FORECASTING COURSES WITH UNCERTAIN STUDENT DEMAND: PROOF OF SUB-PROBLEM INDEPENDENCE FOR COURSE LOADING OPTIMIZATION

Cheryl Eisler
Defence Research and Development Canada
101 Colonel By Drive, Ottawa, Canada
cheryl.eisler@drdc-rddc.gc.ca

Dr. Peter Dobias
Defence Research and Development Canada
101 Colonel By Drive, Ottawa, Canada
peter.dobias@drdc-rddc.gc.ca

Dr. Min Jing Liu
Defence Research and Development Canada – Post-Doctoral Fellow
Victoria, Canada
jingjing984@gmail.com

ABSTRACT

Planning and scheduling courses for military training across multiple occupations is a complex task, especially in the face of uncertain levels of demand and rapidly shifting requirements over time. Offering too many courses is resource intensive and may be unjustified; if too few courses are offered either an insufficient number of personnel will be trained or additional costs may be incurred. Under the assumption that the historical attendance in courses is a valid predictor of future demand and looking at the force as a whole allows for statistical distributions for demand to be derived. The objective can then be reformulated to minimize the number of courses offered. In order to allow for more efficient calculations, the sub-problem for each course has been proven independent and solved for in the objective function. This eliminated the use of specialized optimization software for the original problem, and reduced calculation run-time.

Keywords: Course Loading, Scheduling Optimization, Forecasting, Sub-Problem Independence.

1 INTRODUCTION

The problem of class scheduling optimization has been studied in detail throughout the literature (Nakasuwan et al, 1999; Wasfy et al, 2007; Winch, 2013 to name but a few), with variations that factor in resource assignment (Badri, 1996; Gunawan, 2011), timetabling (Schaerf, 1999; Hossain et al, 2007), and various degrees of preferential heuristics (Schniederjans, 1987; Dahiya, 2015). However, the bulk of the research problems are predicated on the knowledge of the student demand. Such an assumption about the incoming training requirement may be valid in settings where deadlines are set well in advance for course session registration and the population levels are sufficiently high for many common courses, thus predictable statistics over time can be generated regarding anticipated attendance.

In the military training setting – especially for small militaries – the problem is quite different. Because of the specialized nature of military occupations, courses may be offered on a regular basis with mandated content (similar to undergraduate level studies in university); however, there may be few students each year in each session (more similar to graduate level classes). Predicting student demand can be difficult,

SpringSim-ANSS 2017, April 23-26, Virginia Beach, VA, USA

as course attendance can fluctuate significantly. While it is possible to track all personnel within a given military service and attempt to predict their upcoming training requirements, there are many variables that increase the level of uncertainty in the prediction as one extrapolates further forward in time. Health, personal and family commitments, and career changes can all abruptly impact the decisions that factor into a single individual's requirements. Unlike the university setting, health issues may force withdrawal of students in courses with a physical component, and leave/postings/deployments on short notice are common; these are just a few reasons for sudden withdrawals within the military training system that do not exist in the civilian world. Further considerations include rapid policy changes that may mandate a change in course content, and the unique availability restrictions on multiple types of resources that may be required for military training (such as warships, aircraft, ground vehicles or other major platforms, qualified trainers, classrooms, virtual or mock training environments, and the supporting logistics chain to supply materiel such as fuel and ammunition necessary to run the course). Therefore, the military training problem is highly constrained, but the methods developed herein could be applied in any context (such as new course offerings for universities). However, the level of constraints precludes common methodologies developed for civilian applications from being directly applicable to the small military training system problem.

If the military force structure is examined as a whole, attendance in past sessions (vice registration, which would neglect withdrawals) can be used to derive probabilistic distributions for the future demand. This assumption eliminates the need to factor in all of the individual elements that affect attendance, and provides the training planners with a simple function that can be tailored to reflect other stressors (such as a major upcoming deployment requiring a significant portion of the force, a known shift in demographics or population size, or the introduction of a new asset type that requires common training). Once the student demand is forecasted, the number of sessions for each course to offer can be calculated such that the expectation for cancelled sessions (due to too few students) and surplus sessions (due to not enough offered originally) are reduced. Gurvich et al (2010), Zambelli et al (2009), and Eliashberg et al (2009) propose similar variations on the problem, except only one type of resource is required to handle any given demand. In this case, a variety of resources (trainers, classrooms, simulators, distance learning units, and other military assets) must be evaluated and assigned to each session in order to accommodate all of the training requirements and minimize asset allocation.

