22nd IMEKO TC4 International Symposium & 20th International Workshop on ADC Modelling and Testing
SUPPORTING WORLD DEVELOPMENT THROUGH ELECTRICAL&ELECTRONIC MEASUREMENTS
IASI, ROMANIA, September 14-15, 2017

On the Assessment of Slow Voltage Variations in Electric Distribution Networks using K-Means Clustering Algorithm

[1]Gheorghe Grigoras, [1]Bogdan-Constantin Neagu, and [1]Alexandra Adăscăli□ei

[1]*"Gheorghe Asachi" Technical University, Iasi, Romania,*
ggrigor@tuiasi.ro,Phone: +40232278683

***Abstract* – The paper presents a clustering based analysis for the assessment of slow voltage variations from the electric distribution networks. The analysis uses the K-means clustering algorithm for classifying the electric substations (Medium Voltage/Low Voltage – MV/LV) in patterns characterized by linguistic terms from the point of view of slow voltage variations. The patterns were obtained using as input data in the clustering process the voltage quality indices (average deviation, average square deviation, and voltage irregularity degree). The analysis was made on a large supply zone with 722 electric distribution substations divided in three areas (remote, close, and suburb). The results confirmed the robustness and efficiency of proposed approach in the analysis of databases with a high amount of information insured by measurement devices.**

Keywords* – *voltage quality,slow voltage variation, electric distribution networks, K-means clustering.

I. INTRODUCTION

The time variation of voltage in the nodes of electrical networks is related to the variations of voltage drop in time. They appear on the components of the electrical networks (lines and transformers) due to change in their operation schemes, as well as the change in time of the active power and reactive flows, as a result of the power variations corresponding to the consumers as according to their active and reactive loads and the power generation in power plants [1], [2]. From the point of view of production mode of disturbances occurring in the electrical network (long or short duration disturbances) and the change over time, voltage variations can be classified as: slow variations, sudden variations, and voltage dips [3] - [5].

The slow voltage variations occur as a result of overlapping of some periodic variation due to the change in the structure and power required by consumers, between the maximum and minimum values of load, or between the working and weekend days and some random variation arising as a result of the connected or disconnected operations of consumers to network.

The sudden voltage variations have a repetitive, cyclical or random character occurring with a gradient $\Delta U/\Delta t$ greater than 1%. These are caused by the intermittent operation of some installations producing the power shock with passenger character in the network (at the LV level - welding equipment, at the medium voltage level - pumps or rolling mills, at the high voltage level – electric arc furnaces, etc.), changes in the operation schemes of electric networks, short disappearances of voltage due to the working of automation systems.

The voltage dips represent rapid decreases with a very short duration (maximum 3 sec.), due to the transient disturbances that may occur in the electrical networks, such as asymmetrical and symmetrical short circuits, connections of some electric motors with high absorbed currents at starting etc.

When the short duration disturbances appear in the operation of electrical networks and lead to the occurrence of sudden voltage variations or voltage dips, the return at the initial values of voltages in the nodes of network or at the values close to it is made basically through the disappearance of cause or the intervention of automation system. The time limit that divides the short duration disturbances (sudden variations or dips) by the long disturbances (slow variations) is generally the required time for the protection devices, automation systems and switching equipment to restore the voltage in the electrical networks, if it is possible. In Romania, this limit is considered as 3 seconds. Short duration disturbances, caused by the sudden and transient changes of transversal impedances from the electric networks due to short circuits or ground faults occurring accidentally, lead at voltage drops belonging to the category of voltage dips. The short duration disturbances caused by the sudden and transient changes of longitudinal impedances from the electric networks which lead at the disappearances of voltage belonging to the category of micro interruptions [6] – [8].

Considering the above aspects, it is necessary that the power system, as a whole, to ensure that these long or short duration disturbances to have features supported easily by consumers. In terms of the consumers, they must have electric equipment and devices insensitive to these phenomena, such as voltage variations which are inevitable in the electrical networks.

