*Research Article*

# A Parameter Tuning Methodology for Metaheuristics Based on Design of Experiments

Mehdi Fallahi [a,*], Somayeh Amiri[b], Masoud Yaghini [c]

[a] Department of Industrial Engineering, Kermanshah University of Technology, Kermanshah, Iran
[b] Department of Industrial Engineering, Bu-Ali Sina University, Hamedan, Iran
[c] School of Railway Transportation Engineering, Iran University of Science and Technology, Tehran, Iran
* Corresponding author. Tel.: 00989189515003;
E-mail address: mehdi.fallahi@kut.ac.ir

*A b s t r a c t*

**Keywords:**

*Parameter Tuning Methodology, Metaheuristic Algorithms, Optimization, Design of Experiments.*

Many parameters have to be tuned for any metaheuristics. Parameter tuning may permit a superior flexibility and robustness, but requires a careful initialization. Those parameters may have a large influence on the efficiency and effectiveness of the search. The optimal values for the parameters mainly depend on the problem. In order to let a project to be replicated, a standard procedure as a methodology is required. In this paper, a parameter tuning methodology for metaheuristics based on design of experiments is proposed. The proposed methodology comprises five phases, namely, Problem Characteristics Screening, Clustering, Parameter Screening, Response Surface Modeling and Optimization. The proposed methodology is applied to the Ant Colony System algorithm for solving 47 traveling salesman problem instances. For validation of the proposed methodology, the different alternative approaches for parameter tuning are compared and it is concluded that, the methodology presents better results than the other alternative approaches.

## 1. Introduction

In order to perform metaheuristic algorithms, many parameters need to be tuned. These parameters are not only numerical values, but may also engage the use of the search components. Parameter tuning may allow a superior flexibility and robustness, but requires a careful initialization. Those parameters may have a large influence on the efficiency and effectiveness of the search. The optimal values for the parameters mainly depend on the problem and on the search time that the user wants to spend on solving the problem. Universally, optimal parameter values that one set for a specified metaheuristic does not exist [1].

There are two special strategies for parameter tuning: the online parameter initialization and the off-line parameter initialization. In the online approach, the parameters are controlled and updated

dynamically or adaptively throughout the execution of the metaheuristic, whereas in the off-line parameter initialization, the values of different parameters are fixed before the execution of the metaheuristic [1].

Usually in off-line parameter initialization, the metaheuristic designer tune one parameter at a time, and it's the best value is determined empirically. In this situation, no interaction between parameters is studied. This sequential optimization strategy does not assurance to find the optimal setting, even if an exact optimization setting is performed. To defeat this problem, experimental design is used.

Design of Experiment (DOE) offers efficiency in terms of the quantity of data that needs to be gathered. This is vital when attempting to understand vast design spaces. All DOE conclusions are based on statistical analyses and so are supported with mathematical precision. This allays every concern regarding the subjective understanding of results [2].

In this paper, the proposed methodology for parameter tuning of metaheuristics is developed on the basis of the method given by Ridge [2] and Ridge & Kudenko [3-6]. There are several shortcomings in the Ridge's approach. This paper aims to present a methodology that proposes a simple and more practical approach than provided by Ridge. In the proposed methodology the unnecessary steps are removed and some related steps are integrated. A clustering phase to cluster the instances based on their actual differences is added. By using this clustering phase, parameter tuning is performed in the relation to problem characteristics of instances.

The paper is organized as follows. In Section 2, the important works on off-line parameter initialization is reviewed. In Section 3, some basic concepts of DOE is described. In Section 4, the proposed methodology is presented. In Section 5, as a case study, the proposed methodology is applied to the ACO algorithm for solving the symmetric TSP problem. In Section 6, the methodology is compared with different alternative strategies to validate it. Section 7, includes conclusions and some suggestions for continued researches on this subject.

## 2. Literature Review

In this section, the most important works on off-line parameter initialization are reviewed.

Coy et al. [7] presented a process for finding heuristic parameter settings that applied to two local search heuristics with 6 tuning parameters and some Vehicle Routing Problems (VRP). Theirs method does not investigate the relationship between instances, parameter settings and performance.

Caserta and Rico [8] applied a Circumscribed Central Composite design to make observations and used a higher degree polynomial to estimate the response surface of a set of parameters. Parsons and Johnson [9] employed a full factorial design to screen the finest parameter settings. Breedam [10] attempted to find significant parameters for a Genetic Algorithm (GA) and a Simulated Annealing (SA) algorithm applied to the VRP using an Analysis of Variance (ANOVA) method. Seven GA and

eight SA parameters are examined. Park and Kim [11] developed a non-linear response surface method to find parameter settings for an SA algorithm.

Adenso-Diaz and Laguna [12] applied a factorial design combined with a local search technique to systematically find the finest parameter values for a heuristic. Their method, CALIBRA, was demonstrated on six different combinatorial optimization applications. The restrictions of CALIBRA are that it can only tune five parameters and is unable to examine interactions between parameters.

Fığlalı et al. [13] investigated the parameters of Ant System on different sized and randomly generated job-shop scheduling problems using DOE.

Ridge [2] and Ridge & Kudenko [3-6], used a two-stage nested design to test whether problem characteristics have an effect on the performance of metaheuristics. Then, by a fractional factorial design all metaheuristic parameters are tested and those parameters that have not significant effect are screened out. In the next stage, a response surface model is constructed, and the parameters are optimized. This approach is applied to Ant Colony System (ACS) and MAX-MIN, two versions of Ant Colony Optimization (ACO) [14]. The disadvantages of this approach are mentioned in Section 1.

Hutter et al. [15] and Hutter [16] proposed an automatic algorithm configuration framework in which the settings of discrete parameters were optimized to yield maximal performance of a target algorithm for a specified class of problem instances.

Birattari et al. [17] and Birattari [18-19] applied F-races to the configuration of stochastic local search algorithms. The inputs of their algorithm consist of a finite set of algorithm configurations, and an instance distribution. It iteratively runs the target algorithm with all surviving parameter configurations on a number of instances sampled from instance distribution. After each iteration, first the non-parametric Friedman test is applied to verify whether there are significant differences between the configurations. If this is the case, and based on the results of Friedman post-tests, the inferior-performing configurations are eliminated. This procedure is iterated until only one configuration survives or a given cutoff time is reached. A related work presented by Balaprakash et al. [20] iteratively performs F-races on differently defined subsets of parameter configurations.

Srinivasan and Ramakrishnan [21] used a two level fractional factorial design for screening Inductive Logic Programming (ILP) system parameters and then optimized effective parameters by response surface methodology.

## 3. Design of Experiment (DOE)

In this section, some basic concepts of DOE are presented. For more detailed DOE and response surface methodology refer to Montgomery [22] and Myers and Montgomery [23].

### 3.1. Definitions

The experiment is defined as a test or series of tests in which intended changes are effected on the input variables of a process or system such that observe and identify the reasons for changes that may be observed in the output response. Statistical design of experiments denotes the process of planning the experiment such that suitable data that can be analyzed with statistical techniques will be gathered, resulting in valid and objective conclusions [22].

A factor is an independent variable manipulated in an experiment since it is hypothesized to influence one or more of the response variables. In the heuristic performance study, the factors cover both the heuristic tuning parameters and the most important problem characteristics. The levels are the different values at which a factor is set. In experimental design, a treatment is a specified combination of factor levels. The particular treatments are dependent on the particular experimental design and on the ranges over which factors are varied [24].

An effect is a change in the response variable as a result of a change in one or more factors. The main effects can be defined as follows: The main effect of a factor is the amount of the change in the response variable to changes in the level of the average factor over all levels of all the other factors. In factorial experiments, two factors are said to interact if the effect of one variable is different at different levels of the other variables. In general, when variables operate independently of each other, they do not demonstrate interaction [24].

### 3.2 Factorial Experimental Designs

Full factorial design is a crossing of total levels of all factors. This is a very powerful but costly design. A more useful type of factorial for DOE applies k factors, each at two levels. The 2k factorial design represents the smallest number of runs with which k factors can be considered in a full factorial design. If there are more than four or five factors, it is usually not required to run all possible combinations of factor levels. A fractional factorial experiment is a variation of the basic factorial design in which only a subset of the runs is performed [22].

