

Performance of a Genetic Algorithm applied to robust design in multiobjective systems under different levels of fractioning

Desempeño de un Algoritmo Genético aplicado a diseño robusto en sistemas multiobjetivo bajo diferentes niveles de fraccionamiento

Enrique Canessa-Terrazas^{1}, Héctor Allende-Olivares²*

¹ Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez. Av. Padre Hurtado 750. C.P. 2540201. Viña del Mar, Chile.

² Departamento de Informática, Universidad Técnica Federico Santa María. Av. España 1680. C.P. 2390123. Valparaíso, Chile.

(Received July 31, 2014; accepted November 19, 2014)

Abstract

This paper studies the performance of a Genetic Algorithm (GA) to find solutions to problems of robust design in multiobjective systems with many control and noise factors, representing the output vector in a single aggregation function. The results show that the GA is able to find solutions that achieve a good adjustment of the responses to their corresponding target values and with low variability, even with highly fractional experimental designs, which provide a limited number of data points to be fed to the GA. This conclusion is important for the practical application of the GA to robust design studies. Generally, such studies are carried out using scarce resources and dealing with other limitations, which force the engineer to use few experimental treatments and gather a limited amount of data. Thus, knowing that the GA performs well under such situation expands its applicability.

-----*Keywords:* Taguchi methods, parameter design, genetic algorithms, performance analysis

Resumen

Este trabajo estudia el rendimiento de un Algoritmo Genético (AG) para encontrar soluciones a problemas de diseño robusto en sistemas multiobjetivo, con muchos factores de control y ruido, representando el vector de salida

* Corresponding author: Enrique Canessa Terrazas, e-mail: ecanessa@uai.cl
DOI: 10.17533/udea.redin.n75a09

en una sola función de agregación. Los resultados muestran que el AG es capaz de encontrar soluciones que entregan un buen ajuste de la media de las respuestas a sus respectivos valores objetivo y con baja variabilidad, incluso con diseños experimentales altamente fraccionados, los cuales proveen de un número limitado de datos que se ingresan al AG. Esta conclusión es importante para la aplicación práctica del AG a estudios de diseño robusto. Generalmente, dichos estudios son llevados a cabo usando recursos escasos y lidiando con otras limitaciones, lo que obliga al ingeniero a usar pocos tratamientos y recoger una cantidad limitada de datos. Por eso, saber que el AG se comporta bien bajo esas situaciones, expande su aplicabilidad.

-----*Palabras clave:* métodos de Taguchi, diseño de parámetros, algoritmos genéticos, análisis de rendimiento

Introduction

Robust design (RD) is a methodology developed by Genichi Taguchi [1] that tries to adjust the mean of the responses of a system as close as possible to their corresponding target values and simultaneously attempts to achieve low variability in those responses, even in the presence of noise factors. To do so, Taguchi proposes a two stage methodology labeled Parameter Design (PD). First, an engineer identifies and sets values of controllable factors of the system to achieve mean adjustment and then, he/she sets other control factors to get low variability under different noise conditions [1]. The application of PD has had a profound worldwide effect on enhancing the quality of products and services in many industries [1-6]. To facilitate parameter design studies, the work presented in [7, 8] developed a tool based on Genetic Algorithms (GA), which automatically finds the levels of control factors that achieve mean adjustment and low variability in all responses. Given that other researchers have also explored the application of GA to robust design (e.g. [2, 9, 10] to name just a few), it was sensible to apply the same optimization algorithm for developing such tool. Some newer optimization algorithms might also have been applied, but there is no evidence that they could always and consistently outperform GAs [11, 12]. Moreover, the GA developed in [8] is readily generalizable to multiobjective systems with many outputs, control and noise factors, and

uses the parameter design approach, which is frequently applied in real studies of robust design [3-6]. In following Taguchi's idea that many quality problems may be tackled using simple methods [1], the GA handles multiobjective systems by collapsing the mean and variance of the outputs in a single aggregate fitness function (similar to what Taguchi does in PD by using the so called Signal to Noise Ratio (SNR) [1]). Although this approach is simple, it is a valid one and the GA has been applied to different real RD problems at many firms with good results [6, 8]. Examples range from adjusting the width of the painted strip of a car painting system to enhancing copper-mining operations [6, 8]. However, one main concern regarding the application of the GA to robust design is that an engineer will normally have limited time and economic resources to conduct the study [1, 5, 6]. Thus, generally, the engineer uses highly fractional experimental designs to gather the data for the analyses [1, 5, 6]. Hence, as the fractioning level increases, the amount of collected data decreases, which might lessen the performance of the GA. Therefore, the main focus of this paper is to analyze the ability of the GA to find good solutions to problems of robust design under different fractioning levels of the experiments used in gathering the data. In line with Taguchi's quality approach [1], we acknowledge that this work addresses rather specific type of problems, leaving out more encompassing issues, which are much harder to solve [11, 12]. However, to the best of our

knowledge, such analysis has not been elsewhere conducted and still for many developing countries, enhancing specific production and/or service processes is worthwhile doing. Finally, a secondary objective of this paper is to experiment with some parameters of the GA, which might also impact its performance, but keeping the original mechanisms implemented in the GA.

