Simulating NHL Games to Motivate Student Interest in OR/MS

Armann Ingolfsson
School of Business
University of Alberta
Edmonton, Alberta T6G 2R6, Canada
Armann.Ingolfsson@UAlberta.Ca

Abstract

Monte Carlo simulation models of regular season and playoff games in the National Hockey League can be used to motivate students who are sports fans to study OR/MS models. I describe how such models can be implemented in a spreadsheet, how my colleagues and I have used them in our courses, and how students have reacted to them.

Editor's note: This is a pdf copy of an html document which resides at http://ite.pubs.informs.org/Vo5No1/Ingolfsson/

1. Introduction

This article describes Monte Carlo simulation models of games in the regular season and playoffs of the National Hockey League (NHL). These models can be used as examples in management science courses, to demonstrate to students the wide scope of applicability of the models they are studying. These simulation examples will not save a management science course that is in trouble, but I believe they have the potential to make a good management science course better, by motivating more students to take a personal interest in how OR/MS models function. My colleagues and I at the University of Alberta School of Business (in particular, Terry Daniel and Erhan Erkut) have successfully used these examples in introductory OR/MS courses at the undergraduate and MBA levels.

Maehr and Meyer (1997), in a survey article on motivation and schooling, suggest that it may be useful in educational contexts to replace the word "motivation" with "personal investment" and state that the primary question is "when and how ... individuals invest time, talent, and energy in a particular activity." Obviously, instructors cannot themselves make this investment for students; instructors can only provide opportunities for students to make such investments. I believe that by providing a variety of examples of how OR/MS tools can be used to answer questions that students care about, instructors can significantly expand students' investment and understanding of OR/MS tools, and thereby increase the number of students who find an opportunity that they are personally interested in taking advantage of.

At the University of Alberta, and elsewhere in Canada, many students are passionate ice hockey fans and any examples involving this sport are likely to catch their attention. In March and April, enthusiasts of NHL teams are often preoccupied with the issue of whether "their team" will qualify for the playoffs. This issue has been of particular interest to fans of the Edmonton Oilers, because the answer to the question "will the Oilers make they playoffs?" has not been revealed until close to the end of the 82-game regular season in each of the last seven seasons prior to the 2003-2004 season (see Table 1). In years when the Oilers have made the playoffs, fan interest shifts to the question of "how far will they get?" I present separate Monte Carlo simulation models to help provide answers to these two questions. Optimization models provide an alternative way to answer the former question (see Adler et al., 2002 and the RIOT web site[1]). Simulation and optimization provide different and complementary answers to the question of whether a team will make the playoffs: optimization can sometimes provide

(1) http://riot.ieor.berkeley.edu/~baseball/

 © INFORMS ISSN: 1532-0545

a definite "yes" or "no" answer (sometimes before newspaper sports writers reach such conclusions), while simulation can provide an estimate of the probability that the answer will be "yes." In addition, simulation can provide estimates such as:

- How many points will most likely be needed to qualify for the playoffs?
- If the Oilers make the playoffs, which team will they most likely play?
- Suppose the Oilers get X points in their remaining Y games. How likely are they to make the playoffs?

Table 1: Date when the Edmonton Oilers either clinched a playoff spot or were eliminated during the last seven seasons prior to the 2003-2004 season (Downloaded from http://www.edmontonoilers.com/pdf/2003_oilers_seasonreview.pdf)

Season	Date	Result	Oilers game #
1996-97	April 5th	Clinched with a 5-5 OT tie vs Vancouver	79
1997-98	April 15th	Clinched with a 5-3 win vs Anaheim	81
1998-99	April 12th	Clinched with a 5-4 OT win at San Jose	80
1999-00	April 7th	Clinched with a 5-4 OT win at Vancouver	81
2000-01	April 5th	Clinched when San Jose defeated Phoenix 3-0	81
2001-02	April 12th	Eliminated by a 2-0 loss vs Calgary	81
2002-03	March 26th	Clinched with 4-3 win vs Phoenix	77

A growing number of articles describe how OR/MS instructional purposes can be served with examples from a variety of sports, for example, soccer (Chu, 2003) and golf (Tiger and Saltzer, 2004). The other articles in this special issue provide additional examples. A series of articles in Interfaces (Morrison and Wheat, 1986, Erkut, 1986, Nydick and Weiss, 1989, and Washburn, 1991) present various ways to estimate the optimal time to "pull the goalie" near the end of a hockey game. The models in these papers can provide interesting exercises for courses on stochastic processes. Mullet (1977) discusses the use of Poisson processes to model various statistics generated by the NHL.

