THE IMPORTANCE OF THE OBJECTIVE FUNCTION DEFINITION AND EVALUATION IN THE PERFORMANCE OF THE SEARCH ALGORITHMS

Javier Otamendi
E-mail: jotamendi_30@yahoo.com

KEYWORDS
Simulation and optimization, search methods, capability index, cost functions.

ABSTRACT

A qualitative and quantitative comparison of simulation optimization methodologies is presented to specifically study the importance of the objective function on the election of the algorithm to generate the alternatives to be simulated. If a capability index is used in conjunction with a powerful rejection algorithm and a efficient metaheuristic, the average number of repetitions per simulated alternative is to be decreased significantly, facilitating the simulation of more alternatives. The probability of the convergence of the algorithm increases then considerably. The credibility in the solution is also raised since the decision is made both in terms of average behavior and variability.

INTRODUCTION/PROBLEM DEFINITION

The problem is to optimize a simulation vector of responses Ψ over a region $\Theta \subset \Re^P$ with respect to an ordered p-tuple of input factor settings $X = (x_1, x_2, \ldots, x_p, \ldots x_P) \in \Theta$ (Jacobson and Schrueben 1989). In other words, a p-tuple (or alternative or combination) must be selected among a full set in terms of several responses (or criteria or fitness).

The difficulty to solve the problem is triple. First, each of the J individual responses $Y_j \in \Psi$ is a random variable $Y_j = f(X)$, which must be estimated. Second, the individual responses must be aggregated in an overall response or objective function $Y = f(\Psi)$, which is then also a random variable. Third, the total number of p-tuples to evaluate, I, might be very large.

The solution comes in the form of a simulation optimization methodology (SOM) which combines a generation algorithm of a subset of p-tuples to be simulated and the evaluation of an objective function.

Figure 1 summarizes a simulation optimization methodology within the framework of a generic optimization methodology, particularizing it for a simulation study. The system is first represented in a simulation model and then an iterative procedure is followed: "simulations are normally used in a scenario-by-scenario base, with the designer generating a solution and subsequently having the computer evaluating it" (Gama and Norford 2002).

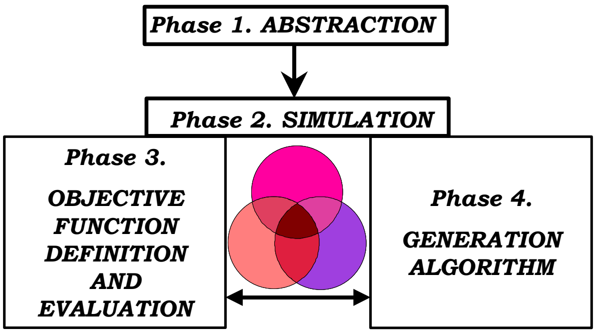

Figure 1. Simulation Optimization Methodology

The origin of any SOM resides in the impossibility of analyzing the huge set of alternatives that are under study. The size of the subset of the alternatives that a methodology is able to simulate basically depends on the calculation speed of the computer.

If the number of alternatives is not excessive and each and every one of them might be simulated in the required time, the WHOLE SET should be analyzed and no generation algorithm is needed. The problem is reduced to the evaluation of each alternative.

If, as it is usually the case, the set is too large because the execution of one repetition is too slow or because there is a continuous variable under consideration, then a SUBSET is to be analyzed. The generation algorithm is crucial and will be dependent on the shape Ψ over the region Θ, which depends on the objective function that has been defined. "The success of the optimization procedure depends on the choice of the objective function and its functional relationship to the control parameters" (Hilgers and Boersma 2001).

What follows is the qualitative and quantitative study of the available SOMs, grouped not only in terms of their different formulations but also on their desirable characteristics. The study is performed first in terms just of the objective function to explain the way it is defined and evaluated. Then, the focus is in the combination of an objective function and generation algorithms and how they affect the final solution.

Proceedings 16th European Simulation Symposium
György Lipovszki, István Molnár © SCS Press, 2004
ISBN 1-56555-286-5(book) / ISBN 1-84233-106-x(CD)

DESIRABLE CHARACTERISTICS OF THE SIMULATION METHODOLOGY

Many SOMs have been developed. They all try to find a satisficing alternative, with a fitness value close to that of the optimum alternative, without having to analyze all the available alternatives, that is, in a reasonable execution time.

A desirable search methodology must accelerate then the selection of satisficing alternatives. The optimum simulation search methodology will be one in which the defined multicriteria objective function allows for the simulation of only a few alternatives, among which the subpar alternatives are run only one time or not simulated at all, and the real candidates are run a small, preset number of times (R_{max}). In summary, and since the response surface Ψ is a random variable, the objectives of a SOM are to "maximize statistical power and psychological validity" (Wager and Nichols 2003).

Let's separate the above objectives into three characteristics, namely, efficacy, efficiency and credibility (Law and Kelton 1991).

Efficacy and credibility

The two characteristics are closely related. If efficacy is a measure of the distance to the ideal objective function value, the credibility is the psychological measure that the solution obtained is the optimum one.

Efficacy is to be calculated as follows:

$$\text{EFFICACY} = 100\% * \left(1 - \frac{|Y - Y^*|}{|Y^*|} \right)$$

where:

- Y = objective function value for a particular combination or p-tuple
- Y^* = objective function value for the optimum combination, which is not known, unless all of the combinations are analyzed.

Credibility is the confidence of the decision maker in the solution obtained, that is, the degree of belief that the optimum solution has been obtained. It will be quantified as a number between 0 and 100.

Efficacy and credibility are affected by two factors. The first is the percentage of simulated alternatives. The more alternatives are run, the higher the credibility will be. If the whole set is evaluated, efficacy of the selected alternative is 100%, since Y^* is known and $Y_{selected} = Y^*$. If only a subset is evaluated, there is no certainty that the selected alternative is the optimum one. That is why a good methodology should search for $P(Y_{selected} = Y^*) \rightarrow 1$.

The second is the number of runs made per alternative (R). The more runs performed, the higher the credibility, since the estimation of the individual objective functions is improved, unless a small number guarantees that the alternative is not optimal but subpar, and therefore is subject to be rejected.

Efficiency or the total number of runs, NTOT

Efficiency is the time spent to select an alternative among those included in the available set.

Quantitatively, the efficiency is measured as the total number of simulation runs (*NTOT*), which might be split into the following factors:

$$\text{NTOT} = \frac{\text{Runs}}{\text{Alternative}} * \text{Alternatives}$$

It is possible to improve the efficiency by reducing either factor of the expression. Depending on the objective function to estimate, the number of runs might be different (first factor) and also the method to select the subset of alternatives to simulate (second factor).

