

Integrated Greedy-Tabu-Search Method for the Economic Statistical Design of Shewhart Charts

Santiago-Omar Caballero-Morales

Technological University of the Mixteca, Postgraduate Division
Huajuapán de León, Oaxaca, Mexico
scaballero@mixteco.utm.mx

Abstract. Shewhart charts, also known as control charts, are important Statistical Process Control (SPC) techniques used for prompt detection of failures in a manufacturing process and minimization of production costs. Techniques have been used to find the chart's parameters that best comply with economic and statistical requirements. In this paper a method that integrates a Greedy and a Tabu-Search (TS) algorithm is presented to estimate these parameters for \bar{X} control charts which are used for controlling process mean. When tested with different cost models and statistical restrictions with general failure distribution, and constant and variable sampling intervals, the performance of the proposed method was more accurate when compared to TS alone and a Genetic Algorithm.

Keywords: Control Charts, Economic Statistical Design, Greedy Algorithms, Tabu-Search

1 Introduction

Shewhart or control charts are tools used to determine whether or not a production process is in a state of statistical control, and thus, the entities being produced are within quality requirements. These requirements are set by Upper and Lower Control Limits (UCL, LCL). If the quality attribute or feature of sampled entities is not within these limits, then the process is in an out-of-control state (non-conforming entities are being produced). In this case is necessary to find and correct the assignable cause that originated this state (e.g., a failure) [9]. As presented in Figure 1 a control chart requires three main parameters: n , the size of the sample; h , the length of the sampling interval between samples; and k , the coefficient of the control limits [9]. These parameters are selected based on economic and statistical restrictions because there are costs and times associated to sampling and searching of assignable causes.

The Economic Statistical Design (ESD) [2,4,13,14] of control charts is the approach developed to design control charts considering the interactions between the chart's parameters and the time and money costs associated to the sampling

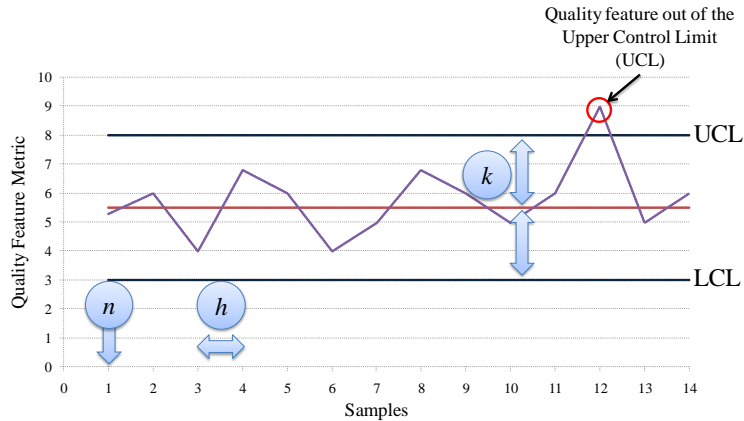


Fig. 1. General Control Chart.

and searching/repairing of an assignable (or multiple) causes of failure. The ESD considers the statistical requirements set by the error probabilities α (detecting an out-of-control state when the process is fine) and β (not detecting an out-of-control state when the process is not fine). Finding the most suitable values of n , h , and k considering the economic and statistical restrictions is not an easy task. This is because of the number of variables and the complexity of the cost function. The ESD of control charts is considered a problem for combinatorial optimization [8].

Among the techniques to find a suitable solution for the ESD of control charts, Hooke and Jeeves (HJ) [11], Genetic Algorithms (GAs) [2,4,13,8,7], Tabu Search (TS) [1] and Combinatorial Methods (CB) [14] can be mentioned. An advantage of TS is that hybridization (integration with other meta-heuristics) is possible to improve performance on combinatorial problems [10].

