IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 25, 2019, accepted November 9, 2019, date of publication November 14, 2019, date of current version December 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953528

A Design of Experiments Comparative Study on Clustering Methods

NATÁLIA MARIA PUGGINA BIANCHESI, ESTEVÃO LUIZ ROMÃO, MARINA FERNANDES B. P. LOPES, PEDRO PAULO BALESTRASSI, AND ANDERSON PAULO DE PAIVA

Industrial Engineering Institute, Federal University of Itajubá, Itajubá 37500-015, Brazil

Corresponding author: Pedro Paulo Balestrassi (ppbalestrassi@gmail.com)

This work was supported in part by the Brazilian Agencies of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

ABSTRACT Cluster analysis is a multivariate data mining technique that is widely used in several areas. It aims to group automatically the n elements of a database into k clusters, using only the information of the variables of each case. However, the accuracy of the final clusters depends on the clustering method used. In this paper, we present an evaluation of the performance of main methods for cluster analysis as Ward, K-means, and Self-Organizing Maps. Differently from many studies published in the area, we generated the datasets using the Design of Experiment (DOE) technique, in order to achieve reliable conclusions about the methods through the generalization of the different possible data structures. We considered the number of variables and clusters, dataset size, sample size, cluster overlapping, and the presence of outliers, as the DOE factors. The datasets were analyzed by each clustering method and the clustering partitions were compared by the Attribute Agreement Analysis, providing invaluable information about the effects of the considered factors individually and about their interactions. The results showed that, the number of clusters, overlapping, and the interaction between sample size and number of variable significantly affect all the studied methods. Moreover, it is possible to state that the methods have similar performances, with a significance level of 5%, and it is not possible to affirm that one outperforms the others.

INDEX TERMS Clustering methods, design of experiments, K-means, self-organizing maps, ward.

I. INTRODUCTION

Cluster analysis, also known as unsupervised classification, is a multivariate statistical data mining technique [1]–[3], based only on variable information that aims to separate a set of objects into different clusters in which each one must contain similar objects according to some distance function statistics and, at the same time, dissimilar to the objects of other clusters. In other words, the result obtained from the application of this method is a set of clusters with internal cohesion and external isolation [4].

The first published record on a clustering method was made by Sorensen in 1948 [5]. Since then, methods of clustering analysis have been developed due to the need to analyze the large amount of data collected in various areas of knowledge [4], e.g.: marketing, identifying market share; medicine, identifying patients with a common disease cause; education, measuring psychological characteristics to identify groups of students that need special attention; biology, building a taxonomy of groups and subgroups of similar plants; climatology, providing new insights into climatological and environmental trends.

Clustering applications examples found in the literature also demonstrate this diversity. There are papers about the use of clustering algorithms to identify characteristics of people with attempted suicide [6]; to facilitate the diagnosis and treatment of cancer [7]; to identify residential and social patterns of homeless adults [8]; and also in applications in the field of production engineering, e.g., clustering method for production planning [9], [10], and for analyzing product portfolios [11].

In general, whenever large amounts of information need to be classified into a small number of categories, clustering analysis can be useful. Researchers are often faced with the problem of clustering data into meaningful structures. In this context, an important aspect to be considered is the method used to measure the similarity among the elements in order to identify if they should be in a same cluster.

The associate editor coordinating the review of this manuscript and approving it for publication was Qi Zhou.

The accuracy of the final categories depends on the method used to classify the objects. An improper choice of the clustering method may compromise the results obtained. Then, there is a growing concern to make the methods suitable for certain situations and also fewer complexities.

Some studies were conducted to evaluate the performance of some methods for cluster analysis as Self-Organizing Methods (SOM), hierarchical, and non-hierarchical clustering methods. A deeper theory about these method can be found in [12], [13], and [4], respectively.

Therefore, the purpose of this paper is to present a comparative study among the performance of main clustering methods. However, differently from many studies published in the area, this study includes some factors and interactions that have not been studied before, and presents an innovative approach to generalize the datasets.

Unlike the other papers, this one does not consider an usual approach of trial and error to generalize datasets, what could lead to restricted conclusions without a formal and larger analysis. Instead, we considered the DOE technique that can achieve more reliable results.

Using the DOE technique it is possible to simulate synthetic datasets and to evaluate the performance of each method, identifying some parameters that most affect their results, and verifying which one presents the best result.

In terms of the rest of this paper, Section 2 demonstrates some related works where traditional clustering methods are compared with SOM or used together to achieve a better result. Section 3 consists on a brief background review about cluster analysis methods and neural network. Section 4 contains a brief explanation about DOE and how was constructed the experimental design. In section 5 we presented the results of the experimental design analysis. Finally, section 6 presents the conclusion of this paper and some discussions.

II. RELATED WORK

As aforementioned, clustering methods are very important in different research areas and are explored by several authors with the goal of demonstrate which method has a better performance.

Mangiameli *et al.* [14] compared SOM with non-hierarchical (Ward) algorithm and several agglomerative hierarchical methods, including centroid, single, complete and average. The data were generated in a normal distribution without correlation between the variables and considering the factors: number of cluster, number of variables, overlapping, the presence of irrelevant variables, and outliers. A total of 252 data sets were generated, containing 50 observations in each cluster. In general, the results showed that the mean recovery rate decreases as the number of clusters and the degree of intracluster dispersion increase. In addition, in all tested situations the SOM algorithm presents a superior result to the hierarchical methods.

Mingoti and Lima [15] compared SOM with hierarchical methods (single, complete, centroid, average and Ward), K-means and Fuzzy methods. The considered factors were number of clusters, number of variables, correlation degree, overlapping, and the presence of outliers. The dataset size was defined as 500 observations for all experiments. In general, the cluster analysis was more affected by the overlapping than by the outliers and the SOM method suffered the greatest performance variations.

Balakrishnan *et al.* [16] compared SOM with K-means method using a data simulation procedure. The data were simulated according to a normal distribution without correlation between the variables and considering three factors: numbers of clusters, number of variables, and an error in the distance matrix. The authors generated a total of 108 data sets, containing 50 observations per cluster. In general, the k-means method presented a better result than SOM for all factors. For both methods the best results were for small number of cluster and high number of clusters.

Waller *et al.* [17] compared SOM with hierarchical (Ward) and non-hierarchical (K-means) methods. The data were generated in a normal multivariate distribution, considering the number of cluster, number of variables, intracluster correlation, degree of non-overlapping, and the level of data dispersion in each cluster. They generated 480 data with five replicates and the number of samples in each cluster was determined by a uniform distribution between 10 and 50. The results showed that intracluster correlation and the data dispersion had little influence on clustering accuracies. Overall, all methods showed a similar performance and the SOM was the most stable. None of the methods performed well on high number of clusters and all methods presented better results as the number of variables increases.

In general, the studies present that the performance of the clustering methods is worse as the number of clusters, the number of variables and the degree of overlapping increases. The results also show that the intracluster correlation [17], [15], the data dispersion [17], and the error in the distance matrix [16] had very little influence on clustering precision.

In addition, the results reveal contradictory opinions about the presence of outliers. Some authors suggest that outliers had little influence on the performance of the clustering methods [14]. Others imply that the presence of outliers affects the clustering performance [15].

Another divergence in the research is in relation to the performance of the SOM method. For some authors, SOM is the clustering method which presents the best performance [14], [17]. However, others disagree and defend the idea that SOM is the method that is the most affected by the variations of data structures and the one that presents the worst performance [15], [16].

Nevertheless, 1) no study evaluated the effect of the dataset sample size; 2) no study evaluated the effect of the clusters being the same or different size; 3) the results did not consider the effects of interactions among the factors; and 4) none of the studies used DOE to simulate datasets and to evaluate the performance of the clustering methods.

TABLE 1. Clustering methods.