The remainder of this paper is organized as follows. Section 2 describes the rules that determine which teams qualify for the NHL playoffs and outlines a Monte Carlo simulation model for games played during the regular season. Section 3 briefly describes a corresponding model for games during the playoffs. The final section discusses how my colleagues and I have used the simulation models in the past, how the models have stimulated student interest, and some of the generalizable lessons that students have learned from immersing themselves in these models.

2. Regular Season Simulation

The regular season of the National Hockey League (NHL) runs from October to April. Each of thirty teams play 82 games. The teams are divided into an Eastern and a Western conference, and each conference is divided into three divisions with 5 teams each. Each game consists of "regulation time" and, possibly, "overtime." If one of the teams wins in regulation time, then it earns 2 points and the other team earns no points. If the game is tied at the end of regulation time, then both teams get 1 point, and one overtime period is played. If one of the teams scores in overtime, then it earns an extra point and the game ends. The total number of points earned by both teams in one game is therefore either 2 or 3.

Newspapers and various internet sites[(2)(3)], provide abundant statistics on play in the NHL. Tables 2 and 3 show information that is particularly relevant to constructing a simulation model of the remainder of the regular season. Table 2 shows the standings in the Western Conference while Table 3 shows the schedule of remaining games, as of March 29, 2004. The focus will be on the Western Conference and the Edmonton Oilers, to reduce the size of the model and to make the illustrations more specific, but the approach could be used for either conference or both conferences together and any team. By focusing on the Western Conference,

(2) http://www.nhl.com
(3) http://sports.yahoo.com/nhl

one can eliminate all games between Eastern Conference teams, and this is done in Table 3 and in the simulation model.

Table 2: Western Conference standings (downloaded from NHL website http://nhl.com/onthefly/standings/conf_stand.html on March 29, 2004). Legend: RK = rank, GP = games played, W = wins, L = losses, T = ties, OTL = overtime losses, PTS = points, GF = goals for, GA = goals against, HOME = home record (W-L-T-OTL), AWAY = away record, L10 = record for last 10 games.

Western Conference

RK		GP	W	L	T	OTL	PTS	GF	GA	HOME	AWAY	L10
1	y - DETROIT*	78	45	20	11	2	103	243	178	29-6-4-0	16-14-7-2	6-3-1-0
2	x - SAN JOSE*	79	41	20	12	6	100	211	176	23-7-7-2	18-13-5-4	8-1-1-0
3	x - COLORADO*	78	38	22	13	5	94	225	188	18-14-6-1	20-8-7-4	4-5-1-0
4	x - DALLAS	79	40	24	13	2	95	186	166	25-6-8-0	15-18-5-2	5-1-2-2
5	x - VANCOUVER	78	39	24	10	5	93	218	189	19-13-7-0	20-11-3-5	3-4-2-1
6	CALGARY	79	40	29	7	3	90	195	172	20-14-5-1	20-15-2-2	6-3-1-0
7	ST LOUIS	78	37	28	11	2	87	184	191	22-10-7-0	15-18-4-2	6-2-2-0
8	EDMONTON	79	35	27	12	5	87	216	201	22-12-4-3	13-15-8-2	7-0-0-3
9	NASHVILLE	78	35	28	11	4	85	205	209	21-9-7-2	14-19-4-2	2-3-3-2
10	LOS ANGELES	78	28	26	16	8	80	199	205	15-14-9-1	13-12-7-7	2-7-0-1
11	MINNESOTA	78	27	28	20	3	77	173	172	16-13-7-2	11-15-13-1	6-2-1-1
12	ANAHEIM	80	28	34	10	8	74	181	210	18-10-7-4	10-24-3-4	3-6-1-0
13	PHOENIX	79	21	34	18	6	66	185	237	10-19-7-4	11-15-11-2	1-6-2-1
14	COLUMBUS	78	24	42	8	4	60	169	224	17-16-4-2	7-26-4-2	4-6-0-0
15	CHICAGO	78	20	40	11	7	58	182	244	13-16-6-5	7-24-5-2	2-5-2-1

Table 3: Remaining NHL regular season games (downloaded from NHL website http://nhl.com/onthefly/schedules/index.html on March 29, 2004).

