An Approach for Calculation of Distribution Energy Losses Using Clustering Technique

Dragan S. Tasić[1] and Miodrag S. Stojanović[2]

Abstract – **This paper presents a method for calculation of annual distribution (electrical energy) losses based on clustering technique. Annual energy losses can be found calculating power losses for each hour during the year. However, this require large number of calculations (8760 i.e. 8784 for leap year). Clustering technique can be utilized to reduce the number of load flow calculations. In order to show applicability of clustering technique, annual simulation of distribution network has been made at the first. Power losses are calculated for each hour of the simulation. After that, number of clusters needed to obtain satisfying results is found.**

Keywords – **Distribution losses, Clustering technique, Estimation, Load category.**

I. Introduction

Power and energy losses are inevitable consequence of energy transmission and distribution from generation to consumer points. The total losses sometimes make ten or more percents of delivered energy. Therefore, it is important to have right estimate of losses, as well as to find ways for their reduction. Basic items of mentioned problems are: dispense technical and non-technical losses, determine structure of losses (distribution of losses throughout the network elements), locate critical elements from aspect of losses, and select optimal methods for losses reduction. The importance of technical losses becomes even higher for distribution utilities in deregulated environment since the non-technical losses will become out of concern (retail companies will take care of them).

Energy losses determination is more complex problem than determination of power losses, for which only conditions of the system at the specific moment are necessary. Namely, for determining energy losses in some network element, within specified time interval, it is necessary to know the curent curve of the element. Since this is very detailed work even in the case of only a few elements, it is usually simplified so that the load curve is divided into segments during which the load can be considered as constant. Calculation of energy losses, in this case, is reduced to a certain number of power losses calculations. Of course, this approach is correct only in the case when the load curve is not much variable. In other way, the approximation of load curve with small number of segments leads to great errors and, of course, bed final results.

In this paper clustering technique is applied to select characteristic states, based on which we can determine distribution energy losses with the satisfying accuracy. Using this technique, selection of characteristic states is made quite objectively. Presented method does not require to much of knowledge about a distribution network that has being analyzed, which enables its application even in the cases of sudden changes in circumference and structure of consumption as well as in configuration of network.

II. Simulation of Load Curves

Knowledge of load curves for each particular element is needed for exact calculation of distribution energy losses. That is not possible, because of measurements are made only at some locations in the network. Therefore, it is imposed problem to estimate load curves of network elements based on the available data. This paper only discusses application of clustering technique to annual distribution losses calculation problem. Therefore, load curves of network elements are simulated without considering estimation aspects. Two simulation methods, with quite different load curves, are applied in order to point out general applicability of the clustering technique.

The first simulation method requires knowledge of the following data for each element:

- annual peak loading (winter maximum $S_{\max}^{\mathrm{W}}$),
- annual minimal loading (summer minimum $S_{\min}^{\mathrm{S}}$),
- minimal daily loading of the day when annual peak loading appears ($S_{\min}^{\mathrm{W}}$),
- maximal daily loading of the day when annual minimal loading appears ($S_{\max}^{\mathrm{S}}$),
- hour when annual peak loading appears ($T_{\max}^{\mathrm{W}}$),
- hour when maximal daily loading appears on the day of annual minimal loading ($T_{\max}^{\mathrm{S}}$).

The hours when daily peaks on the days of annual minimum and maximum appear are considered in order to comprehend different load types, i.e. non simultaneous of daily peak loadings for load nodes.

