Enhancing Business Process Management With Simulation Optimization

**Jay April , Marco Better, Fred Glover,
James P. Kelly, Manuel Laguna
OptTek Systems, Inc.**

Introduction

A growing number of business process management software vendors are offering simulation capabilities to extend their modeling functions and enhance their analytical proficiencies. Simulation is positioned as a means to evaluate the impact of process changes and new processes in a model environment through the creation of "what-if" scenarios. Simulation is promoted to enable examination and testing of decisions prior to actually making them in the "real" environment. Since simulation approximates reality, it also permits the inclusion of uncertainty and variability into the forecasts of process performance. This paper explores how new approaches are significantly expanding the power of simulation for business process management.

Less than a handful of business process software vendors offer optimization to supplement their simulation capability. However, the need for optimization of simulation models arises when the process analyst wants to find a set of model specifications (i.e., input parameters and/or structural assumptions) that leads to optimal performance. On one hand, the range of parameter values and the number of parameter combinations is too large for analysts to simulate all possible scenarios, so they need a way to *guide* the search for good solutions. On the other hand, without simulation, many real world problems are too complex to be modeled by mathematical formulations that are at the core of pure optimization methods. This creates a conundrum; pure optimization models alone are incapable of capturing all the complexities and dynamics of the system, so one must resort to simulation, which cannot easily find the best solutions. Simulation Optimization resolves this conundrum by combining both methods.

The merging of optimization and simulation technologies has seen remarkable growth in recent years. A Google.com search on "Simulation Optimization" returns more than four thousand pages where this phrase appears. The content of these pages ranges from articles, conference presentations and books, to software, sponsored work, and consultancy.

Until relatively recently, however, the simulation community was reluctant to use optimization tools. Optimization models were thought to over-simplify the real problem, and it was not always clear why a certain solution was the best (Barnett 2003). However, a vast body of research in the area of metaheuristics, coupled with improved statistical methods of analysis, has reduced this resistance considerably. In 1986, Dr. Fred Glover coined the term metaheuristic to describe a master strategy that guides and modifies other heuristics to produce solutions beyond those that are normally generated in a quest for local optimality. The heuristics guided by such a meta-strategy may be high-level procedures or may embody nothing more than a description of available moves for transforming one solution into another together with an associated evaluation rule.

Today, there exist very powerful algorithms to guide a series of simulations to produce high quality solutions in the absence of tractable mathematical structures. Furthermore, we are now able to precisely compare different solutions in terms of quality. Nearly every commercial discrete-event or Monte Carlo simulation software package contains an optimization module that performs some sort of search for optimal values of input parameters (April, et al., 2003). OptQuest[®], a leading optimization tool for commercial simulation software, employs

metaheuristics such as scatter search and tabu search, and techniques such as neural networks, to provide optimization capabilities to users. Among the many simulation software products that deploy the OptQuest® technology, SIMPROCESS and SIMUL8 are two examples of available products that are being used in business process software applications.

In this article, we present two examples of simulation optimization using OptQuest® .to illustrate how to optimize simulation models. In the first case, we construct a discrete event simulation model of a hospital emergency room to determine a configuration of resources that results in the shortest average cycle time for patients (DeFee, 2004). In the second case, we develop a simulation model to minimize staffing levels for personal claims processing in an insurance company. We then summarize some of the most relevant approaches that have been developed for the purpose of optimizing simulated systems. Finally, we concentrate on the metaheuristic black-box approach that leads the field of practical applications, and we provide some relevant details on how this approach has been implemented and used in commercial software.

Optimization of Simulation Models

Once a simulation model has been developed to represent a system or process, we may want to find a configuration that is best, according to some performance measure, among a set of possible choices. For simple processes, finding the best configuration may be done by trial-and-error or enumeration of all possible configurations. When processes are complex, and the configuration depends on a number of strategic choices, the trial-and-error approach can be applied with only very limited success. In these cases, we use an optimization tool to guide the search for the best configuration.