Some details of the Genetic algorithm used in robust design

This part of the paper introduces some aspects of the developed GA necessary for understanding the analysis that was performed in the present study. More details of the GA may be found in [7, 8]. The GA developed in [8] represents the combinations of k control factors that may take s different levels (values) of a robust design experiment using an integer codification. One chromosome will be composed by a combination of different levels for each factor, which corresponds to a particular treatment of the experiment. Let f_{ij} be the factor j of chromosome l , with $j = 1, 2, \dots, k$ and $l = 1, 2, \dots, N$. Each f_{ij} can take the value of a given level of the factor j , that is $1, 2, \dots, s$. One chromosome (or solution) is expressed as a row vector (see Eq. 1). The matrix representing the total population of solutions X will be composed of N chromosomes (see Eq. 2).

$$x_l = [f_{l1}, f_{l2}, \dots, f_{lk}] \quad (1)$$

$$X = [x_1, x_2, \dots, x_N]^T \quad (2)$$

Each of the chromosomes (solutions) x_l will generate a different response y of the system when the control factors are set to the corresponding levels specified in the chromosome x_l . The GA searches through the space of possible treatment combinations, finding the combinations that

minimize the variance of the responses and adjust their means as close as possible to their corresponding target values. The fitness function used to guide the GA for each response is the following (Eq. 3):

$$\text{Fitness function} = \phi(x_l) = -\{s^2(x_l) + [T - \bar{y}(x_l)]^2\} \quad (3)$$

where T is the target value for Nominal the Better (NTB) quality characteristic, and $\bar{y}(x_l)$ and $s^2(x_l)$ are the corresponding mean and variance of the solution represented by chromosome x_l .

A multiobjective system will have more than one response (in general, $r = 1, 2, \dots, R$ responses) and hence, one fitness function for each of them. Following the work presented in [13, 14], the total fitness function for multiobjective systems will be expression (4), which consists of a desirability function $D(\phi(x_l))$ and a penalty function $P_l(y(x_l))$:

$$D(x_l) = D_l(\phi(x_l)) - P_l(y(x_l)) \quad (4)$$

where $y(x_l)$ is the generic form of designating all the responses of the treatment combination x_l . Thus, the penalty function will depend on all the replications of all the responses of the system. Each element of expression (4) can be decomposed into expressions (5) and (6) according to [14]:

$$D_l(\phi(x_l)) = (d_{1l}(\phi_1(x_l)) \times d_{2l}(\phi_2(x_l)) \times \dots \times d_{Rl}(\phi_R(x_l)))^{\frac{1}{R}} \quad (5)$$

$$P_l(y(x_l)) = \left[(p_{1l}(y_1(x_l)) \times p_{2l}(y_2(x_l)) \times \dots \times p_{Rl}(y_R(x_l)))^{\frac{1}{R}} - c \right]^2 \quad (6)$$

Moreover, each element of the penalty function (6) can be expressed as expression (7) shows [14]:

$$p_{r1}(y_l(x_l)) = \begin{cases} c + \left(\frac{L_r - \bar{y}_r(x_l)}{H_r - L_r} \right), & \bar{y}_r(x_l) \leq L_r \\ c, & L_r \leq \bar{y}_r(x_l) \leq H_r \\ c + \left(\frac{\bar{y}_r(x_l) - H_r}{H_r - L_r} \right), & \bar{y}_r(x_l) \geq H_r \end{cases} \quad (7)$$

Each response y_r has a target value T_r and a lower and upper limit given by L_r and H_r respectively, in which $L_r < T_r < H_r \forall r, r = 1, 2, \dots, R$. For chromosome x_l to be feasible, the corresponding response must be within those limits ($L_r \leq \bar{y}_r(x_l) \leq H_r$). The constant c avoids p_r from becoming zero if infeasible cases arise, and thus ensures that a non-zero P is calculated for a non-feasible solution (see (6)). A value of 0.0001

is assigned to c , which does not influence the value of the final solution (see [14]).

The desirability function must also have a lower and upper limit for each of the elements $d_{rl}(\phi_r(x_l))$ (see expression (5)), called the desirability limits [14]. Expression (8) presents the desirability function, where b_r corresponds to the most desirable case and a_r to the least desirable case:

$$d_{rl}(\phi_r(x_l)) = \begin{cases} 0, & \phi_r(x_l) \leq a_r \\ \frac{\phi_r(x_l) - a_r}{b_r - a_r}, & a_r \leq \phi_r(x_l) \leq b_r \\ 1, & \phi_r(x_l) \geq b_r \end{cases} \quad (8)$$

Parameter b_r may be expressed by equation (9), using the minimum variability attained in the treatment combinations of the initial experiment performed and assuming that response r is at its target value T_r :

$$b_r = -s_{\min}^2(x_l), \forall r, l \quad (9)$$

The value of a_r corresponds to the least desirable case, where the variance of the response is a maximum and the mean of it is at any of the two limits (H_r or L_r). Thus, parameter a_r may be calculated using expression (10):

$$a_r = -\left(s_{\max}^2(x_l) + (\max\{(H_r - T_r), (T_r - L_r)\})^2\right), \forall r, l \quad (10)$$

From expressions (3) through (8), one can see that in the fitness calculation, the GA needs to know the responses corresponding to the experimental

treatment, which each chromosome represents. However, some of those treatments might not have been part of the experiment that the engineer conducted to gather the data. Thus, the GA needs to estimate those responses. For estimating the mean of a response for a non-tried chromosome (treatment), the GA calculates the main effect of each of the treatment levels on the response and a grand mean using all the observations corresponding to the experiment that was carried out. Then, the GA adds to the grand mean, the corresponding main effects of the levels indicated by the chromosome. For estimating the variance of the response for a non-tried chromosome (treatment), the GA uses a similar procedure. The GA first computes a global variance considering all the replications of all the treatments tried in the original experiment. Then the GA calculates the main effect of each control factor on the variance.

Finally, the GA sums the main effects of the levels indicated in the chromosome to the global variance. Those procedures correspond to a linear estimation usually applied in the Taguchi method [1]; for a worked out numerical calculation, see e.g. [1] page 16. In the search process, the GA uses roulette selection, a uniform crossover and a bit by bit (factor by factor) mutation operator. The GA employs a stopping criterion of reaching 3000 chromosomes in the population, a crossover probability of 0.3 and a mutation probability of 0.05.

Experimental design for the GA's performance analysis

The main objective of this study was to evaluate the performance of the GA under different fractioning levels of the experimental design used to gather the data. However, one can also change the value of other parameters of the GA that were kept constant in previous studies [7, 8]. That entails identifying the most relevant factors to be manipulated, and the experimental design and performance measures to be used.