Thus, it is apparent that the voltage variations from the MV and LV distribution networks which supplies directly consumers determines the quantitative and qualitative changes of production, the efficiency of installations, wear of equipment, and disturbance of technological processes. All these effects are reflected in additional costs or economic damage produced the different consumers' categories due to the inadequate values of voltage from the supply node. Also, the voltage variations influence the operation of distribution networks changing the active and reactive power flows and power and energy losses which occur due to mainly the change of absorbed power by electric equipment and devices according with the power static characteristics [6], [7].

From the three types of voltage variations, only the slow variations are treated in this paper. It is proposed a classification of the electric substations from different distribution networks in function by the indices which characterize the slow voltage variations based on an Artificial Intelligence technique (clustering). The main reason is represented by fact that the voltage measurements in each substation is possible, but leads to quickly growing costs due to the huge amount of substations and a large time period for the data management. Thus, in LV and MV networks are necessary new and powerful tools for classifying the substations in terms of the voltage quality indices. The proposed approach was tested for a large supply zone with 722 electric distribution substations divided in three areas (remote, close, and suburb) determined in function by distance between MV/LV substation and HV/MV supply electric station. The clustering process led to five clusters for each area according to the indices of slow voltage variation. The analysis of results highlighted that the clusters correspond some levels of voltage quality often encountered in the operation of electric distribution systems.

II. EVALUATION OF SLOW VOLTAGE VARIATIONS

The slow voltage variations describe the variations of the supply voltage around the nominal voltage (U_n) under steady state regime within a range of U_n ±10 % (the mean value during 95% of the time shall be between 90% and 110% of the nominal voltage) [2], [5]. These values are defined in standards that exist in most countries. In Europe, the standard EN 50160 has been released in 2007 and is adopted also in Romania. Thus, the slow voltage variations are usually evaluated by the determination of root mean square (RMS) value of the supply voltage and the 95 percentile over one week of 10-minute mean RMS values is considered as site index. Regarding to the supply voltage levels, these are different for every node (electric substation) in the electric distribution network. The voltage variations in nodes are affected by the following factors: customers' load structures and characteristics of supply network (transformer impedance, line impedance, voltage level, etc.).

The voltage variation represents a disturbance with an increase or decrease of supply voltage normally due to the variation of load in the electric distribution system or a part of it, which actually is different from rapid voltage change [9], [10]. This is often measured by the voltage deviation, defined as:

$$\Delta U = \frac{U - U_n}{U_n} \cdot 100\,[\%] \qquad (1)$$

where: U *is* the value of measured voltage in a node and U_n is nominal voltage.

A smaller voltage deviation from the nominal voltage represents the better operation conditions of the system. Considering the voltage variation in a node of the analysed system, the mean deviation of the voltage to the nominal value within a time period T can be calculated using the relation:

$$\overline{\Delta U} = \frac{1}{T}\sum_{t=0}^{T}\frac{U(t) - U_n}{U_n} = \frac{100}{T}\sum_{t=0}^{T}\frac{U(t) - U_n}{U_n}[\%] \qquad (2)$$

$$\overline{\Delta U} = (\overline{u} - 1) \cdot 100 \quad [\%] \qquad (3)$$

where: $\overline{u}$ is the average level of voltage in a node, in time period T.

The index $\overline{u}$ (dimensionless) characterizes electric distance of a node from power reactive source. If is higher than 1, then the node is close to source and if this is less than 1 then the node is afar, located at the remote ends of system.

The variance of voltage deviation is the second index used in the assessment of slow voltage variations. This can be calculated with the following relation:

$$\sigma_{\Delta U}^2 = \frac{1}{T}\sum_{t=0}^{T}\left(\Delta U(t) - \overline{\Delta U}\right)^2 = \frac{100^2}{T}\sum_{t=0}^{T}\left(u(t) - \overline{u}\right)^2 [\%] \qquad (4)$$

From this index, it can calculate the standard deviation, $\sigma_{\Delta U}$, which is the square root of variance.

