the set of individual events

- R is the set of relay events, where $I \cap R = \emptyset$ and $I \cup R = A$.

- t_{ij} is the estimated time for swimmer i in event $j, j \in A$.

- $a(1)_j$, $a(2)_j$ and $a(3)_j$ are the best, second best and third best times, respectively, for the opponent's swimmers in event j. Because of increased accuracy in timing, it is reasonable to assume that there are no ties in a race, so that no two times are equal and $a(1)_j < a(2)_j < a(3)_j$ for all j. For convenience in notation, we also define (fr all j) $a(s)_j = 0$ if $s \leq 0$ and $a(s)_j = M$ (an arbitrarily large number) if $s \geq 4$.

- x_{ij}, y_{ij}, and z_{ij} are assignment variables that indicate whether swimmer i on the squad competes in event j and, if so, has the best, the second best, or the third best time on the squad, respectively; so that

  $$x_{ij} \equiv \begin{cases} 1 & \text{if } i \text{ is assigned to event } j \text{ and has the best time on the squad in event } i \\ 0 & \text{otherwise} \end{cases}$$

  $$y_{ij} \equiv \begin{cases} 1 & \text{if } i \text{ is assigned to event } j \text{ and has the } 2^{nd} \text{ best time on the squad in event } i \\ 0 & \text{otherwise} \end{cases}$$

  $$z_{ij} \equiv \begin{cases} 1 & \text{if } i \text{ is assigned to event } j \text{ and has the } 3^{rd} \text{ best time on the squad in event } i \\ 0 & \text{otherwise} \end{cases}$$

- $r(1)_j$, $r(2)_j$, and $r(3)_j$ are the "realized" times for the best, second best and third best swimmers on the squad in event j. Since four swimmers are assigned to each relay event, their "realized" time for the event is the sum of their individual times.

Natural Constraints and Relations Among the Variables

Since each event must have three entries from a squad,

$$\sum_i x_{ij} = 1 \quad j \in I \qquad \sum_i x_{ij} = 4 \quad j \in R \tag{1}$$

$$\sum_i y_{ij} = 1 \quad j \in I \qquad \sum_i y_{ij} = 4 \quad j \in R \tag{2}$$

$$\sum_i z_{ij} = 1 \quad j \in I \qquad \sum_i z_{ij} = 4 \quad j \in R \tag{3}$$

Each swimmer can enter at most four events, with at most two of those being individual events. Therefore, the following constraints apply

$$\sum_{j \in I} x_{ij} + y_{ij} + z_{ij} \leq 2 \quad \text{for all } i \tag{4}$$

$$\sum_{j \in A} x_{ij} + y_{ij} + z_{ij} \leq 4 \quad \text{for all } i \tag{5}$$

Finally, a swimmer can only place once in an event, leading to

$$x_{ij} + y_{ij} + z_{ij} \leq 1 \quad \text{for all } i, j \tag{6}$$

The realized time for a swimmer in an event is a function of his or her estmated time and whether he or she swims in the event. This is expresssed by the relations

$$r(1)_j = \sum_i t_{ij} x_{ij} \quad j \in A \tag{7}$$

$$r(2)_j = \sum_i t_{ij} y_{ij} \quad j \in A \tag{8}$$

$$r(3)_j = \sum_i t_{ij} z_{ij} \quad j \in A \tag{9}$$

In order to force consistency in the order in which swimmers place we use the constraints

$$r(1)_j + \epsilon \leq r(2)_j \quad j \in A \tag{10}$$

$$r(2)_j + \epsilon \leq r(3)_j \quad j \in A \tag{11}$$

(where ϵ is a very small number).

The "realized" times in each event are compared to the opponent times to determine the outcome of the various events. For example, if

$$r(1)_j < r(2)_j < a(1)_j < a(2)_j < a(3)_j < r(3)_j, \tag{12}$$

then the squad would be rewarded points for first, second and sixth place in event j.

The goal of the squad is to maximize the total points won during the meet. To address this objective, we define the indicator variables w_{jlmn} that specify the outcome of event j, where

$$w_{jlmn} = \begin{cases} 1 & \text{if, in event } j\text{, the best swimmer receives place } l\text{, the second best} \\ & \text{ swimmer receives place } m\text{, and the third best swimmer receives place } n \\ 0 & \text{otherwise} \end{cases}$$

In an event with six swimmers, $l \in \{1, 2, 3, 4\}$, $m \in \{2, 3, 4, 5\}$, and $n \in \{3, 4, 5, 6\}$. Note that w_{jlmn} is defined only when $l < m < n$, i.e., the second best swimmer can not place better than the best swimmer, etc.