DATE	VISITOR		HOME		TIME
29-Mar-04	Blue Jackets	0	Sabres	0	7:00 PM ET
29-Mar-04	Wild	0	Red Wings	0	7:00 PM ET
29-Mar-04	Kings	0	Avalanche	0	9:00 PM ET
29-Mar-04	Coyotes	0	Canucks	0	10:00 PM ET
30-Mar-04	Oilers	0	Blues	0	8:00 PM ET
30-Mar-04	Blackhawks	0	Predators	0	8:00 PM ET
31-Mar-04	Red Wings	0	Blue Jackets	0	7:00 PM ET
31-Mar-04	Avalanche	0	Wild	0	8:00 PM ET
31-Mar-04	Oilers	0	Stars	0	8:00 PM ET
31-Mar-04	Coyotes	0	Flames	0	9:00 PM ET
31-Mar-04	Canucks	0	Mighty Ducks	0	10:30 PM ET
31-Mar-04	Sharks	0	Kings	0	10:30 PM ET
01-Apr-04	Red Wings	0	Blues	0	8:00 PM ET
01-Apr-04	Predators	0	Blackhawks	0	8:30 PM ET
02-Apr-04	Avalanche	0	Blue Jackets	0	7:00 PM ET
02-Apr-04	Stars	0	Wild	0	8:00 PM ET
02-Apr-04	Flames	0	Kings	0	10:30 PM ET
02-Apr-04	Canucks	0	Sharks	0	10:30 PM ET
03-Apr-04	Blues	0	Predators	0	3:00 PM ET
03-Apr-04	Blackhawks	0	Coyotes	0	4:00 PM ET
03-Apr-04	Blue Jackets	0	Red Wings	0	7:30 PM ET
03-Apr-04	Oilers	0	Canucks	0	10:00 PM ET
04-Apr-04	Blues	0	Wild	0	3:00 PM ET
04-Apr-04	Blackhawks	0	Stars	0	3:00 PM ET
04-Apr-04	Predators	0	Avalanche	0	4:00 PM ET
04-Apr-04	Flames	0	Mighty Ducks	0	4:00 PM ET
04-Apr-04	Kings	0	Sharks	0	4:00 PM ET

The regular_season.xls[4] workbook contains a complete simulation model for games remaining in the Western

(4) http://ite.pubs.informs.org/Vol5No1/Ingolfsson/regular_season.xls

Conference regular season, including and after March 29, 2004. The following color coding is used in the model:

- **Bold green font** indicates input data.
- Cells with red background indicate random numbers.
- Cells with yellow background indicate output measures.

The model is implemented using only native Excel features, i.e., the built-in RAND() function is used to generate pseudo − random numbers, and data tables are used to replicate the simulation and to collect statistics. The color coding should make it straightforward to convert the model to use a simulation add-in such as Crystal Ball (2004) or @Risk (2004) to generate random numbers and collect statistics. One should keep in mind that Excel's RAND() function apparently uses a random number generator that leaves something to be desired (L'Ecuyer, 2001). The generator was rewritten for Excel 2003(5), but unfortunately the revision introduced a serious bug that causes the generator to occasionally generate negative numbers. A "hotfix" package from Microsoft(6) corrects this bug

I begin by describing the simulation of an individual game. Table 2 lists, for each team, the number of games played (GP), won (W), lost in regulation time (L), tied (T), and lost in overtime (OTL). This information is sufficient to calculate the number of points (PTS), using PTS = 2*W + T + OTL. From this information, one can calculate a team's win, loss, and tie percentages, using WP = W/GP, LP = (L+OTL)/GP, and TP = T/GP. The model uses the following approach to simulate each remaining regular season game:

1. Estimate the probability P that the visitor team will win the game as the average of the visitor team's win percentage and the home team's loss percentage. Estimate the probability Q that the game will result in a tie as the average of the two team's tie percentages.
2. Generate a uniformly distributed random number U1. If U1 ≤ P, then the visitor team wins in regulation time. If U1 > P + Q, then the visitor team loses in regulation time. Otherwise, the game is tied at the end of regulation time.
3. If the game was tied at the end of regulation time, then an overtime period will be played. Generate a second uniformly distributed random number U2 to determine the outcome of the overtime period. If U2 ≤ P, then the visitor team wins in overtime. If U2 > P + Q, then the visitor team loses in overtime. Otherwise, the game remains tied.