### 3.3. The Two-Stage Nested Design

In multifactor experiments, the levels of one factor (e.g., factor B) are similar but not identical to different levels of another factor (e.g., A). Such an arrangement is called a hierarchical or nested design, with the levels of factor B nested in the levels of factor A. Factor A is used in the model as the parent. Factor B is nested within Factor A. Factor B is set as a random factor since the unique instances are randomly constructed. Factor A is a fixed factor since the experimenter selects its levels [22].

### 3.4. Response Surface Methodology

Response Surface Methodology (RSM) is a group of mathematical and statistical methods, which are helpful for the modeling, and analysis of problems in which a response of concern is affected by a number of variables and the objective is to optimize this response. Designs for fitting response surfaces are called response surface designs. Central Composite Design (CCD) is the most known type of designs used for fitting these models [25-27]. There are two parameters in the design that must be determined: the distance of the axial runs from the design center and the number of center points $n_c$. An existing factorial or resolution V and higher design from the screening stage can be augmented with appropriate star points to produce the CCD or FCC (Face Centered Design) designs [22].

### 4. The Proposed Methodology

### 4.1. Overall Structure of the Proposed Methodology

In order to let a project to be replicated, a standard procedure as a methodology for tuning metaheuristic parameters is required. Webster's collegiate dictionary defines methodology as ''A body of methods, rules, and postulate employed by a discipline'' or ''The analysis of the principles or procedures of inquiry in a particular field'' [28].

The proposed methodology is developed on the basis of the sequential experimentation method. The main advantage of this method is its resource efficiency.

The DOE is used in the proposed methodology for tuning the parameter of metaheuristic algorithms. There are five phases in the proposed methodology. Figure 1 shows a schematic plan of phases and steps of the proposed methodology. In the first phase, the problem characteristics that may affect metaheuristic algorithm are screened. In the second phase, the instances on the basis of affecting problem characteristics are clustered. Clustering phase provides a set of clusters based on the most important problem characteristics. Parameter screening phase produces a reduced set of the most important tuning parameters. The response surface modeling phase, results the mathematical functions relating the tuning parameters to each response of interest. In the last phase, the parameters in relation to the importance of each response are optimized. In the following subsections, the phases of the proposed methodology are explained.
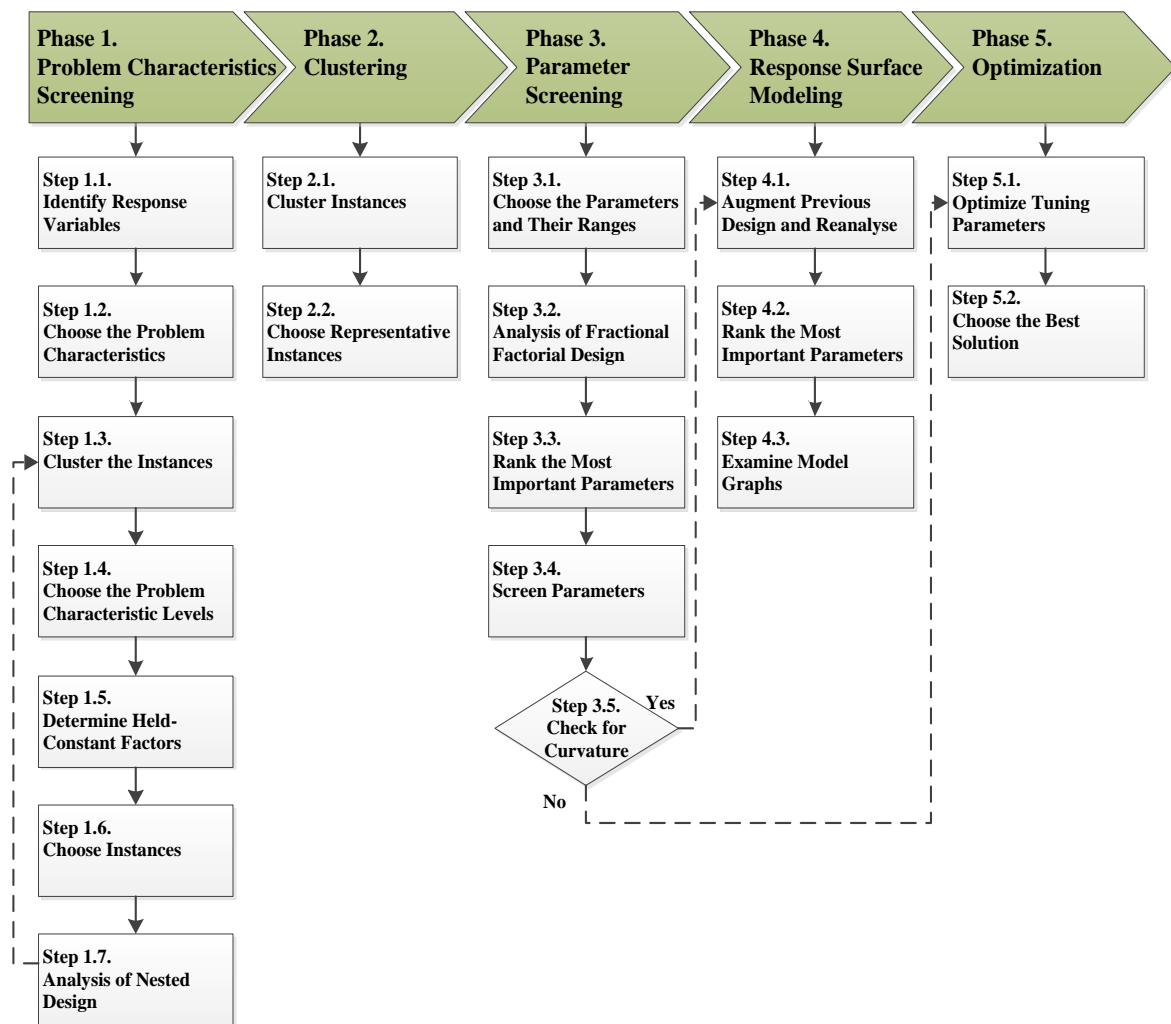
Fig. 1: A schematic plan of the proposed methodology

## 4.2. Phase 1: Problem Characteristic Screening

In order to tune parameters of metaheuristics to solve a set of instances, it must be noted that in addition to tuning parameters, some problem characteristics may affect performance of metaheuristic algorithms. In most research, this important element is ignored. In the proposed methodology, the parameters are tuned with respect to the problem characteristics. The instances are clustered on the basis of problem characteristics. Prior to clustering the instances, it must be tested and decided whether problem characteristics are effective and incorporated into clustering instances.

The steps of Phase 1 are described as here below.


*Step 1.1. Identify Response Variables*

In this step, the responses to be measured are recognized. These responses must measure solution quality, solution time or each other concerned performance measure.

*Step 1.2. Choose the Problem Characteristics*

In Step 1.2, following the inspection of the instances, the problem characteristics hypothesized to affect the interested response are chosen.

The Steps 1.3 to 1.7 need to be repeated for each problem characteristic.

*Step 1.3. Cluster the Instances*

In this step, the instances should be clustered for each problem characteristic.

*Step 1.4. Choose the Problem Characteristic Levels*

The average of each generated cluster of instances from Step1.3 could be taken as the problem characteristic levels.

*Step 1.5. Determine Held-Constant Factors*

All tuning parameters and other problem characteristics should assume constant values during the experiment. For parameter tuning, these values may be usually found in the literature. For the problem characteristics, these values may be the average of central cluster or the average of a cluster with most members.

*Step 1.6. Choose Instances*

Three or more instances should be chosen for each problem characteristic level. The instances, which are most similar to the center of current problem characteristics clusters, must be selected while the values of other problem characteristics are held constant.

*Step 1.7. Analysis of Nested Designs*

An experimental design should be generated and then an ANOVA analysis should be applied to identify whether the current problem characteristics affect the performance of the metaheuristic algorithm.

The most important difficulty faced in trying to experiment problem instances characteristics is that, instances are unique. That is to say, while several instances may have the same characteristics, it is hypothesized to affect the response; these instances are unique. As an example, there are an infinite number of potential instances that have the same characteristic of problem size. The uniqueness of instances will therefore cause different values of the response, although the instances have identical levels of the hypothesized characteristic. The experimenter's difficulty is unraveling the effect of the hypothesized characteristic from the inescapable variability between unique instances. To defeat this problem, a two-stage nested design should be generated, where the hypothesized characteristic is the parent factor, and the unique instances are nested within a given level of the parent [2].