In this paper a hybrid TS metaheuristic is presented for the ESD of \bar{X} control charts. This method is presented with the following characteristics: (1) integration of TS with a Greedy algorithm for initialization of the search space; (2) diversification of the search space for the TS algorithm performed with random uniform moves. Two case studies were considered to validate the performance of the integrated method defined as G-TS (Greedy-TS). Significant improvements were obtained with the G-TS method across different cost functions, restrictions, and settings when compared to TS alone and a GA.

2 Integrated Approach

The proposed approach integrates two algorithms as presented in Figure 2. It consists of a “Greedy” algorithm which generates a set of random solutions

that comply with the economic and statistical restrictions of the cost model of the control chart. With this algorithm an initial solution is produced for the Tabu-Search (TS) algorithm which performs improved search to find a better solution. The details of both algorithms are presented in the following sections.

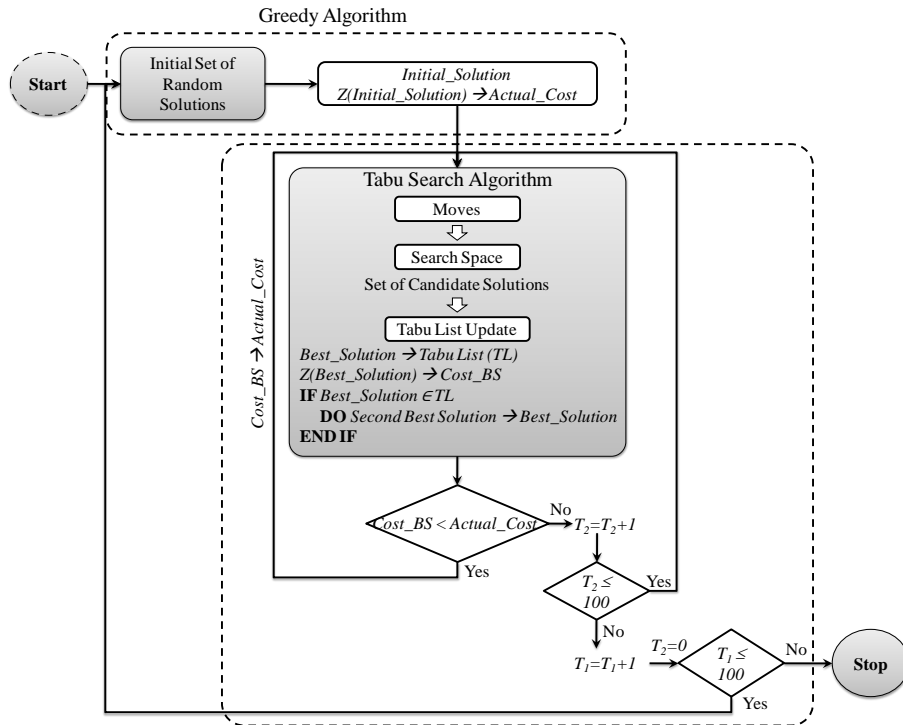


Fig. 2. Structure of the Integrated Greedy-TS (G-TS) Metaheuristic.

2.1 Greedy Algorithm

A greedy algorithm follows the problem solving heuristic of making the locally optimal choice at each stage with the hope of finding a global optimum. Although a greedy strategy does not in general produce an optimal solution, it may lead to locally optimal solutions that approximate a global optimal solution in a reasonable time [3]. In Figure 3 the structure of the greedy algorithm is presented where the following variables are initialized:

- *Total_Solutions* = 0, this is the counter for the total number of solutions (up to 100) in the “Initial Set of Random Solutions” (see Figure 2).

- $Best_Cost = Inf$, this is the reference for the best cost found on each iteration of the greedy algorithm. Initially, this value is set equal to Inf (infinite). Each time that a solution is found with a cost better than the reference (in this case, a minimum cost) this variable is updated.
- $No_Best = 0$, if the best solution found in the present iteration of the greedy algorithm has a better cost than the reference ($Z(Best_Initial_Solution) < Best_Cost$) then the algorithm is iterated again and the reference is updated ($Z(Best_Initial_Solution) \rightarrow Best_Cost$). Otherwise the No_Best variable is increased by one. The algorithm is iterated until $No_Best = 100$.