The above formula might be further specified if the number of runs per alternative depends on the objective function of an alternative. If the estimation of the value of the objective function makes an alternative feasible, more repetitions should be made than if the estimation makes an alternative clearly infeasible or subpar.

Then, the distribution of the repetitions per alternative, Ω, might be defined as:

$$p(\Omega = R) = p(R) \qquad \forall R \in [0, R_{max}]$$

where:

- R is the number of repetitions that are to be performed per alternative
- R_{max} is the maximum number of repetitions
- $p(R) = I_R/I$; probability of running an alternative R times
- I = total number of alternatives
- I_R = number of alternatives that are run R times.

Therefore, the total number of runs is the sum of the total number of runs for each R:

$$NTOT = \sum_{R=0}^{R_{max}} NTOT_R$$

which is the number of alternatives times the number of runs for each R:

$$NTOT = \sum_{R=0}^{R_{\max}} R\ I_R$$

which might be calculated also as the total number of alternatives times the average number of runs per alternative:

$$NTOT = \sum_{R=0}^{R_{\max}} R\ I\ p(R) = I \sum_{R=0}^{R_{\max}} R\ p(R) = I * R_{mean}$$

NTOT might be lowered then by either reducing I (simulating only a small subset), decreasing Rmax per evaluated feasible alternative, or altering Ω (shifting the weight towards the lower values of R). Any of the factors are dependent on the objective function used.

OBJECTIVE FUNCTION DEFINITION AND EVALUATION

An important step of the SOM is to define the objective function that is going to evaluate the fitness of an alternative, that is, the capability of an alternative to meet requirements. This objective function is an aggregation of several individual criteria, and must include an idea of risk.

It is necessary, first, to define a fitness value per criterion that allows for a satisfaction and aspiration analysis following the multicriteria decision making theory (Zeleny 1982). An alternative will be satisfactory if it fulfills certain satisfaction requirements and it will be ideal if it reaches certain aspiration levels. The fitness value must include the information provided for each criterion (satisfaction limits - or worst permissible values - and aspiration level - or ideal value) and the random variables Y_j obtained by the repetitive execution of simulation models.

The fitness value must quantitatively reflect the degree of fulfillment of requirements. Therefore, it must include in the analysis the whole output distribution per criterion, and not only its mean behavior. The proposed alternatives must simultaneously generate values inside the satisfaction limits, avoiding adverse situations, and a mean behavior close to the ideal value or aspiration level.

Then, a multicriteria fitness value must be defined to aggregate individual fitness value, in order to perform an optimization or search process with a single objective and select among the alternatives that are within the satisficing subset. This multicriteria fitness value summarizes the degree of satisfaction of all the individual criteria.

Individual criterion

The main characteristics of one criterion must be:

1. Subjective: comes from someone's mind.
2. Give an idea of preference: alternatives are to be compared and ordered.

Each criterion must be optimized, either by:

1. Maximizing its value, for example, service level.
2. Minimizing its value, for example cost.
3. Searching for a target value, for example budget.

In all three cases, both an aspiration level T_j and upper and lower satisfaction levels (LSL_j and USL_j) are subjectively set (Barba-Romero and Pomerol 1997), levels that define the acceptability or feasibility of a given alternative.

Then, it is necessary to estimate the performance of criterion *j* for each alternative *i* in a measure $v_{ij} = f(Y_{ij})$ and compare it with the desirable values. The probability distribution of v_{ij} is to be estimated from R repetitions of the simulation model.

Aggregated criterion

It is not trivial to develop a multicriteria measure $v_i = f(v_{ij})$ that aggregates the individual criteria. The individual measure for each criterion has to be selected and then combined all together into a single measure. The bigger problem is to put all the measures in the same unit.

Even in the same units, priorities or ranking between the criteria might be set. Therefore weights that combine priorities and change of units are to be determined.

Setting the subjective values

As already mentioned, the determination of the objective function is subjective in nature, since the weights, the targets and the specification limits are subjectively assigned.

In terms of the weights, several procedures are available (for example, Analytic Hierarchy Process (AHP) (Saaty 1980)) to give some sense to the assignment process.

Regarding the other parameters (USL, T, LSL), it is usually possible to correctly assign those values since the decision maker knows the system in hand, especially for the target.

To conclude, it should be mentioned that it is usually easier to set the specification limits than the weights (Barba-Romero and Pomerol 1997), since each criterion is studied independently.

QUALITATIVE ANALYSIS OF OBJECTIVE FUNCTIONS

In this section, the two main types of multiple objective functions are introduced.

Average

Usually, the individual average value for any measure, $\overline{v_{ij}}$, is used, and its confidence interval given to provide an idea of variation.

The multicriteria fitness value is generally a linear combination of the averages for each criterion, where the weights are used to change units and measure the relative importance of each criterion:

$$\overline{v_i} = \sum_J w_j \overline{v_{ij}}$$

The weaknesses of this fitness value is that it only includes mean behaviour and that several criteria with different units are aggregated.

The following are two possible aggregations and how they are optimized.

Economic
Both individual criteria must be converted to the same monetary units. If a criterion is not monetary, it is incorporated as a penalty for not fulfilling requirements (Hilgers and Boersma 2001):

Economic = Profit - Penalty

This objective value is to be maximized.

Deviation
Each objective function j has its own target value, T_j. The deviation z_j is calculated as the difference between that target and the average value for the criterion, $\overline{Y}_j$:

$$z_j = - |\overline{Y}_j - T_j|$$

If subjective numeric weights are assigned to each criterion, w_j, the aggregated objective function is then:

$$Deviation = \sum_{j=1}^{J} w_j z_j$$

This objective value is to be minimized.

Process capability index (PCI)

The second option comes from the field of quality control. The whole distribution Y_{ij} is compared with both the specifications and the target and a summary measure estimated. This measure is called Process Capability Index (PCI).

There are several capability indices (Kotz and Lovelace 1998). However, two have come up as the ones more commonly used in industry. One is C_{pk}, whose value is above 1 if the whole distribution is between the specification limits regardless of where the average or the median lie with respect to the target, and C_{pm}, which aggregate the satisfaction of the limits and the closeness to the target.