Some applications of simulation optimization may include the goal of finding:

- the best configuration of machines for production scheduling
- the best integration of manufacturing, inventory, and distribution
- the best layouts, links, and capacities for network design
- the best investment portfolio for financial planning
- the best utilization of employees for workforce planning
- the best location of facilities for commercial distribution
- the best operating schedule for electrical power planning
- the best assignment of medical personnel in hospital administration
- the best setting of tolerances in manufacturing design
- the best set of treatment policies in waste management

The optimization of simulation models deals with the situation in which the analyst would like to find which of possibly many sets of model specifications (i.e., input parameters and/or structural assumptions) lead to optimal performance. In the area of design of experiments, the input parameters and structural assumptions associated with a simulation model are called *factors.* The output performance measures are called *responses.* For instance, a simulation model of a manufacturing facility may include factors such as number of machines of each type, machine settings, layout, and the number of workers for each skill level. The responses may be cycle time, work-in-progress, and resource utilization.

In the world of optimization, the factors become decision variables, and the responses are used to model an objective function and constraints. Whereas the goal of experimental design is to find out which factors have the greatest effect on a response, optimization seeks the combination of factor levels that minimizes or maximizes a response (subject to constraints imposed on factors and/or responses). Returning to our manufacturing example, we may want to formulate an optimization model that seeks to minimize cycle time by manipulating the number of workers and machines, while restricting capital investment and operational costs as well as maintaining a

minimum utilization level of all resources. A model for this optimization problem would consists of decision variables associated with labor and machines as well as a performance measure based on a cycle time obtained from running the simulation of the manufacturing facility.   The constraints are formulated both with decision variables and responses (i.e., utilization of resources).

When changes are proposed to business processes in order to improve performance, the projected improvements can be simulated and optimized artificially. The sensitivity of making the changes on the ultimate objectives can be examined and quantified, reducing the risk of actual implementation. Changes may entail adding, deleting, and modifying processes, process times, resources required, schedules, work rates within processes, skill levels, and budgets. Performance objectives may include throughput, costs, inventories, cycle times, resource and capital utilization, start-up times, cash flow, and waste. In the context of business process management and improvement, simulation can be thought of as a way to understand and communicate the uncertainty related to making the changes, while optimization provides the way to manage that uncertainty.

Selecting the Best Configuration for a Hospital Emergency Room Process

The following example is based on a model provided by CACI, and simulated on SIMPROCESS. Consider the operation of an emergency room (ER) in a hospital. Figure 1 shows a high-level view of the overall process. The process begins when a patient arrives through the doors of the ER, and ends when a patient is either released from the ER or admitted into the hospital for further treatment. Upon arrival, patients sign in, are assessed in terms of their condition, and are transferred to an ER room. Depending on their condition, patients must then go through the registration process and through the treatment process before being released or admitted into the hospital.
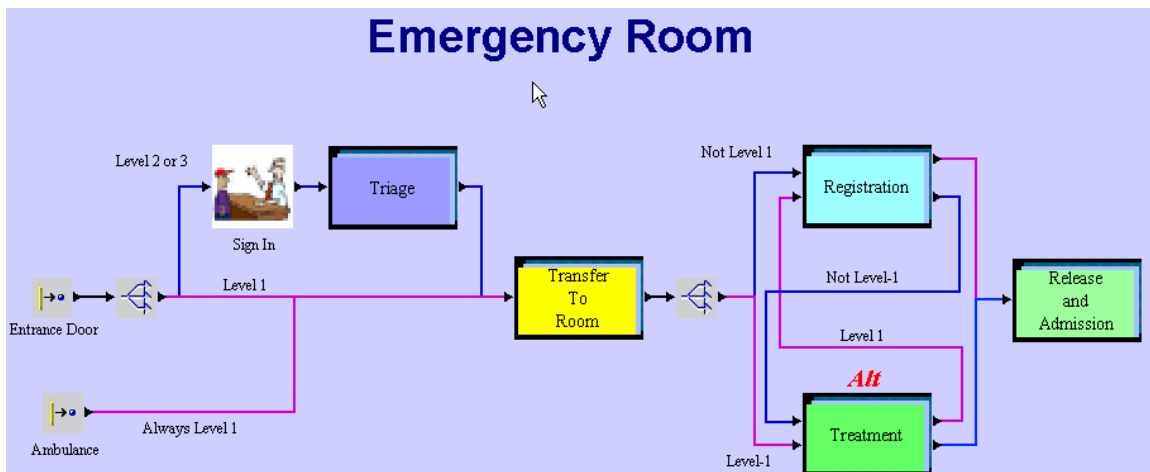

Figure 1. High-level process view

Patients arrive either on their own or in an ambulance, according to some arrival process. Arriving patients are classified into different levels, according to their condition, with Level 1 patients being more critical than Level 2 and Level 3.