It is possible to combine steps 2 and 3 and use one random number instead of two, but I have separated the steps for ease of explanation. Figure 1 illustrates how one can implement this approach in Excel. The Figure shows part of a worksheet that simulates the outcome of every remaining game in the regular season. In the first of these games, Columbus visits an Eastern Conference opponent (the model focuses on the Western Conference and therefore suppresses information about the particular team in the Eastern Conference that Columbus plays). We see that Columbus has a win percentage of 0.308 and Eastern Conference teams have a win percentage of 0.423 and a tie percentage of 0.156; hence a loss percentage of 0.421. The probability P is therefore calculated to be the average of 0.308 and 0.421, or 0.365. The tie probability Q is calculated to be the average of the tie percentage for Columbus and the tie percentage for Eastern Conference teams, resulting in Q = 0.129. In Figure 1, the random number used to simulate regulation time was 0.485, which is between P and P + Q, so the simulated outcome of regulation time is a tie. Consequently, overtime is simulated as well. The overtime random number is 0.225, which is less than P + Q, so Columbus wins the simulated game in overtime.

			Visitor		Home		Game		Regular time		Overtime	
Date	Visitor	Home	Pr{Win}	Pr{Tie}	Pr{Win}	Pr{Tie}	Pr{Win}	Pr{Tie}	RAND()	Outcome	RAND()	Outcome
29-Mar-04	Columbus	Eastern	0.308	0.103	0.423	0.156	0.365	0.129	0.485	T	0.225	W

Figure 1: Simulation of one regular season game.

(5) http://support.microsoft.com/default.aspx?scid=kb;en-us;828795&Product=xlw
(6) http://support.microsoft.com/default.aspx?kbid=833855

The approach used to simulate a single game is simple, but it is also simplistic in many ways: it assumes the win, loss, and tie probabilities are the same in regulation time and overtime, it ignores the possibility of a "home ice advantage," it ignores variations between different Eastern Conference teams, and it ignores the possibility of temporal variation in the outcome probabilities due to changes in team personnel or "momentum." The simplicity makes it easier to explain the approach to students, and I believe the simplistic modeling also has an advantage, because it may motivate students to criticize the assumptions of the model and to attempt, on their own, to make it more realistic. This is beneficial because it is often difficult to get students to critically assess the assumptions in models that are typically covered in OR/MS courses, because they lack experience with the reality that is being modeled (a production facility, for example). If the example were used in a statistics class, then it could provide an opportunity to discuss ways to judge whether home ice advantage or perceived streaks are real rather than random (see Smeeton, 2003, for example). If a student wanted to calculate separate outcome probabilities for regulation time and overtime, she would need to collect more data than is available in Table 2, because the standings table gives only partial information about how many games required overtime. Specifically, the standings table does not indicate how many of a team's wins occurred in overtime. To obtain this information, a student would need to obtain and tabulate data on the outcomes of individual games.

After simulating every remaining game in the regular season, various Excel functions are used to aggregate the results and generate simulated standings at the end of the regular season, as shown in Figure 2. The top eight teams from each conference qualify for the Stanley Cup Playoffs, but the process is complicated by several tie-breaking rules. The official rules are as follows[7]

Playoff Tie-Breaking Formula
At the conclusion of the regular season, the standing of the teams in each Conference shall be determined in accordance with the following priorities in the order listed:
a) First place in each of the three Divisions seeded 1, 2 and 3.
b) The higher number of points earned by the Club.
c) The greater number of games won by the Club.
d) The higher number of points earned in games against each other among two or more Clubs having equal standing under priorities (b) and (c).

NOTE: For the purpose of determining standing under priority (d) for two or more Clubs that have not played an even number of games with one or more of the other tied Clubs, the first game played in the city that has the extra game (the "odd" game) shall not be included. When more than two Clubs are tied, the percentage of available points earned in games among each other (and not including any "odd games") shall be used to determine the standing.
e) The greater positive differential between goals scored for and against by Clubs having equal standing under priority (d).

The simulation model implements items a), b), and c) but ignores the remaining tie-breaking rules. The approach to implement rules a), b), and c) is motivated by goal programming. Specifically, the model calculates "pseudo points" for each team, as follows:

For division winners: Pseudo points = (M1 + M2) * TP + W

For all other teams: Pseudo points = M2 * TP + W

The numbers M1 and M2 are chosen to be sufficiently large so that the three division winners will be ranked 1, 2, and 3 in their Conference, followed by the five other teams in the conference with the highest point totals (TP), and the number of wins (W) is used to break ties. I used M1 = 100 and M2 = 10. The top eight teams, ranked by pseudo points, are then assumed to qualify for the playoffs.

The procedure to calculate pseudo points could be improved by accounting for the possibility that two or more teams are tied for the highest point total in a division. For our specific purpose of determining whether the Edmonton Oilers will make the playoffs, this is unlikely to make much difference, because the Oilers have, unfortunately, not been in any danger of winning their division for the past several years!