The $C_{pm}F$ index has also been proposed (Otamendi 2001) to combine the two previous indices in to a single measure:

$$\mathrm{CpmF_{ij}} = \begin{cases} \mathrm{Cpm_{ij}} & \forall \mathrm{MCpk_{ij}} \geq 1 \\ 0 & \forall \mathrm{MCpk_{ij}} < 1 \end{cases}$$

where:

$$\mathrm{Cpm_{ij}} = \frac{\mathrm{USL_j} - \mathrm{LSL_j}}{6\sqrt{s_{ij}^2 + \left(\overline{Y_{ij}} - T_j\right)^2}}$$

$$\mathrm{MCpk_{ij}} = \min\left[\frac{\mathrm{USL_j} - \mathrm{P50_{ij}}}{\mathrm{P99.865_{ij}} - \mathrm{P50_{ij}}}, \frac{\mathrm{P50_{ij}} - \mathrm{LSL_j}}{\mathrm{P50_{ij}} - \mathrm{P0.135_{ij}}}\right]$$

USL_j	= upper specification limit (criterion *j*)
LSL_j	= lower specification limit (criterion *j*)
T_j	= target value (criterion *j*)
$\overline{Y_{ij}}$	= average of the data (alternative *i*, criterion *j*)
s_{ij}	= standard deviation of the data
$P_{k\,ij}$	= k% percentile of the data

and MC_{pk} is just the generalization of the C_{pk} index for non-normal distributions.

The $C_{pm}F$ index takes a value of 0 if the alternative does not fulfill requirements, and C_{pm} if all the requirements are met. In other words, subpar alternatives are given a value of 0 and the rest a value which combines the distance to the ideal and the requirements. This objective value is to be maximized.

In the multicriteria case, besides the linear combination option presented for the averages, there is also a possibility that makes sense to develop the MPCI (Multicriteria PCI). Since the capability indices are unitless, and the philosophy is to fulfill requirements, a conservative option is to assign the fitness value of the worst criterion to the alternative:

$$CpmF_i = \min\,(CpmF_{ij}) \quad \forall j = 1...J$$

The strength of this type of aggregation is that each criterion is first evaluated on its own, and then combined into the unique measure. Each and every criterion must be fulfilled for the alternative to be valid and the aggregation performed.

The weakness of the capability indices is that extreme percentiles are used for their calculation. These P_k are estimated with more repetitions than the ones needed to estimate average behaviour.

QUANTITATIVE COMPARISON OF OBJECTIVE FUNCTIONS

A case study is presented to quantitatively compare the objective functions mentioned in the previous section. This case study will also be used later in the article to compare SOMs that include a search algorithm.

The system

The inventory system under study has been adapted from (Taha 1988), by changing the given random variables to non-normal, asymmetric distributions. Four products are stored and sent to customers.

Stock level is revised periodically (*review interval, RI*). If it falls below the *reorder point* (*RO*), a call is made to the supplier to ask for the needed quantity to reach the *maximum inventory level* (*ML*).

The possible values for each factor are:

RI ∈ {2, 3, 4, 5, 6, 7} 6 levels
RO ∈ {250, 275, 300, ..., 725, 750} 21 levels
ML ∈ {3500, 3600, ..., 5400, 5500} 21 levels

for a total of 21*21*6 = 2646 possible combinations or alternatives (*RI, RO, ML*) to choose from.

The decider looks to minimize cost without reducing the level of service. The target for the cost is 7% with an upper limit of 12%. For the service level, the target is 99.5% and the lower limit is 95%.

The estimation of the aggregated objective function for each alternative is based on R_{max} = 100, so the total number of runs performed for the WHOLE SET are:

$$NTOT = R_{max} * I$$
$$= 100 * 2646 = 264600 \text{ runs.}$$

Economic
It is calculated as:

Economic = Profit – Penalty

where the profit is a monetary value in currency units:

Profit = Gross income - Costs

and penalty is a subjective monetary value assigned to the lack of service, for example, 10000 m.u. (monetary units) per day of lateness:

Penalty = 10000 * Late Days

Table 2 includes the results for the best 20 alternatives.

Table 2. Results for *Economic* Objective Function

FACTORS / PARAMETERS			OBJ. FUNCTION	
REVIEW INTERVAL	*REORDER POINT*	*MAXIMUM LEVEL*	*Economic*	
			Fitness	*Efficacy*
3	675	3500	373.44	100.00%
4	700	3500	373.31	99.96%
3	700	3500	373.30	99.96%
4	725	3500	373.16	99.93%
3	725	3500	373.16	99.93%
4	750	3500	373.00	99.88%
3	750	3500	372.94	99.87%
2	675	3600	372.59	99.77%
2	700	3600	372.08	99.64%
2	725	3600	371.90	99.59%
2	750	3600	371.69	99.53%
2	750	3700	371.60	99.51%
2	750	3800	371.24	99.41%
2	700	3900	370.82	99.30%
2	425	4000	370.82	99.30%
2	450	4000	370.73	99.28%
2	725	3900	370.71	99.27%
2	675	3900	370.70	99.27%
2	750	3900	370.70	99.27%
2	475	4000	370.66	99.26%
		Optimum	**373.44**	

The results are very similar for all of the first 20 p-tuples, with the best alternative presenting a review interval of 3. For comparison purposes, in the last column, a measure of efficacy is included.

Deviation
Table 3 includes the results for the best 20 alternatives, for subjective weights of 90% and 10% for the service level and cost, respectively.

Table 3. Results for *Deviation* Objective Function

FACTORS / PARAMETERS			OBJ. FUNCTION	
REVIEW INTERVAL	*REORDER POINT*	*MAXIMUM LEVEL*	*Deviation*	
			Fitness	*Efficacy*
4	575	3600	0.13	100.00%
2	600	3900	0.14	88.17%
4	575	3500	0.15	87.36%
4	625	3600	0.15	86.36%
2	650	3900	0.15	81.94%
4	700	3500	0.16	77.39%
2	625	3800	0.16	75.65%
3	700	3500	0.16	75.20%
4	625	3500	0.16	74.53%
4	675	3700	0.16	74.36%
3	625	3500	0.16	73.76%
2	575	3800	0.16	73.11%
2	650	3800	0.17	69.94%
2	600	4000	0.17	69.83%
2	700	3900	0.17	69.64%
4	675	3600	0.17	68.79%
2	675	3900	0.17	68.66%
2	600	3800	0.17	68.11%
2	650	4000	0.17	67.65%
4	675	3500	0.17	66.74%
		Optimum	**0.13**	

For this objective function, the best alternative is one in which the review interval is set to 4, and a low 66.74% efficacy for the 20th alternative is obtained.

Process capability index $C_{pm}F$
Table 4 includes the results for the best 20 alternatives.