Level 1 patients are taken to an ER room immediately upon arrival. Once in the room, they undergo their treatment. Finally, they complete the registration process before being either released or admitted into the hospital for further treatment.

Level 2 and Level 3 patients must first sign in with an *Administrative Clerk*. After signing in, their condition is assessed by a *Triage Nurse*, and then they are taken to an ER room. Once in the room, Level 2 and 3 patients must first complete their registration, then go on to receive their

treatment, and, finally, they are either released or admitted into the hospital for further treatment.

The treatment process consists of the following activities:

1. A secondary assessment performed by a nurse and a physician.
2. Laboratory tests, if necessary, performed by a patient care technician (PCT).
3. The treatment itself, performed by a nurse and a physician.

The registration process consists of the following activities:

1. A data collection activity performed by an Administrative Clerk.
2. An additional data collection activity performed by an Administrative Clerk, in case the patient has Worker's Compensation Insurance.
3. A printing of the patient's medical chart for future reference, performed by an Administrative Clerk.

Finally, 90% of all patients are released from the ER, while the remaining 10% are admitted into the hospital for further treatment. The final release/hospital admission process consists of the following activities:

1. In case of release, either a nurse or a PCT fills out the release papers (whoever is available first).
2. In case of admission into the hospital, an Administrative Clerk fills out the patient's admission papers. The patient must then wait for a hospital bed to become available. The time until a bed is available is handled by an empirical probability distribution. Finally, the patient is transferred to the hospital bed.

The ER has the following resources:

- Nurses
- Physicians
- PCTs
- Administrative Clerks
- ER Rooms

In addition, the ER has one Triage Nurse and one Charge Nurse at all times.

Due to cost and layout considerations, hospital administrators have determined that the staffing level must not exceed 7 nurses, 3 physicians, 4 PCTs, and 4 Administrative Clerks. Furthermore, the ER has 20 rooms available; however, using fewer rooms would be beneficial, since other departments in the hospital could use the additional space more profitably. The hospital wants to find the configuration of the above resources that minimizes the total asset cost. The asset cost includes the staff's hourly wages and the fixed cost of each ER room used. We must also make sure that, on average, Level 1 patients do not spend more than 2.4 hours in the ER. This can be formulated as an optimization problem, as follows:

Minimize the expected **Total Asset Cost**
Subject to the following constraints:
        Average Level 1 Cycle Time is less than or equal to **2.4 hours**
        # Nurses are greater than or equal to 1 and less than or equal to **7**
        # Physicians are greater than or equal to 1 and less than or equal to **3**
        # PCT's are greater than or equal to 1 and less than or equal to **4**
        # Admin. Clerks are greater than or equal to 1 and less than or equal to **4**
        # ER Rooms are greater than or equal to 1 and less than or equal to **20**

This is a relatively simple problem in terms of size: 6 variables and 6 constraints. However, if we were to rely solely on simulation to solve this problem, even after the hospital administrators have

narrowed down our choices to the above limits, we would have to perform 7x3x4x4x20=6,720 experiments. If we want a sample size of, say, at least 30 runs per trial solution in order to obtain the desired level of precision, then each experiment would take about 2 minutes[1]. This means that a complete enumeration of all possible solutions would take approximately 13,400 minutes, or about 70 working days. This is obviously too long a duration for finding a solution.

In order to solve this problem in a reasonable amount of time, we called upon the OptQuest® optimization technology integrated with SIMPROCESS. As a base case we decided to use the upper resource limits provided by hospital administrators, to get a reasonably good initial solution. This configuration yielded an Expected Total Asset Cost of $ 36,840, and a Level 1 patient cycle time of 1.91 hours.

Once we set up the problem in OptQuest, we ran it for 100 iterations (experiments), and 5 runs per iteration (each run simulates 5 days of the ER operation). Given these parameters, the best solution, found at iteration 21 was:

Nurses	Physicians	PCTs	Admin Clerks	ER Rooms
4	2	3	3	12

The Expected Total Asset Cost for this configuration came out to $ 25,250 (a 31% improvement over the base case), and the average Level 1 patient cycle time was 2.17 hours. The time to run all 100 iterations was approximately 28 minutes.

After obtaining this solution, we redesigned some features of the current model to improve the cycle time of Level 1 patients even further. In the proposed model, we assume that Level 1 patients can go through the treatment process and the registration process in parallel. That is, we assume that, while the patient is undergoing treatment, the registration process is being done by a surrogate or whoever is accompanying the patient. If the patient's condition is very critical, than someone else can provide the registration data; however, if the patient's condition allows it, then the patient can provide the registration data during treatment.