In this Step, the following actions should be carried out: (a) a randomized two-stage nested design is generated; (b) the designed treatments are run to gather required data; (c) an ANOVA analysis is perform; (d) the model plots are examined so that to check that model assumptions not violated; (e)

the responses are transformed (if required); (f) the outliers are identified and removed; (f) the statistical power is checked; (h) the ANOVA table is interpreted; and (i) model graphs are examined.

## 4.3. Phase 2: Clustering

In this phase, the instances are clustered based on all detected significant problem characteristics of the previous phase. According to the definition of clustering in Han and Kamber [26], it is the process of grouping the data into classes or clusters so that objects within a cluster have high similarity in comparison to one another but are very dissimilar to objects in other clusters. Dissimilarities are assessed with respect to the value of attributes describing the objects.

The steps of Phase 2 are described as follows.

*Step 2.1. Cluster Instances*

Here, the instances are clustered on the basis of the problem characteristics obtained in the previous phase.

*Step 2.2. Choose Representative Instances*

In this step, three or more instances on the basis of the average values of each generated clusters in Step 2.1 must be selected.

The Phases 3 and 4 should be carried out for every generated cluster and response separately.

## 4.4. Phase 3: Parameter Screening

The parameter screening phase aims to determine parameters with statistically significant effect on each response. As a result, the parameters with no effect on performance are identified saving experimental resources in subsequent modeling experiments. Parameter screening phase also provides the importance ranking of the parameters. The steps of Phase 3 are described as follows.

*Step 3.1. Choose the Parameters and Their Ranges*

In this step, the algorithm tuning parameters are going to be screened are chosen as well as the ranges over which these parameters will vary are determined.

*Step 3.2. Analysis of $2^{k-n}$ Fractional Factorial Design*

In this step, in order to analyze the effects of tuning parameters a $2^{k-n}$ fractional factorial design with resolution V or VI is generated and data is gathered. Then, an ANOVA test to determine which tuning parameters are affected in the performance of metaheuristic are applied.

In this step, the following actions should be undertaken: (a) a 2k-n fractional factorial design is generated; (b) the designed treatments to gather required data are run; (c) the effects to be included in ANOVA test are selected by backward regression; (d) an ANOVA analysis is performed; (e) the model assumption's plots are examined; (f) responses are transformed (if required); (g) the outliers are

identified and removed; (h) the statistical power is checked; (i) the model fit information are examined.

*Step 3.3. Rank the Most Important Parameters*

The terms in the model should be ranked on the basis of their sum of square values in the ANOVA table. These ranks can then be studied together with the related p-values for the model terms. The large sum of square values corresponds to the most important model terms and the p-value showing they are statistically significant.

*Step 3.4. Screen Parameters*

Those parameters that are not significant statistically and having a relatively low ranking could be removed from the subsequent experiments, since they lack any important influence on the response. NB, if a parameter is important for even one response, then it should be kept in the subsequent experiments.

*Step 3.5. Check for Curvature*

In this phase, the screening design yields a planar relationship between the parameters and the response. This relationship is often a higher order than planar. Therefore, it is important to find out whether such curvature exists to prove the need to use a more complicated response surface and its related experimental design in subsequent experiments.

Adding center points to a design allows determination of whether the response surface is not planar or actually contains some type of curvature. A center point is a treatment combining of all factors' values at the center of their ranges. The average response value of the actual data at the center points is compared to the predicted value of the center point that obtained from averaging of all the factorial points. With the existence of the response surface curvature in the region of the design, the actual center point value will be either higher or lower than that estimated by the factorial design points. If no curvature exists, the screening experiment's planar models should be sufficient to predict responses. In this case, skip Phase 4 and go to Phase 5.

**4.5. Phase 4: Response Surface Modeling**

A simple linear model is applied in the screening design to decide whether there is a significant difference between high and low levels of the factors. For this purpose, only the edges of the design space are of interest. By contrast, a response surface model needs a more complicated design since it endeavors to construct a model of the factor-problem-performance relationship over the whole design space. The steps of Phase 4 are described as follows.

*Step 4.1. Augment Previous Design and Reanalyse*

Constructing a response surface needs a specific type of experimental design. The FCC design is suggested in the proposed methodology. Where design factors are restricted within a certain range, the FCC is the most fitting design. In this step, the previous constructed design is augmented with FCC axial points, ANOVA test is applied, and a quadratic model is generated.

In Step 4.1, the following actions should be carried out: (a) the previous constructed design is augmented; (b) the new treatment with the random order to gather required data is run; (c) the effects to be included in the ANOVA test by stepwise regression are selected; (d) an ANOVA test is performed; (e) the model assumption's plots are examined; (f) the responses are transformed (if required); (g) the outliers are identified and removed; (h) the model fit information is checked.

*Step 4.2. Rank the Most Important Parameters*

The terms in the model equations should be ranked on the basis of their ANOVA F values. These ranks can then be considered together with the p values of the model terms.

*Step 4.3. Examine Model Graphs*

The graphs of the responses for each parameter should be examined. This indicates whether statistically significant and highly ranked parameter practically affects the response significantly. It also shows the possible location of optimal response values. At this stage, the experimenter has a model of all responses over the total design space. These models are confirming the accuracy of the metaheuristic predictors. At this stage, it is possible to employ the model for tuning the metaheuristic. The Phase 5 should be separately done for every generated cluster.

## 4.6. Phase 5: Optimization

In this phase, the functions constructed for all responses from the previous phases are optimized. This optimization phase leads to parameter setting for each of the created clusters from Phase 2. Here, there are a number of possible optimization aims. One may want to attain a response with a specified value (target value, maximum or minimum). More typically, due to the heuristic compromise, one may want to optimize a number of responses. The multiple responses are given in terms of desirability functions. The total desirability of responses is the geometric average of the individual desirability. The well-known Nelder-Mead Downhill Simplex is employed to the response surface model's equations with intention maximizing the desirability. The steps of Phase 5 are described as follows.

*Step 5.1. Optimize the Tuning Parameters*

A numerical optimization of desirability is performed for each of clusters created in Phase 2 by using the Nelder-Mead Simplex. The optimization objectives in most metaheuristics are the solution error and the solution run time responses minimization. These objectives can have various priorities.

*Step 5.2. Choose the Best Solution*

Following the completion of the optimization, the maximum desirability solution is selected. It should be noted that there may exist a number of solutions of extremely similar desirability with the different factor setting. This reflects the multiobjective nature of optimization and the possibility of several regions of interest [2].

## 5. Case Study

### 5.1. Introduction

In this section, the proposed methodology is applied to ACS algorithm, a version of ACO Dorigo and Stützle [14], to solve the Travelling Salesman Problem (TSP). For this purpose, 47 instances with size less than 500 cities are chosen. The design generation and statistical calculations can be performed with statistical software packages such as Design-Expert, JMP and Minitab. In this paper, Design-Expert 8 and Minitab 16 are used. SPSS Clementine 12 is applied to cluster instances. ACS algorithm is implemented by Java programming language. All experiments are run on a computer with Intel Core 2 Duo 2.53 GHz CPU and 4.00 GB RAM.

### 5.2. Phase 1: Problem Characteristic Screening

*Step 1.1. Identify Responses Variables*

To analyze the performance of ACS, a number of measures can be utilized. In this example, the relative gaps of known optimum (R-Gap) and solution time (Time) are selected.

*Step 1.2. Choose the Problem Characteristics*

To solve TSPs by using the ACS algorithm, several problem characteristics may affect metaheuristic algorithm performance. The two problem characteristics are considered as Coefficient of Variation (CV) of distance between cities and the number of cities (Size). The standard deviation, expressed as a percentage of the mean is CV. For improved clustering, Size and CV of the problems are normalized as N-Size and N-CV, respectively in the range of zero and one. The instances are selected from TSPLIB1 and TSP2 websites. The instances information is summarized in Table 1.