Non-Hierarchical	Hierarchical
Self-Organizing Maps	Single Linkage
Fuzzy clustering algorithms	Complete Linkage
	Average Linkage
K-medoids	Centroid Linkage
K-means	Ward

III. CLUSTERING METHODS

A formal definition of the clustering problem can be found in [18]. Given a set of p elements $X = \{X_1, X_2, \ldots, X_p\}$, the problem of clustering consists of obtaining a set of k clusters, $G = \{G_1, G_2, \ldots, G_k\}$. The set G is considered a clustering with k clusters if the conditions in (1), (2), and (3) are fulfilled:

$$\bigcup_{i=1}^{k} G_i = X \tag{1}$$

$$G_i \neq 0, \quad \text{para } 1 \leq \text{i } \leq \text{k} \tag{2}$$

$$G_i \cap G_j = 0, \quad \text{para } 1 \leq \text{i, j} \leq \text{k e i} \neq \text{j} \tag{3}$$

It is emphasized by the conditions that an element cannot belong to more than one cluster and each cluster must have at least one element.

In a k-cluster problem, in which k is given as parameter for the solution, the total number of different forms of clustering p elements of a set in k clusters is shown in (4):

$$N(p,k) = \frac{1}{k!}\sum_{i=0}(-1)^i \binom{k}{i}(k-i)^p \tag{4}$$

Considering (4), the number of possible solutions to a k-cluster problem has an exponential growth. For example, for a combination of 10 elements in 2 clusters and 100 elements in 2 clusters, there are respectively 511 and 6.34×10^{29} different ways of combining the elements. Thus, it demonstrates the complexity in finding the best clustering solution within the available solutions and methods.

The existing methods for the solution of clustering problems can be classified, in general, in hierarchical and non-hierarchical methods [19], [20], as shown in Table 1.

Among the methods in Table 1, we chose to study a traditional hierarchical and non-hierarchical method, and the SOM method, which is a more recent ANN approach for clustering problems.

The hierarchical method chosen is the Ward. It is seen as one of the best techniques for measuring distances between clusters [21]. Mangiameli *et al.* [14] suggest that the Ward method should always be employed because it presents better results among the hierarchical methods. Mingoti and Lima [15] add that the Ward method is more stable and easier to implement.

The traditional non-hierarchical method that will be used in this paper is the K-means. It is the probably most well-known non-hierarchical method [22], [23], and it is also the most commonly due to the ease of implementation, simplicity, efficiency, and empirical success [24].

In the following sections, the Ward, K-means and SOM methods will be explained respectively, in which will be presented their concepts and procedures.

A. THE HIERARCHICAL METHOD WARD

The hierarchical methods are simple techniques where the data are partitioned successively, producing a hierarchical representation of the clustering of each stage, that can be represented by a graphical called dendogram [4].

Hierarchical techniques are the most widely disseminated clustering method [25], and involve, basically, two steps. The first one refers to the estimation of a measure of similarity between individuals, whereas the second one is related to clustering the objects into subgroups based on a joining technique [26].

The most common metrics used to measure the similarity between individuals are the Euclidean distance, the Mahalanobis distance and the Manhattan distance. Considering the efficiency of the metrics and the viability of application [27], we will use the Euclidean distance for calculate the similarity matrix. The Euclidean distance is the geometric distance in the dimensional space and is it can be written as shown in (5) [1].

$$d_{ij} = \left[\sum_{k=1}^{p}\left(x_{ik} - x_{jk}\right)^2\right]^{1/2}, \quad \text{i} \neq \text{j} \tag{5}$$

where x_{ik} and x_{jk} are, respectively, the k^{th} variable value of the p-dimensional observations for elements i and j.

The joining technique that we will use, as aforementioned, is the Ward method. It is an agglomerative clustering method that searches for partitions that minimize the error associated with each cluster [26]. This error is defined in (6):

$$ESS_k = \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2 \tag{6}$$

where k is the cluster, n is the total number of objects in the cluster k, and x_i is the ith object of the cluster k. Thus, it is possible to perform the complete hierarchical clustering analysis.

In general, the procedure of the hierarchical agglomerative clustering method can be described in a few steps [4]:

1) Start: each of n elements is considered as a unique cluster.
2) Clusters are compared to each other by using a distance measure.
3) A new clustering is formed by joining the clusters with smaller distance.
4) This procedure is repeated several times until all the elements are grouped according to the desired number of clusters.
5) Only two clusters can be joined at each stage and they cannot be separated later.

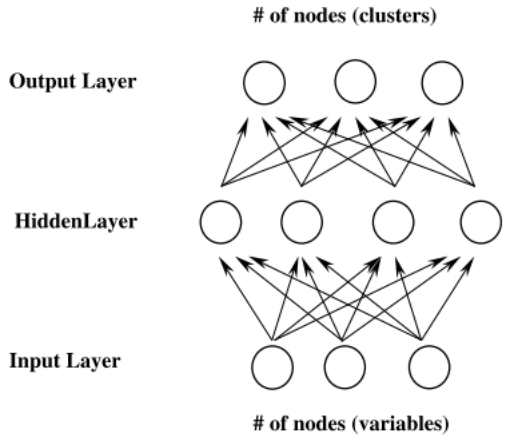

FIGURE 1. Illustration of a neural network for clustering.

B. THE NON-HIERARCHICAL METHOD K-MEANS

Non-hierarchical methods seek to find the best partition of n elements in a number of clusters k pre-defined.

In addition to the pre-definition of the number of clusters k, the K-means algorithm requires the initial definition of the seed of each cluster, which can be performed automatically by the algorithm or be indicated by the user.

The basic steps of the K-means algorithm are [4]:

1) Select n elements to be the initial seed of the k clusters.
2) Each element is associated to the cluster that has the smallest Euclidean distance (5) between its seed and the element.
3) The seed of each cluster is recalculated by using the average vector of the elements belonging to the cluster;
4) The procedure is then repeated until the seeds of the clusters stabilize.

Therefore, the accuracy of the K-means procedure is very dependent upon the choice of the initial seed [21].

C. THE ARTIFICAL NEURAL NETOWORKS SOM

One of the more important ANN is the Self-Organization Maps (SOM) developed by Finn Teuvo Kohonen in the early 1980s for solving clustering problems [12].

The SOM constitutes a class of artificial neural networks based on competitive learning that uses competition as a way of learning to make the adjustments of the weights. Another important feature of this algorithm is that it uses unsupervised training, in which the clustering algorithm is based only on the similarity of data without using historical data.

A map of Kohonen is usually restricted to a one-dimensional or two-dimensional arrangement of neurons that provide a topology-preserving mapping from the input space to the clusters. The number of input elements depends on the database that are being used and each output unit represents a cluster. Between the input and the output layer there is a hidden layer. The output of each layer is an input of the next one as shown in Fig. 1 [15], [28].

In clustering problems, the ANN input elements can be divided into two main stages [15]. The first one is called training, in which a specific dataset is used to train the network with the training set to adjust the weights. During the training process, an optional validation set can be used for tuning the parameters of a model and avoiding over-fitting problem. Once the training stage is completed and the ANN weights are adjusted, the testing stage begins, where the ANN is used to classify unseen data and evaluate the ANN performance.

The SOM training algorithm can be decomposed in four phases [12]:

i The map initialization.
ii The competitive process.
iii The cooperative process.
iv The synaptic adaptation.

The map initialization consists of assigning a vector of weights to the connections between neurons of the input and output layers. The choice of the vector of weights can be done by assigning small random values, in this way no previous order is imposed on the map. Thus, an input vector is presented to the network and each neuron receives this object and calculates its activation level. The activation level is represented by the proximity value between the input vector and each output neuron j. It is measured by the Euclidean distance d_j in (7).

$$d_j = \left[\sum_{j=1}^{N-1} \left(x_i(t) - w_{ij}(t) \right)^2 \right]^{1/2} \tag{7}$$

where $x_i(t)$ is the input vector for the neuron i at time t, $w_{ij}(t)$ is the weight vector between the input neuron i and the output neuron j at time t, and t is the number of iteration that can at least 1000.