Figure 2 shows the model with this change. By implementing this change in the optimized model, we now obtain an average Level 1 patient cycle time of 1.98 (a 12% improvement).
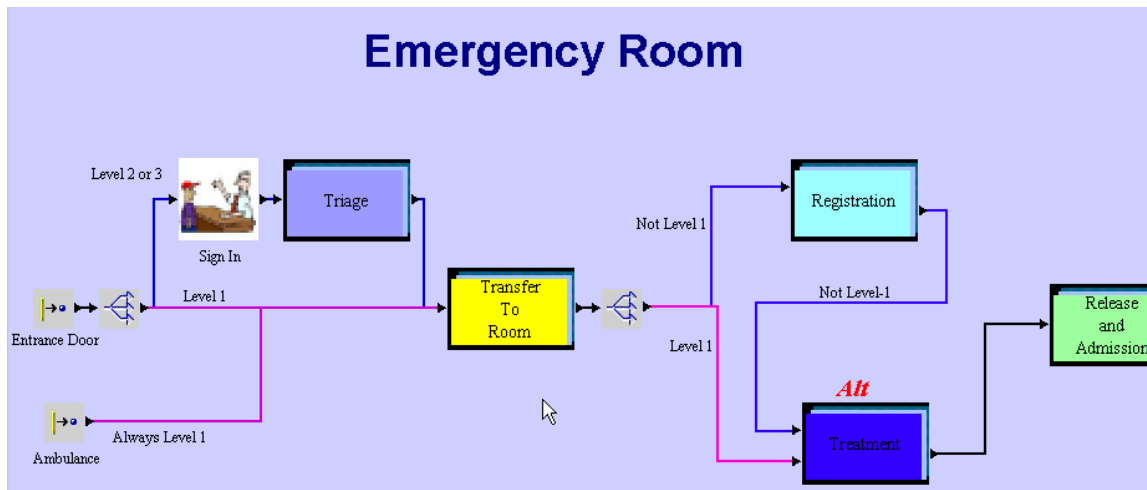

Figure 2. Proposed Process

---

[1] We timed one experiment with 30 runs on a Dell Dimension 8100, with an Intel Pentium 4 processor @ 1700 MHz.

If we now optimize this new model, given the change that we implemented, we obtain a new optimal solution in 28 iterations, as follows:

Nurses	Physicians	PCTs	Admin Clerks	ER Rooms
4	2	2	2	9

This configuration yields an Expected Total Asset Cost of $ 24,574, and an average Level 1 patient cycle time of 1.94 hours.

By using optimization, we were able to find a very high quality solution in less than 30 minutes. In addition, we were able to make changes to improve the model and re-optimize to find a better configuration. It is quite unlikely that this solution would be found relying solely on simulation.

Selecting the Best Staffing Level for a Personal Claims Process at an Insurance Company

The following example is based on a model provided by SIMUL8 Corporation. We used SIMUL8 for the simulation, and OptQuest for SIMUL8 for the optimization.

A personal claims department in an insurance company handles claims made by their clients. Claims arrive according to a Poisson process, with a mean inter-arrival time of 5 minutes. Figure 3, on the next page, is a process map depicting the personal claims process in terms of swim lanes.

The first lane corresponds to work done by a claims handler (CH) located at the client's local service center. Upon arrival of a claim, the assessor determines if the client has a valid policy. If no (5% of all cases), then the case is terminated; otherwise (95% of all cases), the assessor enters the appropriate information in the system.

In the second lane, an assessor located at the service center (ASC) receives the information from the claims handler. The assessor first determines if the claim is covered by the client's policy. If not (5% of all cases), the case is terminated; otherwise (95% of all cases), the assessor approves the preliminary estimate of the damage. If the damage exceeds $2,000 (35% of all cases), the claim is sent to an assessor at headquarters for approval; otherwise (65% of all cases), it is sent directly to a Senior Assessor.

Lane 3 corresponds to the assessor at headquarters (AHQ). The assessor first authorizes the on-site investigation of the accident. If the investigation determines that the incident is not covered by the client's policy (2% of all cases), then the case is terminated; otherwise (98% of all cases), a final price is determined and the case is approved.