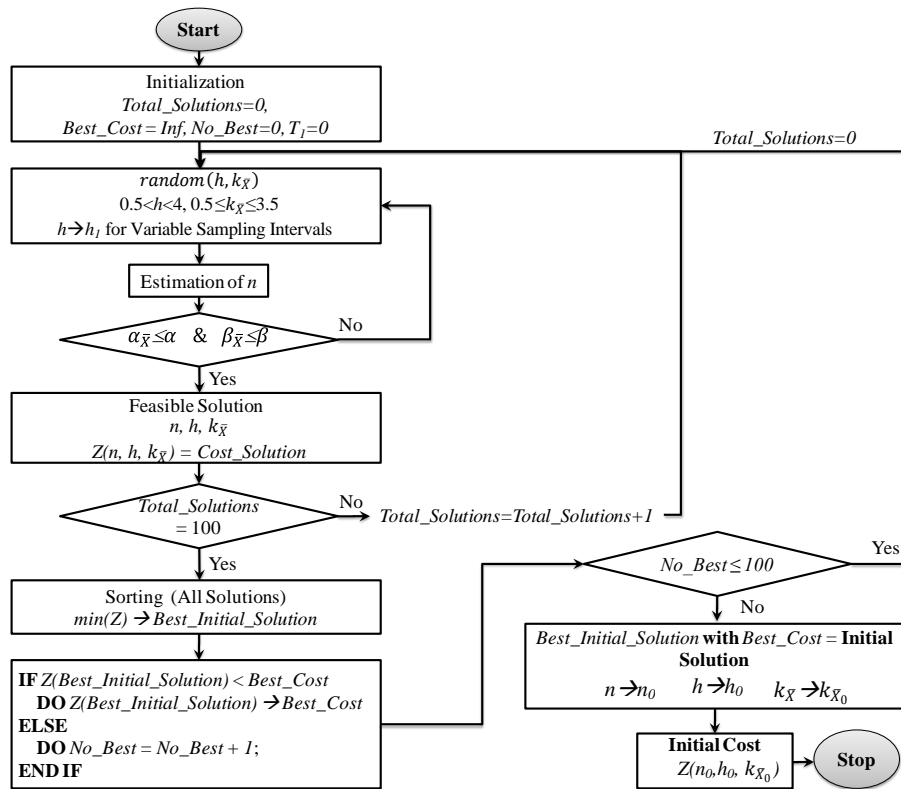


Fig. 3. Structure of the Greedy Algorithm.

The algorithm begins by estimating values for the parameters h and $k_{\bar{x}}$ of the \bar{X} chart. These values are randomly generated with a uniform probability distribution with the following limits: $0.5 < h < 4$, and $0.5 \leq k_{\bar{x}} \leq 3.5$. Then, for

each pair $h, k_{\bar{X}}$, n is estimated by using the expression $n = \left(C + \frac{k_{\bar{X}}}{\delta}\right)^2$, where C is a value from a standardized normal distribution for a required value of β or α [?]. Then, for the solution formed by the parameters n, h and $k_{\bar{X}}$, the values for $\alpha_{\bar{X}}$ and $\beta_{\bar{X}}$ are computed. If these values are within given limits ($\alpha_{\bar{X}} \leq \alpha, \beta_{\bar{X}} \leq \beta$) then the solution complies with statistical restrictions, otherwise the set is re-estimated.