In the competitive process, the neurons compete with each other through the activation levels, being only one neuron the winner. The winning neuron is the one that vector of weights has the smallest Euclidean distance, i.e., the winning neuron is the one that has a minimum value d_j.

The cooperative process implies in the influence that the winning neuron exerts on the neighboring neurons. The winning neuron determines the spatial location of a topological neighbor of excited neurons, providing the basis for cooperation between neighboring neurons. The winning neuron tends to excite more the immediate neighboring neurons than those neurons that are more distant. Thus, the topological neighborhood around the winning neuron decays smoothly with lateral distance.

For example, let $h_{j,i}$ be the topological neighborhood centered on the winning neuron i and surrounded by a set of cooperative excited neurons of which a typical neuron is denoted by j, and $d_{j,i}$ be the lateral distance between the winner neuron i and the neuron j. Then, it can be assumed that the topological neighborhood $h_{j,i}$ is a unimodal function of the lateral distance $d_{j,i}$, satisfying two requirements [29]:

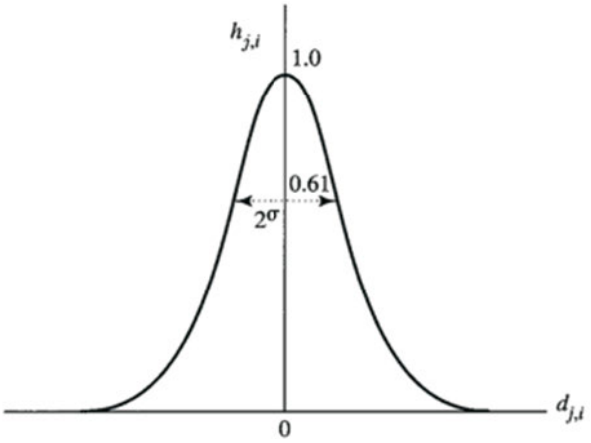

FIGURE 2. Gaussian function.

1) The topological neighborhood $h_{j,i}$ is symmetric and maximum around the winning neuron defined by $d_{j,i} = 0$;
2) The amplitude of the topological neighborhood $h_{j,i}$ decreases monotonically with increasing lateral distance $d_{j,i}$ and effective width, decaying to 0 for $d_{j,i} \to \infty$, which is a necessary condition for convergence. So, if $d_{j,i} = 0$, then $h_{j,i}$ is maximum and as $d_{j,i} \to \infty$, $h_{j,i}$ becomes 0.

Another important feature of this algorithm is that the width of the neighborhood decreases in function of the number of iterations t. Therefore, a good option for $h_{j,i}$ could be the Gaussian function variant with time as shown in (8) and in Fig. 2 [29]. When $d_{j,i} = 0$, $h_{j,i}$ maximum value is 1 and as it goes to this direction and the other direction, its value finally decreases to 0.

$$h_{j,i}(t) = \exp\left(\frac{d_{j,i}^2}{2\sigma^2(t)}\right), \quad t = 0, 1, 2, \ldots \tag{8}$$

where σ(t) is the "effective width" of the topological neighborhood defined in (9).

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\tau_1}\right) \tag{9}$$

where σ_0 is the initial value of σ, and τ_1 is a constant of time.

Thus, as the number of iterations increases the width σ(t) decreases at an exponential rate and the topological neighborhood is reduced.

The neighborhood function $h_{j,i}(t)$ should initially include almost all network neurons centered on the winning neuron i and then shrink slowly over time until be only the neuron itself. Assuming the use of a two-dimensional map, it is possible to adjust the initial size σ_0 equal to the "radius" of the map.

The last phase is the synaptic adaptation. Once the winning neuron and its neighbors are determined, their weight vectors (w) are updated and approximated of the input object x. However, this approximation decreases as the corresponding neuron is further from the winner neuron i. For every neuron within the topological neighborhood of the winner i, the weight vectors are modified according to (10).

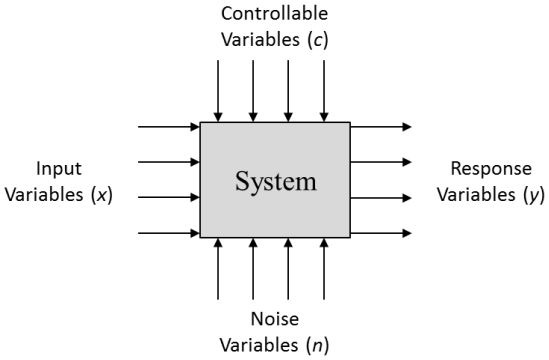

FIGURE 3. Model of experimental system.

$$w_j(t+1) = w_j(t) + \eta(t)\, h_{j,i}(t)\,(x - w_i(t)) \tag{10}$$

where η is the learning rate of the algorithm. It has its value changed at each iteration t as shown in (11):

$$\eta(t) = \eta(0) \exp\left(-\frac{t}{\tau_2}\right), \quad t = 0, 1, 2, \ldots \tag{11}$$

where $\eta(0)$ is the initial value of η, and τ_2 is another constant of time.

Haykin [29] suggests that the parameter $\eta(t)$ should start with a value close to 0.1, decreasing gradually but remaining above 0.01. These values are reached through the following parameters in (12) and (13).

$$\eta(0) = 0.1 \tag{12}$$

$$\tau_2 = 1000 \tag{13}$$

The stages 2 to 4 are repeated as a new object is presented to the map. As the network trains the neurons with the same characteristics begin to approach each other. Thus, similar elements are positioned close to each other forming a topology with a gradient of characteristics.

IV. EXPERIMENTAL DESIGN

A. DESIGN OF EXPERIMENTS

DOE is an invaluable technique where experiments are planned and the data are analyzed by statistical methods, resulting in valid and objective conclusions [30].

An experiment can be defined as a series of tests in which a set of input variables or factors (x) are changed by the experimenter in a controlled way (c) to observe and identify how the responses (y) of that system are affected due to these changes [31], as shown in Fig. 3 [30]. It allows an understanding of which factors are significant and how they interact with each other.

The simplest experimental design model is the 2^m factorial, with m factors at 2 levels each. These levels, -1 and $+1$, represent the lower and the upper limits for the interval in which the variable is analyzed [31].

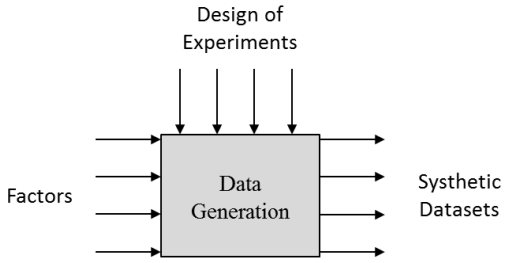

FIGURE 4. Model of experimental system for simulation.

The relation between factors and response can be established according to some modeling algorithm by combining techniques such as regression model, analysis of variance and hypothesis testing. The Ordinary Least Square (OLS) is an algorithm typically used for estimating coefficients of a regression model in order to minimizing the sum of square differences between the observed and predicted values, which means, minimizing the error ε_i.

A linear regression model estimated by using this algorithm can be viewed in (14) [30].

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{21} + ... + \beta_k x_{ik} + \varepsilon_i \qquad (14)$$

where y_i is the responses for the experiment i; $x_i, x_{i2}, \ldots, x_{ik}$ are input variables; β_0 is a constant; $\beta_1, \beta_2, \ldots, \beta_k$ are coefficients to be estimated; and ε_i is the error.