In lane 4, the senior assessor (SA) receives the claim, checks it, completes it, and provides the final approval. Once the claim is approved, it is sent to documentation control.

Documentation control (DC), in lane 5, is in charge of processing the payment to the client, closing the case, and, finally, filing the claim.
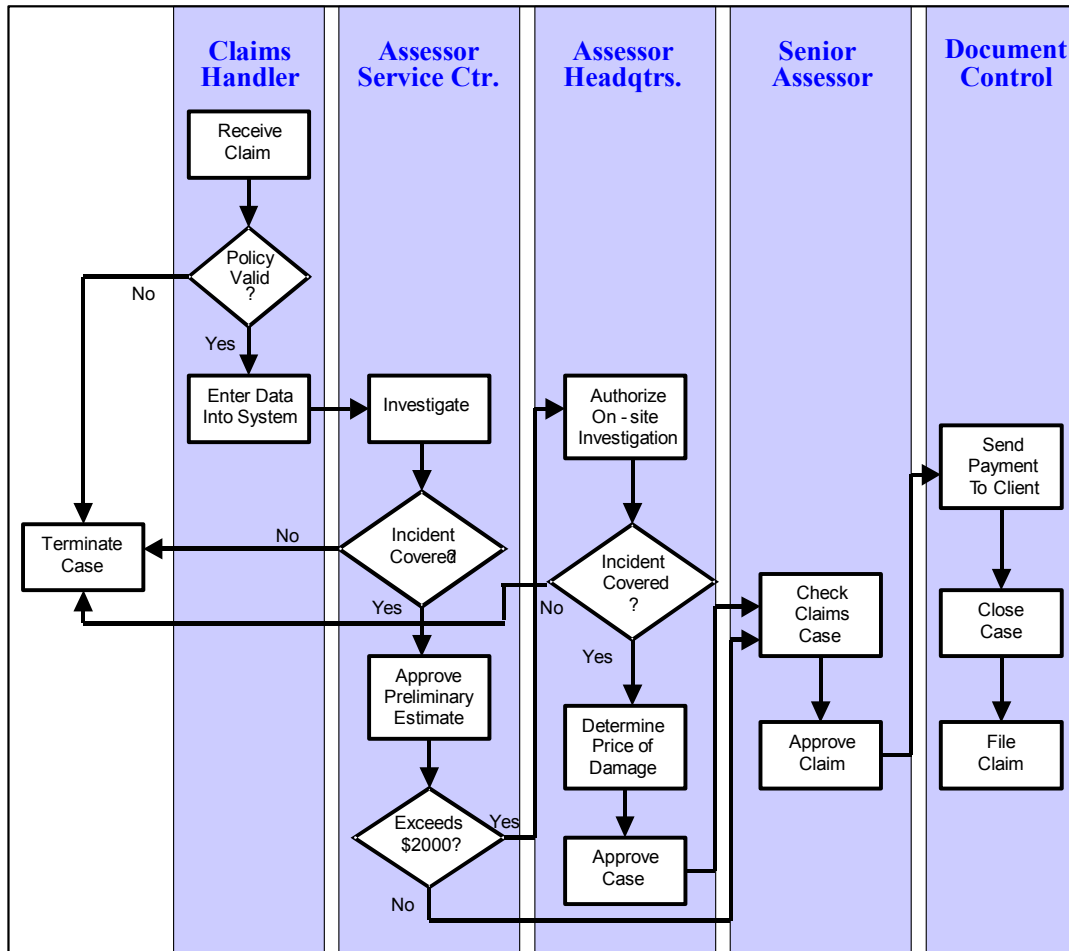
Figure 3. Map of Personal Claims Process

The problem in this example is to find staffing levels for each of the five resource types, in order to minimize headcount, while keeping average throughput above 1500 claims during 4 weeks. Each resource type has a maximum limit of 20 people, and the overall headcount in the process cannot exceed 90. The formulation of the optimization problem is as follows:

Minimize the **Headcount**
Subject to the following constraints:
Average Throughput is equal to or greater than **1500**
    # Claims Handlers are greater than or equal to 1 and less than or equal to **20**
    # Service Center Assessors are greater than or equal to 1 and less than or equal to **20**
    # Headquarter Assessors are greater than or equal to 1 and less than or equal to **20**
    # Senior Assessors are greater than or equal to 1 and less than or equal to **20**
    # Document Controllers are greater than or equal to 1 and less than or equal to **20**
    The overall headcount cannot exceed 90.