This study proposed a means of modelling the supply-demand relationship as it strived to answer two questions: 1) Considering uncertain student demand, how many sessions of each specific course should be offered in order to minimize late cancellations or additions, while also maintaining a low risk of not meeting the demand? 2) How should the session schedule be organized in order to maximize the effectiveness of the resource usage? The objective for answering these two questions was to ensure that resources would be allocated in an efficient manner, meeting but not significantly exceeding demands (i.e., avoiding the need to add or cancel sessions). By extension, this would simplify resource management including enabling more efficient longer-term acquisition and maintenance planning.

The optimization considered the scheduling interdependencies between courses, students, and teaching resources within the system in order to improve the quality of the estimate of resource requirements. The model was developed in two stages. At first, a simplified model using Palisade Corporation's @RISK add-on for Microsoft Excel® was developed using Monte Carlo simulation optimization. The model was then revised to enumerate through a bounded solution for sub-problems of the demand estimates. This was done in order to remove the need for specialized add-on software and reduce computational complexity and run-time. This paper describes the improved methodology, and the results of the initial test runs of the scheduling optimization model.

The paper is organized as follows. In Section 2 the modelling approach is summarized, and a brief description of the original @RISK model is given for comparison. In Section 3, the new model, including optimization methodology, is discussed. Finally, an example of a trial run is provided in Section 4, and recommendations for follow-on work are identified.

2 PRIOR MODEL DEFINITION

The original optimization model (Eisler et al, 2015) was built in using Palisade Corporation's @RISK add-on for Microsoft Excel®. This model took as an input a list of offered courses with the required resources and constraints, and optimized the session schedule with respect to a predefined scoring function. The optimization problem was separated into two parts. 1) Estimate the required session demand (the number of students vary from year to year, from session to session); and 2) Develop a session schedule and resource assignment plan based on the identified interdependencies and constraints. The first part utilized the historical data for the student demand and Monte Carlo Simulation Optimization (MCSO) was used to obtain the total quantity of sessions to offer to minimize expected session cancellations and additions.

The second was based on a modification to the classical Job Shop Scheduling Problem (JSSP) (Graham, 1966), defined as the task of scheduling N tasks of varying durations to M identical machines in a manner that will complete those tasks in the minimum possible time. The military training problem differs somewhat from the original JSSP, since courses require multiple resources that are generally of very different natures, with different availability schedules. Furthermore, because the availability of military resources is limited, the model had to be able to accommodate the possibility of insufficient resources.

2.1 Session Quantity Optimization

Historical course session attendance was used to forecast student demand through the fit of a probability distribution on a course-by-course basis. The distribution functions were then used in a MCSO to calculate the optimum number of sessions to offer for each course with respect to a predefined objective function while subject to a constraint.

2.1.1 Generating Probability Distributions for Expected Student Demand

Course session attendance was drawn from a historical database over a five-year period. Restricting the data to the last five years reduced the likelihood of including courses that were no longer of relevance to the military training staff. While this may initially underestimate demand for newly introduced courses, the procedure can be run on an annual basis to update the values used to feed the distributions. In cases where student demands are known or can be better predicted in advance, the distributions can either be updated to include these values (to ensure a level of stochastic variability in the model) or replaced.

One key limitation from the source data was that attendance was measured based on the number of students that completed each course in the past five years. The number of students that had originally registered, and either withdrew or failed to complete the course (for any variety of reasons) was not measured. Ordinarily, the cancelled sessions would count as having zero students. However, the purpose of the study was to reduce or eliminate cancelled sessions, not to mimic historical cancellation levels. Therefore, it was assumed that only the non-zero student loading was valid.

In the original model, Palisade Corporation's @RISK Professional 6.3 add-in for Microsoft Excel® was used to generate the total number of students for each course based on the historical data. The statistical distribution for the Monte Carlo model of the student demand was selected manually. The statistical distribution was selected according to following rules (Eisler et al, 2015):

- 1 data point: No distribution used; the single data point was treated as a constant and used directly in the scheduling module;
- 2 data points: A uniform distribution between the two sample values was used;
- 3 data points: A triangular distribution was used; the three data points were used as the distribution parameters (min, max, and most likely – peak);

- 4 data points: A triangular distribution was used, where the most likely value was selected using @RISK's graphical distribution fitting function from a plot of the sample values; and
- 5 data points: A Poisson distribution was used, fit from @RISK's distribution fitting function.

2.1.2 Estimating Session Quantities

Monte Carlo Simulation Optimization was employed to generate the student demand for t = 1,000 iterations, each run for 1,000 trials. The session quantities for each course were then estimated based on the expected demand.