This outcome indicator can be used to construct constraints on the realized times $r(i)_j$. For the example given in (12) above, in which the squad places first, second and sixth in event j, then $w_{j126} = 1$, and the condition $r(2)_j < a(1)_j$ (the squad's second best swimmer is faster than the opponents best swimmer) must hold. This can be enforced by using the constraint

$$r(2)_j + \epsilon \leq a(1)_j w_{j126} + (1 - w_{j126})M, \tag{13}$$

where M is a very large number. Thus, if $w_{j126} = 1$, then $r(2)_j < a(1)_j$ as desired. However, if the outcome does not occur, so that $w_{j126} = 0$, then $r(2)_j$ is essentially unconstrained (it is bounded above by a very large number.) Similarly, for the example in (12) to hold $a(3)_j < r(3)_j$ must be true for the squad's third best swimmer to place sixth. The constraint

$$r(3)_j \geq a(3)_j w_{j126} + \epsilon \tag{14}$$

ensures this, since $r(3)_j$ is either bounded from below by zero or by $a(3)_j$, depending on the outcome.

For each event j every feasible combination of l, m and n has a pair of constraints similar to those in (13) and (14) for each of the three swimmers in the event.

Finally, the constraint

$$\sum_{1 \leq l < m < n \leq 6} w_{jlmn} = 1 \quad j \in A \tag{15}$$

guarantees that each event will have exactly one outcome.

Objective Function

The reward for each outcome is determined by adding the points for each place for each type of event. The point structure used for this model is that used in Michigan High School competition, as shown in Table 1.

Place	1	2	3	4	5	6
Points (Individual Event)	6	4	3	2	1	0
Points (Relay Event)	8	4	2	0	0	0

Table 1. Points for individual and relay events

Let g_{jlmn} be the reward for receiving the lth, mth and nth places in event j. For example, $g_{j126} = 10$ if j is an individual event and $g_{j126} = 12$ for a relay event. This variable incorporates the fact that a team cannot be awarded points for more than two finishing places in a relay event. For example, $g_{j123} = 12$ for a relay event, since the points for the third place finish are not included.

The total reward for the entire meet is

$$T = \sum_{j,l,m,n} g_{jlmn} w_{jlmn}. \tag{16}$$

The problem then is to maximize T with respect to the variables x_{ij}, y_{ij}, z_{ij}.

Integer Programming Formulation

The integer program formulation of the rostering problem can be now stated as:

$$\max_{x_{ij}, y_{ij}, z_{ij}} \quad \sum_{j,l,m,n} g_{jlmn} w_{jlmn} \tag{17}$$

subject to

$$\sum_i t_{ij} x_{ij} = r(1)_j \quad j \in A$$
$$\sum_i t_{ij} y_{ij} = r(2)_j \quad j \in A$$

$$\sum_i t_{ij} z_{ij} = r(3)_j \quad j \in A$$

$$r(1)_j + \epsilon \leq r(2)_j \quad j \in A$$

$$r(2)_j + \epsilon \leq r(3)_j \quad j \in A$$

$$\sum_{1 \leq l < m < n \leq 6} w_{jlmn} = 1 \quad j \in A$$

$$\sum_i x_{ij} = 1 \quad j \in I$$

$$\sum_i x_{ij} = 4 \quad j \in R$$

$$\sum_i y_{ij} = 1 \quad j \in I$$

$$\sum_i y_{ij} = 4 \quad j \in R$$

$$\sum_i z_{ij} = 1 \quad j \in I$$

$$\sum_i z_{ij} = 4 \quad j \in R$$

$$\sum_{j \in I} x_{ij} + y_{ij} + z_{ij} \leq 2 \quad \text{for all } i$$

$$\sum_{j \in A} x_{ij} + y_{ij} + z_{ij} \leq 4 \quad \text{for all } i$$

$$x_{ij} + y_{ij} + z_{ij} \leq 1 \quad \text{for all } i, j$$

$$\epsilon + r_j(1) \leq a(l)_j w_{jlmn} + M(1 - w_{jlmn})$$

$$\epsilon + a(l-1)_j w_{jlmn} \leq r(1)_j$$

$$\epsilon + r_j(2) \leq a(m-1)_j w_{jlmn} + M(1 - w_{jlmn})$$

$$\epsilon + a(m-2)_j w_{jlmn} \leq r(2)_j$$

$$\epsilon + r_j(3) \leq a(n-2)_j w_{jlmn} + M(1 - w_{jlmn})$$

$$\epsilon + a(n-3)_j w_{jlmn} \leq r(3)_j.$$

$$x_{ij}, y_{ij}, z_{ij} \in \{0, 1\} \quad \text{for all} \quad i, j$$

$$w_{jlmn} \in \{0, 1\} \quad \text{for all } j \in A \text{ and } 1 \leq l < m < n \leq 6$$