It is supposed maximal and minimal daily loadings are changed during the year following the cosine functions. Beside this, it is supposed that annual peak demand appear on the January 1. Maximal and minimal daily loadings of node *k* on the day *i* are:

[1]Dragan S. Tasić is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Yugoslavia, E-mail: dtasic@elfak.ni.ac.yu

[2]Miodrag S. Stojanović is with the Faculty of Electronic Engineering, University of Niš, Aleksandra Medvedeva 14, 18000 Niš, Yugoslavia, E-mail: miodrag@elfak.ni.ac.yu

$$S^{i}_{\max k} = S^{\mathrm{W}}_{\max k} - (S^{\mathrm{W}}_{\max k} - S^{\mathrm{S}}_{\max k})\cos(2\pi \frac{i}{n_d}) \ , \qquad (1)$$

$$S^{i}_{\min k} = S^{\mathrm{W}}_{\min k} - (S^{\mathrm{W}}_{\min k} - S^{\mathrm{S}}_{\min k})\cos(2\pi \frac{i}{n_d}) \ . \qquad (2)$$

where n_d is number of days (365 or 366 for a leap year).

The hour of the day *i* when daily peak loading appears is:

$$T^{i}_{\max k} = T^{\mathrm{W}}_{\max k} - (T^{\mathrm{W}}_{\max k} - T^{\mathrm{S}}_{\max k})\cos(2\pi \frac{i}{n_d}) \ . \qquad (3)$$

When maximal and minimal powers as well as hour of daily peak are determined for day *i*, we can calculate power for each hour of day *i*. Power of load node *k* for hour *j* of day *i* is:

$$S^{i,j}_{k} = S^{i}_{\min k} + (S^{i}_{\max k} - S^{i}_{\min k})\cos(2\pi \frac{j - T^{i}_{\max k}}{24}) \ . \qquad (4)$$

A form of supposed annual loading curve for the node *i* is shown on the Fig. 1.

The second simulation method starts from the assumption that each load node can put into one of load categories. Hourly load patterns for different load categories (Fig. 2) are known as well as annual peak loading of each load node. Beside that, it is known variation of maximal and minimal relative daily loading during the year (Fig. 3).

Maximal and minimal loading of load node *k* for day *i* are:

$$S^{i}_{\max k} = S^{\mathrm{W}}_{\max k} \cdot p_{\max}(i) \ , \qquad (5)$$

$$S^{i}_{\min k} = S^{\mathrm{W}}_{\min k} \cdot p_{\min}(i) \ . \qquad (6)$$

When powers $S^{i}_{\max k}$ and $S^{i}_{\min k}$ are determined, we can calculate power of load node *k* for *j*-th hour of *i*-th day considering curves on Fig 3:

$$S^{i,j}_{k} = S^{i}_{\min k} + (S^{i}_{\max k} - S^{i}_{\min k})\, p^{d}(j) \ . \qquad (7)$$

III. CLUSTERING TECHNIQUE

Clustering is one of the methods for analyzing and processing large and not well-known amount of data. This is the method of classifying the data set into subsets, clusters, based on a defined similarity measure [1, 2]. On this way, a set of characteristic states that describe analyzed problem can be generated.

The main characteristic of the cluster is its center. The center of cluster or centroid is an average of all examples that belong to this cluster. In other words, it is representative of all examples that belong to the cluster.

In many cases, there is no clear criterion under which "real" number of clusters can be determined. That is the reason for experimenting with different number of clusters in order to determine what way of clustering is appropriate for specific application.