Once again, a what-if analysis of all the possible solutions to this problem would require about 800,000 scenarios. Optimization is necessary to find a good solution efficiently. A good starting point can probably be established by consulting with experienced managers in the insurance claims area, based on the expected demand of claims.

We use OptQuest to optimize the staffing levels of this system. We run OptQuest for 100 iterations and 5 simulation runs per iteration. Table 1 shows a summary of the results, and

Figure 4 shows the graph depicting the search of OptQuest for Simul8 towards improving solutions. The performance graph shows the value of the objective (in this case Headcount) on the *y-axis*, and the number of iterations on the *x-axis*. The line shows only improving solutions.

						Throughput (# claims)				
Solution	**Claims Handler**	**Assessor (SC)**	**Assessor HQ**	**Senior Assessor**	**Doc. Control**	**LSL**	**Avg.**	**USL**	**Headcount**	**Avg. Cycle Time (min.)**
1	9	17	17	15	16	1563.00	1568.00	1573.00	74	639.00
2	9	17	17	14	16	1559.00	1564.00	1570.00	73	658.00
3	8	17	16	15	16	1562.00	1567.00	1573.00	72	646.00
4	9	18	12	15	11	1611.00	1622.00	1633.00	65	503.00
5	9	18	11	15	11	1610.00	1621.00	1632.00	64	510.00

Table 1 header: Simul8 Results / Personal Claims Process - 20 Trials

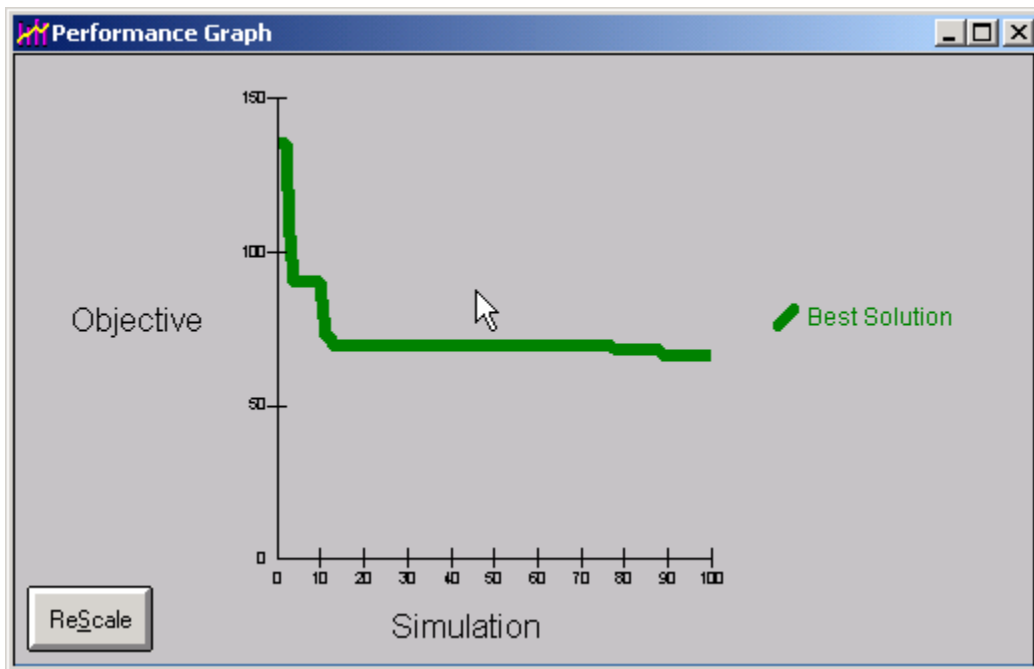**Table 1. Summary Results**

Figure 4. Performance Graph for the Optimization Of the Personal Claims Process

Since some of the solutions obtained from our optimization are relatively close in terms of throughput and cycle time, an analyst may want to reevaluate a set of the best *n* solutions to assess the precision of the results. In Table 1, we present the best 5 solutions obtained from our OptQuest run, by conducting an experiment of 20 trials for each solution. The information can now be given to a process manager. The manager can analyze the trade-offs between headcount and throughput or cycle time, to decide which configuration best aligns with service levels and process goals. For example, we can see that solutions 1, 2, and 3 are statistically the same. Solutions 4 and 5 are significantly better than 1, 2, and 3 in terms of headcount, throughput, and cycle time, so the analyst should pick one of these. Which one is better? We re-ran 60 trials for each of these two solutions, to obtain a 95% confidence interval for each of these measures. In both cases the confidence intervals overlap. In fact, the resulting confidence intervals for throughput are almost identical; the intervals for cycle time are also very close, with the expected cycle time for solution 4 (503 minutes) about 1.4% lower than that for solution 5 (510 minutes). The analyst should consider if the savings obtained from having one assessor less justifies such a small difference in cycle time. If so, then solution 5 should be chosen.