The third index is represented by the voltage irregularity degree. This can be determined using the following relation:

$$\varepsilon_q{}^2 = \frac{1}{T}\sum_{t=0}^{T}(\Delta U(t))^2 = \frac{100^2}{T}\sum_{t=0}^{T}\left(\frac{U(t)-U_n}{U_n}\right)^2 = \\ = \frac{100^2}{T}\sum_{t=0}^{T}(u-1)^2 \ [\%^2] \tag{5}$$

The normalized values for the voltage irregularity degree are presented in [9]:

- ≤ 10 [%] – the voltage quality is *very good*;
- 10 – 20 [%] - the voltage quality is *good*;
- 20 – 50 [%] - the voltage quality is *moderate*;
- ≥ 50 [%] - the voltage quality is *unsatisfactory.*

III.CLUSTERING TECHNIQUES

The clustering techniques represent special techniques for arranging the input data based on spatial arrangement of feature vectors. To study the similarity / difference between the elements of a data set, in order to be grouped, each element is defined by a vector whose components are the representative features / attributes of the vector. Establishing the number of components / attributes and their definition is based on the deeper analysis, with available data, and taking into account the expert' experience. In the next stage, to determine similarity / difference between the involved elements, the distance between feature vectors is calculated. Inside of such process, input vectors are grouped according to the distance between representative points. At the final of clustering process, one or more groups (clusters) are obtained, depending on the problem, which represent the spatial placement of the considered features for the involved elements in clustering process. Inside of such groups (clusters), the points are close together in relative to a joint centre [11], [12].

The criterion is characterized by a measured distance between vectors corresponding these elements. Thus, two or more elements belonging to the same cluster that are closer in relation to a certain distance. This clustering is based on the calculation of a distance.

Given a database X formed by x^j vectors, $j = 1, \ldots, m$, with n characteristics, $x_i^j = [x_1^j \ \ x_2^j \ \ldots \ x_n^j]$, different distances between vectors can be defined. Thus, if it considers two vectors x^r and x^s, the Euclidean distance is calculated [11]:

$$d(x^{(r)}, x^{(s)}) = \sqrt{(x^r - x^s)(x^r - x^s)^t} \tag{6}$$

One of the simplest unsupervised learning algorithms that solve the well-known clustering problem is K-means. The process follows a quick and easy way to classify the input data set in a number of K clusters (K is fixed a priori). The basic idea is to define K centroids, one for each cluster.

The centroids must be rationally chosen because different locations lead to different results. The best choice is to fix as further possible towards one another. The next step is to take each element from input data and linked to the nearest centroid. The first stage of clustering process is finished when there aren't elements ungrouped. A recalculation of the new positions for the centroids of resulted clusters from the previous step will be made. The process will continue until the new positions will not change significantly [11] –[13].

The aim of this algorithm is to minimize the squared error function:

$$J = \sum_{k=1}^{K}\sum_{h=1}^{n_k}\left\| x_l^k - c_k \right\|^2 \tag{7}$$

where $\left\| x_l^k - c_k \right\|^2$ is distance between the element $x_l{}^k$, $l = 1, \ldots, n_k$ (the total number of elements from cluster k) and the centroid of cluster c_k , $k = 1, \ldots, K$.

The steps of K-means clustering algorithm are presented in Fig. 1.