Identification of experimental factors

A previous analysis suggested that the performance of the GA may be influenced by the experimental design that the robust design study uses [7]. Specifically, the level of fractioning in such design will determine the number of data points that the GA will have for beginning the search process. Since a GA's long-term performance is also influenced by population size [15, 16], which depends on its initial size, one may conclude that the level of fractioning should impact the GA's ability to find good solutions. Moreover, remember that the GA developed for robust design needs to estimate missing values and thus, the number of missing values that the GA will need to estimate depends on the number of data points fed to the GA. Since those missing values are only approximate estimates of the true unknown ones, it is reasonable to think that the more missing values the GA has to estimate, the

lower its performance will be. Thus, the higher the fractioning level of the experimental design, the larger the number of missing values that the GA will need to estimate and, the poorer its performance may be. The level of fractioning is also an important parameter to be considered, since in robust design, the experimenter generally tries to use highly fractional designs to lower the cost of the experiments [1, 5, 6]. Thus, from a practical point of view, assessing how well the GA behaves under different levels of fractioning will set a limit on the experimental designs that may be used with the GA in robust design studies. With regard to other parameters of the GA, a revision of relevant literature shows that there exist many studies that consistently find that the crossover probability (p_c) and the mutation probability (p_m) have an important effect on a GA's performance [15-21]. Thus, it is sensible to include both parameters in the experiment. Additionally, the same above-mentioned studies found that values of parameters associated to the calculation of the fitness function were also relevant, because they allow to more finely or coarsely discriminate among solutions. In the present study, since the desirability limits a_r and b_r play an important role in the calculation of the value of the desirability functions, the experiment should consider them.

Finally, one should note that there exist other parameters and mechanisms that might influence the performance of the GA [15-19, 21], but some of them are not applicable to the present situation (e.g. size of generation gap) and other would substantially alter the GA (e.g. parent selection strategy, cross-over and mutation mechanisms), which if changed, would not permit the comparison with previous results.

Experimental design and levels of factors

Having identified the four factors that will be used in the experiments, the study needs to establish the experimental design and the levels of the factors. To be able to appreciate whether non-linearity exists in the effects, the experimental design will

correspond to an Orthogonal Array (OA) $L_9(3^4)$. This OA allows analyzing main effects using a convenient number of experimental conditions. However, it omits interaction effects. Given that in preliminary experiments with the GA, no interaction effects were found, it is more important to focus the analysis on non-linearity of main effects. The OA $L_9(3^4)$ specifies three levels for each factor. In order to be able to use previous studies [7, 8] as a base scenario, one of the levels of the identified factors will be the same values used in those research papers. Regarding the fractioning levels, in [7, 8], the authors applied an inner OA $L_{64}(4^{10})$ for the 10 control factors and an outer OA $L_{16}(4^5)$ for the 5 noise factors. Thus, the product array has 64 times 16, 1024 design points. This study will keep the outer OA, since robust design literature suggests that a small number of noise factors and levels for them be considered [1, 5, 6]. Regarding the inner array, a $L_{48}(4^{10})$ and a $L_{80}(4^{10})$ were selected as alternatives for the present analysis. Using those OAs, each experiment will have 768 treatment combinations ($L_{48}(4^{10}) \times L_{16}(4^5)$), 1024 combinations ($L_{64}(4^{10}) \times L_{16}(4^5)$) and 1280 combinations ($L_{80}(4^{10}) \times L_{16}(4^5)$). That implies a percentage difference in data points among experimental designs of - 40% when comparing the extreme cases (1280 with 768), - 20% (1280 with 1024) and - 25% (1024 with 768). Note that we selected OA's with high fractioning levels, given that in robust design that is a common practice [5, 6]. As robust design interventions are normally carried out in a short time, due to restrictions in the downtime of processes and economic resources to conduct experiments, only a limited number of treatments is tried [5, 6]. Though using lower levels of fractioning in this work might have better allowed exposing the effect of fractioning level on the performance of the GA, that would have been irrelevant from a practical viewpoint, i.e. we would have used low fractioning levels that are almost never used in robust design interventions.

In the case of crossover and mutation probability, the study will employ the values used in previous

analyses [7, 8], which are $p_c = 0.3$ and $p_T = 0.05$. The other two levels will correspond to values similar to the ones used in previous studies [15, 22], $p_c = 0.9$ and 0.6 , and $p_T = 0.01$ and 0.001 . For parameters a_r and b_r , the analysis will consider a value equal to the one used in [7, 8], and the other two levels will be equal to 0.8 and 1.2 times a_r and b_r .

Responses to be measured

In previous studies of the GA, the analysis has used different case studies. One of such case studies involved the optimization of the four responses of a system simulator independently from each other. Thus, the GA was dealing with four different single objective systems, each with ten control factors and four noise factors. Additionally, a second case study used the same simulator, but the GA simultaneously optimized the four responses, which means that the GA was handling a four-dimensional multiobjective system. The present study will use the same two case studies. More details of the simulator and case studies may be found in [7, 8]. Note that these simulated systems are representative of real processes subject to robust design studies, and thus our conclusions are applicable to such situations [3-6]. For the single objective systems, the experiment will measure the desirability function of each response $D((x_j))$ for the chromosomes delivered by the GA (see Eq. (4)), since that function aggregates both mean adjustment and variance reduction. Additionally, the analyses will consider the value of the mean of the response and its corresponding adjustment to its target value and the standard deviation of the response. In the case of the multiobjective system, the study will analyze the overall aggregated desirability for the four responses (see eqs. (4), (5) and (6)), and also mean adjustment and standard deviation for each of the four responses. Table 1 shows the experimental design. Each treatment combination will be run 30 times in the GA, so that the ANOVA may achieve a statistical power of around 80% to 85%, which was calculated using the table provided in [23], and data collected in previous work.