DOE is a commonly used technique for process to find the optimal and robust solution y through the combination of variables x [32], [33]. However, the DOE technique can be used for others purpose, e.g., it can be applied in a simulation problem. It increases the transparency of simulation model behavior and the effectiveness of reporting simulation results [34]. Furthermore, it allows controlling the factors that will be used in the simulation and present better and faster results than trial and error simulation. Therefore, DOE is a useful and necessary part of analysis of simulation [32].

In this paper, DOE is used to simulate synthetic datasets through the combination of different factors that is described in the next section. In this context, Fig. 3 can be represented analogously by Fig. 4.

B. DATA GENERATION

In order to generate distinct situations to be solved by the clustering methods, initially we must select the factors that can influence the performance of these methods. Considering [14]–[17], we define the six following factors: number of variables (b), number of clusters (k), dataset size (z), sample size (n), overlapping (ov) and outliers (ot).

The number of variables (b) indicates the number of characteristics measured at each observed element with levels 4 and 6. Since the literature suggests that the performance of clustering methods increases with greater number of variables [21], these levels are enough to determine the results.

The number of clusters (k) indicates the number of final partitions in which elements will be separated. Data were generated considering situations in which the problem requires the partition of 2 and 4 clusters.

TABLE 2. Sample size generation.

Same sample size				Different sample size			
k	Sample Size	k	Sample Size	k	Sample Size	k	Sample Size
1	50%	1	25%	1	30%	1	10%
2	50%	2	25%	2	70%	2	20%
		3	25%			3	30%
		4	25%			4	40%

The dataset size (z) indicates the number of observations present in each dataset. They were generated from continuous data with normal distribution N (μ; 1) containing 200 or 400 samples. It is necessary a minimum of 20 observations for each sample to have a reliable result [30].

The sample size (n) represents the sample size of each cluster. Then, clusters can have the same sample size or they can have different sample size. Since the clustering levels are k = 2 and 4, the samples for each cluster were generated as detailed in Table 2.

Overlapping (ov) has the purpose of creating situations in which two or more clusters have similar elements. Then, we used the concept of ''effect size'' to generate samples presenting low (20%) and high (80%) levels of overlapping.

Effect size is a concept related to the confidence of statistical tests. Effect size is defined as the estimation of the magnitude of the relation between variables, the effect that one variable exerts on the other, or even as the difference between two samples [35].

To calculate the effect size usually is used the ''Cohen's d'' metrics as shown in (15) and (16) [36].

$$d = \frac{x_1 - x_2}{s_p} \qquad (15)$$

where,

$$s_p = \sqrt{\frac{(n_1 - 1)\, s_1^2 + (n_2 - 1)\, s_2^2}{n_1 + n_2 - 2}} \qquad (16)$$

Cohen [36] classified the effect size as small (d = 0.2) when the difference between two sample is difficult to see with naked eye, and as large (d = 0.8) when the difference between two sample is evident to see with the naked eye.

According to [36], for d = 0.2, 58% of cluster 2 will be above the mean of cluster 1; 92% of the two clusters will be overlapped; and there is a 56% of chance that a randomly chosen element from cluster 2 will be greater than a randomly chosen element from cluster 1. For d = 0.8, 79% of cluster 2 will be above the mean of cluster 1; 69% of the two groups will be overlapped and there is a 71% of chance of a randomly chosen element from cluster 2 will be greater than a randomly chosen element from cluster 1. That is depicted in Fig. 5.

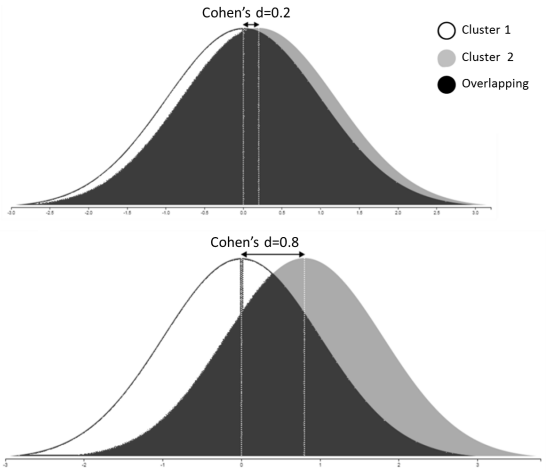

FIGURE 5. Cohen's d representation.

TABLE 3. Experimental factors.

	b	*z*	*k*	*n*	*ov*	*Ot*
(-1)	4	200	2	Same (S)	d=0.8	0%
(+1)	6	400	4	Different (D)	d=0.2	20%

The insertion of outliers (*ot*) in the samples aims to simulate a discrepant error in the measurement of the variables. For this purpose, a "contamination" was introduced on 20% of the samples in all variables. This contamination consists of a normal distribution with a standard deviation five times the original observations N (μ; 5).

A resume of the factors and their levels are detailed in Table 3.

Considering the six factors in Table 3, we constructed the design matrix by using Minitab®statistical software.

The design matrix used was the 2^{m} fractional factorial experiments with m = 6, resolution IV and two replications, resulting in 32 experiments presented in Table 4.

The fractional factorial experiments is a class of DOE widely used in experiments involving several factors, where it is necessary to study the significance of the factors and the joint effect of the factors in a response [30], and when the resource available for the experiment is scarce [37].

A particular feature of fractional factorial design is the fact that it does not present a complete experimental arrangement, so it presents a confounding between the main effects and the interactions. The intensity of confounding is called resolution [38]. The higher is the resolution, the smaller is the confusion.

In this paper, the matrix generated has resolution IV with 3^{rd} order interactions. That means that at least some main effects will be confused with three-factor interaction effects, and at least some two-factor interaction effects are confused with other two- factor interaction effects. This confounding is generally weak, therefore, negligible.

TABLE 4. Design matrix.

run	*b*	*z*	*k*	*n*	*ov*	*ot*
1	4	200	2	S	0.8	0%
2	6	200	2	S	0.2	0%
3	4	400	2	S	0.2	20%
4	6	400	2	S	0.8	20%
5	4	200	4	S	0.2	20%
6	6	200	4	S	0.8	20%
7	4	400	4	S	0.8	0%
8	6	400	4	S	0.2	0%
9	4	200	2	D	0.8	20%
10	6	200	2	D	0.2	20%
11	4	400	2	D	0.2	0%
12	6	400	2	D	0.8	0%
13	4	200	4	D	0.2	0%
14	6	200	4	D	0.8	0%
15	4	400	4	D	0.8	20%
16	6	400	4	D	0.2	20%
17	4	200	2	S	0.8	0%
18	6	200	2	S	0.2	0%
19	4	400	2	S	0.2	20%
20	6	400	2	S	0.8	20%
21	4	200	4	S	0.2	20%
22	6	200	4	S	0.8	20%
23	4	400	4	S	0.8	0%
24	6	400	4	S	0.2	0%
25	4	200	2	D	0.8	20%
26	6	200	2	D	0.2	20%
27	4	400	2	D	0.2	0%
28	6	400	2	D	0.8	0%
29	4	200	4	D	0.2	0%
30	6	200	4	D	0.8	0%
31	4	400	4	D	0.8	20%
32	6	400	4	D	0.2	20%

To compose this arrangement, it was computed a complete factorial 2^{m-2}. Let m be the number of factor equal 6, so the arrangement will have complete 24 experiments. Thus, the factors b, z, k, n compose a complete factorial, and $ov = b \times z \times k$ and $ot = z \times k \times n$ were assumed.

In this case, we did not randomize the experiments because the experiments are used only as a data simulation. The design matrix is a guide to indicate the combination of factors to generate the datasets that will be used for clustering analysis.

Afterwards, according to each run of the design matrix, we generated a dataset containing different levels of factors.