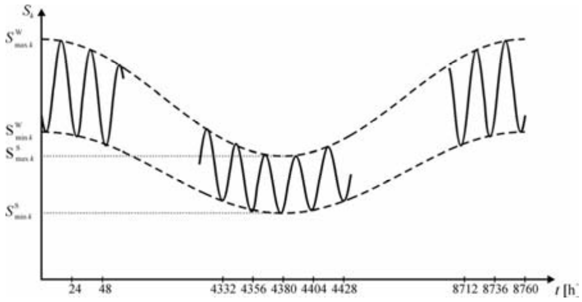

Fig.1. Cosine form of annual load curve

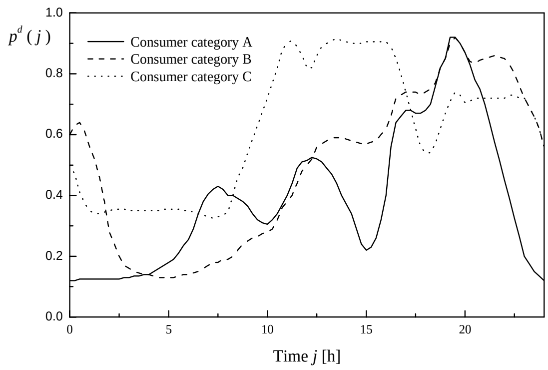

Fig.2. Hourly load patterns for different consumer categories

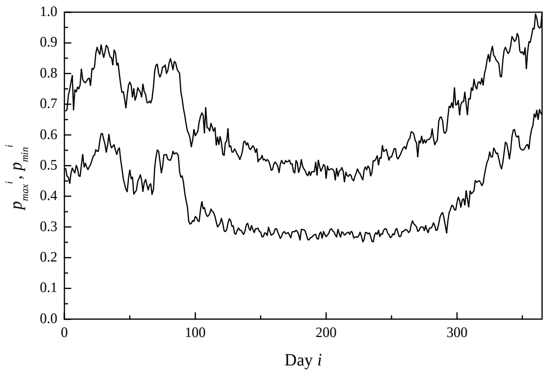

Fig.3. Variation of maximal and minimal daily loading during the year

In process of preparing data for clustering, there is a need for normalization of data. This is done because in many cases the task of classifying data which represent different kinds of variables is faced (for example powers, voltages, currents,..). Normalization enables computation of the classifying centers that are completely independent of the physical units of variables (V, A,…). With this procedure value range of every attribute is brought to the range 0÷1. Therefore, normalized maximal value of every attribute will be 1, and normalized minimal value 0. It should be noticed that normalization at the same time mean increasing the influence of variables witch change in a narrow range. Otherwise, it should be pointed out that the normalization is not necessary in cases where analyzed problem is one-dimensional.

There are a large number of algorithms for solving the clustering problem. Most of them are based on methods of "joining", "cutting" and "rearranging". These methods are used in popular algorithms: clustering through minimum graph tree, algorithm of maximum cross-section and reclassification based on the nearest average.

In this paper clustering is made on the complete state vector that includes active and reactive powers of all load nodes and root node voltage for every hour during the year. Therefore, dimension of the state vector is:

$$DVS = 2 \cdot N_{\ln} + 1 \ , \tag{8}$$

where $N_{\ln}$ is number of load nodes.

Firstly, the normalization of attributes using the general relations has been done:

$$x_{ji} = \frac{X_{ji} - X_{j\min}}{X_{j\max} - X_{j\min}} \ , \tag{9}$$

where $X_{j\max}$ is maximum and $X_{j\min}$ minimum of j-th attribute by all examples.

Then the clustering of data on subset of hours is done. As a similarity measure (degree of association) Euclidian distances has been utilized. In general case for two vectors X and Y of dimension N Euclidian distance is determined by the relations:

$$ED(XY) = \left[\frac{1}{N} \sum_{i=1}^{N} (X_i - Y_i)^2 \right]^{1/2} \ . \tag{10}$$

The clustering is done in the way that distance between the center of the each cluster and every example that belongs to that cluster, is smaller than a priory defined boundary Euclidian distance ε_k. The process of clustering begins with the first considered example that is set to be the center of first cluster. Then, by the determined order, the classifying of second example is done. If the Euclidian distance is smaller then the boundary one, it is put in the first cluster, otherwise it forms the second cluster. In the case when second example is put in the first cluster, a new center of that cluster should be recalculated. Then, the distance between the third example and the center of the first and second cluster, if it exists, is calculated. This example is put in the nearest cluster, i.e. in the cluster from which center has the smallest distance, under condition that the distance is smaller then specified boundary Euclidian distance. If that is not the case, it forms a new cluster. The procedure is carried out until all the examples are classified. According to this, it is clear that the number of cluster depends on the boundary value of Euclidian distance, and is not known at the beginning of clustering process.