Table 1 Experimental design used in the performance analysis of the GA

Experimental design: OA $L_9(3^4)$ with 30 replications			
Factor A: mutation probability (p_r)	Levels: $a_1 = 0.001$	$a_2 = 0.01$	$a_3 = 0.05$
Factor B: crossover probability (p_c)	Levels: $b_1 = 0.3$	$b_2 = 0.6$	$b_3 = 0.9$
Factor C: desirability limits a_r and b_r	Levels: $c_1 = 0.8 a_r, b_r$	$c_2 = 1.0 a_r, b_r$	$c_3 = 1.2 a_r, b_r$
Factor D: level of fractioning of inner array	Levels: $d_1 = L_{48}$	$d_2 = L_{64}$	$d_3 = L_{80}$
Response: $D(x_i), \bar{y}(x_i), \frac{\bar{y}(x_i) - T}{T}, s(x_i)$			

Experimental results and analysis

Given that this study needs to conduct controlled experiments with specific levels of fractioning, to test the GA a simulator was built, which is described in detail in [7, 8]. This simulator has four responses, ten control factors and four noise factors and interactions among the factors. Thus, compared with real studies of RD, this simulator represents difficult to optimize systems [1, 4-6]. It should also be noted that we have applied the GA to solve RD problems in real systems, ranging from adjusting the width of the painted strip of a car painting system to enhancing copper-mining operations [6-8], which lends credibility to the application of the GA to real systems.

Results obtained for the single objective systems

For the following case studies, the four responses of the simulator are optimized independent from each other, so that the GA deals with four single objective systems. For each experimental treatment shown on Table 1, the algorithm was run 30 times. Table 2 presents the results for response one of the single objective system. This table shows the experimental design corresponding to OA $L_9(3^4)$, along with the average value of the desirability function and the average value of response y_i , the average percentage difference of y_i with regard to its corresponding target value ($T = 200$) and the average standard deviation of y_i . These averages were calculated using the results

of the 30 runs for each experimental treatment and considering the best five solutions delivered by the GA in each run. Additionally, Table 2 presents the statistically significant effects of the factors on those outputs. The effects were calculated using the following contrasts: effect $F1$ = (average of the corresponding response for level 3 of factor F) - (average of the corresponding response for level 1 of factor F), and effect $F2$ = (average of the corresponding response for level 2 of factor F) - (average of the corresponding response for level 1 of factor F). For example, for the effect of factor C on the average desirability function (DF) of y_i : $C1 = (.95 + .97 + .97) / 3 - (.91 + .90 + .92) / 3 = 0.0533$, and $C2 = (.95 + .95 + .94) / 3 - (.91 + .90 + .92) / 3 = 0.0367$. Note that in the case of the effects on average percentage difference with regard to the target value, these were calculated using the absolute value of such percentages. That was done, since the percentage difference may be positive or negative, thus the positive values might be cancelled out by the negative ones. The results presented on Table 2 indicate that there are no large differences in the performance of the GA under the nine different treatments. Only effects $C1$ and $C2$ on the value of the desirability function (DF) are partially statistically significant. Figure 1 (a) also shows that factor C has the most important impact on DF. Remember that level $C1$ corresponds to $0.8 a_r$ and b_r , $C2$ is equal to a_r and b_r , and $C3$ represents $1.2 a_r$ and b_r . Thus, the larger the values of a_r and b_r , the less strict the calculation of DF will be. Hence, from a practical point of view, although DF might be larger, that

does not mean that the adjustment of the response to its target value will necessarily be tighter and/or that the standard deviation of the response will be smaller. Figure 1 (a) also suggests that

factor D (level of fractioning) may affect DF and that its effect may be non-linear. However, the corresponding ANOVA does not confirm the visual conclusion (for effect D , p-value = 0.70).

Table 2 Results for the single objective system response one

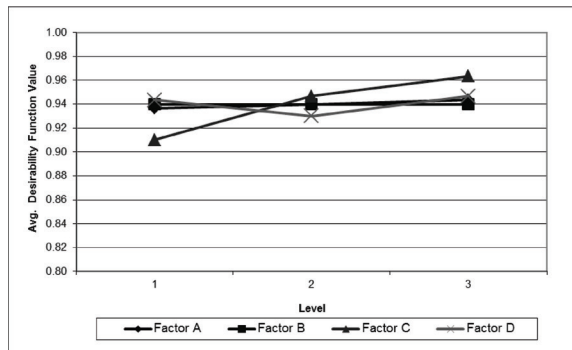
Treatment	Combination $L_9(3^4)$				Y1 response ($T = 200$)			
	Level of Factor				Avg. DF	Avg. y_1	Avg. difference (%)	Avg. std. deviation
	A	B	C	D				
1	1	1	1	1	0.91	205.1	2.54	2.818
2	1	2	3	2	0.95	193.5	-3.24	2.822
3	1	3	2	3	0.95	199.5	-0.25	2.788
4	2	1	3	3	0.97	199.5	-0.25	2.788
5	2	2	2	1	0.95	205.1	2.54	2.818
6	2	3	1	2	0.90	193.5	-3.24	2.822
7	3	1	2	2	0.94	193.5	-3.24	2.822
8	3	2	1	3	0.92	198.5	-0.73	2.770
9	3	3	3	1	0.97	199.3	-0.34	2.984
Statistically significant Main Effects					Avg. DF	Avg. y_1	Avg. diff. (%)	Avg. std. dev.
C1					0.0533*	-1.6000	-0.8945	0.0615
C2					0.0367*	0.3333	-0.1615	0.0060

* = partially statistically significant at the 0.067 level

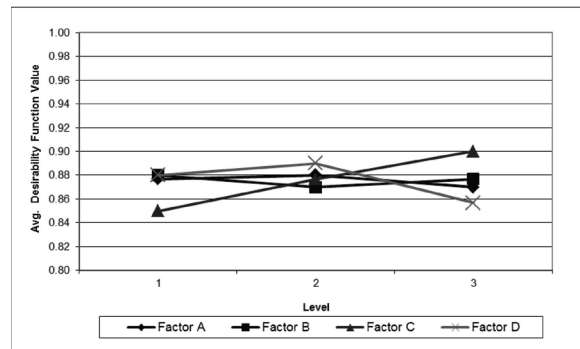
Table 3 shows the same figures, but for response two of the single objective system. The treatment combinations are not repeated in this table, since they are the same already presented in Table 2. For response two, Table 3 indicates that factor D significantly impacts the average standard deviation attained by response two. In general, for level d_3 (L_{80}) the GA obtains a larger average standard deviation of response two than for level d_1 (L_{48}). This means that the GA obtains a

good variance reduction even for high levels of fractioning, which is a positive aspect for robust design studies.

