A Combination of K-Mean Clustering and Elbow Technique in Mitigating Losses of Distribution Network

Minh Quan Duong*, Le Hong Lam, Bui Thi Minh Tu, Giap Quang Huy, and Nguyen Huu Hieu

Abstract— *The mitigation of losses in the distribution network has become the main target of the power companies using both technique and non-technique solutions. Today, two common solutions which are: (i) economic compensation and (ii) reconfiguration of the network. Here, most use is economic compensation using the capacitor to compensate reactive power. However, due to the significant expansion of the distribution network, the available capacitors in the distribution network might not bring maximum economic efficiency. Therefore, the main aim of this paper is to identify the number of capacitors in terms of minimizing loose and maximizing economic. Currently, the commercial software, PSS/ADEPT, are commonly used, but the main drawback is that it uses the consumption of a random day to represent as a typical day for either one session or one year. In fact, the consumption is always variability day by day, thus it is a cause of the incorrect compensation of reactive power. This paper proposes a solution to create a typical load graph for different levels of consumption in one year. The proposed solution is also tested with the real distribution network of Quang Ngai province. The result shows that the proposed solution works properly in terms of reducing losses and enhance the economic aspect.*

Keywords— **K-Mean clustering, Elbow technique, reactive power compensation, power losses, distribution network, and Vietnam power system.**

1. INTRODUCTION

According to the declaration of the Vietnamese power company [1] which presents the planning of the power system from 2011-2020 and considering 2030, Vietnam will import 2300 MW in 2020 (occur 3.1% total consumption), and 7100 MW in 2030 (occur 4.9% total consumption). This shows that the rapid industrialization process has made energy supply fail to meet demand. Therefore, the main target of the Vietnamese power company is reducing the power losses in distribution networks. Every year, a huge budget is allocated to repair and buy equipment in order to reduce the losses. However, the drawback of this activity is the limitation of budget and too difficult to investigate the efficiency of the investment on reducing losses.

In general, the distribution network has specific designs and different operation modes in comparison to the transmission network. Here, the distribution network is distributed in a wide area, thus usually operate with the unbalance mode and has dramatically losses. It should be noted that the losses of the distribution network have a tidy relationship with technical issues of the network from the design stage to the operation stage.

Therefore, new solutions applied to reduce the losses of the network to a reasonable value are the new target of the distributed operators. This mission is very essential since the resources and demand are unbalance. The rate of losses depends on the characteristic of network, power flow, and the management of the operator. It should be noted that the losses are divided into two types: (i) technical losses and (ii) commercial losses.

The technical losses are considered as the power losses when the power is transmitted from sources to demand. Meanwhile, the commercial losses are considered as the power losses cause of the illegal use of electric power, i.e. stealing power, the error of meters. These losses are hard to compute but it has a large impact on the network, consequently, it increases the power losses level of the system.

However, this topic is lack of the consideration of researchers since many publications focused on renewable energy [2, 3, 4, 5] in Vietnam. The paper [2] focuses on the impact of the wind turbines on the stability of the West of the Vietnamese power system in cases of small disturbances. Additionally, papers [3] and [4] investigate deeply about the impact of some wind projects on Ninh Thuan city and Quang Ngai city, respectively. On another direction, paper [5] proposes a model for the wind farm to participate in the ancillary service, thus the wind farm can be optional to support the system.

The level of power losses depends on the characteristics of the network, the amount of power transmission, the supply capacity of the system and the management and operation of the power system. Researches and applications of new solutions to reduce power losses to a reasonable level have been and will be a major goal of the electricity industry. Previous studies have shown many methods to reduce power losses on the distribution grid. Among them, the most commonly used and proven solution is to use capacitor devices placed in parallel with the line to compensate the reactive power [6-8]. As a result, the reactive power line transmitted on

Minh Quan Duong, Le Hong Lam , Bui Thi Minh Tu, Giap Quang Huy, Nguyen Huu Hieu are with The University of Danang – University of Science and Technology, Faculty of Electrical Engineering, Power System Research Group, Vietnam, 54 Nguyen Luong Bang St., Danang city, Vietnam.