Since complete state vector is considered, attributes of the example which represents the center of cluster are, of course, active and reactive powers of load nodes, as well as the voltage of root node. The center of cluster is determined by general equation:

$$b_{jk} = \frac{\sum_{i=1}^{r_k} x_{ji}}{r_k} \ , \tag{11}$$

where b_{jk} is j-th attribute of center of k-th cluster, and r_k number of examples that belong to k-th cluster.

After finished clustering process, a detailed calculation of characteristic states of network is made by procedure presented in [3]. As the result of calculations are obtained node voltages, power flows and power losses for each network element, as well as the total power losses. It is considered that loads are constant during the set of hours that belong to the cluster. Because of that, annual energy losses can be calculated as sum of power losses of cluster centers multiplied by the number of hours that belong to that cluster.

IV. Test Example

The presented procedure is used for determining annual energy losses of the test network shown in Fig. 4. The data about lines are given in [4], while data about maximal and minimal node powers as well as rated powers of distribution transformers are shown on Table I. This table also shows the time when appear winter and summer peak loading for each load node. Load nodes are classified into three consumer categories marked generally as A, B and C category (for example residence, commerce and industry).

Table I
Maximum powers and participation of customer categories

	$P^W_{\max}$ [MW]	$P^W_{\min}$ [MW]	$P^S_{\max}$ [MW]	$P^S_{\min}$ [MW]	$Q^W_{\max}$ [MVAr]	$Q^W_{\min}$ [MVAr]	$Q^S_{\max}$ [MVAr]	$Q^S_{\min}$ [MVAr]	$t^W_{\max}$ [h]	$t^S_{\max}$ [h]	category	S_n [MVA]
1	0.30	0.24	0.18	0.06	0.18	0.10	0.096	0.0240	14	17	A	0.25
2	0.27	0.15	0.15	0.12	0.12	0.024	0.06	0.03	15	19	A	0.25
3	0.36	0.24	0.30	0.12	0.24	0.18	0.18	0.06	12	18	A	0.4
4	0.18	0.12	0.12	0.06	0.09	0.06	0.075	0.03	16	20	B	0.25
5	0.18	0.12	0.15	0.06	0.06	0.03	0.045	0.015	14	12	B	0.25
6	0.60	0.30	0.48	0.21	0.30	0.18	0.21	0.09	15	16	A	0.4
7	0.60	0.39	0.36	0.15	0.30	0.24	0.24	0.06	16	19	A	0.4
8	0.18	0.12	0.12	0.06	0.06	0.03	0.045	0.015	13	19	A	0.25
9	0.18	0.12	0.12	0.06	0.06	0.03	0.045	0.021	14	19	C	0.25
10	0.135	0.075	0.075	0.045	0.09	0.03	0.06	0.03	15	18	A	0.25
11	0.18	0.12	0.12	0.06	0.105	0.075	0.075	0.045	15	15	A	0.25
12	0.18	0.12	0.12	0.06	0.105	0.045	0.075	0.045	11	20	B	0.25
13	0.36	0.24	0.24	0.12	0.24	0.18	0.21	0.06	10	18	B	0.4
14	0.18	0.12	0.12	0.06	0.03	0.018	0.021	0.009	12	16	A	0.25
15	0.18	0.12	0.12	0.06	0.06	0.03	0.045	0.021	13	15	C	-
16	0.18	0.12	0.12	0.06	0.06	0.03	0.045	0.021	14	17	A	0.25
17	0.27	0.21	0.18	0.09	0.12	0.06	0.06	0.052	15	17	A	0.25
18	0.27	0.18	0.18	0.09	0.12	0.06	0.09	0.052	15	19	A	0.25
19	0.27	0.18	0.12	0.09	0.12	0.06	0.09	0.036	13	19	B	0.25
20	0.27	0.18	0.18	0.09	0.12	0.06	0.09	0.036	16	19	C	-
21	0.27	0.18	0.18	0.09	0.12	0.06	0.09	0.03	12	18	A	0.4
22	0.27	0.21	0.18	0.06	0.15	0.12	0.09	0.03	13	15	A	0.4
23	0.84	0.60	0.56	0.30	0.40	0.30	0.30	0.15	14	16	A	0.63
24	0.84	0.60	0.56	0.27	0.40	0.22	0.30	0.15	12	17	A	0.63
25	0.18	0.12	0.12	0.03	0.075	0.045	0.045	0.015	13	17	A	0.25
26	0.18	0.12	0.12	0.06	0.075	0.045	0.045	0.018	13	17	A	0.25
27	0.18	0.12	0.12	0.03	0.075	0.03	0.03	0.021	13	18	A	0.25
28	0.36	0.24	0.24	0.12	0.15	0.12	0.15	0.036	14	18	A	0.4
29	0.60	0.24	0.42	0.18	0.18	0.12	0.12	0.03	15	18	B	0.4
30	0.45	0.39	0.30	0.15	0.21	0.15	0.15	0.06	16	17	A	0.4
31	0.63	0.33	0.33	0.15	0.30	0.21	0.21	0.06	14	18	B	0.63
32	0.18	0.12	0.12	0.06	0.12	0.03	0.12	0.03	15	20	A	0.25