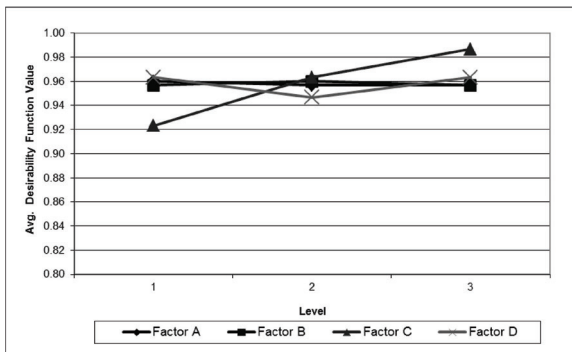
Turning to Figure 1 (b), though the lines representing factors C and D might indicate that the effects of those two factors on DF may be important, the ANOVA indicates that the corresponding effects on DF are not statistically significant (p-value = 0.70).



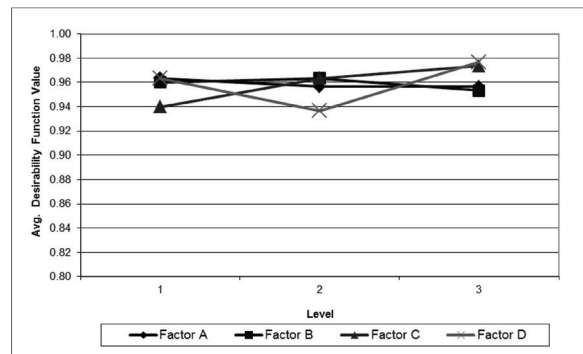
(a)



(b)



(c)



(d)

Figure 1 Average desirability function value (DF) vs. levels of experimental factors for single objective systems: (a) y_1 (b) y_2 (c) y_3 (d) y_4

Table 3 Results for the single objective system response two

Treatment	Y2 response (T = 50)			
	Avg. DF	Avg. y_2	Avg. difference (%)	Avg. std. deviation
1	0.86	51.70	3.39	1.674
2	0.91	50.65	1.30	1.770
3	0.86	54.35	8.70	2.271
4	0.89	54.35	8.70	2.271
5	0.88	51.70	3.39	1.674
6	0.87	50.65	1.30	1.770
7	0.89	50.65	1.30	1.770
8	0.82	54.35	8.70	2.271
9	0.90	51.70	3.39	1.674
Stat. sig. Main Effects	Avg. DF	Avg. y_2	Avg. diff. (%)	Avg. std. dev.
D1	-0.0198	2.6500	5.310	0.5968*

* = statistically significant at the 0.001 level

Table 4 shows the results for response three, which reveals that factor *C* has a significant effect on DF and factor *D* a partially significant effect on the average standard deviation. Although the effects of those two factors on the average response y_3 seem to be also important, the ANOVA shows the contrary, reaching p-values equal to 0.438 and 0.110 respectively. The effects of the factors on DF are visually shown

on Figure 1 (c). Inspection of Figure 1 (c) shows the significant effect of factor *C* on DF. As with response one, this effect is attributed to the looser desirability bounds that those larger values of a_r and b_r bring about. Regarding the effect of factor *D* on the average standard deviation, it can be concluded that higher levels of fractioning may not be detrimental to variance reduction.

Table 4 Results for the single objective system response three

<i>Treatment</i>	<i>Y3 response (T = 1000)</i>			
	<i>Avg. DF</i>	<i>Avg. y_3</i>	<i>Avg. difference (%)</i>	<i>Avg. std. deviation</i>
1	0.93	998.47	-0.15	36.05
2	0.98	967.00	-3.30	40.67
3	0.97	998.82	-0.12	38.61
4	0.99	998.82	-0.12	38.61
5	0.97	1076.47	7.65	37.75
6	0.91	967.00	-3.30	40.70
7	0.95	943.70	-5.63	40.70
8	0.93	998.82	-0.12	38.61
9	0.99	1024.99	2.50	39.01
<i>Stat. sig. Main Effects</i>	<i>Avg. DF</i>	<i>Avg. y_3</i>	<i>Avg. diff. (%)</i>	<i>Avg. std. dev.</i>
<i>C1</i>	0.0637**	8.8400	0.78	0.9786
<i>C2</i>	0.0416**	18.2333	3.27	0.5670
<i>D1</i>	-0.0008	-34.4900	-3.31	1.0038*
<i>D2</i>	-0.0209	-74.0767	0.64	3.0870*

* = partially statistically significant at the 0.093 level

** = statistically significant at the 0.001 level

Finally, Table 5 and Figure 1 (d) present the averages and statistically significant effects for response four. Table 5 shows that for response y_4 , the effects of factor *D* on all of the outputs are statistically significant. For factor *C*, the effect *C1* on DF is partially significant and *C2* is significant. As before, a larger value of a_r and b_r augments DF. This can be clearly seen on Figure

1 (d). Regarding factor *D*, Figure 1 (d) visually corroborates that the effect *D1* on DF is positive, which means that comparing level d_3 with d_1 , DF increases as the level of fractioning decreases. However, effect *D2* on DF indicates that the contrary happens for levels d_2 and d_1 . A negative effect *D2* implies that DF diminishes as the level of fractioning decreases.