---

[1] http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95
[2] http://www.tsp.gatech.edu/data/index.html

Table 1: Problem instances information

| Name | Size | N-Size | Edge | Edge Std | CV | N-CV |
|---|---|---|---|---|---|---|
| dj38 | 38 | 0.000 | 727.744 | 374.294 | 51.432 | 0.358 |
| eil51 | 51 | 0.032 | 32.400 | 15.070 | 46.531 | 0.000 |
| st70 | 70 | 0.079 | 52.260 | 24.390 | 46.671 | 0.010 |
| eil76 | 76 | 0.094 | 33.200 | 15.760 | 47.479 | 0.069 |
| pr76 | 76 | 0.094 | 7558.700 | 3914.080 | 51.783 | 0.384 |
| rat99 | 99 | 0.151 | 85.000 | 46.800 | 55.054 | 0.623 |
| rd100 | 100 | 0.153 | 555.660 | 262.820 | 47.298 | 0.056 |
| kroD100 | 100 | 0.153 | 1631.100 | 867.210 | 53.167 | 0.485 |
| kroA100 | 100 | 0.153 | 1710.700 | 916.030 | 53.547 | 0.513 |
| kroC100 | 100 | 0.153 | 1700.550 | 910.740 | 53.555 | 0.514 |
| kroE100 | 100 | 0.153 | 1732.150 | 933.530 | 53.896 | 0.538 |
| kroB100 | 100 | 0.153 | 1687.540 | 912.900 | 54.096 | 0.553 |
| eil101 | 101 | 0.156 | 33.910 | 16.350 | 48.213 | 0.123 |
| lin105 | 105 | 0.166 | 1177.350 | 670.850 | 56.980 | 0.764 |
| pr107 | 107 | 0.171 | 5405.157 | 3104.762 | 57.441 | 0.798 |
| pr124 | 124 | 0.213 | 5623.350 | 2848.450 | 50.654 | 0.301 |
| ch130 | 130 | 0.228 | 356.220 | 170.000 | 47.718 | 0.087 |
| xqf131 | 131 | 0.230 | 35.480 | 20.682 | 58.293 | 0.860 |
| pr136 | 136 | 0.243 | 6074.000 | 2945.430 | 48.493 | 0.143 |
| pr144 | 144 | 0.262 | 5639.510 | 2813.440 | 49.888 | 0.245 |
| ch150 | 150 | 0.277 | 359.300 | 170.000 | 47.131 | 0.044 |
| kroA150 | 150 | 0.277 | 1717.350 | 919.030 | 53.514 | 0.511 |
| kroB150 | 150 | 0.277 | 1711.610 | 922.400 | 53.891 | 0.538 |
| pr152 | 152 | 0.282 | 6914.830 | 3668.430 | 53.052 | 0.477 |
| u159 | 159 | 0.300 | 2827.621 | 1477.498 | 52.252 | 0.418 |
| qa194 | 194 | 0.386 | 479.420 | 263.534 | 54.969 | 0.617 |
| rat195 | 195 | 0.389 | 116.520 | 63.950 | 54.884 | 0.611 |
| kroB200 | 200 | 0.401 | 1664.180 | 892.350 | 53.622 | 0.518 |
| kroA200 | 200 | 0.401 | 1701.170 | 917.360 | 53.925 | 0.541 |
| ts225 | 225 | 0.463 | 7080.030 | 3321.760 | 46.917 | 0.028 |
| tsp225 | 225 | 0.463 | 183.580 | 95.210 | 51.863 | 0.390 |
| pr226 | 226 | 0.465 | 7503.010 | 3708.920 | 49.432 | 0.212 |
| xqg237 | 237 | 0.493 | 52.470 | 29.195 | 55.641 | 0.666 |
| gil262 | 262 | 0.554 | 101.920 | 48.250 | 47.341 | 0.059 |
| pr264 | 264 | 0.559 | 4248.446 | 2557.952 | 60.209 | 1.000 |
| a280 | 280 | 0.599 | 121.800 | 62.620 | 51.407 | 0.356 |
| pr299 | 299 | 0.646 | 2540.610 | 1464.160 | 57.630 | 0.811 |
| lin318 | 318 | 0.693 | 1849.040 | 901.900 | 48.776 | 0.164 |
| bcl380 | 380 | 0.847 | 67.330 | 33.946 | 50.415 | 0.284 |
| pbl395 | 395 | 0.884 | 50.310 | 25.767 | 51.209 | 0.342 |
| rd400 | 400 | 0.896 | 528.870 | 250.550 | 47.374 | 0.062 |
| pbk411 | 411 | 0.923 | 54.450 | 27.410 | 50.343 | 0.279 |
| fl417 | 417 | 0.938 | 1189.510 | 685.070 | 57.593 | 0.809 |
| pbn423 | 423 | 0.953 | 53.589 | 26.384 | 49.235 | 0.198 |
| pbm436 | 436 | 0.985 | 53.230 | 26.642 | 50.043 | 0.257 |
| pr439 | 439 | 0.993 | 4328.390 | 2597.690 | 59.878 | 0.976 |
| pcb442 | 442 | 1.000 | 1748.000 | 830.680 | 47.522 | 0.072 |

*Step 1.3. Cluster the Instances*

The TSP instances are separately clustered for each problem characteristic. In this case, the Two-Step algorithm [29] is used for clustering the instances. For N-Size characteristic three clusters with average 0.178, 0.501 and 0.935 are detected and for N-CV characteristic three clusters with average 0.163, 0.523 and 0.86 are obtained.

*Step 1.4. Choose the Problem Characteristics Levels*

The averages of the generated clusters from the previous step are used as the levels of each problem characteristic.

*Step 1.5. Determine Held-Constant Factors*

For an experiment on N-Size, N-CV is fixed at the low level (0.163) and for experiment on N-CV, N-Size is fixed on the high level (0.935). All tuning parameters are fixed in all experiments based on the recommended values in the literature as in Table 2.

Table 2: Fixed values of tuning parameters

| Tuning parameters | Fixed values |
|---|---|
| Number of Iterations | 250 |
| Number of Ants | 10 |
| Percentage of Candidate List | 0.15 |
| Alpha | 1 |
| Beta | 2 |
| Coefficient of Local Evaporate | 0.1 |
| Coefficient of Global | 0.1 |
| q0 | 0.9 |
| Coefficient of Initial Pheromone | 1 |

*Step 1.6. Choose Instances*

Three instances in the vicinity of the center of clusters for each problem characteristic are chosen. The selected instances are listed in Table 3.

Table 3: The selected instances for experiment on N-Size and N-CV

| For experiment on | N-C V fixed value | Levels of N-Size | Selected instance | N-Size | N-CV |
|---|---|---|---|---|---|
| N-Size | 0.163 | 0.178 | eil101 | 0.156 | 0.123 |
| | | | pr136 | 0.243 | 0.143 |
| | | | ch130 | 0.228 | 0.087 |
| | | 0.501 | pr226 | 0.465 | 0.212 |
| | | | gil262 | 0.554 | 0.059 |
| | | | ts225 | 0.463 | 0.028 |
| | | 0.935 | pbn423 | 0.953 | 0.198 |
| | | | pbm436 | 0.985 | 0.257 |
| | | | rd400 | 0.896 | 0.062 |

| For experiment on | N-Size fixed value | Levels of N-CV | Selected instance | N-Size | N-CV |
|---|---|---|---|---|---|
| N-C V | 0.935 | 0.163 | pbn423 | 0.953 | 0.198 |
| | | | pbm436 | 0.985 | 0.257 |
| | | | rd400 | 0.896 | 0.062 |
| | | 0.523 | pbl395 | 0.884 | 0.342 |
| | | | pbk411 | 0.923 | 0.279 |
| | | | bcl380 | 0.847 | 0.284 |
| | | 0.860 | fl417 | 0.938 | 0.809 |
| | | | pr439 | 0.993 | 0.976 |
| | | | pr299 | 0.646 | 0.811 |

*Step 1.7. Analysis of Nested Design*

Two-stage nested designs are generated for effect analysis of N-Size and N-CV. In each problem, characteristic levels are the average of the clusters. For each level, three instances are selected. Figure 2 and Figure 3 respectively illustrate schematic two-stage nested design for N-Size and N-CV.
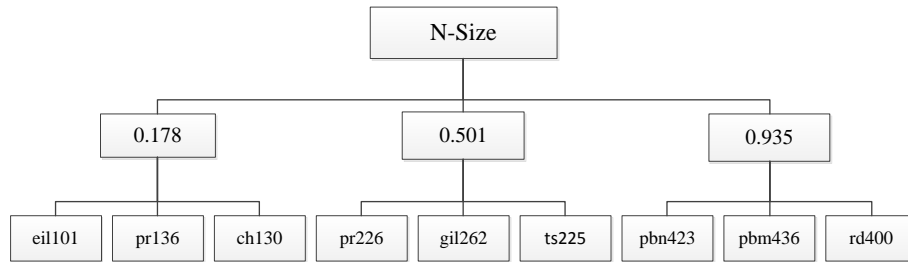