Denote $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{z}$, $\boldsymbol{p}$ to be vectors of length N as per (1), with $\boldsymbol{x}$ representing the total number of scheduled sessions for all N courses (where x_i is the number of sessions for each course i = 1,2,…,N), $\boldsymbol{y}$ representing the minimum number of sessions required to enroll all students, $\boldsymbol{z}$ representing the maximum number of sessions that could be offered without cancellations due to insufficient student load, and $\boldsymbol{p}$ representing the total number of students requiring courses from the probability distributions.

$$\mathbf{x}=\begin{pmatrix}x_1\\ .\\ x_i\\ .\\ x_N\end{pmatrix},\mathbf{y}=\begin{pmatrix}y_1\\ .\\ y_i\\ .\\ y_N\end{pmatrix},\mathbf{z}=\begin{pmatrix}z_1\\ .\\ z_i\\ .\\ z_N\end{pmatrix},\mathbf{p}=\begin{pmatrix}p_1\\ .\\ p_i\\ .\\ p_N\end{pmatrix} \text{ where } x_i, y_i, z_i, p_i \in \mathbb{W}=\{0,1,2,\ldots\} \text{ for } i=1,2,\ldots,N \tag{1}$$

Denote n_{min_i} and n_{max_i} as the minimum required and maximum allowed number of students for every session of course i. The formulae for computing $\boldsymbol{y}$ and $\boldsymbol{z}$ are given as:

$$y_i=\begin{cases}\left|\dfrac{p_i}{n_{\max_i}}\right|+1; & \text{for } \left(p_i \bmod n_{\max_i}\right)\geq n_{\min i}\\ \left|\dfrac{p_i}{n_{\max_i}}\right|; & \text{for } \left(p_i \bmod n_{\max_i}\right)< n_{\min i}\end{cases}, \text{ and } z_i=\left|\dfrac{p_i}{n_{\min_i}}\right|$$

Let $\boldsymbol{h}$ be a vector of length N, with its elements representing the number of extra sessions that must be scheduled for each course in order to enroll all students, and A similar vector of cancelled sessions, $\boldsymbol{k}$:

$$h_i(x_i,y_i)=\begin{cases}x_i-y_i; & \text{for } z_i\geq y_i\\ 0; & \text{for } x_i<y_i\end{cases}, \text{ and } k_i(x_i,z_i)=\begin{cases}z_i-x_i; & \text{for } z_i\geq x_i\\ 0; & \text{for } z_i<x_i\end{cases} \tag{2}$$

A scalar penalty function c can then be introduced as the sum of the extra required and cancelled sessions:

$$c(\mathbf{x},\mathbf{y},\mathbf{z})=\sum_{i=1}^{N}\left(h_i(x_i,y_i)+k_i(x_i,z_i)\right) \tag{3}$$

Ideally, $c = 0$ when $z_i \leq x_i \leq y_i$, if $\boldsymbol{y}$ and $\boldsymbol{z}$ are fixed; meaning there are no extra or cancelled sessions when the number of sessions falls between the minimum required to enroll all students and the maximum that can be offered without cancellations. However, the values of $\boldsymbol{y}$ and $\boldsymbol{z}$ depend on the value of student demand $\boldsymbol{p}$, which is obtained by Monte Carlo simulation. Thus, the selection of a single set of values for $\boldsymbol{x}$ for all simulated values of $\boldsymbol{y}$ and $\boldsymbol{z}$ becomes a penalty function minimization problem.

Denote $h_{\mu,i}$,$k_{\mu,i}$, $y_{\mu,i}$ and $z_{\mu,i}$ as the values of h, k, y, z in μ^{th} iteration, $\mu = 1,2,\ldots,t$. Then, the first optimization objective, C, over all trials is:

$$C = \min\left(\bar{c}\right) = \min\left(\frac{1}{t}\sum_{\mu=1}^{t}\left(c_{\mu}\left(\mathbf{x},\mathbf{y},\mathbf{z}\right)\right)\right) = \min\left(\frac{1}{t}\sum_{\mu=1}^{t}\left(\sum_{i=1}^{N}\left(h_{\mu,i}\left(x_i, y_{\mu,i}\right) + k_{\mu,i}\left(x_i, z_{\mu,i}\right)\right)\right)\right) \tag{4}$$

Denote the fractional capacity of the i^{th} course, f_i, as the total number of students scheduled for that course divided by the product of the maximum allowable number of students per session and the total number of sessions to be held for course i,

$$f_i = p_i / (n_{\max} \times x_i) \tag{5}$$

Thus, the second optimization objective, G, over all trials can be expressed as (6):

$$G = \max\left(\bar{f}\right) = \max\left(\frac{1}{t}\sum_{\mu=1}^{t}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\frac{p_{u,i}}{n_{\max_{u,i}} \times x_i}\right)\right)\right) \tag{6}$$

where G is the maximum overall fractional capacity, which is obtained by averaging the fractional capacity of all courses, under the constraint defined in (4).