Table 5 Results for the single objective system response four

<i>Treatment</i>	<i>Y4 response (T = 500)</i>			
	<i>Avg. DF</i>	<i>Avg. y_4</i>	<i>Avg. difference (%)</i>	<i>Avg. std. deviation</i>
1	0.95	427.6	14.47	31.17
2	0.96	567.7	13.54	36.08
3	0.98	461.9	7.62	28.01
4	0.99	461.9	7.62	28.01
5	0.97	427.6	14.47	31.17
6	0.91	567.7	13.54	36.08
7	0.94	567.7	13.54	36.08
8	0.96	461.9	7.62	28.01
9	0.97	423.0	15.40	31.68

<i>Stat. sig. Main Effects</i>	<i>Avg. DF</i>	<i>Avg. y_4</i>	<i>Avg. diff. (%)</i>	<i>Avg. std. dev.</i>
C1	0.0367*	-1.533	0.31	0.1721
C2	0.0257**	0.000	0.00	0.0000
D1	0.0148**	35.833 [†]	-7.16 [†]	-3.3269 [†]
D2	-0.0274**	141.63 [†]	-1.24 [†]	4.7418 [†]

* = partially statistically significant at the 0.070 level

** = statistically significant at the 0.05 level

† = statistically significant at least at the 0.01 level

Thus, the effect of the level of fractioning on DF seems to be highly non-linear. The effects of factor *D* on the average response y_4 and mean adjustment imply that as the level of fractioning decreases, mean adjustment gets tighter. This conclusion seems intuitively correct. However, since it is not the same for each of the four responses, it must be treated with caution. Regarding variance reduction, the effect *D1* is negative, which at first might imply that the smaller the fractioning level, the smaller the standard deviation that response y_4 gets. Nevertheless, a positive effect *D2* shows that the effect of factor *D* on standard deviation is highly non-linear. The standard deviation of response y_4 increases from level d_1 to d_2 and then decreases from level d_2 to d_3 . Thus, one cannot say that a higher level of fractioning is necessarily detrimental to variance reduction.

In summary, for the four single objective systems as a whole, the previous analyses show that there

are no big effects of crossover and mutation probability on the performance of the GA. On the other hand, the desirability limits impact DF. As a_r and b_r increase, DF also augments. The other factor that influences the value of DF is the level of fractioning of the experimental design. However, there is no clear conclusion regarding whether high levels of fractioning necessarily decrease DF. This is a positive conclusion, since it indicates that the performance of the GA is not strongly influenced by high levels of fractioning.

Results obtained for the multiobjective system

This case study used the same simulator and the same experimental design as before, but the GA optimized the four responses at the same time. This means that the GA is optimizing a four-dimensional multiobjective system. As in the

previous analysis, 30 runs were carried out for each of the nine experimental treatments. Table 6 shows the figures for the multiobjective system, where only the statistically significant effects are displayed. The results shown in Table 6 are almost the same as the ones shown on the tables corresponding to the single-objective systems. However, Table 6 presents the average DF, which corresponds to the aggregation of the DF's of each of the four responses of the multiobjective system (see expression (4)). Additionally, the average response, percentage difference with regard to the corresponding target value and standard deviation is presented for each response. The same is done for the corresponding effects.

Here again, those averages are computed using the 30 runs for each treatment and taking into account the best five solutions delivered by the GA in each run. A first point to note is that the DF's are generally smaller than the ones of the one-objective systems (to easily appreciate that, compare the levels of DF on Figure 1 with those on Figure 2). This is expected, since the GA has a tougher task when trying to simultaneously adjust the mean of the four responses and decrease their variances. Regarding the effects of the factors on DF, only the effect *C* is significant. As with the single objective systems, the larger the desirability limits, the larger DF becomes. This can be clearly seen on Figure 2.

Table 6 Results for the multiobjective system

Treatment	Avg. DF	Avg. response				Avg. difference (%)				Avg. std. deviation			
	All resps.	y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
1	0.749	164.5	48.13	1058.2	555.8	-17.75	-3.73	5.820	11.15	3.308	0.810	36.60	32.16
2	0.852	177.3	44.90	1096	697.3	-11.36	-10.20	9.643	39.45	2.806	0.820	40.26	31.32
3	0.847	206.0	43.34	922.8	674.8	2.98	-13.32	-7.720	34.96	2.100	0.858	38.52	28.00
4	0.881	209.9	43.16	924.4	618.8	4.97	-13.69	-7.564	23.76	2.268	0.843	39.05	28.80
5	0.820	164.5	48.13	1058	555.8	-17.75	-3.73	5.820	11.15	3.308	0.810	36.60	32.16
6	0.752	255.0	47.17	968.6	508.4	27.5	-5.66	-3.143	1.67	2.582	0.873	38.42	31.27
7	0.809	230.3	41.21	1059	475.6	15.14	-17.59	5.979	-4.89	2.530	0.935	37.17	30.08
8	0.775	198.3	41.59	943	543.4	-0.86	-16.81	-5.696	8.67	2.232	0.847	37.58	31.30
9	0.860	170.5	47.72	1019	454.2	-14.73	-4.55	1.900	-9.16	2.879	0.940	39.37	29.37

Stat. sig. Main Effects	Avg. DF	Avg. response				Avg. diff. (%)				Avg. std. dev.			
	All resps.	y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4	y_1	y_2	y_3	y_4
C1	0.1056*	-20.01	-0.374	23.33	54.27	-5.017	-0.748	2.333	10.85	-0.056	0.024	2.027	-1.747
C2	0.0669*	-5.68	-1.405	23.66	32.88	-3.413	-2.809	2.366	6.577	-0.061	0.024	-0.104	-1.495
D1	0.0247	38.21	-5.30*	-115*	90.42	-13.81	-10.6*	-11.5*	18.08	-0.97*	-0.004	0.857	-1.869
D2	-0.0053	54.34	-3.57*	-3.53*	38.48	1.257	-7.15*	-0.35*	7.697	-0.53*	0.023	1.094	-0.346

* = statistically significant at least at the 0.05 level

Table 6 and Figure 2 also show that the effects *D1* and *D2* on the average response of y_2 and y_3 , and their corresponding effects on percentage difference with regard to their target values are significant. In both cases, the effects are negative, which implies that a larger level of fractioning entails a tighter mean adjustment. However, note

that this conclusion is not consistent among all of the four responses. Regarding standard deviation, those two effects are significant for response y_1 . Since effects *D1* and *D2* on the average standard deviation of y_1 are negative, it may be concluded that the smaller the level of fractioning, the smaller the standard deviation of y_1 will be.