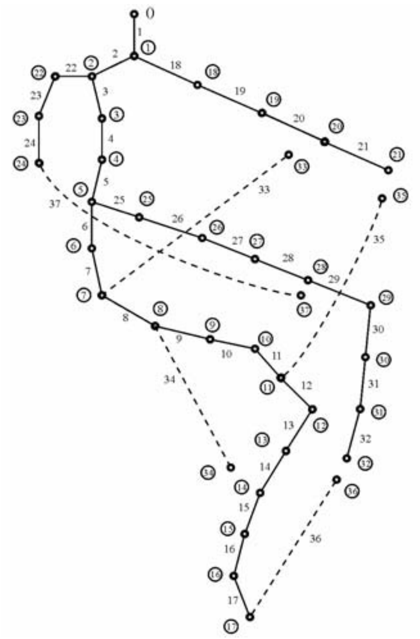

Fig.4. Test network with 32 nodes

TABLE II
RESULTS OF THE FIRST SIMULATION METHOD

number of clusters	ΔW_{a_p} [MWh]	$\Delta\%$	ΔW_{a_q} [MVArh]	$\Delta W^{L}_{a_p}$ [MWh]	$\Delta W^{L}_{a_q}$ [MVArh]	$\Delta W^{T_{Fe}}_{a_p}$ [MWh]	$\Delta W^{T_{Cu}}_{a_p}$ [MWh]	$\Delta W^{T_{Cu}}_{a_q}$ [MVArh]
1	4585.3	10.8	4716.4	3706.4	2465.2	195.64	683.27	2251.2
2	4896.7	3.49	5044.8	3972.3	2643.6	195.5	728.88	2401.2
3	4997.4	1.5	5151.5	4058	2701	195.45	743.91	2450.5
5	5037.5	0.71	5194.2	4092.2	2723.9	195.43	749.90	2470.2
10	5042.4	0.61	5205.1	4094.1	2725.2	195.43	752.88	2479.9
20	5057.6	0.32	5223.6	4106.1	2733.3	195.42	756.05	2490.2
25	5060.5	0.26	5227.4	4108.3	2734.8	195.42	756.79	2492.6
30	5064.7	0.175	5232.3	4111.8	2737.1	195.42	757.55	2495.2
50	5067.9	0.115	5236	4114.2	2738.8	195.42	758.18	2497.2
100	5070	0.07	5238.6	4116	2740	195.42	758.63	2498.7
8760	5073.6		5243	4118.9	2741.9	195.41	759.37	2501.1