*Corresponding author: Minh Quan Duong; E-mail: dmquan@dut.udn.vn.

the line decreases, resulting in reduced power losses.

However, it is necessary to have an accurate load graph, also known as a typical load graph, to work optimally. But in practice, choosing a typical load graph is done in two directions: (i) select load graphs any days of the year or (ii) use the maximum load graph. Selecting any load graph as a typical load graph will cause errors in reactive power compensation such as wrong capacity, incorrect placement, continuity switching of the capacitor causing a shortening of the life of capacitor as well as switchgear.

In this paper, the authors propose an advanced method to create a typical load graph. The proposed method is a combination of the Elbow and K-means algorithms to select some representative days in one year. Finally, the proposed method is applied to the actual grid model of Quang Ngai province through PSS/ADEPT software to optimize the number of capacitors in this distribution network.

2. CREATING A TYPICAL LOAD GRAPH

The proposed method includes two steps: (i) arrange days in one year into some groups which contain similar consumption and (ii) creating typical load graphs of one day from these groups. In the first step, the combination of K-means and Elbow method explained in Section 2.2 and Section 2.3 is used. Here, the Elbow method will select the number of groups and the K-mean method will generate the typical load associated with the number of groups. Meanwhile, the second step is explained clearly in Section 2.4.

2.1 Common methods to create the typical load graph [9]

2.1.1 Partitionting Clustering Algorithms

This method conducts data partitioning. Firstly, it creates an initial data set of partition *k* (number of partitions to build) or centroids (central value of that region), then using iterative migration techniques to improve partitioning by moving objects from one group to another. This method consists of two algorithms that are popular with *K-means* and *K-methods*. Usually *k* is randomly selected and then repeats the clustering of data points into partitions *k* such that the clustering criteria are optimized. This means that the sum of squared errors between objects in a partition compared to centroids is minimal.

2.1.2 Hierarchical Algorithms

This method conducts dividing data sets of objects into a hierarchy of a group. It can be represented by a tree structure diagram called dendrogram. The root node representing the entire data set for each branch is a unique object of the data set. Clustering results can be obtained by dendrogram at different levels. There are two general approaches to decentralized methods: (i) the *combination* (from bottom up) and (ii) *division* (top down). The combination or division ends when the desired number of clusters has been formed. Usually the number of repetitions is based on a certain criterion and is usually the distance between clusters. The algorithms that represent this method are *cure*, *birch* and *chameleon*.

2.1.3 Density-Based Methods

To detect clusters of any shape, density-based clustering methods have been developed. This method often considers clusters as dense regions of objects in the data space separated by low density areas that exhibit noise and have two main approaches. The first method is pinning to a training data point and representative algorithms including *dbscan* and *optics*. The second approach is to pin to a point in the attribute space and it includes the *denclue* algorithm.

Currently, the algorithm is widely used to study clustering problems as *K-Means* algorithm because of its simplicity and ease of implementation, it can handle quite a large dataset so the author chooses this algorithm. to solve the above problem.

2.2 K-means algorithm [10]

The K-means algorithm is implemented with the purpose of dividing data into different clusters so that the data in the same cluster has the same properties.

The simplest idea about a cluster is to gather points close to each other in a certain space. Fig. 1 shows a typical example of 3 data clusters (abbreviated as cluster).

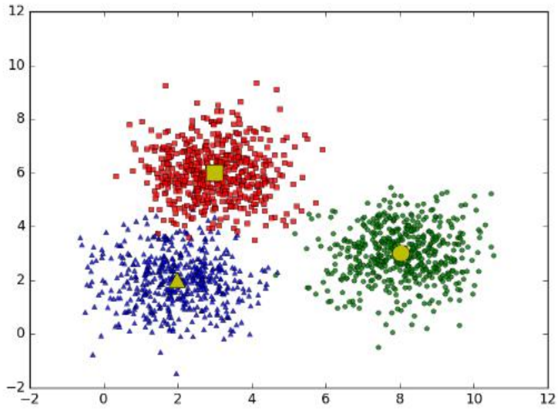

Fig. 1. An example of clustering data.