Here again, this conclusion must be treated with caution given that it is not consistent among the four responses.

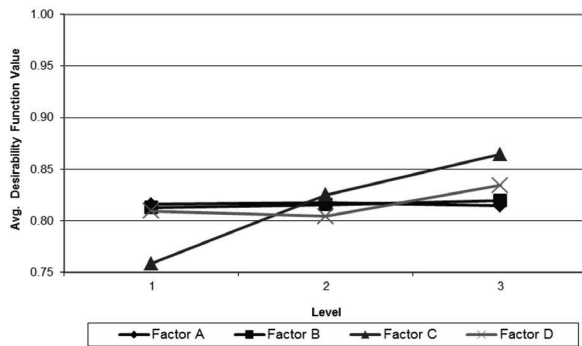


Figure 2 Average desirability function value (DF) vs. levels of experimental factors for the responses of the multiobjective system

In summary, the previous analyses allow to reach almost the same conclusions that were already mentioned for the single objective systems. Crossover and mutation probability do not significantly influence the performance of the GA and the desirability limits a_r and b_r affect DF. Higher values for a_r and b_r tend to increase DF. Contrary to what happened with the single objective systems, the level of fractioning impacts the average of some of the responses of the multiobjective system and their corresponding standard deviations. In general, a lower level of fractioning entails a tighter mean adjustment and smaller variance. However, those effects are not consistently found for all the four responses, and thus must be treated with caution.

Effect of crossover and mutation probabilities

Regarding the effect of the crossover and mutation probability on the performance of the GA, the results show that they do not significantly affect the capability of the GA to find the best solutions. However, it should be noted that the analysis of all of the solutions found by the GA, not only the five best ones, reveals that p_c and p_T might influence the diversity of about 20% of

the worst solutions found by the GA. From all the solutions delivered by the GA, 80% were consistently found in all 30 runs and correspond to the best ones, i.e. those that achieve a good mean adjustment and low variation for the responses. The other 20% are relatively worse solutions compared with the other 80%. For that 20% of worst solutions, an increase in p_c and p_T augmented the diversity among them, i.e. the GA found slightly different combinations of treatments among runs. That result is expected and it is consistent with previous findings [15, 16], given that a higher p_c and p_T generally increases the searched volume of the parameter space. Nevertheless, since the engineer is interested only in the best solutions and all of them were consistently found by the GA, that result is unimportant from a practical point of view. Finally, it should be noted that the average computational time it took the GA to reach the stopping condition was 41.5 [s] with a standard deviation of 3.7 [s], using Matlab v. 6.5.0.180913 and an HP PC with Intel Core i5-2500S CPU @ 2.70GHz, 3.2 GB RAM, running under OS MS Windows 7 Enterprise, v.6.1.7601, SP 1.

Conclusions

The most important result of the performance analysis suggests that high levels of fractioning of the experimental designs used to gather the data for the robust design study do not strongly and detrimentally influence the capability of the GA to find good solutions. In the experimental treatments, the level of fractioning or correspondingly, the number of data points fed to the GA were 768 ($L_{48}(4^{10}) \times L_{16}(4^5)$), 1024 ($L_{64}(4^{10}) \times L_{16}(4^5)$) and 1280 ($L_{80}(4^{10}) \times L_{16}(4^5)$). The percentage difference among those numbers are - 40% when comparing the extreme cases (1280 with 768), - 20% (1280 with 1024) and - 25% (1024 with 768). Thus, it may be concluded that significantly changing the levels of fractioning (within high levels) does not negatively impact the performance of the GA. Though this result may seem rather counterintuitive, it is plausible given that the experimental designs were all

highly fractionated to begin with. Thus, decreasing the original level of fractioning by even 40% still entails a high level of fractioning. We must note that the use of low fractioning levels might have allowed us to better assess the impact of fractioning on the performance of the GA. However, as already explained in Experimental design and levels of factors, we did not use low fractioning levels, given that in robust design experiments those designs are almost never used [5, 6].

The practical implication of the above discussion is that an engineer can apply the GA to robust design studies, even when using highly fractional designs. This is an important aspect, since engineers generally face low budget studies and other constraints, such as short periods of time to conduct experiments, need to keep the production and/or service process running when executing the study and short downtime of processes. Hence, generally engineers require running rather small experiments and gathering a small amount of data, which entail using highly fractional designs, like the ones used in this study [5, 6]. Also, the short computational time it takes the GA to find the solutions, makes its application highly practical. It should be noted that the claim regarding the practical application of the GA-based tool applies only to this type of problem, i.e. robust design studies, but there are no guarantees that it will work in all classes of quality improvement studies. However, given that we have applied the GA to a diverse range of real life systems with good results [6-8] and that the simulator used in testing the GA represents difficult to optimize systems [1, 3-6], we consider that the claim is not overly optimistic. In addition, based on the above, this work shows that it is possible to reduce the gap between theory and practice in some evolutionary computation algorithms, such as recommended in [12]. We are aware that we are solving local and specific quality problems and more holistic approaches are still needed to tackle more important and challenging problems [11, 12], but for many developing countries, such restricted efforts are still worth pursuing.