Figure 2 shows how pseudo points are used to rank the teams and determine the eight teams that qualify for the playoffs. In addition to the playoff ranks, various other statistics can be calculated for each replication of the simulation, for example, the number of points needed to qualify for the playoffs, number of

[7] NHL website http://nhl.com/nhlhq/faq/go_figure.html

points earned by Edmonton, and Edmonton's opponent in the first round of the playoffs, if applicable.

		Total								
City	Team	GP	W	L	T	OTL	PTS	Division rank	Pseudo points	Playoff rank
Chicago	Blackhawks	82	21	43	11	7	60	5	621	15
Columbus	Blue Jackets	82	25	44	8	5	63	4	655	14
Detroit	Red Wings	82	46	22	12	2	106	1	11706	**1**
Nashville	Predators	82	36	30	11	5	88	3	916	9
St Louis	Blues	82	40	28	11	3	94	2	980	**7**
Calgary	Flames	82	42	30	7	3	94	3	982	**6**
Colorado	Avalanche	82	41	22	13	6	101	1	11151	**3**
Edmonton	*Oilers*	*82*	*36*	*29*	*12*	*5*	*89*	*4*	926	**8**
Minnesota	Wild	82	27	30	22	3	79	5	817	11
Vancouver	Canucks	82	42	25	10	5	99	2	1032	**5**
Anaheim	Mighty Ducks	82	28	36	10	8	74	4	768	12
Dallas	Stars	82	42	24	14	2	100	2	1042	**4**
Los Angeles	Kings	82	30	27	16	9	85	3	880	10
Phoenix	Coyotes	82	23	35	18	6	70	5	723	13
San Jose	Sharks	82	43	21	12	6	104	1	11483	**2**
Eastern									# ranked <= 8	8
									PTS needed	89
									EDM PTS	89
									EDM qualifies?	1
									EDM opponent	Detroit

Figure 2: Simulated standings at end of regular season.

One can use an Excel data table to collect results from multiple replications (see Winston and Albright, 2000 for instructions on how to use data tables to replicate simulation results). Figure 3 shows part of a table of 500 replications. According to the simulation, Edmonton had a 0.61 chance of qualifying for the playoffs, given the standings on March 29, 2004. If they were to qualify, their opponent would most likely be either Detroit (0.61 probability) or San Jose (0.30 probability). The expected number of points needed to qualify for the playoffs was 90.2.

Replication	Chicago	Columbus	Detroit	Nashville	St Louis	Calgary	Colorado	Edmonton	# ranked <= 8	points needed	EDM PTS	EDM qualifies	EDM opponent
	15	14	1	9	7	5	3	8	8	91	91	1	Detroit
1	15	14	1	9	8	6	3	7	8	91	93	1	San Jose
498	15	14	1	9	8	5	3	7	8	90	91	1	San Jose
499	15	14	1	9	6	7	4	8	8	91	91	1	Detroit
500	15	14	1	8	6	7	5	9	8	87	87	0	FALSE
Average	14.8	14.2	1.0	8.2	7.4	6.1	3.6	8.1	8.014	90.19	90.13	0.61	
Min	13	13	1	6	5	3	2	6	8	87	87	0	
Max	15	15	2	10	9	8	6	10	9	93	93	1	

EDM opponent	Prob.	Prob2.
Chicago	-	-
Columbus	-	-
Detroit	0.37	0.61
Nashville	-	-
St Louis	-	-
Calgary	-	-
Colorado	0.03	0.05
Edmonton	-	-
Minnesota	-	-
Vancouver	0.02	0.03
Anaheim	-	-
Dallas	-	-
Los Angeles	-	-
Phoenix	-	-
San Jose	0.18	0.30

Conditional on EDM qualifying.

Figure 3: Table of 500 replications of the simulation.

The "# ranked <= 8" statistic was included to evaluate how often the tie-breaking rules that were left out of the model were needed. This statistic equaled either 8 or 9 in every replication, and its average was 8.014. One can find the number X of replications in which the additional tie-breaking rules were needed by solving the equation (8*(500-X) + 9*X)/500 = 8.014. The result is X = 7, so the estimated probability that the additional tie-breaking rules are needed is 7/500, or about 1.5%.