The @RISK Optimizer was used to generate the expected course demand, $\boldsymbol{p}$, by Monte Carlo simulation in order to maximize the mean fractional course capacity (5) and minimize the scalar penalty function (3). Given an infeasible starting solution of zero sessions for all courses, a solution for 1,000 trials with 1,000 iterations could be found in approximately 12 minutes on a single core computer with an Intel® i7 640 2.13 GHz processor.

2.2 Session Scheduling and Resource Optimization

Given a number of sessions to hold (with defined durations) and pools of resources, the schedule optimization problem can be summarized as finding the best choice of start dates, while considering resource assignment limitations, with respect to some objective function. The session schedule optimization is constrained by a sub-problem of resource assignment. In order to reduce required computational time, resource selection was separated from the main scheduling problem. Viable resources are ranked and good candidates are selected from a feasible set.

Once the resource selection was finalized, the optimization loop continued using the Frontline Systems' evolutionary algorithm (Frontline Systems Inc., n.d.) until predefined termination conditions were met (e.g., maximum number of sub-problems or the time spent on computation with no improvement to the objective function). Such termination parameters must be explicitly specified for any genetic algorithm because, unlike in stochastic optimization methods, there is no notion of derivatives or convergence. In addition, due to the large size of the explored solution space, genetic optimization problems are computationally expensive. A practical issue with this approach is that Microsoft Excel's® native Solver has an upper limit of 200 variables for non-linear problems. If resource assignments were implemented as independent, single-digit binary variables, the size of the problem would quickly become too large for the Solver to handle; this would impose an artificial limitation on the manageable problem size (Frontline

Systems Inc., n.d.). The resource selection algorithm aims to map a session schedule to one particular resource mix, which means that only one objective function evaluation is necessary.

2.2.1 Resource selection

For the sub-problem of resource selection for a given session *s* of a given course, the preferred state of the individual resource schedule is characterized the session capacity fraction, $cf_s = T_s/T_{Total}$. This is the fraction of schedule slots, *T*, that are currently assigned to a particular session.

Given a start date proposed by an iteration of the session scheduling optimization algorithm, resources are scheduled based on how full the overall resource schedule is. The process is then repeated for every session. The loop is recursive in nature, as the schedule for each resource is updated and used as input to schedule resources for the next session. This algorithm can be thought of as an attempt to move the resource schedules near a local optimum.

In order to help the Solver select from feasible resource assignment solutions, the concept of localized conflicts (*LC*) was introduced. *LC*s are the number of days of conflicts that are created by assigning more than one session to a resource. The function for scoring the overall resource schedule *R* is given in (7). *R* is an artificial function that was developed utilizing a modified variable transformation on the Weibull probability density function (α, β, γ, δ as tuning parameters), which utilizes the localized conflicts and session capacity fraction. The value of *R* enables the preferred selection of the more heavily utilized resources, but not so heavily utilized near 100% that there is no flexibility in the schedule (e.g., to accommodate last-minute schedule changes, unavailability of assets due to illness, unexpected maintenance, or weather, etc.) nor room for time off (i.e., leave for teachers).

$$R = \alpha\beta^{-\alpha}\left(-\gamma\left(cf_s - 1\right)\right)^{\alpha-1} e^{-\left(\frac{-\gamma\left(cf_s-1\right)}{\beta}\right)^{\alpha}} \left(\frac{1}{LC}\right)^{\delta} \tag{7}$$

2.2.2 Session Scheduling

The session scheduling algorithm utilizes the Frontline Systems' evolutionary algorithm, where the primary decision variables are the session start dates. The initial solution value selected for each date was 1 January of the year to be modelled. The start dates were subject to minimum and maximum value constraints (i.e. must fall within the year to be modelled), as well as to be integer values. The tuning parameters that affect the population size, mutation rate and random seed for the solver were left at their default values. The objective function was defined as the schedule capacity fraction for one day of the schedule, *cf*, as a function of the session capacity fraction for all courses and sessions:

$$cf = \begin{cases} \sum_{i=1}^{N}\sum_{s}^{x_i} cf_s - 1; & \text{for } \sum_{i=1}^{N}\sum_{s}^{x_i} cf_s \geq 1 \\ 1; & \text{for } \sum_{i=1}^{N}\sum_{s}^{x_i} cf_s = 0 \end{cases} \tag{8}$$