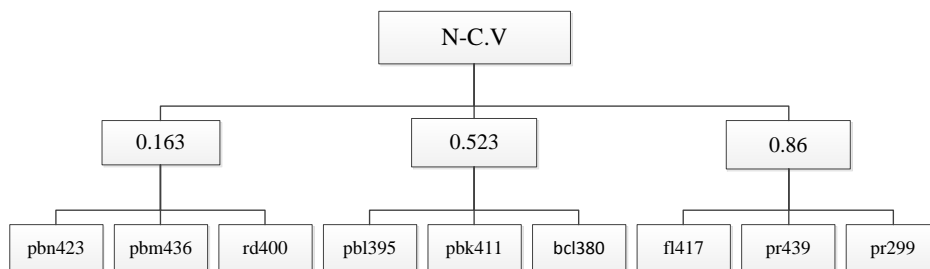Fig. 2: Two-step nested design to experiment on N-Size



Fig. 3: Two-step nested design to experiment on N-CV

ACS algorithm is run for each treatment to collect the data in the randomized order. Then, the collected data are analyzed and model assumption plots are examined. For both responses, transformation for N-Size and N-CV are needed. Power transformation based on the recommended power by Box-Cox plot is performed. Several outliers are detected and replaced with the results of new runs. Sufficient power has been obtained with 10 replicates. With respect to ANOVA test results, the effects of both N-Size and N-CV problem characteristics are statistically significant for Time and R-Gap responses.

Figure 4 (a, b, c, d) illustrates the effect of N-Size and N-CV on Time and R-Gap responses. As can be seen in the plot (a), while N-Size is increased, Time is extremely enlarged; in plot (b), when N-Size is increased, R-Gap is extremely enlarged; in plot (c), when N-CV is increased, Time is extremely declined; in plot (d), in two extreme levels, R-Gap is high and in middle level, R-Gap is extremely declined.
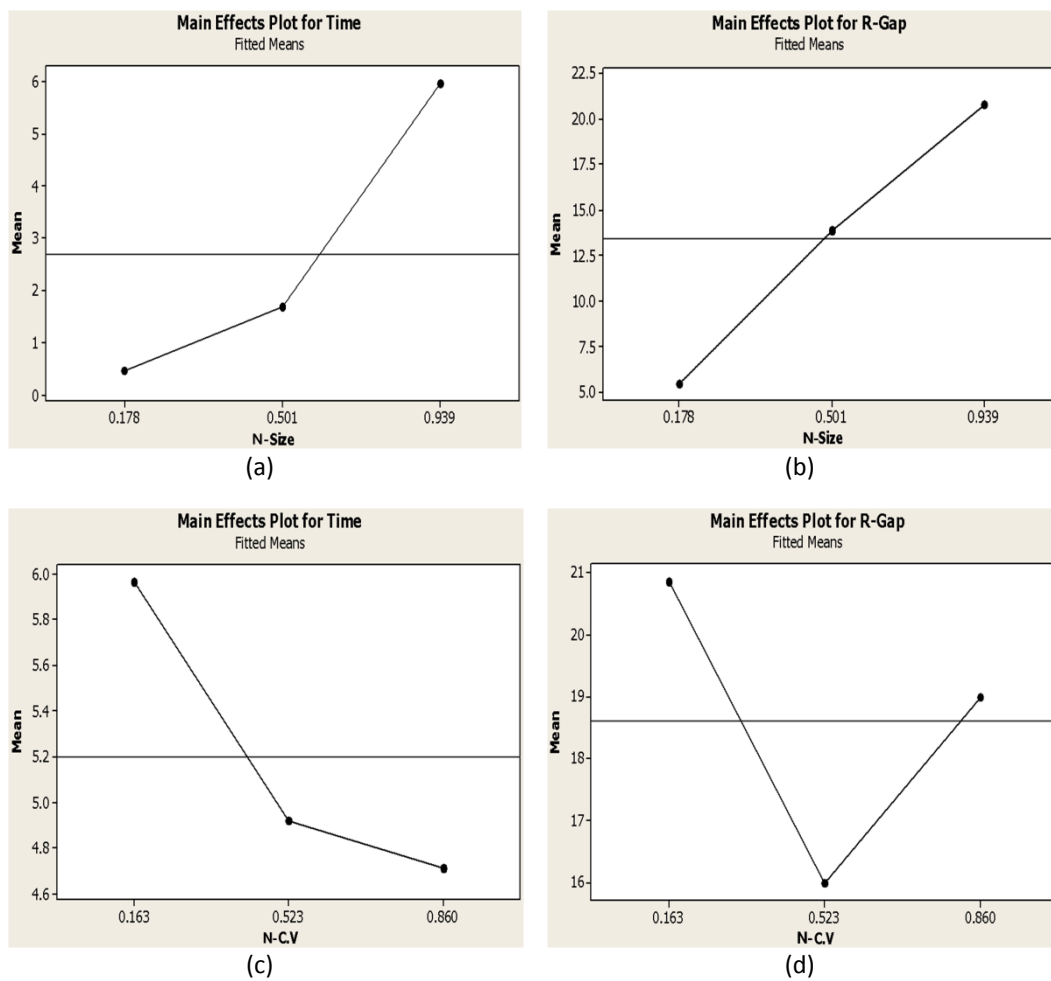
Fig. 4: Main effects of problem characteristics, on the responses

## 5.3. Phase 2: Clustering

*Step 2.1. Cluster Instances*

Figure 5 shows the instances in each the generated clusters based on N-Size and N-CV. In this case, the Two-Step algorithm is used for clustering the instances. For example, in Figure 5, one can see Cluster 1 having high value N-Size and low value N-CV, and Cluster 2 with high value for both problem characteristics.
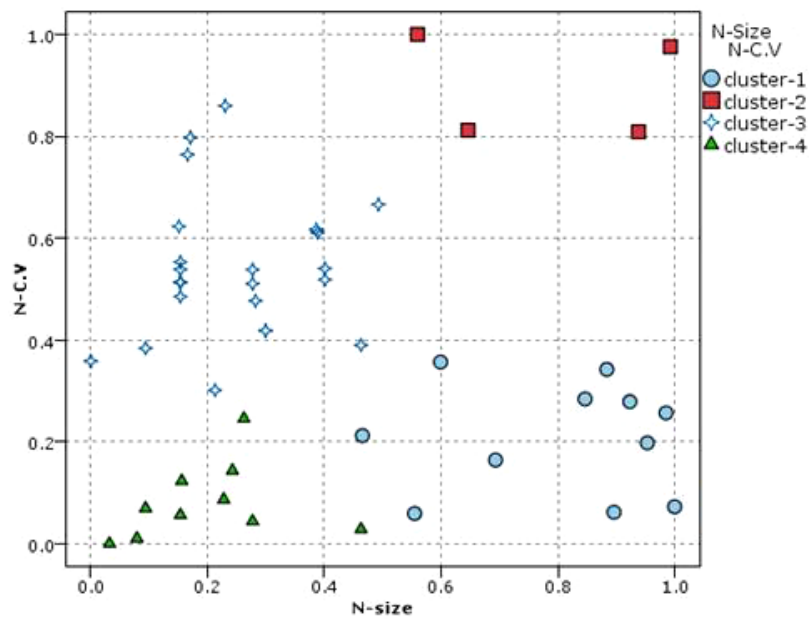
Fig. 5: Clustering the instances based on problem characteristics of N-Size and N-CV

Table 4 summarizes the information of the created clusters. Cluster 3 has the most members and Cluster 2 has the least members. The averages of N-Size and N-CV for each problem characteristics are shown in this table.