For example, the first dataset was composed by 4 variables, 200 samples, 2 clusters with 100 samples each one, low overlapping and without outlier. In total, we generated 32 datasets that were analyzed by the implementation of the clustering methods.

V. ANALYSIS OF RESULTS

Ward and K-means were implemented using Minitab®, and SOM network was implemented by using Statistica®. For each generated population the network was trained by using 60% randomly observations from the original dataset, in which 20% was the validation set. So, the remaining 20% observations were allocated to the testing set.

In the training process, we used the ''seed for sampling'' equal to 1000. The dimensions of the topological map were defined as the number of clusters for each dataset. From the topology, we defined the neighborhood. We specified the initial neighborhood size as the ''radius'' of a square neighborhood centered on the winning neuron and the final neighborhood size we defined equal to zero. The neighborhood is scaled linearly from the Start value to the End value. We initialized the learning rate equal to 0.1 and it was linearly reduced to 0.02 during 1000 training cycles. The weights were initialized as a normal randomization with mean set to zero and variance 0.1. It helps the network to gradually grow from its linear (small weight values) to nonlinear (large weight values) mode for modeling the data when necessary during the training process. Training stops when any one of the termination criteria (maximum number of cycles, maximum number of iterations, or convergence criterion) is satisfied.

After clustering methods have been implemented, the clustering final partitions were submitted to an evaluation test, which compares the results with the original clusters. The statistical test used was the Attribute Agreement Analysis, which is calculated by the proportion of correctly allocated observations to the originally simulated clusters. The results are shown in Table 5.

TABLE 5. Attribute agreement analysis of clustering.

run	Attribute Agreement (%)		
	Ward	*K-means*	*SOM*
1	72.00	74.50	76.00
2	54.50	51.00	53.00
3	50.25	50.50	50.50
4	50.25	73.50	72.00
5	23.50	27.00	28.50
6	44.50	44.50	48.00
7	56.50	59.25	50.50
8	36.50	34.50	34.50
9	68.50	70.50	54.50
10	62.50	65.50	69.50
11	50.25	50.50	57.25
12	79.75	82.25	79.25
13	29.00	36.50	32.50
14	61.50	56.00	52.00
15	46.50	50.00	38.50
16	35.50	38.50	37.25
17	72.50	84.50	75.50
18	50.00	53.00	50.50
19	50.00	50.50	50.50
20	50.50	79.50	52.75
21	26.50	22.50	24.00
22	44.50	47.50	45.50
23	56.75	59.50	56.50
24	33.00	32.50	35.50
25	63.50	67.50	73.00
26	62.00	68.00	68.50
27	50.50	50.00	51.75
28	79.00	80.25	78.00
29	33.00	28.00	35.00
30	62.00	54.50	46.00
31	45.50	47.50	26.25
32	35.00	37.50	37.00

In view of the results, we first performed a paired-t test to verify if the differences among the clustering methods were statistically significant with level of significance $\alpha = 5\%$. In this test is normally used the statistic called ''*p-value*''. If the p-value is less than the level of significance α, the null hypothesis is rejected. A deeper explanation can be found in [30].

The null hypothesis is the differences between samples equal to zero, and the alternative hypothesis is the differences between samples greater than zero. Comparing Ward with K-means and SOM the p-values obtained was equal to 0.478 and 0.973 respectively, and comparing K-means with SOM the p-value was equal to 0.513.So, as all p-values are greater than 0.05, we can accept the null hypothesis and conclude that the both method has similar performance, with a 5% of significance level.

In order to better understanding which factors and interactions influence the performance of each method, we applied the results of the Attribute Agreement Analysis as the response in the DOE.

Thus, we could establish mathematical relationships between the analyzed responses and the input parameters by using the Ordinary Least Squares Method (OLS) and Analysis of Variance in the Minitab®. The estimated coefficients in OLS are indicated in Table 6. The coefficients are coded by letters A, B, C, D, E and F. All the following analysis considered a significance level of 5%.

After, we performed a residual analysis to ensure that the regression model obtained by OLS has a good fit. The residual should be uncorrelated, normally and randomly distributed [30].

TABLE 6. Estimated coefficients.

Code	Coefficients	Response		
		Ward	*K-means*	*SOM*
	Constant	**51.117**	**53.976**	**51.250**
A	*b*	**1.445**	**2.180**	**2.453**
B	*z*	**-0.758**	0.789	-0.750
C	*k*	**-9.258**	**-11.742**	**-12.031**
D	*n*	**2.883**	**1.211**	1.016
E	*ov*	**-8.492**	**-10.477**	**-6.516**
F	*ot*	**-3.680**	**-1.445**	**-2.734**
AB	$b \times z$	**-1.867**	0.367	0.328
AC	$b \times k$	**0.758**	**-1.227**	0.297
AD	$b \times n$	**4.211**	**2.945**	**3.719**
AE	$b \times ov$	**2.055**	**1.883**	1.031
AF	$b \times ot$	**-0.789**	**2.102**	**2.844**
BD	$z \times n$	-0.492	**-1.414**	-0.860
BF	$z \times ot$	**-1.242**	0.117	**-2.172**
ABD	$b \times z \times n$	**0.773**	-0.430	0.719
ABF	$b \times z \times ot$	**-1.414**	-0.836	-1.469

Coefficients in bold indicate significant terms (p-values<0.05).

In the first analysis, the results showed that all methods presented uncorrelated residues (Low *Pearson* and *p-value* > 0.05). The residues of SOM and Ward did not have a normal distribution (*p-value* < 0.05 and *Anderson-Darling (AD)* >1), what implies in an inaccurate confidence intervals and imprecise p-values. The residues of SOM also did not have a random distribution, presenting a special cause of 'clustering' (*p-value* <0.05).

Then, we reduced some not significant terms to achieve the residual conditions. The Table 7 presents the residual analysis.

Thereafter, all models presented good adjustments, since the values of R^2 (adj.) and R^2 (pred.) were higher than 80%, indicating great reliability and predictability as shown in Table 8.

Therefore, we could obtain the general equations of each model that are described in (17) (18) (19).

$$\begin{aligned} Y_{Ward} &= 51.117 + 1.445A - 0.758B - 9.258C + 2.883D \\ &\quad -8.492E - 3.680F - 1.867AB + 0.758AC + 4.211AD \\ &\quad +2.055AE - 0.789AF - 1.242BF - 1.414ABF \end{aligned} \tag{17}$$

$$\begin{aligned} Y_{K-means} &= 53.977 + 2.180A + 0.790B - 11.742C \\ &\quad +1.211D - 10.477E - 1.445F + 0.367AB - 1.227AC \\ &\quad +2.945AD + 1.883AE + 2.102AF - 1.414BD \\ &\quad +0.117BF - 0.430ABD - 0.836ABF \end{aligned} \tag{18}$$

TABLE 7. Residual analysis.

Residual Analysis		Ward		K-means		SOM	
Test		*before*	*after*	*before*	*after*	*before*	*after*
(1)	*AD*	**1.52**	**0.34**	0.39	0.39	**1.20**	**0.58**
	P-value	**<0.05**	**0.49**	0.36	0.36	**<0.05**	**0.12**
(2)	*Pearson*	-0.17	-0.13	-0.02	-0.02	-0.19	-0.17
	P-value	0.35	0.47	0.90	0.90	0.30	0.35
(3)	*Clustering*	0.51	0.36	0.08	0.08	**0.04**	**0.64**
	Mixtures	0.49	0.64	0.92	0.92	0.96	0.36
	Trends	0.19	0.19	0.50	0.50	0.19	0.67
	Oscillating	0.81	0.81	0.50	0.50	0.81	0.33

(1) Normality tests; (2) Correlation tests; (3) Randomness tests.

TABLE 8. Of the regression model.