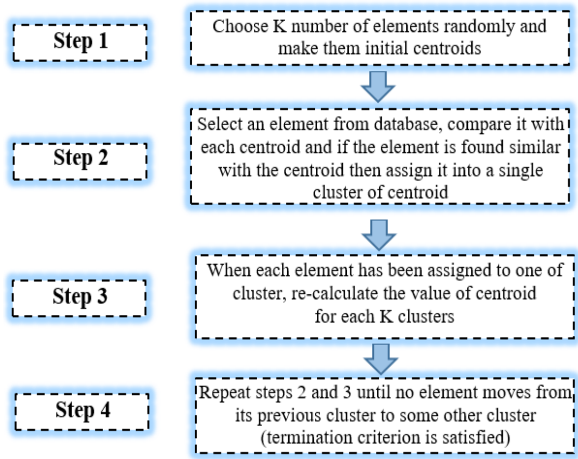

Fig. 1. K-Means Clustering Algorithm

The evaluation of obtained results represents the main subject of cluster validity. The following properties of clusters are examined in an analysis: density, sizes and form of cluster, separation of clusters, robustness of classification. There are three main approaches to cluster validation [12], [14]: external tests, internal tests, and relative tests. The internal validation tests are more popular in practice in a cluster analysis. From these, the test based on the Silhouette Coefficient (SC) calculation is one of the most used. This calculates the silhouette width for each sample, average silhouette width for each cluster and overall average silhouette width for a total data set. An interpretation of SC is the following: 0.71 – 1.0 (a strong structure has been found); 0.51 – 0.7 (a reasonable structure has been found); 0.26 – 0.5 (the structure is weak and could be artificial); < 0.25 (no substantial structure has been found).

IV. CASE STUDY

The clustering based analysis for the assessment of voltage quality from the view of point of slow variations in a real electric distribution system belonging to a Distribution Network Operator from Romania is based on the algorithm presented in Section III.

In the first phase of analysis, the data regarding a number by 722 electric MV/LV substations, where the voltage measurements were made, were updated to be used in a pre-classification process in three areas depending on the electric distance from the electric supply station (HV/MV): remote (117 substations), close (452 substations), and suburb (153 substations).

In the second phase, for each electric substation were calculated the indices for evaluation of slow voltage variations. In every substation (on LV side of distribution transformer), the current and voltage measurements (CVM) with recorder at 10 minutes were performed along aday. The measurements were made using the power quality analysers.

In the third phase, the electric substations belonging each areas (remote, close, and suburb) were classified using the K-means clustering algorithm according to the indices presented in Section II. The best results were obtained for case K= 5 (for each area). The value for SC is between 0.6 and 0.7, which means that a good structure for each cluster has been found.

The graphical representation of clusters for each area is presented in Figs. 3 – 5.The characteristics of clusters represented by the indices of slow voltage variations are presented in Table 1. Following theresults, some linguistic terms can be assigned in an assessment of voltage quality in the distribution networks: *Very Good*, *Good*, *Moderate*, *Satisfactory*, and *Unsatisfactory.*

Thus, Cluster 1 correspond to the substations with a unsatisfactory voltage quality, Cluster 2 is at limit between a unsatisfactory and moderate voltage quality, Clusters 3 belongs to a moderate voltage quality, Cluster 4 is at limit between a good and very good voltage quality and Cluster 5 has a very good voltage quality.

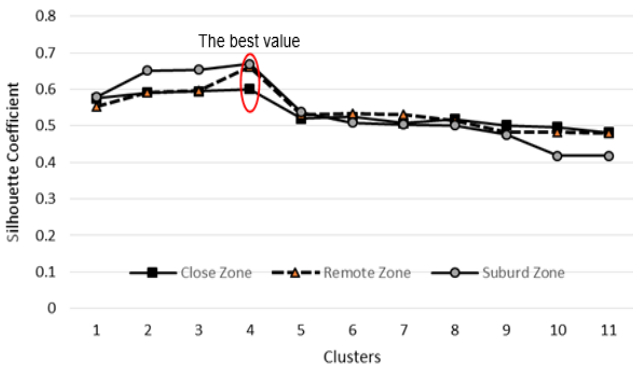

Fig. 2. The variation of SC for each analysed area (close, remote, and suburb)