Table 4:   Information of the created clusters

| Clusters | No of instances | N-Size average | N-CV average |
|----------|-----------------|----------------|--------------|
| Cluster | 11 | 0.800 | 0.208 |
| Cluster | 4 | 0.784 | 0.899 |
| Cluster | 22 | 0.248 | 0.544 |
| Cluster | 10 | 0.199 | 0.081 |

*Step 2.2. Choose Representative Instances*

In this step, new instances should be chosen on the basis of the average values of each created clusters in Step 2.1. In this case, three instances are selected from each cluster. Table 5 presents the selected instances and their N-Size and N-CV values. It is attempting to select those instances nearest to the both N-Size and N-CV problem characteristics values.

The next phases are performed only for Cluster 3. For other clusters, the steps and phases are the same.

Table 5:   The selected instances

| For cluster | Average N-Size in Cluster | Average N-CV in Cluster | Selected instances | N-Size | N-CV |
|---|---|---|---|---|---|
| **Cluster 1** | 0.800 | 0.208 | bcl380 | 0.847 | 0.284 |
| | | | lin318 | 0.693 | 0.164 |
| | | | pbk411 | 0.923 | 0.279 |
| **Cluster 2** | 0.784 | 0.899 | pr299 | 0.646 | 0.811 |
| | | | fl417 | 0.938 | 0.809 |
| | | | pr439 | 0.993 | 0.976 |
| **Cluster 3** | 0.248 | 0.544 | kroB150 | 0.277 | 0.538 |
| | | | kroA150 | 0.277 | 0.511 |
| | | | pr152 | 0.282 | 0.477 |
| **Cluster 4** | 0.199 | 0.081 | ch130 | 0.228 | 0.087 |
| | | | rd100 | 0.153 | 0.056 |
| | | | eil101 | 0.156 | 0.123 |

### 5.4. Phase 3: Parameter Screening

*Step 3.1. Choose the Parameters and Their Ranges*

Table 6 shows ACS algorithm parameters and their ranges. For each parameter, low and high levels are determined. These ranges are extracted from the literature and pilot studies. These nine parameters are selected for tuning the algorithm by the proposed methodology.

Table 6:  The ACS algorithm parameters and their ranges

| | Tuning parameters | Low level | High level |
|---|---|---|---|
| **A** | Number of Iterations | 100 | 400 |
| **B** | Number of Ants | 4 | 30 |
| **C** | Percentage of Candidate List | 0.1 | 0.8 |
| **D** | Alpha | 1 | 13 |
| **E** | Beta | 1 | 13 |
| **F** | Coefficient of Local Evaporate | 0.01 | 0.99 |
| **G** | Coefficient of Global | 0.01 | 0.99 |
| **H** | q0 | 0.01 | 0.99 |
| **J** | Coefficient of Initial Pheromone | 1 | 1.5 |

*Step 3.2. Analysis of $2^{k-n}$ Fractional Factorial Design*

A fractional factorial design with resolution VI and three blocks (each block for one of the selected instances) is generated. ACS algorithm is run for each treatment to collect the data in the randomized order. Then the collected data are analyzed and model assumption plots are examined. For both responses, transformation for N-Size and N-CV based on recommended power by Box-Cox plot is performed. The outliers are detected and replaced with the results of new runs and sufficient power has been reached with 10 replicates. In ANOVA analysis, for both responses, power transformation is needed and performed based on the recommended power by Box-Cox plot.

*Step 3.3. Rank the Most Important Parameters*

In Tables 7 and 8, the main effective terms are ranked with respect to their sum of squares for both responses based on ANOVA table. Two-interaction terms are removed from these tables. For Time response, Beta has the highest importance. For R-Gap response, Number of Ant has the highest importance. For two of the responses, Coefficient of Initial Pheromone and Global Evaporation have the lowest importance.

Table 7:  Ranked main effects for Time response in Cluster 3

| Rank | | Tuning parameters | Sum of squares | P-Value |
|---|---|---|---|---|
| 1 | E | Beta | 1.405203 | < 0.0001 |
| 3 | H | q0 | 0.407872 | < 0.0001 |
| 4 | B | Number of Ants | 0.198977 | < 0.0001 |
| 6 | C | Percentage of Candidate List | 0.10576 | < 0.0001 |
| 8 | A | Number of Iterations | 0.069171 | < 0.0001 |
| 10 | F | Coefficient of Local Evaporate | 0.048048 | 0.0008 |
| 13 | D | Alpha | 0.028656 | 0.0095 |
| 23 | J | Coefficient of Initial Pheromone | 0.006782 | 0.2065 |
| 43 | G | Coefficient of Global | 9.79E-05 | 0.8793 |

Table 8:  Ranked main effects for R-Gap response in Cluster 3

| Rank | | Tuning parameters | Sum of squares | P-Value |
|---|---|---|---|---|
| 1 | B | Number of Ants | 3.592266067 | < 0.0001 |
| 2 | C | Percentage of Candidate List | 3.217295487 | < 0.0001 |
| 3 | D | Alpha | 1.912241857 | < 0.0001 |
| 4 | A | Number of Iterations | 1.697503602 | < 0.0001 |
| 9 | H | q0 | 0.003790664 | < 0.0001 |
| 20 | F | Coefficient of Local Evaporate | 3.66297E-05 | 0.0254 |
| 23 | E | Beta | 1.73746E-06 | 0.6257 |
| 24 | J | Coefficient of Initial Pheromone | 7.21429E-07 | 0.7533 |
| 25 | G | Coefficient of Global | 3.87085E-07 | 0.8179 |

*Step 3.4. Screen Parameters*

In both responses, Coefficient of Initial Pheromone, Coefficient of Global Evaporation, and Beta for Time are not significant, but they are added to the hierarchical model. In addition, some of the two-interaction terms of these factors are significant.

*Step 3.5. Check for Curvature*

In this step, some the center points are added to the design and the model is checked for lack of fit for the need to quadratic model. In Cluster 3, both responses are quadratic. In the next phase, the results of constructing a quadratic model are presented for Cluster 3 by using response surface modeling.

### 5.5. Phase 4: Response Surface Modeling

*Step 4.1. Augment Previous Design and Reanalyse*

Previously constructed design is augmented with FCC axial points, applying ANOVA test and then fitting a quadratic model to the data.

*Step 4.2. Rank the Most Important Parameters*

The terms in the ANOVA table must be ranked on the basis of their F values. These ranks can be examined together with the related p values for the model terms. For each response, ten most important effects on the basis of F values are ranked in Tables 9 and 10. In Table 9, for R-Gap, the most important interaction effect is EH (Beta*q0). In Table 10, for Time, the most important interaction effect is the CD (Percentage of Candidate List*Alpha).

Table 9:  Ranked important effects of R-Gap response in Cluster 3

| rank | | Term | F-Value | P-Value |
|------|-----|-------------------------------|-----------|-----------|
| 1 | E | Beta | 572.6373 | < 0.0001 |
| 2 | EH | Beta* q0 | 310.4234 | < 0.0001 |
| 3 | H | q0 | 165.6959 | < 0.0001 |
| 4 | B | Number of Ants | 82.97454 | < 0.0001 |
| 5 | F^2 | (Coefficient of Local Evaporate)^2 | 59.85296 | < 0.0001 |
| 6 | E^2 | (Beta)^2 | 47.36844 | < 0.0001 |
| 7 | DF | Alpha* Coefficient of Local | 42.27011 | < 0.0001 |
| 8 | C | Percentage of Candidate List | 41.33599 | < 0.0001 |
| 9 | FH | Coefficient of Local Evaporate* q0 | 39.76342 | < 0.0001 |
| 10 | A | Number of Iterations | 28.93076 | < 0.0001 |

Table 10:  Ranked most important effects of Time response in Cluster 3

| rank | | Term | F-Value | P-Value |
|------|-----|-------------------------------|-----------|-----------|
| 1 | B | Number of Ants | 628080.10 | < 0.0001 |
| 2 | C | Percentage of Candidate List | 570018.30 | < 0.0001 |
| 3 | D | Alpha | 340205.10 | < 0.0001 |
| 4 | A | Number of Iterations | 296963.20 | < 0.0001 |
| 5 | D^2 | (Alpha)^2 | 17930.68 | < 0.0001 |
| 6 | C^2 | (Percentage of Candidate List)^2 | 5529.97 | < 0.0001 |
| 7 | B^2 | (Number of Ants)^2 | 5450.18 | < 0.0001 |
| 8 | CD | Percentage of Candidate List* | 4247.63 | < 0.0001 |
| 9 | A^2 | (Number of Iterations)^2 | 1345.33 | < 0.0001 |
| 10 | H | q0 | 602.10 | < 0.0001 |