Table 4. Results for $C_{pm}F$ Objective Function

FACTORS / PARAMETERS			OBJ. FUNCTION	
REVIEW INTERVAL	*REORDER POINT*	*MAXIMUM LEVEL*	*CpmF*	
			Fitness	*Efficacy*
2	500	3800	0.65	100.00%
4	575	3600	0.64	99.77%
2	550	3700	0.60	92.52%
2	600	3700	0.58	90.04%
2	600	3900	0.58	89.13%
4	575	3500	0.57	88.59%
4	625	3600	0.57	87.71%
2	600	3800	0.56	86.64%
2	500	3900	0.56	86.48%
2	575	3800	0.56	86.09%
2	550	3900	0.55	85.90%
2	650	3900	0.54	84.43%
2	475	3800	0.54	83.29%
2	650	3700	0.53	82.84%
2	550	3800	0.53	82.83%
4	700	3500	0.52	81.35%
2	525	3800	0.52	80.88%
2	625	3800	0.52	80.18%
3	700	3500	0.52	79.98%
2	575	3700	0.52	79.93%
		Optimum	**0.65**	

The review interval is found to be 2 for this measure.

Comparison

Table 5 includes a comparison of the results obtained for each objective function, for the best 20 alternatives found feasible according to the $C_{pm}F$ index.

Table 5. Comparison

FACTORS / PARAMETERS			OBJECTIVE FUNCTION					
REVIEW INTERVAL	*REORDER POINT*	*MAXIMUM LEVEL*	*Economic*		*Deviation*		*CpmF*	
			Fitness	*Efficacy*	*Fitness*	*Efficacy*	*Fitness*	*Efficacy*
2	500	3800	319.99	85.69%	0.47	-162.90%	0.65	100.00%
4	575	3600	-4.40	-1.18%	0.13	100.00%	0.64	99.77%
2	550	3700	297.98	79.80%	0.63	-286.86%	0.60	92.52%
2	600	3700	309.78	82.95%	0.17	66.46%	0.58	90.04%
2	600	3900	363.64	97.38%	0.14	88.17%	0.58	89.13%
4	575	3500	-105.52	-28.26%	0.15	87.36%	0.57	88.59%
4	625	3600	-4.82	-1.29%	0.15	86.36%	0.57	87.71%
2	600	3800	349.23	93.52%	0.17	68.11%	0.56	86.64%
2	500	3900	364.27	97.54%	0.37	-84.57%	0.56	86.48%
2	575	3800	342.90	91.82%	0.16	73.11%	0.56	86.09%
2	550	3900	363.98	97.47%	0.19	51.87%	0.55	85.90%
2	650	3900	366.11	98.04%	0.15	81.94%	0.54	84.43%
2	475	3800	264.69	70.88%	0.21	40.46%	0.54	83.29%
2	650	3700	336.57	90.13%	0.18	62.60%	0.53	82.84%
2	550	3800	342.99	91.85%	0.38	-95.17%	0.53	82.83%
4	700	3500	373.31	99.96%	0.16	77.39%	0.52	81.35%
2	525	3800	334.32	89.52%	0.21	34.94%	0.52	80.88%
2	625	3800	351.05	94.01%	0.16	75.65%	0.52	80.18%
3	700	3500	373.30	99.96%	0.16	75.20%	0.52	79.98%
2	575	3700	297.84	79.76%	0.20	44.66%	0.52	79.93%
		Optimum	**373.44**		**0.13**		**0.65**	

It is surprising that the best alternative produces a very low value for *Deviation* and the second a low value for *Economic*. It is clear then that the final decision depends on the objective function used. And each decider might choose a very different response measure.

Even the choice of subjective values for the parameters within the three objective functions might vary the final solution. To study the robustness of the final decision, a variation in the subjective values of the monetary penalty, the criteria targets and specification limits is tested. For the monetary value, a smaller value of 5000 m.u. is proposed instead of the original 10000 m.u. For the cost criterion, the target value is kept at an almost achievable value of 7% whereas the upper satisficing value lowered from 12% to 9%. For the service criterion, the target is set at 100% and the lower satisficing limit raised to 98%. Again, it looks like it is easier to set the values for the targets and limits than the one for the penalty. The results obtained for the $C_{pm}F$ index for the best 20 alternatives are included in Table 6.

Table 6. Sensitivity Analysis

FACTORS / PARAMETERS			OBJECTIVE FUNCTION					
REVIEW INTERVAL	*REORDER POINT*	*MAXIMUM LEVEL*	*Economic*		*Deviation*		*CpmF*	
			Fitness	*Efficacy*	*Fitness*	*Efficacy*	*Fitness*	*Efficacy*
4	575	3600	184.40	49.38%	0.13	100.00%	0.26	100.00%
2	600	3900	367.04	98.29%	0.14	88.17%	0.23	89.34%
4	575	3500	134.28	35.96%	0.15	87.36%	0.23	88.79%
4	625	3600	183.98	49.27%	0.15	86.36%	0.23	87.91%
2	600	3800	360.43	96.52%	0.17	68.11%	0.22	86.84%
2	575	3800	357.30	95.68%	0.16	73.11%	0.22	86.29%
2	650	3900	368.31	98.63%	0.15	81.94%	0.22	84.63%
4	700	3500	373.31	99.96%	0.16	77.39%	0.21	81.54%
2	625	3800	361.25	96.74%	0.16	75.65%	0.21	80.36%
3	700	3500	373.30	99.96%	0.16	75.20%	0.21	80.17%
4	625	3500	290.71	77.85%	0.16	74.53%	0.21	79.73%
4	675	3700	253.80	67.96%	0.16	74.36%	0.21	79.64%
3	625	3500	337.97	90.50%	0.16	73.76%	0.20	79.24%
2	700	3900	370.82	99.30%	0.17	69.64%	0.20	76.58%
2	675	3900	370.70	99.27%	0.17	68.66%	0.20	76.15%
4	675	3600	248.76	66.61%	0.17	68.79%	0.20	76.08%
2	650	4000	369.74	99.01%	0.17	67.65%	0.19	75.54%
4	675	3500	364.04	97.48%	0.17	66.74%	0.19	75.10%
3	675	3500	373.44	100.00%	0.17	65.10%	0.19	74.17%
3	750	3500	372.94	99.87%	0.18	64.56%	0.19	73.86%
		Optimum	**373.44**		**0.13**		**0.26**	

The best option shifts from the p-tuple 2-500-3800 to the p-tuple 4-575-3600. Even four of the previously first ten acceptable alternatives become unacceptable.

Again, the decision changes in terms of the objective function and its subjective parameters. And the choices are going to affect the performance of the generation algorithms of the subset of alternatives to be simulated.