The complying set of values for $n, h, k_{\bar{X}}$ represents a “Feasible Solution” which is then evaluated in the cost function ($Z(n, h, k_{\bar{X}})$) to obtain its associated cost (*Cost_Solution*). This process is performed 100 times leading to 100 feasible solutions (*Total_Solutions* = 100). Then, the solution with the minimum cost ($\min(Z)$) is considered the *Best_Initial_Solution*. If the cost of this *Best_Initial_Solution* is better ($<$) than the reference cost (*Best_Cost*), then this value is updated (*Best_Initial_Solution* \rightarrow *Best_Cost*). If not, the variable *No_Best* is increased by one and the reference cost remains unchanged. After any of these scenarios, a new process is started with *Total_Solutions* = 0. This is performed until *No_Best* = 100. At this point, the final *Best_Initial_Solution* obtained across all iterations of the greedy algorithm is considered to be the **Initial_Solution** for the Tabu-Search algorithm, and its cost the **Initial_Cost** (or reference cost) for the same algorithm.

2.2 Tabu Search (TS) Algorithm

TS is a meta-heuristic that can guide a heuristic algorithm from a local search to a global space to find solutions beyond the local optimality. It can avoid loops in the search process by restricting certain “moves” that would make the algorithm to revisit a previously explored space of solutions. These moves are kept hidden or reserved (are being kept “Tabu”) in a temporal memory (a “Tabu List”) which can be updated with new moves or released with different criteria [5,6]. The TS algorithm which is presented in Figure 2 and is based on the algorithm presented in [1] takes as input the **Initial_Solution** estimated by the greedy algorithm. This solution is diversified in order to explore a search space more likely to contain a better solution (global optimum). The TS algorithm considers the following variables:

- T_L , the number of iterations of the TS algorithm that a solution remains in the **Tabu List** (TL);
- $T_1 \leq 100$, the number of iterations of the integrated G-TS algorithm (see Figure 2);
- $T_2 \leq 100$, the number of iterations of the TS algorithm with no better solutions.

With the **Initial_Solution** the TS algorithm starts the search task, diversifying this solution into a set of new solutions that forms a **Search Space**.

The diversification is performed by means of “moves” over the elements of the **Initial_Solution** $(n_0, h_0, k_{\bar{X}_0})$. For the case study with variable sampling intervals, all j intervals are dependent of the length of the first interval h_1 as defined by $h_j = \rho h_{j-1}$ [11], hence only h_1 is to be estimated. For all new solutions the values of α 's and β 's are estimated to verify compliance of the statistical restrictions. If a solution does not comply, then $Z = Inf$ (infinite), otherwise Z is computed with the estimated parameters. As this is a minimization problem, the solutions are sorted based on their cost value Z in ascendant order, thus the first element would be that with the minimum cost of all new solutions, the **Best_Solution**.

To avoid search loops (revisiting solutions) and finding a local optimum, the **Best_Solution** is kept in a **Tabu List** (TL) for T_L iterations of the TS algorithm. If in the next iteration, the *Best_Solution* is already in TL, then the *Best_Solution* would be the second (or third, or fourth, etc.) best solution in the sorted Search Space. This enables the TS algorithm to produce more diverse solutions even if these are estimated from a “not so good” solution.

The **Actual_Cost** is updated with that of the *Best_Solution*, **Cost_BS**, and the TS algorithm starts again to produce a set of new solutions from the *Best_Solution*. If no *Best_Solution/Cost_BS* is obtained after T_2 iterations of the TS algorithm, the *Best_Solution* is stored and the G-TS algorithm is executed again (for up to T_1 iterations) to generate new initial values for h and $k_{\bar{X}}$. If T_1 reaches a given limit then the G-TS algorithm stops.

In contrast to the TS algorithm presented in [1] uniform random distribution is used to generate the moves for the diversification of n_0, h_0 and $k_{\bar{X}_0}$. For n_0 the upper limit for the uniform random move changes dynamically to $n_0/2$ instead of being fixed to 1. This enabled further exploration of the search space.

3 Results

3.1 First Case Study

The G-TS algorithm was tested to solve the cost model of Rahim *et al.* [12] for the ESD of \bar{X} control charts with Gamma failure distributions and constant sampling intervals. In [13] the cost model of Rahim *et al.* [12] was adapted to consider general failure distribution for the ESD of \bar{X} charts. This allowed the consideration of an additional failure distribution: Weibull.