*Step 4.3. Examine Model Graphs*

In this step, the graphs of the responses for all parameters can be studied. In this section, two samples of these plots are presented. Figure 6 illustrates an interaction contour plot for Time response. In this figure, one can observe that for reducing Time, Percentage of Candidate List and Alpha must be simultaneously reduced. When the Percentage of Candidate List is low, effect of Alpha is negligible.
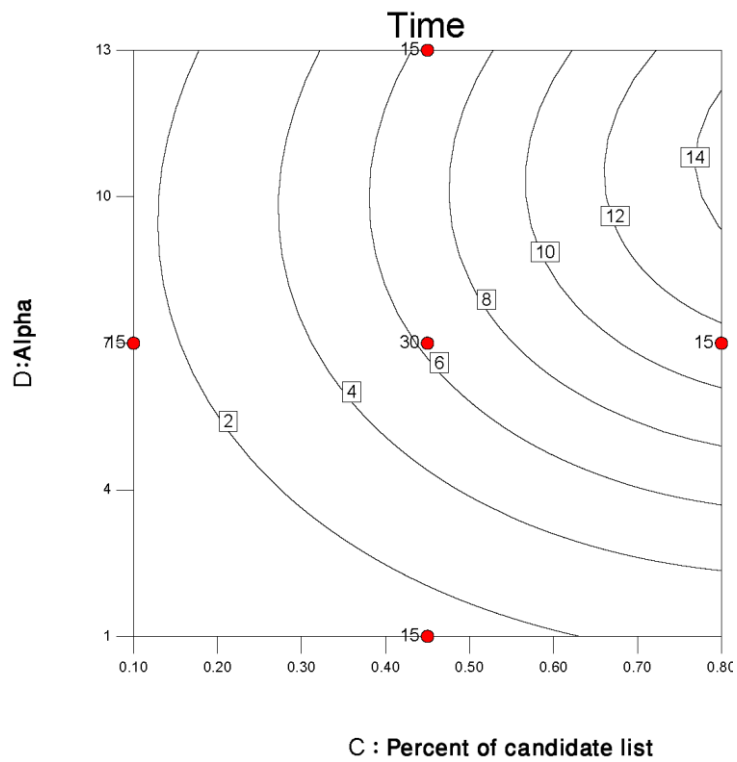
Fig. 6:   The contour plot of Percentage of Candidate List* Alpha (CD) interaction for Time
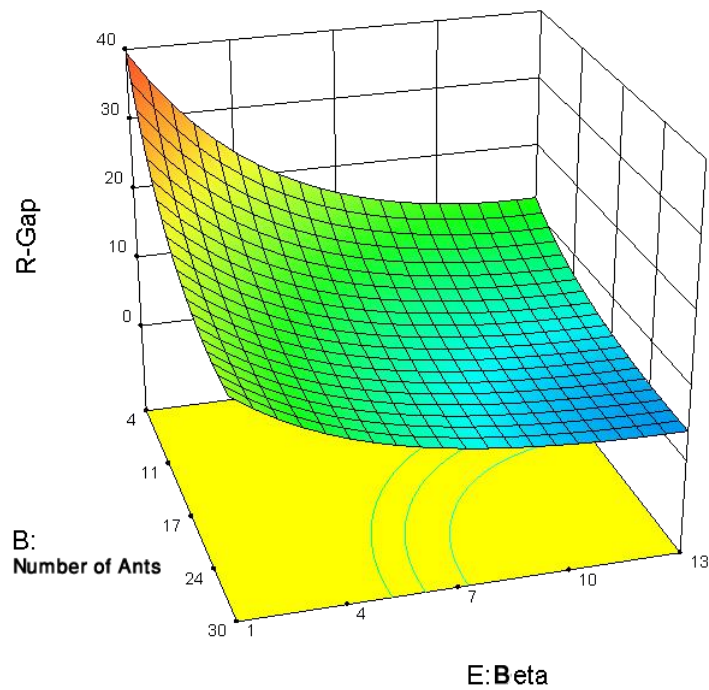


Fig. 7:  The 3D surface plot of Beta*Number of Ants (EB) interaction for R-Gap

In Figure 7, 3D surface plot of Beta*Number of Ants (EB) interaction for R-Gap response is presented. From this plot, it can be concluded to decrease R-Gap, Beta and Number of Ant must be simultaneously increased.

### 5.6. Phase 5: Optimization

*Step 5.1. Optimize the Tuning Parameters*

For each of the created clusters in Phase 2, a numerical optimization of desirability is performed by using the Nelder-Mead Simplex. The optimization aim is to minimize the R-Gap and Time responses. As R-Gap is more important than Time, in desirability function, the importance of R-Gap is set at five and that of Time at one.

*Step 5.2. Choose the Best Solution*

In this step, the solution with the maximum desirability is selected, and the others are discarded. In Table 11, the best parameters for each cluster are presented. It can be seen that for important parameters, different values are concluded for different clusters.

Table 11: Best found parameters for each cluster

| Label | Tuning parameters | Cluster1 | Cluster2 | Cluster3 | Cluster4 |
|-------|-------------------|----------|----------|----------|----------|
| A | Number of Iterations | 264 | 252 | 247 | 269 |
| B | Number of Ants | 25 | 28 | 30 | 30 |
| C | Percentage of Candidate List | 0.11 | 0.1 | 0.1 | 0.79 |
| D | Alpha | 1 | 1 | 2 | 1 |
| E | Beta | 12 | 13 | 13 | 13 |
| F | Coefficient of Local Evaporate | 0.54 | 0.36 | 0.53 | 0.50 |
| G | Coefficient of Global | 0.99 | 0.99 | 0.96 | 0.97 |
| H | q0 | 0.99 | 0.09 | 0.02 | 0.03 |
| J | Coefficient of Initial | 1.48 | 1.01 | 1 | 1 |

Here, a sample plot of the results is presented. Figure 8 shows the 3D surface of overall desirability for different values of Beta and Number of Ant for both responses. As can be observed for getting the best desirability, Beta and Number of Ant must be simultaneously increased.
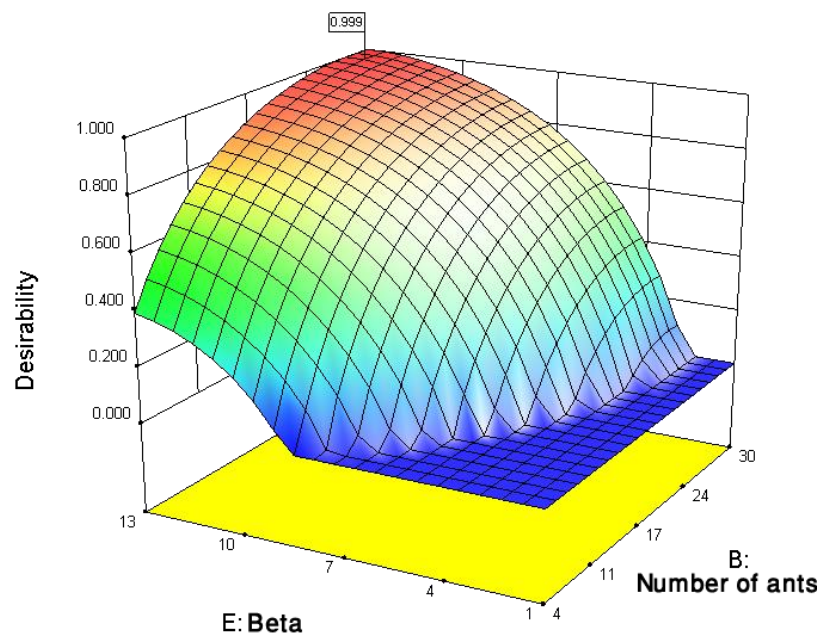
Fig. 8:  The 3D surface of overall desirability for Beta* Number of Ants (EB) interaction

## 6. Validation

For validation of the proposed methodology, ACS is run for solving TSP with 4 parameter settings and then the averages of the ten runs are compared. It is to be noted that twelve problems used for parameter tuning of the four generated clusters are not included in the validation procedure. In addition the three problems that are used in parameter tuning without clustering are removed from problem set. Therefore, 32 instances are applied for validation. The different parameter settings are as follows.