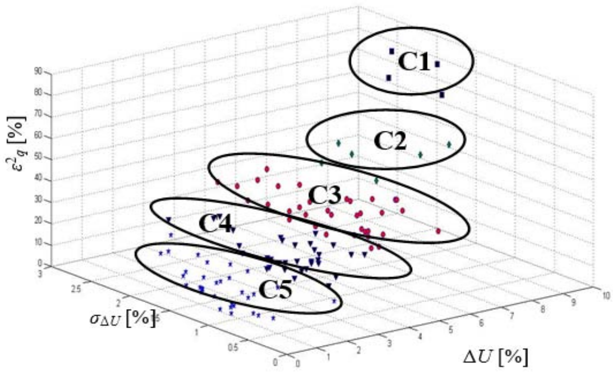

Fig. 3. Clusters for the remote area

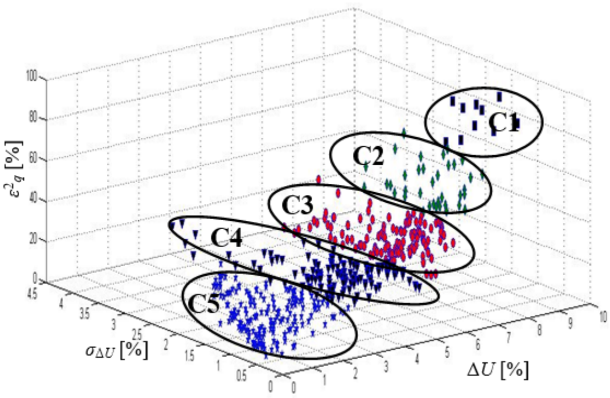

Fig. 4. Clusters for the close area

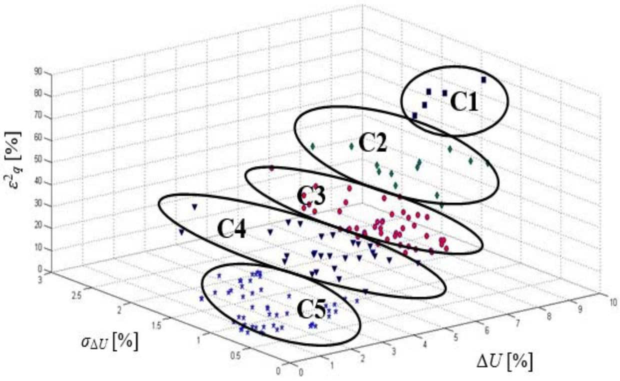

Fig. 5. Clusters for the suburb area

Table 1. The characteristics of clusters for each area

Cluster	Indices	Areas			Voltage quality
		Remote	Close	Suburb	
C1	ΔU [%]	7.77	6.03	6.79	Unsatisfactory
	$\sigma_{\Delta U}$[%]	3.23	5.56	4.16	
	ε^2_q [%]	81.07	82.89	77.73	
C2	ΔU [%]	6.86	5.98	6.84	Satisfactory
	$\sigma_{\Delta U}$[%]	1.54	1.42	1.63	
	ε^2_q [%]	49.79	54.42	50.83	
C3	ΔU [%]	5.27	5.72	5.49	Moderate
	$\sigma_{\Delta U}$[%]	1.46	1.29	1.00	

	ε^2_q [%]	30.33	34.78	31.51	
C4	ΔU [%]	3.72	4.22	4.30	**Good**
	$\underline{\sigma}_{\Delta U}$[%]	1.39	1.48	1.24	
	ε^2_q [%]	16.07	20.25	20.46	
C5	ΔU [%]	1.81	2.12	1.91	**Very Good**
	$\underline{\sigma}_{\Delta U}$[%]	1.54	1.54	1.03	
	ε^2_q [%]	4.98	6.57	5.11	

The linguistic terms are depended principally by the values of average deviation, average square deviation, and voltage irregularity degree. It can be observed that the values of indices increase from cluster C5 to C1, which it is normal in the operation of distribution networks.

V. CONCLUSIONS

In this paper,the slow voltage variations in the nodes (MV/LV electric substations)of distribution networks were analysedfor to assess the voltage quality. These voltage variations in the range [-10%, +10%] from the nominal voltage may lead to additional energy losses, the reduction of life of equipment, the drop in productivity, wastes, etc.