The State-of-the-Art in Simulation Optimization

Four approaches account for most of the academic literature in simulation optimization. They are: (1) stochastic approximation (gradient-based approaches); (2) (sequential) response surface methodology; (3) random search; and (4) sample path optimization (also known as stochastic counterpart). However, none of these approaches have been used to develop optimization for commercial simulation software, mainly because these "methods generally require a considerable amount of technical sophistication on the part of the user, and they often require a substantial amount of computer time as well" (Andradóttir, 1998).

Leading commercial simulation software employs metaheuristics as the methodology of choice to provide optimization capabilities to their users. Like other developments in the Operations Research/Computer Science interface (e.g., those associated with solving large combinatorial optimization problems), commercial implementations of simulation optimization procedures have only become practical with the exponential increase of computational power and the advance in metaheuristic research. The metaheuristic approach to simulation optimization is based on viewing the simulation model as a black box function evaluator.

Figure 5 shows the black box approach to simulation optimization favored by procedures based on metaheuristic methodology. In this approach, the metaheuristic optimizer chooses a set of values for the input parameters (i.e., factors or decision variables) and uses the responses generated by the simulation model to make decisions regarding the selection of the next trial solution.
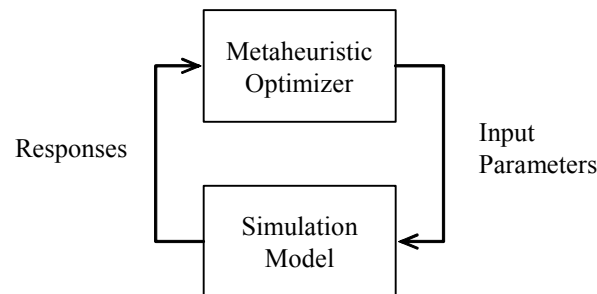

Figure 5. Black box approach to simulation optimization

Most of the optimization engines embedded in commercial simulation software are based on evolutionary approaches. Evolutionary approaches search the solution space by building and then evolving a population of solutions. The evolution is achieved by mechanisms that create new trial solutions out of the combination of two or more solutions that are in the current population. Transformation of a single solution into a new trial solution is also considered in these approaches. Examples of evolutionary optimization approaches are Genetic Algorithms and Scatter Search. The latter, in conjunction with a memory-based approach called Tabu Search, is used in OptQuest.

In the context of simulation optimization, a simulation model can be thought of as a "mechanism that turns input parameters into output performance measures" (Law and Kelton, 1991). In other words, the simulation model is a function (whose explicit form is unknown) that evaluates the merit of a set of specifications, typically represented as set of values. Viewing a simulation model as a function has motivated a family of approaches to optimize simulations based *on response surfaces* and *metamodels*.

A response surface is a numerical representation of the function that the simulation model represents. A response surface is built by recording the responses obtained from running the simulation model over a list of specified values for the input factors. A metamodel is an algebraic model of the simulation. A metamodel approximates the response surface, and therefore optimizers use it instead of the simulation model to estimate performance. Standard linear

regression has been and continues to be one of the most popular techniques used to build metamodels in simulation. More recently, metamodels based on neural networks (Laguna and Martí, 2002), Kriging (van Beers and Kleijnen, 2003), and the Lever Method (April, et al., 2003) have also been developed and used for estimating responses based on input factors. Once a metamodel is obtained, in principle, appropriate deterministic optimization procedures can be applied to obtain an estimate of the optimum (Fu, 2002).