The key feature of the function *cf* in (9) is symmetry about $cf = 1$. The symmetry creates a global minimum at $cf = 1$, and scores an empty schedule slot ($cf = 0$) as equally undesirable to one schedule conflict ($cf(0) = cf(2) = 1$) ; therefore, from the Solver's perspective, they are scored equally, only $cf(1) = 0$ produces a better result. The schedule capacity fraction in (8) is then summed over an entire schedule array (the two dimensional integer array **A,** with *l* rows and *m* columns of time, such as days per week and week number) to obtain the aggregate objective function F_{cf} shown in (9).

$$F_{cf}(\mathbf{A}) = \sum_{d=1}^{l} \sum_{e=1}^{m} cf(\mathbf{A}(d,e)) \tag{9}$$

The function incorporated schedule conflicts (if any) and could include resource costs (in the future), in the form of a weighted sum objective (Kim et al, 2005). The weighted sum approach was selected because the Microsoft Excel® Solver add-in is incapable of handling multiple objective problems.

2.2.3 Sample Performance

A test simulation was run considering 39 military training courses over 50 week time span, with 113 resources separated into 9 different pools. After running the optimization using the standard Solver, a solution was found; however, the number of conflicts was not zero, so the solution was not feasible given the resource pools available in the problem set modelled. The run time was approximately 14 hours

3 REVISED MODEL DEVELOPMENT

The primary driver for development was the need to eliminate the reliance upon specialized software that could not be deployed on internal defence network computer systems. While Microsoft Excel® with the Solver Add-in is readily available on all systems; Palisade Corporation's @RISK is not. To make the course scheduling tool more readily deployable, a restructuring of the problem was undertaken to implement the model solely in Microsoft Excel® and run using Visual Basic for Applications (VBA). Thus, the forecasting of the session quantities and optimization of student loading had to be reformulated such that reasonable computational time frames (similar to previous or better) were obtained in VBA.

3.1 Session quantity optimization

As in the original model, historical course session attendance over a five-year period was used to forecast student demand through the fit of a probability distribution on a course-by-course basis. The distribution functions were then used in a MCSO to derive the maximum overall fractional capacity by enumerating through a bounded set of sub-problems.

3.1.1 Generating probability distributions for student demand

The first step in the model reformulation process was to replace the manual selection for the probability distribution functions using viewing tools within @RISK based on the number of sessions held as data points for each course to obtain attendance values. Instead, VBA functions were created to automatically generate probability distribution functions with following rules:

- 1 data point: No distribution, the single data point was treated as a constant and used directly;
- 2 data points: An integer uniform distribution between the two sample values was used;
- 3 data points: A triangular distribution was used; the three data points were used as the input parameters (min, max, and peak);
- 4 data points: A triangular distribution was used. The minimum (min), peak (most likely), and maximum (max) values were taken over the data points provided. The peak value is given as:

$$peak = 3(mean) - min - max$$
$$\text{*If } peak > max\text{, then } peak = max;\ \text{ If } peak < min\text{, then } peak = min. \tag{10}$$

*Note: This is but one approximation method; another would be to set the min or max to the peak value.

- 5 data points: A Poisson distribution was used, where the mean (M) value of the samples is taken as the event rate (λ).

While the maximum number of data points examined is five, this could easily be modified to accommodate more data points. Table 1 provides a sample data set of the input parameters and selected probability distribution functions from the @RISK and VBA versions of the model. Note that many functions have different calculated peak values; only two courses have different distribution functions, as compared from to the @RISK version of the model (noted with a *). This indicates a reduction in human error by employing a VBA script, rather than relying on a human-in-the-loop to apply the rule set.

Table 1: Comparative model input parameters.