Both failure distributions have a parameter related to the failure rate in a process which is defined as:

$$\lambda = \frac{\text{known number of failures}}{\text{Unit of Time}}. \quad (1)$$

λ is known as the *scale* parameter in the Gamma distribution and represents the known number of failures per unit of time. For the Weibull distribution

$\lambda = 1/c$ where c is defined as the time where the system is likely to fail. In order to explore the performance of the proposed algorithm, four failure rates were considered for λ : 0.5050, 0.2525, 0.1010, 0.0505. Also, for the Weibull distribution, three values for the *form* (f) parameter were considered: 2, 3, 4.

The solving methods considered for this case study were TS alone and GA. TS was implemented as presented in [1]. For the GA, the algorithm presented in [13] was considered. The data set used for this test was: $\mu=182$, $\sigma = \sqrt{10}$, $\delta=0.50$, $Z_0=0.25$, $Z_1=1$, $D_0=50$, $D_1=950$, $a = 20$, $b = 4.22$, $W = 1100$, $Y = 500$, $\alpha_{\bar{X}} \leq 0.15$, $\beta_{\bar{X}} \leq 0.20$. The comparison of performances is presented in Table 1.

For the Gamma distribution, GA, TS, and G-TS provided values for n , h , and $k_{\bar{X}}$ that led to very similar costs Z independently of the failure rates. Significant differences are presented with the Weibull distribution for all parameter settings of *scale* and *form*. The significance of these results is discussed in Section 3.3.

3.2 Second Case Study

For the ESD of \bar{X} control charts with variable sampling intervals the cost model presented in [11,13] was considered. If in a process there are $j = 1, 2, \dots, J$ constant sampling intervals, then $h_j = h$. When the sampling intervals are variable, h_j is different for each period j . In [11] Rahim *et al.* proposed the consideration of a specific number of J samples (sampling intervals) in the production cycle, $j = 1, 2, \dots, J$, so the production cycle could be considered as truncated [13]. A truncated production cycle starts when a new component is installed and ends with a repair or after a fixed number of J sampling intervals. The model with variable sampling intervals considers the following:

- the first interval h_1 is randomly chosen;
- the length of the next sampling intervals are chosen as $h_j = \rho h_{j-1}$, where h_j is the sampling interval for sample j , and ρ is a decrement factor. The sampling intervals h_j are computed by applying the decrement factor to the successor sampling interval, thus $h_1 > h_2 > h_3 > \dots > h_J$, because as time continues the sampling frequency must increase given the natural wear and tear of the components of the process [11].

Hence, for this case study, n , h_1 , and $k_{\bar{X}}$ were estimated. The G-TS algorithm was tested against the GA presented in [13] with Weibull and Gamma distributions with different values of λ and *form*. The number of sampling intervals was $J=5$ and $J=7$ for the experiments with Gamma and Weibull distributions respectively. The results are presented in Table 2. In all cases, the G-TS algorithm found parameters that reduced the cost Z more than the parameters found with the GA. The significance of these results is discussed in Section 3.3.

Table 1. G-TS performance compared to TS and GA: Cost Model of Rahim [12]-Ruvalcaba [13].