	Ward	K-means	SOM
R^2 (adj.) (%)	98.26	97.05	88.54
R^2 (pred.) (%)	96.80	93.92	81.97

$$\begin{aligned} Y_{SOM} &= 1.250 + 2.453A - 0.750B - 12.031C + 1.016D \\ &\quad -6.516E - 2.734F + 3.719AD + 1.031AE + 2.844AF \\ &\quad -2.172BF \end{aligned} \tag{19}$$

By analyzing (17), (18) and (19), we conclude that for all methods the most significant factors are the number of clusters (C) and the overlapping (E), and also the interaction between the number of variable (A) and the sample size (D).

In addition, there are other factors and interactions that are also significant at the 95% confidence level but at a lower intensity. Another interesting result is that the dataset size (B) is not significant for K-means and SOM methods, and the sample size (D) is not significant for SOM. For better visualization of the factors significance, Fig. 6, 7, and 8 present the Pareto Chart for each method.

At this moment, it is necessary to understand how each significant factor influences each model. Then, we developed an analysis of the main effects and the interaction plots for all the responses, as shown in Fig. 9, 10, 11 and 12.

As we can infer from Fig 9, 10, and 11, the increase of the number of variables (A) implies an increase in the clustering performance for all methods, and Ward method is the least sensitive to this factor. Its performance reduced from 49.67% to 52.56%, whereas K-means presented an increasing from 51.80% to 56.16% and SOM from 48.80% to 53.70%.

The dataset size (B) cannot be considered as significant factor for all methods.

The increase of the number of clusters (C) makes the performance of all clustering methods worse. For number of clusters $k = 2$, Ward presented a recovery rate to 60.38%, K-means to 65.72% and SOM to 63.28%. For $k = 4$ Ward had a result of 41.86%, K-means of 42.23% and SOM of 39.22%.

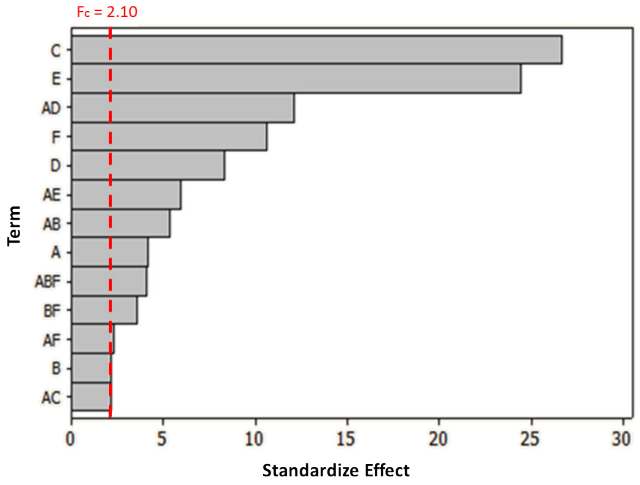

FIGURE 6. Pareto chart for ward.

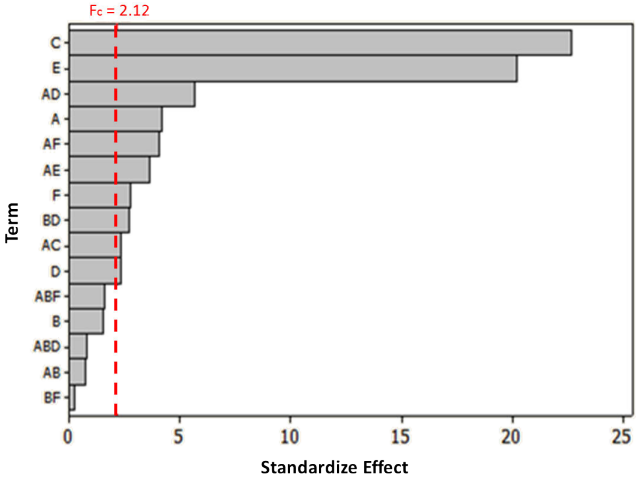

FIGURE 7. Pareto chart for K-means.

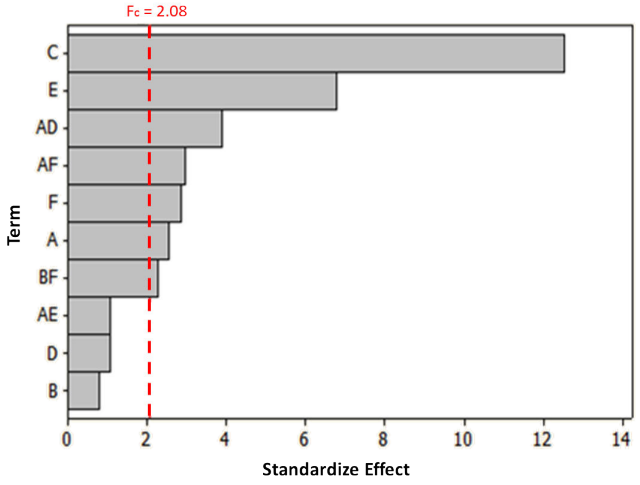

FIGURE 8. Pareto chart for SOM.

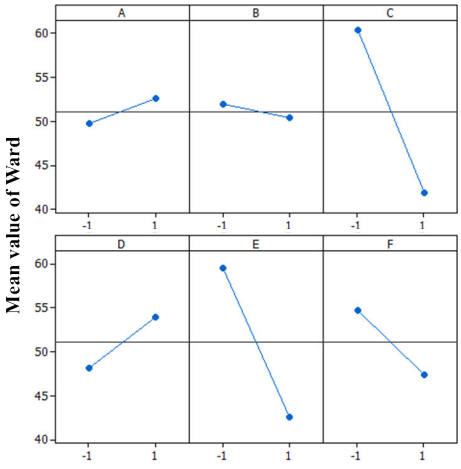

FIGURE 9. Factorial plot for ward response.

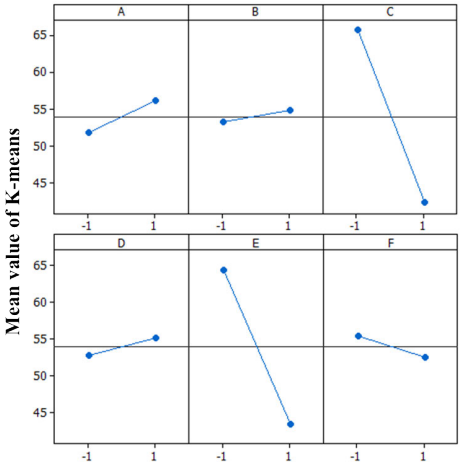

FIGURE 10. Factorial plot for K-means response.

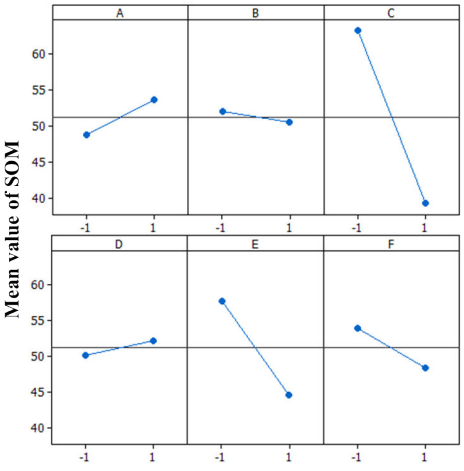

FIGURE 11. Factorial plot for SOM response.

The sample size (D) has a little significance only for Ward and K-means methods, which have better results in situations that clusters have different sample size, 54% for Ward and 55.19% for K-means.