GENERATION ALGORITHMS

There exist several algorithms to generate the subset of alternatives that are going to be simulated. The main first distinction has to be made among the analysis of the WHOLE SET (like the example presented in the previous section) or just part of it, SUBSET, which depends on the size of the set, *I*, and therefore, on the nature and size of the input factors.

If all the factors are discrete, the size of the available set has a combinatorial status. Its size value is the multiplication of the number of the settings for each parameter (f_p):

$$I = \prod_{p=1}^{P} f_p$$

If any factor p is continuous, then I grows to infinity since f_p is also infinity. The only possibility is then to make the possible factor settings discrete so f_p is finite.

Whole Set

If the total number of runs is manageable, meaning that NTOT can be performed, simulating each and every alternative in the feasible set is possible and should be done. Therefore, the probability of not running and alternative is p(R=0)=0.

In terms of the characteristics, efficiency is then not a problem, 100% efficacy is fully achieved and confidence in the solution is total.

Subset

One proposed classification for the SUBSET generation algorithms is the following (Jacobson and Schrueben 1989):

- Path search methods: in which a direction of improvement is determined and followed in repeated steps. The main examples are Response Surface Methodology (RSM) (Rees et al. 1985), Stochastic Approximation (SApprox) (Kiefer and Wolfowitz 1951), Perturbation Analysis (PA) (Ho 1984) and the Likelihood Ratio Method (LRM) (Glynn 1987).
- Pattern search methods: in which a pattern in the behavior of the observations is obtained. The main examples are Hooke-and-Jeeves method (HJ) (Hooke and Jeeves 1961) and the Simplex Method (SM) (Nelder and Mead 1965).
- Random methods: in which a randomly selected subset of alternatives is analyzed.
- Integral methods: in which the analysis is done by space- or region-covering and it is specially designed for global optimization.

Several new methods have arisen in the last decades. They have been named as combinatorial methods, or metaheuristics (April et al. 2003). Three different subgroups might be mentioned:

- Simulated annealing (Eglese 1990), in which a new alternative is selected in the neighborhood of the last simulated alternative.
- Evolutionary algorithms like tabu search (TS) (Karaboga and Kalinli 1997), and genetic algorithms (GA) (Goldberg 1989). They search by building and evolving a subset of alternatives.
- Metamodels , like neural networks (Hopfield and Tank 1985). They are used to algebraically represent the simulation model, facilitating then the optimization procedure.

All these methods have been reported to be very useful when I is very large compared to the other methods mentioned, since they have good convergence properties. The solution of the scheduling problem in flexible manufacturing systems (for example, Fazlollahi and Vahidov 2001) is one of the examples most commonly mentioned.

The main thrust of this group of methodologies is to evaluate the smallest possible subset, that is, the smallest percentage of alternatives, so a satisficing alternative is found in the required time. That means that $p(R=0) > 0$, increasing efficiency, but then 100% efficacy and credibility is not guaranteed. A good compromise between efficacy and efficiency is sought, even knowing that the relationship efficacy-efficiency relies heavily on the initial solution (Shelokar et al. 2004). For that reason, several improvements on the original methodologies are being proposed. For example, in order to generate a good set of initial solutions for a genetic algorithm, a linear program is used (Yokohama 2002).

QUALITATIVE COMPARISON OF REPORTED SOMs

An extensive literature review has been performed to see the combinations of objective functions and generation algorithms that have been reported. Using electronic versions of databases, the keywords “simulation and optimization” provide many references. For example, Science Direct includes 3854 articles, Springer 214 and Kluwer 649.

The vast majority of references include SUBSET SOMs with an aggregate average cost function that is being optimized using simulated annealing, an evolutionary algorithm or a neural network or a combination of them. Obviously, the research focus is on trying to improve the convergence of the algorithms to the optimum value in problems with many alternatives.

Among the reported combinations, genetic algorithms have been used to find a good local area, which is studied using hill-climb algorithms (Hart et al. 1998). Simulated annealing finds a subset of good-quality alternatives which are compared using comparison and selection (Ahmed and Alkhamis 2002). Neural networks are used as a filtering step prior to simulating alternatives generated using Genetic Algorithms (Glover et al. 1996, Johnson and Rogers 2001, Laguna and Martí 2002, Yu and Liang 2001). Scatter search (Glover 1977) follows the same principle but with tabu search and neural networks.

In terms of the WHOLE SET, the number of references is not high, and again an average cost function is what is reported in conjunction with the traditional comparison and selection (CS) method. The major drawback is that the possible number of alternatives to evaluate cannot be too high (usually less than 20) for the selection errors not to be too large (Law and Kelton 1991).

The only combination reported in which the objective function is the capability index is GESAS (Otamendi 1999), with its upcoming natural continuation GESAS II (Otamendi 2004). This SOM has been successfully applied in studies with many alternatives: one with 2646 (Otamendi 1999), other with 288 (Rivera 1998).

What follows is a comparison of SOMs by studying in detail the characteristics of the two opposite leading methodologies: GESAS, which tries to evaluate the WHOLE SET of alternatives (increasing the speed while maintaining 100% credibility), and pure EA, which tries to evaluate a small SUBSET in which the optimum is included (improving credibility while maintaining efficiency). A combination of the two methods is also presented.

Whole Set - GESAS II

It is the first attempt to use a multicriteria process capability index (MPCI) in a WHOLE SET simulation study. The focus is on trying to reduce $NTOT = I * R_{mean}$ via reducing R_{mean}, making it possible to evaluate more alternatives I, while keeping efficacy and credibility at 100%.

The use of a MPCI instead of an average cost function allows for the definition of a rejection algorithm (Otamendi 1999) that discriminates easily between good and subpar alternatives. The number of repetitions R is not fixed anymore but dependent on the fitness of the alternative. The satisfaction limits (USL and LSL) are also updated as better alternatives are found (Otamendi 2004). For feasible alternatives, R_{max} runs are performed, but for subpar alternatives a lower number of executions of the model is needed.

Hence, the distribution of the number of runs Ω in the whole range of possible value of R (R=0, ..., R_{max}) is:

$$p_{GESASII}(R) = \begin{cases} p(0) = 0 \\ p(1) = p(1) \\ p(R) = 0 \quad \forall R = 2,...,R_{\max} - 1 \\ p(R_{\max}) = 1 - p(1) \end{cases}$$

Since all the alternatives are evaluated, p(R=0) is 0. And the performance of the algorithm is such that usually a large percentage of the alternatives are simulated only once, and the rest R_{max}. To facilitate the upcoming comparison, the small percentage that will be run between 2 and R_{max}-1 is not considered.