Once the data table has been generated, the values in it can be "frozen" (by copying the table and pasting it as "values") to facilitate further analysis. In addition to calculating various averages and other summary statistics for each column in the table, Excel's auto filter tool can be used to calculate various conditional probabilities. For a simple example, one could filter the "# ranked <=8" column on the value "9" and count how many rows are displayed, as an alternative to solving the equation given above. For a more interesting example, the column for "EDM PTS" in Figure 3 could be filtered to show only replications in which Edmonton earned 90 points or more, and then one can estimate the probability that Edmonton qualifies for the playoffs, given that the team earns 90 points or more (the result is 252/292 = 0.86.)

3. Playoff Simulation

The eight teams that qualify for the playoffs from each conference are seeded according to the rules described in the previous section. In the first round of the playoffs, the first-seeded team plays the eighth-seeded team, the second-seeded team plays the seventh-seeded team, and so on. Each of these pairs of teams then plays a series of up to seven games, with the series terminated as soon as one of the teams has won four games. The outcome of a game is determined the same way as in the regular season, except that overtime continues until one of the teams scores, so no game results in a tie. The higher-seeded team in each pair enjoys home-ice advantage, which means that four of the (up to) seven games are scheduled in that team's arena, with the usual sequence being Home-Home-Away-Away-Home-Away-Home. The four winners in the first round proceed to the second round, where teams are re-seeded, taking into account results from the first round. The top-seeded team then plays the fourth-seeded team and the second and third teams play the other series in the second round. The winners of the second round then meet in a series that determines the Conference Champion. Finally, the two Conference Champions play the final playoff series to determine the winner of the Stanley Cup.

The basic ingredient in a simulation model of the playoffs is the same as for the regular season, namely, the simulation of a single game. The only difference is that playoff games cannot end in a tie. A simple way to adapt the procedure described in the previous section, for simulating a regular season game, is to renormalize each team's win and loss percentages so that they add up to 100%, calculate the probability P as the average of the visiting team's win percentage and the home team's loss percentage, and then use a single random number to determine which team wins the game.

One added complication in simulating the playoffs, compared to the regular season, is that the set of games to be played is not pre-determined, for two reasons: (1) the number of games in each series can vary from 4 to 7, and (2) the teams that play in rounds 2, 3, and 4 are determined by the outcomes of previous rounds. The first complication can be avoided by simulating seven games in every series, even if one of the teams earns 4 wins in less than seven games. The second complication can be handled using Excel lookup functions.

The workbook playoffs.xls[8] shows a simulation model for the 2004 NHL playoffs, before any playoff games were played. Only the Western Conference is included, for simplicity. The Edmonton Oilers did not qualify for the 2004 playoffs, unfortunately. The workbook shows two examples of statistics that could be collected from the simulation: (1) the winner of the Vancouver - Calgary first round series (Vancouver is estimated to win with probability 0.54) and (2) the Western Conference Champion (Detroit is most likely to win, with probability 0.21).

Observe that the ranking of teams by their probability of winning the Western Conference Championship is the same as the seeding of the teams at the end of the regular season. It is almost inevitable that the simulation results will turn out this way, as long as the probability P is determined, for each game, as described above. However, students could attempt to estimate this probability differently, by taking into account one or more of the following factors:

- Team records during the last part of the regular season (for example, by using exponential smoothing or moving averages).
- Recent changes in team personnel, because of trades or injuries.

(8) http://ite.pubs.informs.org/Vol5No1/Ingolfsson/playoffs.xls

- Home ice advantage.
- Results of regular season games between the two teams in question (this is questionable, because of small sample sizes).

4. Using the Examples to Motivate

My colleagues and I have used the simulation models described above for about eight years, in a core undergraduate OR/MS course, a first-year MBA production and operations management elective, and an undergraduate elective on service operations. In our experience, such examples can be introduced using a minimal amount of lecture time. We usually introduce the examples as "teasers" at the beginning or end of a lecture. It is not difficult to find statements made in the popular press that straightforward modeling shows to be false, and this can be an effective way of capturing student interest. We typically show a recent newspaper article during lecture, show model results that contradict the article, and, ideally, provide insight into why particular statements in the article are false or misleading. The model itself, with instructions, can be made available through the Internet for students to "look under the hood," and interested students can discuss the model and the results of its analysis in course conferences on the Internet.

Our purpose in presenting the simulation examples is not primarily to introduce new modeling techniques, but to stimulate the interest of a large fraction of the students in modeling in general. It is important to keep in mind that some students are not interested in hockey but fortunately we have found it easy to avoid alienating those students. We typically plan on spending less than five minutes on the examples at the beginning or end of a lecture, possibly extending the discussion to ten minutes if the class seems particularly interested. We may show updated results at a later date, particularly if the predictions of the model have become more favorable to the home team. In courses that focus specifically on modeling in sports, such as the course described by Willoughby (2004), one could cover the models in far greater detail and base assignments or projects on them, but so far we have not done this.