Regarding the value of the desirability limits, a_r and b_r , the analysis shows that they impact the value of the DF. However, this result does not necessarily imply that the solutions delivered by the GA will be better in terms of achieving a tighter mean adjustment and/or a smaller variation in the responses of the system. Thus, varying the desirability limits does not have an important practical application.

Additionally, two other conclusions that may be drawn from the experiments are similar to the ones already found in previous analyses of the GA. In general, the performance of the GA is better for single objective than for multiobjective systems and the best solutions are consistently found by the GA in the runs that were executed [7, 8].

Finally, given that the GA has worked rather well in its application to real and simulated systems, despite its simplicity to treat multiobjective systems, a Pareto GA tool is currently under development and testing. This new algorithm may better reveal the trade-off between mean adjustment and variance reduction, and thus allow a more sensible selection of the levels of control factors to be implemented in the system.

Acknowledgements

The authors are grateful to Gonzalo Ureta for his valuable work in preparing this paper.

References

1. G. Taguchi. *Systems of experimental design*. 4th ed. Ed. American Supplier Institute. Dearborn, USA. 1991. pp. 16-511.
2. A. Hajiloo, N. Nariman, A. Moeini. "Pareto optimal robust design of fractional-order PID controllers for systems with probabilistic uncertainties". *Mechatronics*. Vol. 22. 2012. pp. 788-801.
3. K. Ballantyne, R. Oorschot, R. Mitchell. "Reduce optimisation time and effort: Taguchi experimental design methods". *Forensic Science International: Genetics Supplement Series*. Vol. 1. 2008. pp. 7-8.

4. S. Maghsoodloo, G. Ozdemir, V. Jordan, C. Huang. "Strengths and limitations of Taguchi's contributions to quality, manufacturing, and process engineering". *Journal of Manufacturing Systems*. Vol. 23. 2004. pp. 73-126.
5. R. Roy. *Design of Experiments Using the Taguchi Approach*. 1st ed. Ed. J. Wiley & Sons. New York, USA. 2001. pp. 8-513.
6. H. Allende, E. Canessa, J. Galbiati. *Diseño de Experimentos Industriales*. 1st ed. Ed. Universidad Técnica Federico Santa María. Valparaíso, Chile. 2005. pp. 27-200.
7. H. Allende, D. Bravo, E. Canessa. "Robust design in multivariate systems using genetic algorithms". *Quality & Quantity Journal*. Vol. 44. 2010. pp. 315-332.
8. E. Canessa, C. Droop, H. Allende. "An improved genetic algorithm for robust design in multivariate systems". *Quality & Quantity Journal*. Vol. 42. 2011. pp. 665-678.
9. A. Jamali, A. Hajiloo, N. Nariman. "Reliability-based robust Pareto design of linear state feedback controllers using a multi-objective uniform-diversity genetic algorithm (MUGA)". *Expert Systems with Applications*. Vol. 37. 2010. pp. 401-413.
10. O. Köksoy, T. Yalcinoz. "Robust design using Pareto type optimization: a genetic algorithm with arithmetic crossover". *Computers & Industrial Engineering*. Vol. 55. 2008. pp. 208-218.
11. K. Sörensen. "Metaheuristics—the metaphor exposed". *International Transactions in Operational Research*. Vol. 22. 2015. pp. 3-18.
12. Z. Michalewicz. *Quo vadis, evolutionary computation?* Proceedings of the 2012 World Congress Conference on Advances in Computational Intelligence. Berlin, Germany. 2012. pp. 98-121.
13. E. Castillo, D. Montgomery, D. McCarville. "Modified desirability functions for multiple response optimization". *Journal of Quality Technology*. Vol. 28. 1996. pp. 337-345.
14. F. Ortiz, J. Simpson, J. Pignatiello, A. Heredia. "A Genetic Algorithm Approach to Multiple – Response Optimization". *Journal of Quality Technology*. Vol. 36. 2004. pp. 432-449.
15. K. Farhad, D. Hassani. "The effects of parameter settings on the performance of genetic algorithm through experimental design and statistical analysis". *Advanced Materials Research*. Vol. 433-440. 2012. pp. 5994-5999.
16. H. Marziyeh, B. Hossein, K. Farhad. "Evaluating the effects of parameters setting on the performance of genetic algorithm using regression modeling and statistical analysis". *Journal of Industrial Engineering*. Vol. 45. 2011. pp. 61-68.
17. O. Abdul, M. Munetomo, K. Akama. "An adaptive parameter binary-real coded genetic algorithm for constraint optimization problems: Performance analysis and estimation of optimal control parameters". *Information Sciences*. Vol. 233. 2013. pp. 54-86.
18. A. Eiben, S. Smit. "Parameter tuning for configuring and analyzing evolutionary algorithms". *Swarm and Evolutionary Computation*. Vol. 1. 2011. pp. 19-31.
19. M. Kaya. "The effects of two new crossover operators on genetic algorithm performance". *Applied Soft Computing*. Vol. 11. 2011. pp. 881-890.
20. S. Smit, A. Eiben. *Parameter tuning of evolutionary algorithms: generalist vs. specialist*. Proceedings of the International Conference on Applications of Evolutionary Computation. Berlin, Germany. 2010. pp. 542-551.
21. I. Falco, A. Della, E. Tarantino. "Mutation-based genetic algorithm: performance evaluation". *Applied Soft Computing*. Vol. 1. 2002. pp. 285-299.
22. P. Pongcharoen, D. Stewardson, C. Hicks, P. Braiden. "Applying designed experiments to optimize the performance of genetic algorithms used for scheduling complex products in the capital goods industry". *Applied Statistics*. Vol. 28. 2001. pp. 441-455.
23. K. Hinkelmann, O. Kempthorne. *Design and analysis of experiments*. 1st ed. Ed. John Wiley & Sons. New York, USA. 1994. p. 177-181.