Failure Dist.	Parameters	Technique	n	h	$k_{\bar{x}}$	$\alpha_{\bar{x}}$	$1 - \beta_{\bar{x}}$	Z
Gamma	$\lambda = 0.5050$ $f = 2$	GA	24	1.4703	1.4421	0.1493	0.8432	467.83
		TS	24	1.4546	1.4396	0.1500	0.8438	467.81
		G-TS	24	1.4554	1.4396	0.1500	0.8440	467.20
Gamma	$\lambda = 0.2525$ $f = 2$	GA	25	1.8820	1.4725	0.1409	0.8479	347.12
		TS	25	1.8916	1.4719	0.1411	0.8481	347.12
		G-TS	25	1.8808	1.4767	0.1398	0.8469	346.77
Gamma	$\lambda = 0.0505$ $f = 2$	GA	28	3.7413	1.6109	0.1072	0.8496	174.70
		TS	28	3.8588	1.5585	0.1191	0.8615	174.70
		G-TS	28	3.7322	1.6255	0.1041	0.8462	174.70
Weibull	$1/c = 0.5050$ $f = 2$	GA	28	0.6248	1.7025	0.0887	0.8272	549.91
		TS	24	0.6521	1.4447	0.1485	0.8425	547.99
		G-TS	24	0.5857	1.5900	0.1118	0.8050	540.68
Weibull	$1/c = 0.2525$ $f = 2$	GA	31	0.8378	1.9410	0.0522	0.8003	453.43
		TS	25	0.8263	1.4788	0.1391	0.8464	449.41
		G-TS	23	0.7753	1.4975	0.1343	0.8161	442.71
Weibull	$1/c = 0.0505$ $f = 2$	GA	37	1.4401	2.0884	0.0368	0.8297	284.97
		TS	25	1.4165	1.4776	0.1395	0.8467	275.83
		G-TS	24	1.4156	1.6002	0.1095	0.8022	269.99
Weibull	$1/c = 0.5050$ $f = 3$	GA	30	0.5776	1.8717	0.0612	0.8070	477.72
		TS	24	0.6424	1.4458	0.1482	0.8423	480.69
		G-TS	23	0.6000	1.4986	0.1340	0.8158	469.79
Weibull	$1/c = 0.2525$ $f = 3$	GA	29	0.6839	1.7663	0.0773	0.8228	408.44
		TS	24	0.7421	1.4504	0.1469	0.8411	410.97
		G-TS	26	0.7000	1.6977	0.0773	0.8229	399.32
Weibull	$1/c = 0.0505$ $f = 3$	GA	37	1.1217	2.0867	0.0369	0.8301	296.48
		TS	24	1.1295	1.4471	0.1478	0.8419	286.28
		G-TS	27	1.0333	1.7130	0.0369	0.8301	279.20
Weibull	$1/c = 0.5050$ $f = 4$	GA	30	0.6524	1.7662	0.0774	0.8346	435.03
		TS	25	0.6264	1.4685	0.1420	0.8490	435.03
		G-TS	26	0.5527	1.6586	0.0774	0.8346	423.55
Weibull	$1/c = 0.2525$ $f = 4$	GA	37	0.7094	2.0952	0.0361	0.8280	400.11
		TS	26	0.6900	1.4892	0.1364	0.8555	387.85
		G-TS	26	0.7411	1.6682	0.0953	0.8109	371.89
Weibull	$1/c = 0.0505$ $f = 4$	GA	33	1.0000	1.9180	0.0551	0.8300	288.86
		TS	24	1.0432	1.4500	0.1470	0.8413	287.45
		G-TS	25	0.9000	1.6457	0.0998	0.8036	277.80

Table 2. G-TS performance compared to GA: Cost Model of Rahim [11]-Ruvalcaba [13].