As an example, on March 30, 2004, a colleague teaching the core undergraduate OR/MS course presented some of the results shown in Section 2. On March 29, a sports columnist for the Edmonton Journal (MacKinnon, 2004) wrote, commenting on a win by the Oilers the previous night:

> ... it was two points and the Oilers will take them. They now have 87 points and will probably need six more to have a chance to participate in the Stanley Cup playoffs tournament.
>
> There are no guarantees that will be enough.

According to the simulation, if the Oilers were to earn just four more points, their chances of qualifying for the playoffs would be about 89%. Furthermore, in the 59 (out of 500) replications where the Oilers won all three of their remaining games (thus earning six more points), the Oilers qualified every time. A closer look at the remaining games for the Oilers and the two other teams in contention for the last two playoff spots (the Blues and the Predators) revealed that the Oilers play the Blues once, and the Blues and the Predators play each other once, which implied that if the Oilers were to win all three games, then at least one of the other two teams would have fewer points than the Oilers. This analysis can be used to demonstrate how the results of a model become the starting point for developing insight that goes beyond the particular numerical results obtained. In particular, looking at the number of remaining games each team will play is not sufficient to determine the likelihood that a team will qualify for the playoffs; one must also consider how many of those games are against other teams vying for the final few playoff spots.

We have found it relatively easy to find statements in the sports sections of local newspapers that can be used as a lead-in to the simulation models. Here are two examples:

1. A columnist for the Edmonton Sun (Jones, 1996) stated, in March 1996, that "the magic number for the Jets to eliminate the Oilers from the race is four. Any combination of Jets wins and Oilers losses totalling four and the Oilers would be eliminated." A simulation of the remaining regular season games revealed scenarios where the Winnipeg Jets (which no longer exist) won four games, Edmonton lost no games, and Edmonton still qualified for the playoffs.

  ISSN: 1532-0545

2. The coach of the Edmonton Oilers was quoted (MacTavish, 2004) on March 24, 2004, as predicting that 90 points would be sufficient to qualify for the 2004 playoffs. The simulation results in Section 2 suggest that this was a somewhat optimistic but reasonable prediction. (As it turned out, 91 points were needed to qualify for the playoffs in the Western Conference.)

I believe it is most effective to use the examples in March or April, because that is the time when one is most likely to find newspaper articles to help capture student interest and because this is when the models are most likely to produce results that are not obvious from a glance at the current standings table. Earlier in the NHL season, students and journalists are less interested in the playoff race and the estimated probabilities of making the playoffs are likely to vary less between teams than in March and April. Instructors who would like to use similar examples during the fall term might be able to adapt the models in this article to other sports where the season typically ends in the fall, such as soccer or baseball.

Student reactions are typically positive when the examples are presented in class, followed by even more positive responses later on, from students that are "hooked" and explore the models in detail on their own. As an example, four days after the results in Section 2 were presented in class, a female student contacted the instructor to let him know that she had updated the simulation model to incorporate the results of games that had been played since. Readers who have explored the model in detail will realize that this is a nontrivial task, especially for a beginning student. This student had not displayed any particular aptitude for simulation modeling earlier in the term, but she was obviously a hockey fan. For students who have an aptitude for analytical modeling and who are also sports fans, these examples provide an opportunity and an incentive to become more critical model users and developers. They recognize the approximations that are necessary to make a model efficient yet useful and begin to develop an appreciation for which modeling shortcuts compromise the validity of the model (for the purpose that it is intended to serve) and which ones are "not quite realistic yet save a lot of modeling time and effort." We use a variety of "teasers" at the beginning of lectures in addition to these hockey examples, some modeling related and some not. Most students appreciate these attempts to capture their interest. We have never received negative feedback about the hockey examples.

Several past students have been sufficiently interested in the simulation models to attempt to extend the models for their own purposes. A common motivation for this is to do better in "hockey pools" which are common in Canada. A participant in a hockey pool typically picks a number of players from playoff-bound teams to form her own fictional team. The goals and assists scored in the real playoffs determine the performance of each of the fictional teams, and thus, the winner of the pool. A simulation model to help pick such a fictional team would need to be considerably more detailed than the models described in this article, of course.

One past student (Slemko, 2000) developed a more realistic version of the regular season simulation in order to complete the requirements for a Diploma awarded by the Canadian Operational Research Society[9]. Among other things, the student automated the collection of information from a web site that was needed to run the simulation starting from a particular day, programmed the entire simulation in VBA, and implemented all of the tie-breaking rules listed in Section 2.