The resulting quantification of the efficiency is:

$$NTOT = I \sum_{R=0}^{R_{\max}} R\, p_R = I * R_{mean}$$
$$= I * \langle 1 p(1) + R_{\max}[1 - p(1)] \rangle$$

The value of R_{mean} and NTOT for GESASII then depends on R_{max} and p(1). Figure 2 shows the possible values for R_{mean} if p(1) is varied between 0 and 0.999 and R_{max} between 10 and 100.

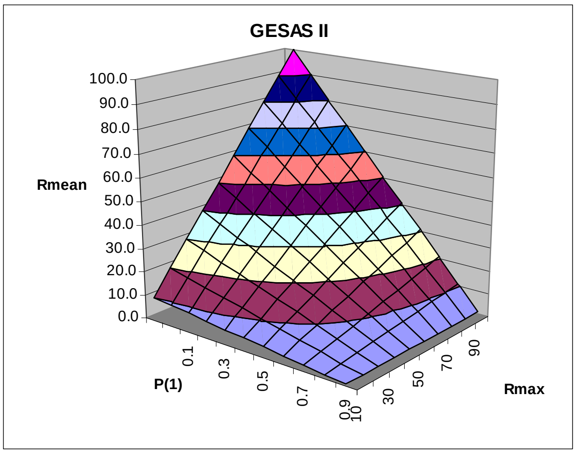

Figure 2. R_{mean} for GESAS II

For common values of R_{max} = 50 and p(1) = 0.80, R_{mean} = 10.8.

Subset - Evolutionary algorithms

It is clear that the main research is being done in the area of evolutionary algorithms. The reason is simple: there are many problems in which only a subset of alternatives is to be simulated, with the corresponding possible loss in efficacy and in psychological credibility.

The focus in this group is then primarily in increasing credibility, via convergence to the optimum solution, and efficiency, via reduction of the simulated subset.

The objective function is an *Average* function, and the number of runs is constant within the simulated set. Hence, the distribution of runs in this case is:

$$p_{SA}(R) = \begin{cases} p(0) = p(0) \\ p(1) = 0 \\ p(R) = 0 \quad \forall R = 2,...,R_{\max} - 1 \\ p(R_{\max}) = 1 - p(0) \end{cases}$$

Many alternatives are not simulated and R_{max} runs are made of the rest. The total number of runs is:

$$NTOT = I \sum_{R=0}^{R_{\max}} R\, p_R = I * R_{mean}$$
$$= I * \langle 0 p(0) + R_{\max}[1 - p(0)] \rangle$$
$$= I * \langle R_{\max}[1 - p(0)] \rangle$$

The value of R_{mean} and NTOT for EA then depends on R_{max} and p(0). Figure 3 shows the possible values for R_{mean} if p(0) is varied between 0 and 0.999 and R_{max} between 10 and 100.

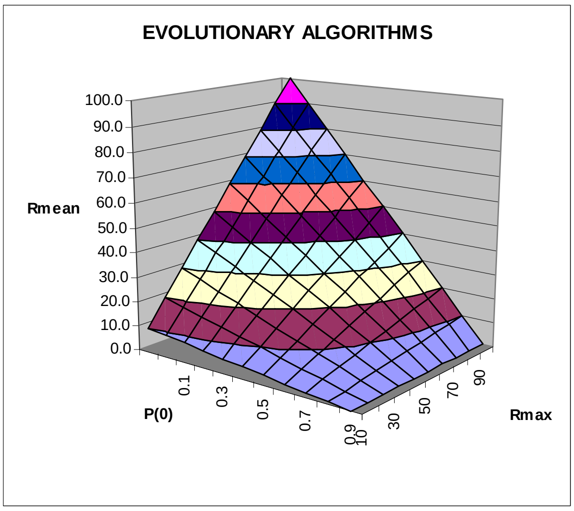

Figure 3. R_{mean} for Evolutionary Algorithms

The graph looks very similar to the one for GESAS II. Using the same values of R_{max} = 50 and p(0)=0.80, R_{mean} = 10.0.

THE NEW METHODOLOGY – GESAS II+EA

As a summary of what has been presented so far is that WHOLE SET methodologies favor credibility and must improve efficiency and SUBSET methodologies favor efficiency and must improve credibility.

In terms of the desirable characteristics of a methodology, credibility is increased as p(0) is reduced and R_{max} is increased, but then efficiency is reduced.

The MPCI and the rejection algorithm of GESASII are powerful tools that must be used to reject subpar alternatives without performing R_{max} runs, so more alternatives are analyzed in the same amount of time (bigger SUBSET). If still the WHOLE SET cannot be analyzed, evolutionary algorithms look like a good choice, but need to improve the convergence.

An improvement on EA might be achieved by the inclusion of $C_{pm}F$ as the objective function instead of the aggregated average cost. Again, the performance of the search algorithm might be dependent on the objective function as well as on the stopping conditions and the initial alternative.

If EA and GESAS are combined into a new SOM, GESAS II + EA, an improvement looks possible. More alternatives will be evaluated per unit time, improving the efficacy for the same number of repetitions.

The distribution of Ω for GESAS II + EA is:

$$p_{GESASii+EA}(R) = \begin{cases} p(0) = p(0)_{EA} \\ p(1) = (1 - p(0)_{EA})\, p(1)_{GESAS} \\ p(R) = 0 \quad \forall R = 2, \ldots, R_{\max} - 1 \\ p(R_{\max}) = (1 - p(0)_{EA})(1 - p(1)_{GESAS}) \end{cases}$$

so the quantification of the efficiency is:

$$NTOT = I \sum_{R=0}^{R_{\max}} R\, p_R = I * R_{mean}$$
$$= I * (1 - p(0)_{EA}) \langle p(1)_{GESAS} + R_{\max}[1 - p(1)_{GESAS}] \rangle$$

Figures 4, 5 and 6 show the possible values for R_{mean} if $p(0)_{EA}$ is varied between 0 and 0.999 and $p(1)_{GESAS}$ is varied between 0 and 0.999 for different values of R_{max}.