Cluster without overlapping (E) presents better results than cluster with overlapping. For overlapping $ov = 20\%$, Ward, K-means and SOM presented performance equals to 59.61%, 64.45% and 57.76%, respectively. For $ov = 80\%$ the Ward's

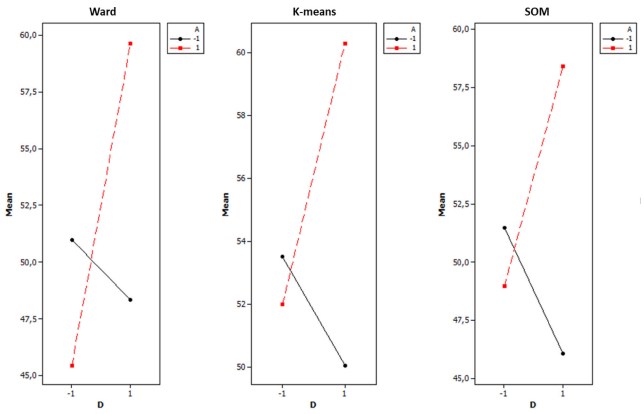

FIGURE 12. Interaction plot for number of variable and sample size.

performance decreased to 42.63%, K-means to 43.5% and SOM to 44.74%.

The presence of outliers (F) makes the performance of all methods worse and the K-means was the most stable. The performance of Ward declined from 59.80 to 47.4%, K-means from 55.42% to 52.53%, and SOM from 53.98% to 48.52%.

Besides the influence of the factors, the influence of the interaction between the number of variable (A) and the sample size (D) is also significant. As shown in Fig. 12, the best performance of all methods occurs when the number of variables is high and the sample size of the clusters is different. For K-means and SOM methods, the effect of A is larger when we have different sample sizes, then the lowest (50.06% and 46.09%) and the highest (60.31% and 58.43%) mean value occurs when A is equals to 4 and 6, respectively, in this context. For the Ward method it does not occur, and its worst performance (45.63%) when dealing with small number of variables and equal cluster sample size, whereas its highest mean value (59.66%).

VI. CONCLUSION

Cluster analysis is widely used in several areas to solve important real problems, and the accuracy of the final solution depends on the clustering method used. Motivated by that, we developed a comparative study among Ward, K-means and SOM clustering to evaluate the performance of each method. The analysis was based on a synthetic dataset, which were created by the DOE technique whose factors were: number of variables, number of clusters, dataset size, sample size, overlapping clusters, and presence of outliers. To solve them by Ward and K-means method we used the Minitab®software and for the SOM method we used the Automated Network Networks Cluster Analysis from Statistica®software. The clustering partitions were compared by the Attribute Agreement Analysis and analyzed by statistic techniques.

The results presented in this paper show that the performance of the clustering algorithm does not have a significant difference, so it is not possible to affirm which method has the best performance.

For number of clusters $k = 2$ all the methods had a good performance and for number of clusters $k = 4$ all methods had an abrupt decrease in their performances. The same is found in [15], [16] and [17]. For situations with low level of overlapping all methods had good performance and for high level of overlapping the performances were worse. That agrees with Mangiameli *et al.* [14] that suggests that all methods present better performance for samples without overlapping and with Waller *et al.* [17] that proposes that SOM is the method that presents the best result for high levels of overlapping. About the number of variables, the best result for all method was found for $p = 6$. This analysis corroborates with [16] and [15] that suggest an improvement in the performance of all methods with the increase of the number of variables. The dataset size and the sample size did not affect very much the performance of the methods. When outliers are introduced, the performances of all methods decrease, as shown in [14], [15]. Mangiameli *et al.* [14] add that the performance of SOM and Ward methods are similar in the presence of outliers, and Mingoti and Lima [15] emphasize that the results of the K-means method are superior to the SOM. The interaction between the number of variable and the cluster sample size was also significant, and the K-means and SOM methods are very affect by the number of variables when the clusters sample sizes are different.

In general, although numerically the k-means method presents better average results, it is not acceptable to say that it has the best performance, since the results of all methods were the same with 95% of confidence. Nevertheless, it is possible to affirm that all methods are significantly affected by the number of clusters, overlapping, and the interaction between sample size and number of variables.

Our study differs from others mainly with respect to the method used to generalize and simulate the datasets. We used the DOE technique, which allows a combination of factors levels that result in an arrangement of experiments with controlled structures. Nevertheless, other papers presented an usual approach of trying and error, what can lead to restricted conclusions without a formal and larger analysis, which can be achieved using DOE.

There is another difference among the mentioned papers. Some of them only consider the interaction when overlapping and outliers are introduced in the data [14], [15], [17]. None studied the interaction among all factors as we did in this paper, as example the interaction between the number of variable and the cluster sample size. The cluster sample size is a factor that was also not studied in the previous papers as well as the dataset size. All papers specified the dataset sample size but they not used it as a controlled factor. Even the results have showed that these factors are not significant, it is an important result because infers that the studied clustering methods do not have their performance affected by the sample size.

Another difference is the application of Ward and K-means methods that were implemented by using the Minitab®software, and the SOM implementation that used

the Statistica®software with the training parameters carefully set.

Another possible reason for different conclusions about what method is the best one is that we used a statistical test, paired-t, to compare the results of the algorithms. Otherwise, if a statistical test is not applied, we may reject the null hypothesis affirming that one method outperforms the other, when actually they have similar performances with 95% of confidence level. Therefore, some authors believe that the SOM is the clustering method that presents the best performance [14], [17], whereas others imply that the best method is the traditional hierarchical or non-hierarchical methods [15], [16].

Many other studies still can be performed as example; 1) Comparison of Artificial Neural Network with others statistical methods by using DOE, 2) Comparison of the clustering algorithms by using other metrics than the Euclidean distance, 3) Dataset generated by a distribution different than normal, and 4) Case study by using the result obtained in this paper.

Further, this study revealed that no matter what method is chosen, we can obtain satisfactory results, since the researcher knows main characteristics of the dataset and applies the method correctly. Besides, this paper may be a reference for researchers that are looking to improve the efficiency and effectiveness in clustering methods.

REFERENCES

[1] X. Ma, S. Chen, and F. Chen, "Multivariate space-time modeling of crash frequencies by injury severity levels," *Anal. Methods Accident Res.*, vol. 15, pp. 29–40, Sep. 2017.

[2] D. McNeish, "Challenging conventional wisdom for multivariate statistical models with small samples," *Rev. Edu. Res.*, vol. 87, no. 6, pp. 1117–1151, Aug. 2017.

[3] M. Trzęsiok, "Measuring the quality of multivariate statistical models," *Acta Universitatis Lodziensis, Folia Oeconomica*, vol. 6, no. 339, pp. 99–110, Jun. 2018.

[4] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*, 5th ed. London, UK: Wiley, 2011.

[5] T. Sorensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons," *Biologiske Skrifter*, vol. 5, no. 4, pp. 1–34, 1948.

[6] H. Kim, B. Kim, S. H. Kim, C. H. K. Park, E. Y. Kim, and Y. M. Ahn, "Classification of attempted suicide by cluster analysis: A study of 888 suicide attempters presenting to the emergency department," *J. Affect. Disorders*, vol. 235, pp. 184–190, Aug. 2018.

[7] Z. Yu, H. Chen, J. You, J. Liu, H. Wong, G. Han, and L. Li, "Adaptive fuzzy consensus clustering framework for clustering analysis of cancer data," *IEEE/ACM Trans. Comput. Biol. Bioinformatics*, vol. 12, no. 4, pp. 887–901, Jul./Aug. 2015.

[8] C. T. Lee, D. Guzman, C. Ponath, L. Tieu, E. Riley, and M. Kushel, "Residential patterns in older homeless adults: Results of a cluster analysis," *Social Sci. Med.*, vol. 153, pp. 131–140, Mar. 2016.

[9] R. Macchiaroli and S. Riemma, "Clustering methods for production planning and scheduling in a flexible manufacturing system," in *Proc. IEEE Int. Conf. Robot. Automat.*, San Diego, CA, USA, May 1994, pp. 3155–3160.