3 Results and Analysis

The integer program (17) was solved using the CPLEX solver (v. 7.500). The model was initially tested with 24 swimmers on our squad and eight events (three relay events and five individual events). The times for the squad and opponents were selected from typical swimmers in the Northwest Conference (collegiate). The assignments, places and points for the eight events are shown in Table 2. The squad gained 78 out of 122 possible points.

Event	Assignment			Places	Points
	1	2	3		
1	{3,10,14,21}	{9,15,19,20}	{2,22,23,24}	1,2,4	12
2	{1,2,7,24}	{4,10,17,19}	{3,5,9,21}	1,5,6	8
3	{1,14,19,22}	{8,16,20,21}	{5,6,10,11}	1,2,6	12
4	3	16	10	2,3,4	9
5	4	2	11	1,3,5	10
6	2	3	1	2,3,4	9
7	6	15	5	1,4,5	9
8	6	14	1	2,3,4	9

Table 2. Results for example with 24 swimmers in eight events

In order to analyze the model's sensitivity to values of the times, we calculated how the squad would perform against an opponent that is faster than expected. Keeping the original squad assignments as shown in Table 2, each opponent's time was lowered by a fixed percent and the points that the squad would receive against the "new" opponent were calculated. If the opponent times are 1% lower, the original roster will only score 63 points, narrowly avoiding a loss (62 points is required for a win in this example). If the opponent times were lowered by 2%, the original roster scores 57 points, thereby losing the meet.

[MACIEK - HERE'S WHERE WE COULD USE AN EXAMPLE OF WHAT HAPPENS

IF WE RE-ROSTER THE SQUAD USING THE 2% LOWER OPPONENT TIMES]

Similarly, we analyzed how the squad would perform if its swimmers were less competitive than predicted. The opponent times were held constant while the squad times were increased. An increase of 1% resulted in a score of 70 points, less than the 78 points that the original roster scored. However, even a 1 or 2% difference from estimated times might not be that common. Competitive swimmers are consistent enough so that even the most inexperienced coaches can predict how well their team will perform within a margin of error below 1%.

The example above is not realistic in terms of the size of a squad and the number of events. A more representative data set was obtained from the Ann Arbor Pioneer High School swim team, consisting of times for 62 swimmers in 11 events (three relay and eight individual). These times were tested against a real opponent (one that the team had previously competed with.)

[MACIEK: WE NEED A MORE "COMPETITIVE" EXAMPLE HERE; IF NECESSARY MAKE ONE UP??]

Using aggressive cutting in CPLEX, this problem was solved in 25 minutes (16 minutes of branch and bound time), a fraction of the time that it required this experienced coach to create his lineup by hand.

4 Conclusions

Our model finds an optimal assignment for swimmers to events. It can be useful as a decision aid for younger coaches overwhelmed by the prospect of creating a full lineup, or as a timesaver for older coaches who would like to spend more time coaching rather than mulling over various possible swimmer lineups. Determining the top swimmer in an event is generally trivial; it is the second and third positions that are most difficult to fill, particularly in relay events. While placing first in an event may be a psychological boost for the team, A meet is often won by the lower tier swimmers. As Pete Higgins, national swim coach of the year at Westminster Academy in Atlanta, GA attests, "we can get first place in an event, but we're not going to be winning anything if our other guys get fifth and sixth." The model can take a lot of the guesswork out of deciding which swimmers to place in the crucial lower positions.

Some future work that can be accomplished using this model includes the addition of certain constraints that may be helpful in creating a lineup. For example, if a swimmer is to compete in two consecutive events, his/her performance in the second event may be degraded. This may be accounted for by increasing the predicted time for the second event by some percentage. As some coaches prefer, the option of allowing a swimmer to compete in consecutive events may be completely prohibited. Also, some states allow for four entries per team in an event. These results would require an obvious but straightforward expansion of the model. Finally, our model provides a tool for the future analysis of strategies that may be used as a rule of thumb when creating lineups, such as what to do with a swimmer who is the best on the team in the majority of events.

[MENTION APPLICABILITY TO TRACK EVENTS ALSO???]