Failure Dist.	Parameters	Technique	J	n	h_1	$k_{\bar{x}}$	$\alpha_{\bar{x}}$	$1 - \beta_{\bar{x}}$	Z
Gamma	$\lambda = 0.5050$ $f = 2$	GA	5	23	2.2430	1.5345	0.1249	0.8061	476.39
		G-TS	5	25	2.6317	1.5450	0.1224	0.8302	472.59
Gamma	$\lambda = 0.2525$ $f = 2$	GA	5	31	3.8033	1.7658	0.0774	0.8457	354.60
		G-TS	5	24	4.0743	1.5059	0.1321	0.8273	352.64
Gamma	$\lambda = 0.0505$ $f = 2$	GA	5	29	10.0290	1.6552	0.0979	0.8502	173.61
		G-TS	5	30	10.4509	1.6009	0.1094	0.8724	173.43
Gamma	$\lambda = 0.5050$ $f = 3$	GA	5	31	3.2457	1.7733	0.0762	0.8439	379.45
		G-TS	5	26	3.1754	1.7027	0.0886	0.8015	377.04
Gamma	$\lambda = 0.2525$ $f = 3$	GA	5	28	5.0753	1.5404	0.1235	0.8655	271.66
		G-TS	5	28	4.9576	1.6702	0.0949	0.8354	271.50
Gamma	$\lambda = 0.0505$ $f = 3$	GA	5	36	14.7360	1.7114	0.0870	0.9012	130.85
		G-TS	5	25	15.2913	1.6382	0.1014	0.8056	130.06
Gamma	$\lambda = 0.5050$ $f = 4$	GA	5	28	3.7286	1.7501	0.0801	0.8148	317.50
		G-TS	5	27	3.8204	1.6122	0.1069	0.8379	316.91
Gamma	$\lambda = 0.2525$ $f = 4$	GA	5	34	6.3295	1.9685	0.0490	0.8282	226.53
		G-TS	5	26	6.2435	1.6992	0.0893	0.8024	223.60
Gamma	$\lambda = 0.0505$ $f = 4$	GA	5	29	17.9552	1.5582	0.1192	0.8717	108.82
		G-TS	5	27	17.9491	1.6731	0.0943	0.8225	107.79
Weibull	$1/c = 0.5050$ $f = 2$	GA	7	56	1.0844	2.0527	0.0401	0.9543	570.79
		G-TS	7	53	1.0852	1.4637	0.1433	0.9852	564.18
Weibull	$1/c = 0.2525$ $f = 2$	GA	7	74	1.4193	2.2541	0.0241	0.9796	451.95
		G-TS	7	58	1.3926	1.4809	0.1386	0.9900	441.30
Weibull	$1/c = 0.0505$ $f = 2$	GA	7	97	2.8885	2.4514	0.0142	0.9933	219.30
		G-TS	7	69	2.8611	1.4562	0.1453	0.9965	208.43
Weibull	$1/c = 0.5050$ $f = 3$	GA	7	44	1.0507	1.8716	0.0612	0.9258	592.15
		G-TS	7	47	1.0398	1.4501	0.1470	0.9760	590.32
Weibull	$1/c = 0.2525$ $f = 3$	GA	7	52	1.3105	1.9964	0.0459	0.9462	526.48
		G-TS	7	58	1.2839	1.4421	0.1493	0.9910	520.93
Weibull	$1/c = 0.0505$ $f = 3$	GA	7	74	2.1194	2.2548	0.0241	0.9796	382.03
		G-TS	7	60	2.1355	1.4895	0.1364	0.9914	373.52
Weibull	$1/c = 0.5050$ $f = 4$	GA	7	46	1.0745	1.5272	0.1267	0.9688	601.08
		G-TS	7	26	1.0286	1.4803	0.1388	0.8575	598.81
Weibull	$1/c = 0.2525$ $f = 4$	GA	7	38	1.2172	1.7637	0.0778	0.9063	552.99
		G-TS	7	47	1.2403	1.4945	0.1351	0.9734	551.20
Weibull	$1/c = 0.0505$ $f = 4$	GA	7	53	1.8217	2.0087	0.0446	0.9486	448.55
		G-TS	7	53	1.7649	1.4684	0.1420	0.9851	443.33

3.3 Analysis

In order to evaluate the statistical significance of the results presented in Tables 1 and 2 a “2-Sample t Test” was performed. For this purpose the hypothesis $H_0 : \mu_x - \mu_{G-TS} = 0$ where $x = \{TS, GA\}$ was considered. It is expected that if significant improvements are obtained then $\mu_x - \mu_{G-TS} \neq 0$ leading to H_0 being rejected. In Table 3 the results of the “2-Sample t Test” are presented for each case study considering a significance level of 0.05. For all cases the improvement achieved with the G-TS algorithm was statistically significant.