TABLE III
RESULTS OF THE SECOND SIMULATION METHOD

number of clusters	ΔW_{a_p} [MWh]	$\Delta\%$	ΔW_{a_q} [MVArh]	$\Delta W^{L}_{a_p}$ [MWh]	$\Delta W^{L}_{a_q}$ [MVArh]	$\Delta W^{T_{Fe}}_{a_p}$ [MWh]	$\Delta W^{T_{Cu}}_{a_p}$ [MWh]	$\Delta W^{T_{Cu}}_{a_q}$ [MVArh]
1	2582.9	10.83	2546.9	2014.1	1341.7	202.57	366.21	1205.2
2	2800.5	3.32	2777.3	2199.4	1465.6	202.47	398.59	1311.7
3	2849.8	1.62	2832.5	2243.2	1494.3	202.45	406.93	1333.6
5	2864.6	1.11	2845	2254.2	1502.2	202.44	408.05	1342.8
10	2873.5	0.8	2854.5	2261.7	1507.3	202.44	409.39	1347.2
20	2886.3	0.36	2870.4	2271.7	1514.1	202.43	412.16	1356.3
25	2889.3	0.26	2874.3	2274	1515.6	202.43	412.87	1358.7
30	2890.1	0.23	2875.3	2274.7	1516.1	202.43	413.03	1359.2
50	2891.1	0.20	2876.2	2275.5	1516.7	202.43	413.25	1359.5
100	2894.5	0.08	2880.7	2278.1	1518.4	202.43	413.95	1362.2
8760	2896.8		2883.4	2279.9	1519.7	202.43	414.41	1363.7

Results of calculation are shown on Table II and Table III. The second and the fourth columns of tables show total active and reactive losses, the fifth and the sixth columns line losses, while columns 7, 8 and 9 show transformer losses. The third column of tables shows errors in percents for total active losses calculated using clustering technique. Exact values of loses are obtained summing power losses for each hour during the year. This corresponds to calculation with 8760 clusters. Differences between energy losses for the first and the second simulation method are obviously result of using the quite different load curves. In other hand, the errors made for same number of clusters is approximate for both simulation methods. This shows that presented method is very robust.

Results given in Table II and Table III also show rather decreasing of error with increasing number of clusters. For the case of three clusters error is less then 2%, and for five clusters error is less then 1%. This error is quite acceptable. Computing time increases with increasing the number of clusters. Therefore authors suggest 5-10 clusters for distribution losses calculation. The range is suggested because of increasing of computing time is not continual. Thus the problem maybe solved quicker with 6 than with 8 clusters. The reason for this lies in the methodology of boundary distance determination for chosen number of clusters.

V. CONCLUSION

The method for distribution losses calculation using clustering technique is presented in this paper. Presented method enables the total energy losses determination as well as determination energy losses for each network element (structural losses analysis). Simulations that have been made by the authors show that satisfying results can be obtained with a quite small number of clusters. Considering the fact that smaller number of clusters requires less of computing time, it is pointed out here that for the need of determining electrical energy losses in distribution network is enough to use up to 10 clusters.

REFERENCES

[1] N. Arsenijević, N. Rajaković, D. Muškatirović, "Proračun gubitaka energije na godišnjem vremenskom horizontu u prenosnoj mreži elektroenergetskog sistema EPS-a", JUKO-CIGRE, Herceg Novi, 16-20. septembar 2001., R 38-04 i R 38-05.

[2] ***, Analiza gubitaka električne energije u prenosnoj mreži EPS-a i iniciranje mera za njihovo sniženje, Studija urađena za EPS, Beograd, 2000.

[3] D. Shirmohammadi, H. W. Hong, A. Semlyen, G. X. Luo, “A Compensation-Based Power Flow Method for Weakly Meshed Distribution and Transmission Networks”, IEEE Trans. on Power Systems, Vol. 3., No. 2, May 1988, pp. 753-762.

[4] S. K. Goswami, S. K. Basu, "A New Algorithm for the Reconfiguration of Distribution Feeders for Loss Minimization", IEEE Trans. on Power Delivery, Vol. 7., No. 3, July 1992, pp. 1484-1491.

[5] Hang-Ching Kuo, Yuan-Yih Hsu, "Distribution System Load Estimation and Service Restoration Using a Fuzzy Set Approach", IEEE Trans. on Power Delivery, Vol. 8, No. 4, October 1993.