	@RISK				**VBA**			
Course Label	**Function**	**Min.**	**Max.**	**Peak**	**Function**	**Min.**	**Max.**	**Peak**
A	Triangle	1	4	2	Triangle	1	4	1.75
B	Poisson	-	-	31.2	Poisson	-	-	31.2
C	Poisson	-	-	21.6	Poisson	-	-	21.6
D*	Uniform	10	28	-	Poisson	-	-	21.4
E	Triangle	6	12	11	Triangle	6	12	11
F	Poisson	-	-	40	Poisson	-	-	40
G	Constant	-	-	7	Constant	-	-	7
H	Triangle	1	3	1.25	Triangle	1	3	1.5
I	Constant	-	-	6	Constant	-	-	6
J	Triangle	14	25	21	Triangle	14	25	22.5
K	Uniform	2	6	-	Uniform	2	6	-
L	Triangle	4	10	5	Triangle	4	10	5.75
M	Triangle	9	13	10.5	Triangle	9	13	9.5
N	Poisson	-	-	13.4	Poisson	-	-	13.4
O	Poisson	-	-	28.8	Poisson	-	-	28.8
P*	Uniform	12	37	-	Poisson	-	-	25.4
Q	Triangle	4	8	7	Triangle	4	8	7
R	Triangle	8	24	12	Triangle	8	24	12
S	Triangle	18	36	30	Triangle	18	36	30.5
T	Triangle	72	183	150	Triangle	72	183	156.75
U	Uniform	8	9	-	Uniform	8	9	-
V	Uniform	6	10	-	Uniform	6	10	-
W	Triangle	8	16	12	Triangle	8	16	12
X	Triangle	12	20	13	Triangle	12	20	13
Y	Triangle	7	24	22	Triangle	7	24	22
Z	Triangle	21	53	24	Triangle	21	53	31
AA	Triangle	15	25	17	Triangle	15	25	17
AB	Constant	-	-	12	Constant	-	-	12
AC	Triangle	2	7	5	Triangle	2	7	5.25
AD	Triangle	11	46	38	Triangle	11	46	38
AE	Constant	-	-	4	Constant	-	-	4
AF	Constant	-	-	7	Constant	-	-	7
AG	Uniform	4	9		Uniform	4	9	-
AH	Constant	-	-	4	Constant	-	-	4
AI	Uniform	9	48	-	Uniform	9	48	-
AJ	Constant	-	-	10	Constant	-	-	10
AK	Constant		-	20	Constant	-	-	20
AL	Uniform		5	14	Uniform	5	14	-
AM	Constant	-	-	12	Constant	-	-	12

3.1.2 Estimating session quantities

Monte Carlo simulation optimization was utilized to derive the average student demand over a large number of iterations. However, the objective function and the constraints were decomposed into a set of sub-problems obtained on a course-by-course basis. As illustrated in (11), the objective function can be re-written such that $max\left(\sum_{\mu=1}^{t}\left(f_{\mu,i}\right)\right)$ represents the maximum sum of the capacity for course *i*. This is averaged over *t* iterations, for all *N* courses. As per (5), each term $f_{\mu,i}$ is dependent only upon the value of p_i, since the expected student demand for each course is independent. Thus each term can be computed separately. Therefore, if a model is evaluating 40 courses, then it could be split into 40 small optimization problems instead of one large sum.

$$\begin{aligned} G = \max\left(\bar{f}\right) &= \max\left(\frac{1}{t}\sum_{\mu=1}^{t}\left(\frac{1}{N}\sum_{i=1}^{N}\left(f_{\mu,i}\right)\right)\right) = \frac{1}{tN}\times\max\left(\sum_{\mu=1}^{t}\left(\sum_{i=1}^{N}\left(f_{\mu,i}\right)\right)\right) \\ &= \frac{1}{tN}\times\max\left(\sum_{\mu=1}^{t}\left(f_{\mu,1}+f_{\mu,2}+,\ldots,+f_{\mu,N}\right)\right) \\ &= \frac{1}{tN}\times\max\left(\sum_{\mu=1}^{t}\left(f_{\mu,1}\right)+\sum_{\mu=1}^{t}\left(f_{\mu,2}\right)+,\ldots,+\sum_{\mu=1}^{t}\left(f_{\mu,N}\right)\right) \\ &= \frac{1}{tN}\times\left(\max\left(\sum_{\mu=1}^{t}\left(f_{\mu,1}\right)\right)+\max\left(\sum_{\mu=1}^{t}\left(f_{\mu,2}\right)\right)+,\ldots,+\max\left(\sum_{\mu=1}^{t}\left(f_{\mu,N}\right)\right)\right) \end{aligned} \tag{11}$$

The same logic applies to the scalar penalty function, so that (4) can be written as:

$$\min\left(c(\mathbf{x},\mathbf{y},\mathbf{z})\right) = \min\left(\frac{1}{t}\sum_{\mu=1}^{t}\left(c_{\mu}(\mathbf{x},\mathbf{y},\mathbf{z})\right)\right) = \frac{1}{t}\min(c_1)+\frac{1}{t}\min(c_2)+,\ldots,\frac{1}{t}\min(c_t) \tag{12}$$

where c_i is the sum of (2) and (5) for course *i* with *t* samples. Each objective to be minimized in (12) can be computed separately, since the terms are independent. The value of the penalty function c_i will increase as x_i increases when $x_i > max(z')$, and increase as x_i decreases when $x_i < min(y')$, where ***y'*** is the set of minimum number of sessions for *t* samples of course *i* and ***z'*** is the set of maximum number of sessions for *t* samples of course *i*. Thus, the optimal solution that satisfies the scalar penalty function must lie between the extrema for all samples when ***x*** is in [*min(**y'**)*, *max(**z'**)*]. Since ***x*** must be integer, it is simpler to enumerate through the set of possible values that could theoretically satisfy all of the samples generated.