1) Tuned parameters by the proposed methodology.
2) Tuned parameters by DOE without clustering for the three problems selected closer to the center of all problems.
3) The parameter values are determined on the basis of the proposed parameters for ACS by Dorigo [14].

The parameters are chosen by random selection of parameters in their specific ranges.

To compare these four parameter settings, the t-paired test is used and the differences between the proposed methodology and the others for two responses are compared. Table 18 summarizes the results of all comparisons. The null hypothesis is that there is no significant difference between two parameter settings, while the alternative hypothesis is that two parameter settings are not equal.

Table 12: Comparisons between parameter settings

| The proposed methodology vs. | Other parameter setting | Response | Mean difference | 95% CI for mean Difference | P-Value | Null hypothesis |
|---|---|---|---|---|---|---|
| | DOE without clustering | R-Gap | -2.349 | (-2.984; -1.714) | 0.000 | Rejected |
| | | Time | 1.097 | (-0.290; 2.484) | 0.117 | Not rejected |
| | Dorigo parameters | R-Gap | -6.272 | (-7.662; -4.881) | 0.000 | Rejected |
| | | Time | 3.932 | (2.173; 5.691) | 0.000 | Rejected |
| | Random parameter | R-Gap | -9.969 | (-11.330; -8.609) | 0.000 | Rejected |
| | | Time | 0.593 | (-0.804; 1.991) | 0.393 | Not rejected |

As can be seen in Table 12, the proposed methodology has better results for the response R-Gap in comparison to the other parameter settings. For the Time response, the proposed methodology is not significantly different with "DOE without clustering" and "Random" parameter setting. Nevertheless, Time response to the proposed parameters by Dorigo is preferred to the proposed methodology. However, with respect to a negligible difference between the proposed methodology and the Dorigo parameters in Time response and low importance of Time in comparison to R-Gap, the proposed methodology has better results in general.

## 7. Conclusions

In this paper, a methodology is presented for parameter tuning of the metaheuristic algorithms based on DOE. The proposed methodology has five phases, namely, Problem Characteristics Screening, Clustering, Parameter Screening, Response Surface Modeling and Optimization. In the first phase, the problem characteristics that may affect metaheuristic algorithm are screened. In the second phase, the instances are clustered based on affecting problem characteristics. Clustering phase provides a set of clusters based on the most important problem characteristics. Parameter screening phase produces a reduced set of the most important tuning parameters. The mathematical functions relating to the tuning parameters to every of the responses of concern are given by response surface modeling phase. In the last phase, the parameter values with respect to the importance of each response are optimized. For evaluation, the proposed methodology is applied to an ACS algorithm for solving 47 TSP problems. For validation of the proposed methodology, the ACS algorithm is run for solving TSP with three other parameter settings including tuned parameters by DOE without clustering, the proposed parameters by Dorigo, and random parameters. Then, the results of the proposed methodology are compared with those of the parameter settings. For the response R-Gap, the proposed methodology has better results in comparison to other parameter settings. For Time response, the proposed methodology has not significant difference with "DOE without clustering" and "Random" parameter settings. Nevertheless, Time response for the Dorigo parameters is better than the proposed methodology. However, with respect to a negligible difference between the results of the proposed

methodology and Dorigo parameters for Time response and low importance of Time in comparison to R-Gap, the methodology has better results in general.

## References

[1] Talbi E-G. Metaheuristics: from design to implementation: John Wiley & Sons, 2009.

[2] Ridge E. Design of experiments for the tuning of optimisation algorithms: University of York, Department of Computer Science, 2007.

[3] Ridge E, Kudenko D, editors. Sequential experiment designs for screening and tuning parameters of stochastic heuristics. Workshop on Empirical Methods for the Analysis of Algorithms at the Ninth International Conference on Parallel Problem Solving from Nature (PPSN), 2006.

[4] Ridge E, Kudenko D. Tuning the Performance of the MMAS Heuristic. Engineering Stochastic Local Search Algorithms Designing, Implementing and Analyzing Effective Heuristics: Springer, 2007: 46-60.

[5] Ridge E, Kudenko D. Determining whether a problem characteristic affects heuristic performance. Recent Advances in Evolutionary Computation for Combinatorial Optimization: Springer; 2008: 21-35.

[6] Ridge E, Kudenko D. Tuning an algorithm using design of experiments. Experimental methods for the analysis of optimization algorithms: Springer, 2010: 265-86.

[7] Coy SP, Golden BL, Runger GC, Wasil EA. Using experimental design to find effective parameter settings for heuristics. Journal of Heuristics, 2001; 7(1):77-97.

[8] Caserta M, Rico EQ. A cross entropy-Lagrangean hybrid algorithm for the multi-item capacitated lot-sizing problem with setup times. Computers & Operations Research. 2009; 36(2): 530-48.

[9] Parsons R, Johnson ME. A case study in experimental design applied to genetic algorithms with applications to DNA sequence assembly. American Journal of Mathematical and Management Sciences. 1997; 17(3-4): 369-96.

[10] Van Breedam A. Improvement heuristics for the vehicle routing problem based on simulated annealing. European Journal of Operational Research. 1995; 86(3): 480-90.

[11] Park M-W, Kim Y-D. A systematic procedure for setting parameters in simulated annealing algorithms. Computers & Operations Research. 1998; 25(3): 207-17.

[12] Adenso-Diaz B, Laguna M. Fine-tuning of algorithms using fractional experimental designs and local search. Operations Research. 2006; 54(1): 99-114.

[13] Fığlalı N, Özkale C, Engin O, Fığlalı A. Investigation of ant system parameter interactions by using design of experiments for job-shop scheduling problems. Computers & Industrial Engineering. 2009; 56(2): 538-59.

[14] Dorigo MS. T. Ant Colony Optimization. Cambridge: MIT Press, 2004.

[15] Hutter F, Hoos HH, Leyton-Brown K, Stützle T. ParamILS: an automatic algorithm configuration framework. Journal of Artificial Intelligence Research. 2009; 36(1): 267-306.

[16] Hutter F. Automated configuration of algorithms for solving hard computational problems. Dissertation, University of British Columbia, 2009.

[17] Birattari M, Stützle T, Paquete L, Varrentrapp K, editors. A Racing Algorithm for Configuring Metaheuristics. GECCO, 2002.

[18] Birattari M. The Problem of Tuning Metaheuristics. Dissertation, Université Libre de Bruxelles, 2006.

[19] Birattari M. Tuning metaheuristics: a machine learning perspective: Springer, 2009.

[20] Balaprakash P, Birattari M, Stützle T. Improvement strategies for the F-Race algorithm: Sampling design and iterative refinement. Hybrid Metaheuristics: Springer; 2007: 108-22.

[21] Srinivasan A, Ramakrishnan G. Parameter screening and optimisation for ILP using designed experiments. The Journal of Machine Learning Research. 2011; 12: 627-62.

[22] Montgomery DC. Design and analysis of experiments: John Wiley & Sons, 2008.

[23] Myers RH, Montgomery DC, Anderson-Cook CM. Response surface methodology: process and product optimization using designed experiments: John Wiley & Sons, 2009.

[24] Montgomery DC. Applied Statistics and Probability for Engineers 6th edition: Wiley, 2013.

[22] S Kakooei, MC Ismaila, B Ari-Wahjoedia, Electrodeposition of Iridium Oxide by Cyclic Voltammetry: Application of Response Surface Methodology, MATEC Web of Conferences, 2014; 13: 04024.

[23] J Rouhi, S Mahmud, SD Hutagalung, S Kakooei, Optimisation of nanooxide mask fabricated by atomic force microscopy nanolithography: a response surface methodology application, Micro & Nano Letters, IET, 2012; 7(4): 325-328.

[24] H Hanaei, FRB Ahmadun, E Mohammadpour, S Kakooei, Optimization of Carbon Nano Tubes Synthesis using Fluidized bed Chemical Vapor Deposition: A Statistical Approach, Caspian Journal of Applied Sciences Research, 2013; 2(3): 138-147.

[25] Merriam-Webster D. Merriam Websters Collegiate Dictionary by Merriam-Webster, 2003.

[26] Han J, Kamber M, Pei J. Data Mining: Concepts and Techniques, (The Morgan Kaufmann Series in Data Management Systems), 2006.