In conclusion, here is a list of valuable lessons that students who delve into the models to this extent have the opportunity to learn. These lessons about model implementation generalize to various other contexts.

- **The data is rarely in the ideal format for analysis.** For example, neither the standings table nor the schedule of games table (which shows the results of past games) allows one to determine which games a particular team won in overtime. As another example, some tables may list teams by city (see Table 2) while others may list teams by their names (see Table 3).
- **The model should be designed for easy updating of input data.** This becomes important for those who want to rerun the model every morning during the last few days of the regular season. Excel allows one to collect data from the Internet either manually, by cutting and pasting, or automatically,

[9] http://www.bus.ualberta.ca/CORS/diploma/english/

with a "web query." While this is a useful feature, web pages with the appropriate statistics may move, change, or disappear from one year to the next.

- **The importance of modular design.** The number of teams, the playoff format, and the tie-breaking rules have all changed several times during the NHL's 86-year history. A well-designed model would need only minor modifications to accommodate such changes.

5. Acknowledgements

Comments from Terry Daniel, Erhan Erkut, an associate editor, and two reviewers have led to substantial improvements to this paper.

References

@Risk, (2004), Palisade Corporation, http://www.palisade.com

Adler, I. A.L. Erera, D.S. Hochbaum, and E.V. Olinick, (2002), "Baseball, Optimization, and the World Wide Web," *Interfaces,* Vol. 32, No. 2, pp. 12-22.

Chu, S. (2003), "Motivating the Poisson Process Using Goals in Soccer," *INFORMS Transactions on Education,* Vol. 3, No. 2, http://ite.pubs.informs.org/Vol3No2/Chu/

Crystal Ball, (2004), Decisioneering, Inc., http://www.decisioneering.com

Erkut, E. (1987), "More on Morrison and What Pulling the Goalie Revisited," *Interfaces,* Vol. 17, No. 5, pp. 121-123.

Jones, T. (1996), "As Big as it Gets," *Edmonton Sun,* Tuesday, April 2, 1996, pp. 55

L'Ecuyer, P. (2001), "Software for Uniform Random Number Generation: Distinguishing the Good and the Bad," Proceedings of the 2001 Winter Simulation Conference, IEEE Press, Dec. 2001, pp. 95-105.

MacKinnon, J. (2004), "It Wasn't Pretty -- In Fact, it Was Coyote Ugly -- But They'll Take It," *Edmonton Journal,* Monday, March 29, pp. D1.

MacTavish, C. (2004), quoted by Matheson, J., "Oil can't Take Foot Off the Gas: Team Has to Keep Winning to Assure Playoff Spot," *Edmonton Journal,* March 24, 2004. pp. D1.

Maehr, M.L. and H.A. Meyer, (1997), "Understanding Motivation and Schooling: Where We've Been, Where We Are, and Where We Need to Go," *Educational Psychology Review,* Vol. 9, No. 4, pp. 371-409.

Morrison, D.G. and R.D. Wheat, (1986), "Misapplications Reviews -- Pulling the Goalie Revisited," *Interfaces,* Vol. 16, No. 6, pp. 28-34.

Mullet, G.M., (1977), "Simeon Poisson and National Hockey League," *American Statistician,* Vol. 31, No. 1, pp. 8-12.

Nydick, R.L. and H.J. Weiss, (1989), "More on Erkut's "More on Morrison and Wheat's 'Pulling the Goalie Revisited'"," *Interfaces,* Vol. 19, No. 5, pp. 45-48.

Slemko, D., (2004), "NHL Simulation," unpublished student project.

Smeeton, N., (2003), "Do Football Teams have Clusters of Wins, Draws and Defeats?" *Teaching Statistics,* Vol. 25, No. 3. pp. 90-92.

Tiger A.A. and D. Salzer, (2004), "Daily Play at A Golf Course: Using Spreadsheet Simulation to Identify System Constraints," *INFORMS Transactions on Education,* Vol. 4, No 2, http://ite.pubs.informs.org/Vol4No2/TigerSalzer/

Washburn, A. (1991), "Still More on Pulling the Goalie," *Interfaces,* Vol. 21, No. 2, pp. 59-64.

Willoughby, K.A. (2004), "Drafts, Dynasties and Dance Cards," *OR/MS Today,* February, pp. 12-13.

Winston, W. and S.C. Albright, (2000), *Practical Management Science,* second edition, Duxbury, Pacific Grove, CA