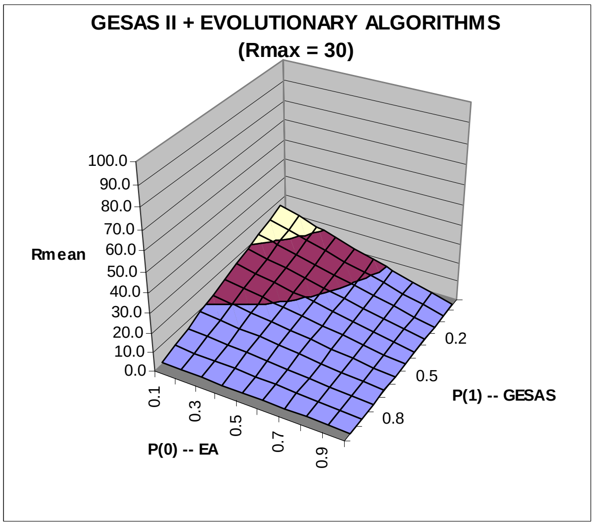

Figure 4. R_{mean} for Combination (R_{max}=30)

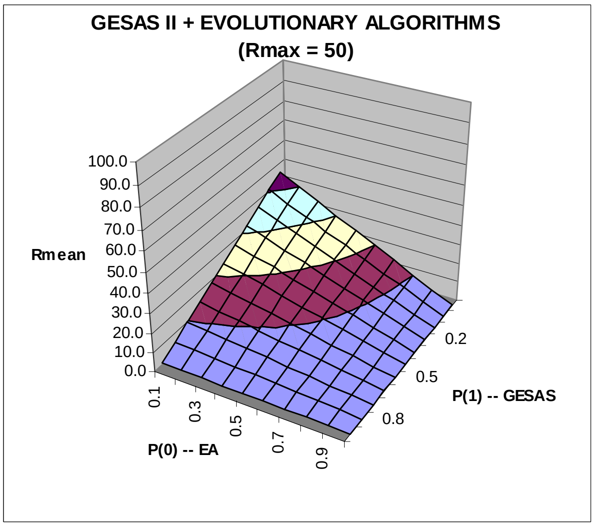

Figure 5. R_{mean} for Combination (R_{max}=50)

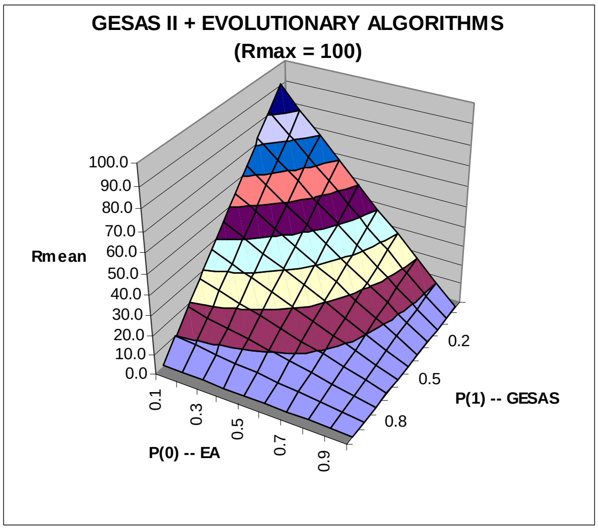

Figure 6. R_{mean} for Combination (R_{max}=100)

Using the same values of R_{max} = 50, $p(0)_{EA}$ = 0.80 and $p(1)_{GESAS}$ = 0.80, R_{mean} = 2.2, five times lower than in the two separated SOMs.

QUANTITATIVE COMPARISON OF SOMs

The inventory problem has been used again but with a small modification. The size of the space has been cut in half by reducing the set of available values of the *review interval* to 2, 3 and 4, eliminating the subpar 5, 6 and 7.

The following methodologies are analyzed (Table 8), all with the $C_{pm}F$ index as the objective function:

- SuperOptimum: infeasible methodology in which one simulation is performed for each alternative except for the selected one, which is run R_{max} times
- Mesh: each alternative is run R_{max} times
- GESAS I: mesh with variable number of repetitions
- GESAS II: GESAS I with updating of satisficing limits
- SA
- GA

Table 8. Comparison of SOMs with PCI fitness

METHODOLOGY			VARIABLES			EFFICACY	EFFICIENCY		
Name	Rmax	I	RI	RP	ML	Efficacy	NTOT	Reps/ Alt.	Speed Ratio
SuperOptimum	30	1323	4	575	3600	100.00%	1352	1.02	1.00
MESH	30	1323	4	575	3600	100.00%	39690	30.00	29.36
GESAS I	30	1323	4	575	3600	100.00%	7548	5.71	5.58
GESAS II	30	1323	4	575	3600	100.00%	3205	2.42	2.37
SA	30	1323	2	425	4000	73.23%	199	0.15	0.15
GA	30	1323	3	675	3800	92.39%	636	0.48	0.47

It looks like the efficiency of GESAS II is very good since uses only 2.39 more time that the super optimal alternative. SA, which is included for comparison, is the quicker methodology but it does not achieve the optimum. GA almost attains the optimum, and its convergence rate is about 5 times greater than that of GESAS II alone.

CONCLUSIONS AND FURTHER RESEARCH

The area of simulation optimization is too complex to develop universal search methodologies. It is very difficult to come up with a multicriteria objective function that includes risk and that might be evaluated with a small number of runs. The selection of the search algorithm is also not easy. So the combination of both objective function and search algorithm is complicated.

It has been shown in this article that the performance of the simulation optimization methodology depends heavily on the definition and evaluation of the objective function, although most of the ongoing research is on the improvement of just the search methodologies over an average response surface.

The analysis has been performed introducing a detailed factorization of the variable NTOT, or total number of repetitions made in the analysis. The separation between runs per alternative, which might be variable, and total number of alternatives, which is usually a just a small subset, helps define and quantify the probability distribution Ω of the number of runs per alternative.

The use of evolutionary algorithms, on one hand, helps reduce the number of runs by reducing the size of the subset to run. The use of a capability index, on the other hand, helps by reducing the number of runs per alternative. The combination of both techniques then reduces drastically the total number of runs without reducing credibility in the solution, as demonstrated quantitatively in the inventory example.

So the future looks promising. Improving each technique independently and coinjontly is the way to go. Studies are being carried in the quality area to define new aggregate indices, so stronger rejection algorithms might be developed, and larger problems simulated exhaustively using GESAS II.

There is research being done in the evolutionary algorithms area to increase the convergence rate and the credibility of the solution.

And tests must be performed to relate both the objective function, crutial part of any SOM, and the search method.

REFERENCES

April, J.; F. Glover; J.P. Kelly; and M. Laguna. 2003. "Practical Introduction to Simulation Optimization." In *Proceedings of the 2003 Winter Simulation Conference* (New Orleans, LA, Dec.7-10). IEEE, Picataway, NJ, 71-78.