Figure 6 depicts the metaheuristic optimization process with a metamodel filter.
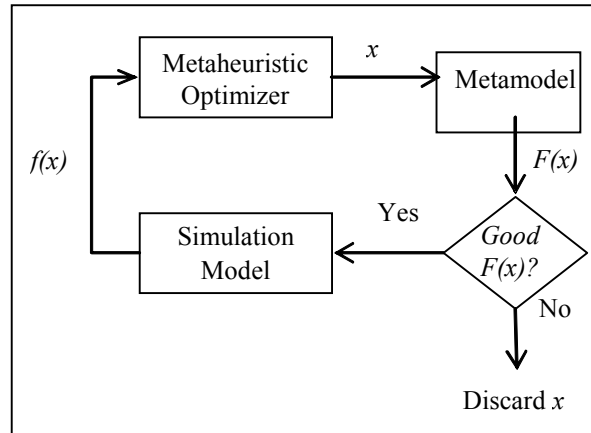

Figure 6. Metaheuristic optimizer with a metamodel filter

An important feature in simulation optimization software is the ability to specify constraints. Constraints define the feasibility of trial solutions. Constraints may be specified as mathematical expressions or as statements based on logic. In the context of simulation optimization, constraints may be formulated with input factors or responses.

If the constraints in a simulation optimization model depend only on input parameters then a new trial solution can be checked for feasibility before running the simulation. An infeasible trial solution may be discarded or may be mapped to a feasible one when its feasibility depends only on constraints formulated with input parameters. OptQuest, for instance, has a mechanism to map infeasible solutions of this type into feasible ones. On the other hand, if constraints depend on responses, then the feasibility of a solution is not known before running the simulation.

Conclusions

Although most commercial simulation software products now have an embedded optimization tool, until recently, the simulation community had not embraced optimization. Optimization models were thought to oversimplify the problem at hand, and there was not a precise way to determine the quality of a solution. Furthermore, traditional approaches for optimizing simulations usually require a considerable level of technical sophistication for the user, as well as considerable computer time.

However, the need for simulation optimization arises frequently in practice, given that most real world systems are too complex to analyze by trial-and-error. This need has resulted in a vast body of research to apply metaheuristic approaches to the optimization of simulations. These approaches help guide the search for configurations of a simulation model that result in optimal or near-optimal performance. The use of metaheuristics, coupled with advanced statistical analysis methods, considerably reduces the required computer time to obtain good solutions, and provides clear and precise measures of solution quality and process performance.

Finally, as demonstrated in our illustrative examples, tools such as OptQuest, along with their simulation software hosts, do not require a high level of technical sophistication from the user.

Simulation models that need to be optimized can be set up rather quickly and intuitively, and results can be analyzed and communicated clearly. As this trend in making the simulation optimization approach easier to use continues, the software is becoming attractive to all levels of management and staff, and, thus, is steadily gaining a wider user base.

Acknowledgements

References

Andradóttir, S. (1998) "A Review of Simulation Optimization Techniques," in *Proceedings of the 1998 Winter Simulation Conference*, D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan (eds.), pp. 151-158.

April, J., J. Kelly, F. Glover and M. Laguna (2003) "Practical Introduction to Simulation Optimization" in *Proceedings of the 2003 Winter Simulation Conference,* S. Chick, T. Sanchez, D. Ferrin and D. Morrice, (eds.) pp. 71-78.

Barnett, M. (2003) "Modeling & Simulation in Business Process Management," BP Trends Newsletter, White Paper & Technical Briefs, pp. 1-10.

DeFee, J. and P. Harmon (2004) "Business Activity Monitoring and Simulation," BP Trends Newsletter, White Paper and Technical Briefs, pp. 1-24.

Fu, M. (2002) "Optimization for Simulation: Theory and Practice," *INFORMS Journal on Computing,* vol. 14, no. 3, pp. 192-215.

Glover, F., M. Laguna and R. Marti (2003) "Scatter Search and Path Relinking: Advances and Applications," in *Handbook of Metaheuristics,* F. Glover and G. Kochenberger (eds.), Kluwer Academic Publishers, Boston.

Law, A. M. and W. D. Kelton (1991) *Simulation Modeling and Analysis,* Second Edition, McGraw-Hill, New York.

Martí, R., M. Laguna and V. Campos (2002) "Scatter Search vs. Genetic Algorithms: An Experimental Evaluation with Permutation Problems," to appear in *Adaptive Memory and Evolution: Tabu Search and Scatter Search*, Cesar Rego and Bahram Alidaee (eds.), Kluwer Academic Publishers, Boston.

––––––––––

Jay April, Marco Better, Fred Glover, James P. Kelly and **Manuel Laguna** are all members of **OptTek Systems, Inc**. for more information, check www.opttek.com .