The algorithm for estimating the session quantities required is given as:

Step 1: Generate *t* samples of expected student demand in a Monte Carlo fashion by using the probability distribution function associated with the i^{th} course (the type of the function depends on the number of data points).

Step 2: For each sample (***x***) of expected student demand, compute ***y***, ***z***, and ***p***.

Step 3: Compute [*min(**y'**)*, *max(**z'**)*] for all *t* samples. For each integer value of ***x*** between [*min(**y'**)*, *max(**z'**)*], compute c_i. The minimum value of c_i over all samples then forms part of the final solution ***c***.

Step 4: Increment to the next course *i* and return to Step 1 for all *N* courses.

This problem is computationally trivial and requires no specialized optimization software to solve. The run time for the initial simulation was less than one minute for 2,000 iterations of 2,000 samples each. For a fourfold increase in simulation, the computational speed was less than 1/12th the original algorithm design, for a 48 times increase in speed and a guarantee of optimality using enumeration. One minor disadvantage of this faster version is that it does not record a simulation log in the same fashion as @RISK, so it is not possible to visually inspect variable values without debugging the code.

3.2 Session scheduling and resource optimization

Using the same case study as for the previous model, 39 military training courses were simulated over 50 week time span, with 113 resources separated into 9 different pools. Since the session scheduling and resource optimization algorithms were not fundamentally changed, the standard Solver was run to obtain the solution found in Table 2. Again, the number of conflicts was non-zero, so the solution was not feasible given the resource pools available in the problem set modelled. For the initial model, the number of additional sessions required and sessions cancelled was not calculated separately from the constraint metric, which is the sum of the two terms.

Table 2: Case study results.

Output	**Value – Initial Model**	**Value – Revised Model**
Constraint Metric	1489	1464
Additional Sessions Required	-	1004
Sessions Cancelled	-	460
Fractional Capacity	0.590	0.498
Total Sessions	88	99

Table 3 provides a detailed summary of the course bounds (minimum and maximum number of students per session), the average quantity of sessions offered annually based on historical data, and comparison of the results from both models. The initial model proposed fewer courses in total, with a higher course loading; however, there were more constraint violations. This meant that additional cost would be incurred to add extra sessions or cancel sessions that were not required as compared to the revised model. Furthermore, since both models were running in MCSO mode with different random seeds and probability distributions for the student demand, the solutions are expected to be visibly different.

4 CONCLUSIONS

A model has been used to analyze the supply and demand relationship, and identify the optimal number of sessions that should be offered in order to meet expected student demand and minimize course cancellations and surplus. This would likely lead to annual cost savings in a large training program. In addition, tailoring the program delivery structure in a way that reduces the number of cancelled sessions, while offering sessions specifically at the times when limited assets (such as ships or major training simulators) are available to support them, would enable more efficient demand and supply forecasting, possibly leading to further savings. Due to the possibility of postings, assignments, and deployments in military careers – in addition to the usual family- or job-related responsibilities, health concerns, or other commitments that may interrupt training – the drop-out rate may be quite high, and should be a major consideration for future work.

ACKNOWLEDGMENTS

The authors would like to thank Owen Lu, University of British Columbia, for his contribution to the work on the original model implementation in @RISK.

Table 3: Session quantity result.

Course Label	Min. Students Required Per Session	Max. Students Allowed Per Session	Historical Avg. Quantity of Sessions Offered (Per Year)	Expected Session Quantity (Per Year) – Initial	Expected Session Quantity (Per Year) – Revised
A	4	6	2	0	0
B	1	12	3	3	5
C	6	24	21	1	2
D	10	25	22	1	1
E	6	12	1	1	1
F	6	12	1	4	4
G	6	10	1	1	1
H	1	6	1	2	1
I	4	8	4	1	1
J	4	8	1	3	3
K	2	6	1	1	1
L	2	4	2	2	2
M	4	8	1	1	2
N	4	8	2	2	2
O	6	18	0	2	3
P	6	12	1	2	3
Q	2	12	1	2	1
R	4	12	1	2	2
S	4	8	0	3	4
T	2	8	21	20	23
U	1	8	3	1	2
V	1	8	4	1	2
W	1	8	2	5	2
X	1	8	2	3	3
Y	1	8	2	2	3
Z	6	16	2	2	3
AA	6	16	1	2	2
AB	6	12	1	1	1
AC	1	6	1	2	2
AD	6	12	1	2	3
AE	6	12	4	0	0
AF	1	6	1	2	2
AG	4	10	5	1	1
AH	2	6	1	1	1
AI	4	12	0	3	4
AJ	6	12	0	1	1
AK	4	8	2	3	3
AL	6	12	2	1	1
AM	12	20	1	1	1
Total			120	88	99

REFERENCES

Badri, M.A. 1996. “A two-stage multiobjective scheduling model for [faculty-course-time] assignments.” *European Journal of Operational Research*, vol. 94, pp. 16-28.