Table 3. 2-Sample t Test.

Case Study	Comparison	Conclusion
First	G-TS vs. TS	μ_{TS} and μ_{G-TS} are significantly different ($p < 0.05$)
	G-TS vs. GA	μ_{GA} and μ_{G-TS} are significantly different ($p < 0.05$)
Second	G-TS vs. GA	μ_{GA} and μ_{G-TS} are significantly different ($p < 0.05$)

In contrast to other TS implementations [1] where only an initial solution is randomly generated in the presented algorithm the greedy process provides an initial solution selected from a whole set which may be of better quality. The importance of the initial solution for the TS process is corroborated by the results reported in Tables 1 and 2.

4 Conclusions

The proposed G-TS algorithm achieved significant benefits when used for the ESD of \bar{X} Shewhart / control charts with different probability distributions when compared with other methods such as GA and TS. These benefits were statistically significant with p -values < 0.05 .

In addition, the integration of a greedy algorithm may lead to increase convergence of other heuristics by providing a more specific search region. This however has to be verified with other integration schemes (e.g., G-GA) and other cost functions.

References

1. Caballero-Morales, S., Trujillo-Romero, F.: Algoritmo alternativo de búsqueda tabú para la solución del problema de diseño económico estadístico de diagramas de control \bar{X} , $\bar{X}-S$. *Research in Computing Science* 55, 81–89 (2012)
2. Chen, F.L., Yeh, C.H.: Economic statistical design of non-uniform sampling scheme \bar{X} -bar control charts under non-normality and Gamma shock using genetic algorithm. *Expert Systems with Applications* 36, 9488–9497 (2009)

3. Cormen, T.H., Leiserson, C.E., Rivest, R.L., Stein, C.: Introduction to Algorithms. The MIT Press, 2nd edition (2001)
4. Franco, B.C., Costa, A.F.B., Machado, M.A.G.: Economic-statistical design of the X chart used to control a wandering process mean using genetic algorithm. Expert Systems with Applications 39, 12961–12967 (2012)
5. Glover, F.: Tabu Search - Part 1. ORSA Journal on Computing 1(2), 190–206 (1989)
6. Glover, F.: Tabu Search - Part 2. ORSA Journal on Computing 2(1), 4–32 (1990)
7. Kethley, R.B., Peters, M.H.: Extending economic design of p charts to handle user specified constraints using a genetic algorithm. IIE Transactions 36, 855–864 (2004)
8. Malaki, M., Akhavan-Niaki, S.T., Javad-Ershadi, M.: A Comparative Study of Four Evolutionary Algorithms for Economic and Economic-Statistical Designs of MEWMA Control Charts. J. of Optimization in Industrial Engineering 4(9), 1–13 (2011)
9. Montgomery, D.C.: Introduction to Statistical Quality Control. John Wiley & Sons, Inc., 7th Edition (2012)
10. Naama, B., Bouzeboudja, H., Lahdeb, M., Ramdani, Y.: A Hybrid Tabu Search and Algorithm Genetic for Solving the Economic Dispatch Problem. Leonardo Journal of Sciences 22, 29–36 (2013)
11. Rahim, M.A : A generalized model for the economic design of \bar{X} -control charts for production systems with increasing failure rate and early replacement. Naval Research Logistics 40, 787–808 (1993)
12. Rahim, M.A , Banerjee, P.K.: The economic design of control chart: A renewal theory approach. Engineering Optimization 12, 63–73 (1987)
13. Ruvalcaba, R.A.: Economic Design of Quality Control Charts. MSc. Thesis, University of the Americas Puebla, Mexico (in Spanish) (2004)
14. Saniga, E., Davis, D.J.: Economic-statistical design of \bar{X} and R or \bar{X} and S charts. J. Quality Technology 33, 234–241 (2001)