The analysis was made using the K-means clustering algorithmfor alarge supply zone with 722 electric distribution substations divided in three areas (remote, close, and suburb) determined in function by distance between MV/LV substation and HV/MV supply electric station. Based on this algorithm the electric substations were classified in patterns described by linguistic terms from the point of view of slow voltage variations. These linguistic terms characterize the voltage quality from distribution networks. The clustering process led to five clusters for each area according to the indices of slow voltage variation. The analysis highlighted that the clusters correspond the levels of voltage quality often encountered in the operation of electric distribution networks. Thus, in the vast majority of the analysed substations(clusters C4 and C5) the voltagequality is *Good* and *Very Good* (about 70 percent)regardless of the electric distance from the electric supply station.

REFERENCES

[1] Bollen, M., Etherden, N., Yang, K., and Chang, G., *Continuity of Supply and Voltage Quality in the Electricity Network of the Future,* Proc. of 15th International Conference on Harmonics and Quality of Power (ICHQP), Hong Kong, June 17-20, 2012, pp. 375-377.

[2] EURELECTRIC, *Power Quality in European Electricity Supply Networks*, 2nd edition, Network of Experts for Standardization, Nov. 2003.

[3] Franken, B., Ajodhia, V., Petrov, K., Keller, K., and Müller, C.,*Regulation of Voltage Quality*, Proc. 9th Int. Conf. Electric Power, Quality and Utilisation, Barcelona, Spain, October 9 -11, 2007.

[4] Vinnal, T., Janson, K., Järvik, J., Kalda, H., and Sakkos, T., *Supply voltage quality in low voltage industrial networks of Estonia*,Estonian Journal of Engineering, vol. 18, no. 2, 2012, pp. 102-126.

[5] EN 50160, *Voltage characteristics of electricity supplied by public distribution systems. A Power Quality Standard*, [Online], Available: http://admin.copperalliance.eu/docs/librariesprovider5/power-quality-and-utilisation-guide/542-standard-en-50160-voltage-characteristicsin.pdf?sfvrsn=4&sfvrs n=4.

[6] Kriukov, A., Grigoras, G., Gavrilas, M., Scarlatache, F., and Vicol, B., *Use of Fuzzy Techniques in Reliability Assessment of Electric Distribution Systems*, Proc. of 16th International Conference on Harmonics and Quality of Power (ICHQP), Bucharest, May 25 – 28, 2014, pp. 689-693.

[7] Neagu, B., and Grigoras, G,*Assessment of Slow Voltage Variations from the Electric Distribution Systems with Fuzzy Techniques*, Proc. of 7th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Bucharest, 25 June -27 June, 2015.

[8] Vinnal, T., Kalda, H., andMõlder, H.,*Measurements and analyses of supply voltage magnitude and voltage variations in Estonian industrial companies*, Proc. 10th International Symposium of Topical Problems in the Field of Electrical and Power Engineering, Pärnu, January 10 – 15, 2011, pp.46-51.

[9] Iordache, M. and Conecini, I., *Electric Power Quality (in Romanian)*, Ed. Tehnica, Bucharest, 1997.

[10] Klaić, Z., Nikolovski, S., and Kraus, Z., *Voltage variation performance indices in distribution network*, Technical gazette, vol. 18, no. 4, 2011, pp. 547-551.

[11] Jain, A.K., Murty, M.N., and Flynn, P.J., *Data Clustering: A Review*, [Online], Available: https://www.cs.rutgers.edu/~mlittman/courses/lightai03/jain99data.pdf.

[12] Grigoras, G., Scarlatache, F., and Neagu, B., *Clustering in Power System. Applications*, Lambert Academic Publishing, Saarbrücken, 2016.

[13] Khalid Imam Rahmani, Md., Pal, N., and. Arora, K,*Clustering of Image Data Using K-Means and Fuzzy K-Means*, International Journal of Advanced Computer Science and Applications, vol. 5, No. 7, 2014, pp. 160 – 163,

[14] Halkidi, M., Batistakis, Y. and Vazirgiannis, M., *On Clustering Validation Techniques*, Journal of Intelligent Information Systems, vol. 17, No. 2/3, 2001, pp. 107 – 145.