Dahiya, S. 2015. *Course scheduling with preference optimization*. M. A. Sc. Thesis, The Pennsylvania State University Graduate School, University Park, Pennsylvania. Available via https://turing.cs.hbg.psu.edu/mspapers/sources/siddharth-dahiya.pdf. Accessed Nov. 25, 2016.

Eisler, C., P. Dobias, and O. Lu. 2015. "Course loading optimisation." In *Proceedings of the 33rd International Symposium on Military Operational Research,* Barrington Digital Library at Cranfield University. Available via http://www.ismor.com/32ismor_archive/papers/pdf/32ismor_course-loading_paper.pdf. Accessed Nov. 25, 2016.

Eliashberg, J, Q. Hegie, J. Ho, D. Huisman, S.J. Miller, S. Swami, C.B. Weinberg, and B. Weirenga. 2009. "Demand-Driven Scheduling on Movies in a Multiplex." International Journal of Research in Marketing, vol. 26, pp. 75-88.

Frontline Systems Inc. Excel Solver, Optimization Software, Monte Carlo Simulation, Data Mining – Frontline Systems, Available at http://www.solver.com/. Accessed Nov. 25, 2016.

Graham, R. 1966. "Bounds for Certain Multiprocessing Anomalies." *The Bell Technical Journal*, pp. 1563-1581, Bell Labs.

Gunawan A., and K.M. Ng. 2011. "Solving the teacher assignment problem by two metaheuristics." *International Journal of Information Management Science*, vol. 22, pp. 73-86, Tamkang University.

Gurvich, I., J. Luedtke, and T. Tezcan. 2010. "Staffing Call Centres With Uncertain Demand Forecasts: A Chance-Constrained Optimization Approach." *Management Science*, vol. 56, no. 7, pp. 1093-1115.

Hossain, S., and M.F. Zibran. 2007. "A multi-phase approach to the university course timetabling problem." In *Proceedings of the 6th Cologne Twente Workshop on Graphs and Combinatorial Optimization*, pp. 73-76, University of Twente, Enschede.

Kim, I. Y., and O.L. de Weck. 2005. "Adaptive weighted sum method for multi-objective optimization: a new method for Pareto front generation." *Structural Multidisciplinary Optimization*, vol. 31, no. 2, pp. 105-116, Springer Science + Business Media.

Nakasuwan, J., P. Srithip and S. Komolavanij. 1999. "Class scheduling optimization." *Thammasat International Journal of Science & Technology*, vol. 4, no. 2, pp. 88-98, Thammasat University Press.

Schaerf, A. 1999. "A survey of automated timetabling." *Artificial Intelligence*, vol. 13, pp. 87-127.

Schniederjans, M.J., and G.C. Kim. 1987. "A goal programming model to optimize departmental preference in course assignments." *Computers & Operational Research*, vol. 14, no. 2, pp. 87-96,.

Wasfy, A., and F.A. Aloul. 2007. "Solving the university class scheduling problem using advanced ILP techniques." In *Proceedings of the 4th IEEE GCC Conference*.

Winch, J.K., and J. Yurkiewicz. 2013. "Student Class Scheduling with Linear Programming." In *Proceedings for the Northeast Region Decision Sciences Institute Annual Meeting*, pp. 374-386.

Zambelli, M.S., I. Luna, and S. Soares. 2009. "Long-Term Hydropower Scheduling Based on Deterministic Nonlinear Optimization and Annual Inflow Forecasting Models." In *Proceedings of the BUCHAREST Power Technical Conference*, IEEE.

AUTHOR BIOGRAPHIES

CHERYL EISLER is a defence scientist at Defence Research and Development Canada's Centre for Operational Research and Analysis (DRDC CORA). She holds an M.A. Sc. in aerospace engineering from Carleton University. Her research interests lie in modelling and simulation for coastal surveillance and force structure analysis.

DR. PETER DOBIAS is also is a defence scientist at DRDC CORA, and holds a Ph.D. in Theoretical Physics from the University of Alberta. His research interests include complex adaptive systems, non-linear systems dynamics, and modeling and simulation of naval warfare-related systems.

Dr. MIN JING LIU received her Ph.D. from McMaster University, and held a Post-Doctoral fellowship with DRDC CORA. Her research interests include graph theory, algorithm design, and large scale integration problems.