[10] A. Nachtwey, R. Riedel, and E. Mueller, "Cluster analysis as a method for the planning of production systems," in *Proc. IEEE Int. Conf. Comput. Ind. Eng.*, Troyes, France, Jul. 2009, pp. 725–728.

[11] J. Hochdorffer, C. Laule, and G. Lanza, "Product variety management using data-mining methods— Reducing planning complexity by applying clustering analysis on product portfolios," in *Proc. IEEE Int. Conf. Ind. Eng. Eng. Manage. (IEEM)*, Singapore, Dec. 2017, pp. 593–597.

[12] T. Kohonen, *Self-Organizing Maps*. Berlin, Germany: Springer-Verlag, 1995.

[13] A. D. Gordon, "A review of hierarchical classification," *J. Roy. Stat. Soc.*, vol. 150, no. 2, pp. 119–137, Mar. 1987.

[14] P. Mangiameli, S. K. Chen, and D. West, "A comparison of SOM neural network and hierarchical clustering methods," *Eur. J. Oper. Res.*, vol. 93, no. 2, pp. 402–417, 1996.

[15] S. S. Mingoti and J. O. Lima, "Comparing SOM neural network with Fuzzy *c*-means, *K*-means and traditional hierarchical clustering algorithms," *Eur. J. Oper. Res.*, vol. 174, pp. 1742–1759, Nov. 2006.

[16] P. V. Balakrishnan, M. C. Cooper, V. S. Jacob, and P. A. Lewis, "A study of the classification capabilities of neural networks using unsupervised learning: A comparison with *K*-means clustering," *Psychometrika*, vol. 59, no. 4, pp. 509–525, Dec. 1994.

[17] N. G. Waller, H. A. Kaiser, J. B. Illian, and M. Manry, "A comparison of the classification capabilities of the 1-dimensional kohonen neural network with two pratitioning and three hierarchical cluster analysis algorithms," *Psychometrika*, vol. 63, no. 1, pp. 5–22, Mar. 1998.

[18] E. R. Hruschka and N. Ebecken, "A Genetic algorithm for cluster analysis," *Intell. Data Anal.*, vol. 7, no. 1, pp. 15–25, Feb. 2003.

[19] D. Fasulo, "An analysis of recent work on clustering algorithms," Dept. Comput. Sci. Eng., Univ. Washington, Washington, DC, USA, Tech. Rep. 01-03-02, 1999.

[20] R. A. Johnson and D. W. Wichern, *Multivariate Statistical Analysis*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1992.

[21] G. W. Milligan, "An examination of the effect of six types of error perturbation on fifteen clustering algorithms," *Psychometrika*, vol. 50, no. 1, pp. 123–127, Sep. 1980.

[22] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2006.

[23] G. Fung, "A comprehensive overview of basic clustering algorithms," Tech. Rep., Jun. 2001. [Online]. Available: http://www.cs.wisc.edu/~gfung/clustering.pdf

[24] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 265–323, Sep. 1999.

[25] K. D. Siegmund, P. W. Laird, and I. A. Laird-Offringa, "A comparison of cluster analysis methods using DNA methylation data," *Bioinformatics*, vol. 20, no. 12, pp. 1896–1904, Aug. 2004.

[26] J. H. Ward, Jr., "Hierarchical grouping to optimize an objective function," *J. Amer. Statist. Assoc.*, vol. 58, no. 301, pp. 236–244, 1963.

[27] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction*. Hoboken, NJ, USA: Wiley, 1990.

[28] H. Kuo and A. Faricha, "Artificial neural network for diffraction based overlay measurement," *IEEE Access*, vol. 4, pp. 7479–7486, 2016.

[29] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1999.

[30] C. D. Montgomery, *Design and Analysis of Experiments*. 6th ed. New York, NY, USA: Wiley, 2005.

[31] D. Staiculescu, N. Bushyager, A. Obatoyinbo, L. J. Martin, and M. M. Tentzeris, "Design and optimization of 3-D compact stripline and microstrip Bluetooth/WLAN balun architectures using the design of experiments technique," *IEEE Trans. Antennas Propag.*, vol. 53, no. 5, pp. 1805–1812, May 2005.

[32] J.-Y. Lee, J.-H. Chang, D.-H. Kang, S.-I. Kim, and J.-P. Hong, "Tooth shape optimization for cogging torque reduction of transverse flux rotary motor using design of experiment and response surface methodology," *IEEE Trans. Magn.*, vol. 43, no. 4, pp. 1817–1820, Apr. 2007.

[33] L. Dascalescu, K. Medles, S. Das, M. Younes, L. Caliap, and A. Mihalcioiu, "Using design of experiments and virtual instrumentation to evaluate the tribocharging of pulverulent materials in compressed-air devices," *IEEE Trans. Ind. Appl.*, vol. 44, no. 1, pp. 3–8, Jan/Feb. 2008.

[34] I. Lorscheid, B.-O. Heine, and M. Meyer, "Opening the 'black box' of simulations: Increased transparency and effective communication through the systematic design of experiments," *Comput. Math. Org. Theory*, vol. 18, pp. 22–62, Mar. 2012.

[35] R. Rosenthal, "Parametric measures of effect size," in *The Handbook of Research Synthesis*, H. Cooper and L. V. Hedges, Eds. New York, NY, USA: Russell Sage Foundation, 1994, pp. 231–244.

[36] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, MI, USA: Lawrence Erlbaum Associates, 1988.

[37] G. F. Giesbrecht and M. Gumpertz, *Planning, Construction, and Statistical Analysis of Comparative Experiments*, Hoboken, NJ, USA: Wiley, 2004.

[38] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experiments*, 1st ed. Hoboken, NJ, USA: Wiley, 1978.

NATÁLIA MARIA PUGGINA BIANCHESI received the B.S. degree in industrial engineering from the Federal University of Itajubá, Brazil, in 2014, where she is currently pursuing the M.Sc. degree in industrial engineering. From 2012 to 2013, she was a Sandwich Student with the Université de Technologie de Troyes, Champagne-Ardenne, France. Her research interests include clustering, neural networks, and statistics.

ESTEVÃO LUIZ ROMÃO received the B.S. degree in industrial engineering from the Federal University of Viçosa, Minas Gerais, Brazil, in 2017. He is currently pursuing the M.Sc. degree in industrial engineering with the Federal University of Itajubá, Brazil. From 2015 to 2016, he was a Sandwich Student with the École Polytechnique Universitaire de l'Université Lyon-1, Roanne, France. His research interests include nonlinear optimization, times series forecasting, ANN, and statistics.

MARINA FERNANDES B. P. LOPES received the B.S. degree in industrial engineering from the Federal University of Itajubá, Minas Gerais, Brazil, in 2016, where she is currently pursuing the M.Sc. degree in industrial engineering. Her research interests include nonlinear regression, artificial neural networks, and statistics.

PEDRO PAULO BALESTRASSI received the D.Sc. degree in industrial engineering from the Federal University of Santa Catarina, Brazil, in 2000. Since 1994, he has been with the Federal University of Itajubá, Brazil, as a Professor of industrial engineering. From 1998 to 1999, he was a Visitor at Texas A&M University, College Station, TX, USA, and from 2005 to 2006, he was a visiting The University of Texas at Austin. His areas of research interest include times series forecasting, ANN, and statistics.

ANDERSON PAULO DE PAIVA received the D.Sc. degree in industrial engineering and mechanical engineering from the Federal University of Itajubá, Brazil, in 2006. He is currently a Professor of quantitative methods with the Industrial Engineering Institute, Federal University of Itajubá. His research interests include multivariate statistical analysis, nonlinear optimization, and neural networks applied to manufacturing processes.