Ahmed, M.A. and T.M. Alkhamis. 2002. "Simulation-based Optimization Using Simulated Annealing with Ranking and Selection." *Computers and Operations Research 29*, No.4, 387-402.

Barba-Romero, S. and J.C. Pomerol. 1997. *Decisiones Multicriterio: Fundamentos Teóricos y Utilización Práctica*. Universidad de Alcalá, Madrid.

Eglese, R.W. 1990. "Simulated Annealing: A tool for Operational Research." *European Journal of Operational Research 46*, No.3, 271-281.

Fazlollahi, B. and R. Vahidov. 2001. "Extending the Effectiveness of Simulation-Based DSS Through Genetic Algorithms." *Information and Management 39*, No.1, 53-65.

Gama Caldas L. and L.K. Norford. 2002. "A Design Optimization Tool Based on a Genetic Algorithm." *Automation in Construction 11*, No.2, 173-184.

Glover, F. 1977. "Heuristics for Integer Programming Using Surrogate Constraints." *Decision Sciences 8*, No.7, 156-166.

Glover, F.; J.P. Kelly; and M. Laguna. 1996. "New Advances and Applications of Combining Simulation and Optimization." In *Proceedings of the 1996 Winter Simulation Conference* (Coronado, CA, Dec.8-11). IEEE, Picataway, NJ, 144-152.

Goldberg, D.E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, New York.

Glynn P.W. 1987. "Likelihood Ratio Gradient Estimation: An Overview." In *Proceedings of the 1987 Winter Simulation Conference* (Atlanta, GA, Dec.14-16). IEEE, Picataway, NJ, 366-375.

Hart, R.P.S.; M.T. Larcombe; R.A. Sherlock; and L.A. Smith. 1998. "Optimisation Techniques for a Computer Simulation of a Pastoral Dairy Farm." *Computers and Electronics in Agriculture 19*, No.2, 129-153.

Hilgers A. and B.J. Boersma. 2001. "Optimization of Turbulent Jet Mixing." *Fluid Dynamics Research 29*, No.6, 345-368.

Ho Y.C. 1984. "Likelihood Ratio Gradient Estimation: An Overview." In *Proceedings of the 1984 Winter Simulation Conference* (Dallas, TX, Nov.28-30). IEEE, Picataway, NJ, 171-173

Hooke R. and T.A. Jeeves. 1961. "A Direct Search Solution of Numerical and Statistical Problems." *Journal of the Association for Computer Machinery 8*, 21-229.

Hopfield J. and D. Tank. 1985. "Neural Computation of Decisions in Optimization Problems." *Biological Cybernetics 52*, 141-152.

Jacobson, S.H. and L.W. Schrueben. 1989. "Techniques for Simulation Response Optimization." *Operation Research Letters 8*, No.1 (Feb), 1-9.

Johnson, V.M. and L.L. Rogers. 2001. "Applying Soft Computing Methods to Improve the Computational Tractability of a Surface Simulation-Optimization Problem." *Journal of Petroleum Science and Engineering 29*, 153-175.

Karaboga, D. and A. Kalinli. 1997. "A Tabu Search Algorithm for Combinatorial Optimisation Problems." In *Proceedings of the 1997 European Simulation Multiconference* (Istanbul, Turkey, Jun.1-4). SCS.

Kiefer J. and J. Wolfowitz. 1952. "Stochastic Approximation of the Maximum of a Regression Function." *Annals of the Mathematical Statistics 23*, 462-466.

Kotz, S. and C.R. Lovelace. 1998. *Process Capability Indices in Theory and Practice*. Arnold, London.

Laguna, M. and R. Martí. 2002. "Neural Network Prediction in a System for Optimizing Simulations." *IIE Transactions 34*, No.3, 273-282.

Law, A. and W.D. Kelton. 1991. *Simulation Model and Analysis*. McGraw-Hill, New York.

Nelder J.A. and R. Mead. 1965. "A Simplex Method for Function Optimization." *Computer Journal 7*, 308-313.

Otamendi, J. 1999. "G.E.S.A.S. Methodology: A Methodology for the Generation, Evaluation and Selection of Alternatives via Simulation." *Simulation 73*, No. 2, 91-99.

Otamendi, J. 2001. *Técnicas de Simulación Avanzada para el Tratamiento de Situaciones de Riesgo*. Ph.D. Dissertation. Universidad Politécnica de Cartagena, Murcia, Spain.

Otamendi, J. 2004. "G.E.S.A.S. II: A Better Relationship Between Efficiency and Efficacy While Experimenting With Simulation Models." In Press, *Simulation*.

Rees, L.P.; E.R. Clayton; and B.W. Taylor. 1985. "Solving Multiple Response Simulation Models Using Modified Response Surface Methodology Within a Lexicographic Goal Programming Framework." *IIE Transactions 17*, No.1, 47-57.

Rivera, J., 1998. *Estudio del Proceso de Clasificación de una Empresa de Paquetería vía Simulación*. Creative Component. Universidad Carlos III de Madrid.

Saaty, T.L. 1980. *The Analytic Hierarchy Process.* McGraw-Hill, New York.

Shelokar, P.S.; V.K. Jayaraman; and B.D. Khulkarni. 2004. "An ant colony approach for clustering." *Analytica Chimica Acta 509*, No.2, 187-195.

Taha, H. 1988. *Simulation Modelling and SIMNET*. Prentice Hall, New York.

Wager, T.D. and T.E. Nichols. 2003. "Optimization of Experimental Design in fMRI: a General Framework Using a Genetic Algorithm." *NeuroImage 18*, No.2, 293-309.

Yokohama, M. 2002. "Integrated Optimization of Inventory-Distribution Systems by Random Local Search and a Genetic Algorithm." *Computers & Industrial Engineering 42*, No.2-4, 175-188.

Yu, H. and W. Liang. 2001. "Neural Network and Genetic Algorithm-Based Hybrid Approach to Expanded Job-Shop Scheduling." *Computers & Industrial Engineering 39*, No. 3-4, 337-356.

Zeleny, M., 1982. *Multiple Criteria Decision Making*, McGraw-Hill, New York.

AUTHOR BIOGRAPHIES

Javier Otamendi Fernández de la Puebla received the B.S. and M.S. degrees in Industrial Engineering at Oklahoma State University, where he developed his interests in Simulation and Total Quality Management. Back in his home country of Spain, he received a B.S. in Business Administration and a Ph.D. in Industrial Engineering. He is currently a simulation and statistics consultant and professor. His e-mail address is jotamendi_30@yahoo.com.