2.2.1. Analysis from mathematical point of view

From the input data and the number of groups which we want to find, the algorithm indicates the center of each group and divides the data points into the corresponding groups. Assume that each data point belongs to only one group.

Suppose that N data points are $X = [x_1, x_2, \ldots, x_N]$ and $k < N$ is the number of clusters to divide. The algorithm should show the centroid points: $m_1, m_2, \ldots, m_k$ and subgroups of each data point. For each data point x_i, we set $y_i = [y_{i1}, y_{i2}, \ldots, y_{ik}]$ as its vector, where if x_i is assigned to cluster k then $y_{ik} = 1, y_{ij} = 0, \forall j \neq k$. If you consider center m_k to be the centroid of each cluster and estimate all points assigned to this cluster, then a poker data point assigned to cluster k will have a margin of error $(x_i - m_k)$. For this error to have the smallest absolute value, the algorithm will find a way to make the equation (1) reaches the minimum value:

$$\left\| x_i - m_k \right\|^2 \quad (1)$$

Furthermore, x_i is assigned to cluster k, so y_{ik} = 1, y_{ij}=0, $\forall j \neq k$. Then, the expression (1) will be rewritten as:

$$y_{ik} \left\| x_i - m_k \right\|^2 = \sum_{j=1}^{k} y_{ij} \left\| x_i - m_j \right\|^2 \quad (2)$$

The entire data error will be:

$$L(Y, M) = y_{ij} \left\| x_i - m_j \right\|^2 \quad (3)$$

In which: $Y = [y1, y2, ..., yN]$, $M = [m1, m2, ..., mk]$ are the vectors of each data point and centroid, respectively, of each cluster.

Since then, it is necessary to take steps to optimize the problem:

$$Y, M = agr\,min_{Y,M} \sum_{i=1}^{N} \sum_{j=1}^{k} y_{ij} \left\| x_i - m_j \right\|^2 \quad (4)$$

a. Fix M, find Y

Assuming that centers are found, the algorithm must find the vector *y* so that the loss function reaches the smallest value. This is equivalent to finding the cluster for each data point. When the centroids are fixed, the problem of finding vector *y* for all data can be subdivided into the problem of finding vector *y* for each xi data point as follows:

$$Y_i = agr\,min_{Y_i} \sum_{j=1}^{k} y_{ij} \left\| x_i - m_j \right\|^2 \quad (5)$$

Since there is only one element of label vector y_i equal to 1, the problem can continue to be written in a simpler form:

$$j = agr\,min \left\| x_i - m_j \right\|^2 \quad (6)$$

Since $\left\| x_i - m_j \right\|^2$ is the square of the distance from the point x_i to centroid m_j, it can be concluded that each point belongs to the nearest centroid cluster. From that we can easily deduce the vector *y* of each data point.

b. Fix Y, find M

Assuming a cluster has been found for each point, the new centroid algorithm is needed for each cluster so that the loss function reaches the smallest value. Once the *y* vector has been defined for each data point, the centroid problem is expressed by the equation:

$$m_i = agr\,min_{m_j} \sum_{i=1}^{m} y_{ij} \left\| x_i - m_j \right\|^2 \quad (7)$$

At this point, the solution of the equation can be found through the method of derivative functions equals to zero, the optimal function is a continuous function and has a definite derivative at every point. Derivative (7) we get:

$$\frac{\partial (m_j)}{\partial m_j} = 2 \sum_{i=1}^{n} y_{ij} (m_j - x_i) \quad (8)$$

Solving the derivative equation equals to 0:

$$mj \sum_{i=1}^{N} y_{ij} = \sum_{i=1}^{N} y_{ij} x_i \quad (9)$$

It can be seen that the main denominator is the counting of the number of data points in cluster *j*, and the numerator is the sum of the data points in cluster *j*. In a simpler way: m_j is the arithmetic mean of points in cluster *j*.

2.2.2. Algorithm

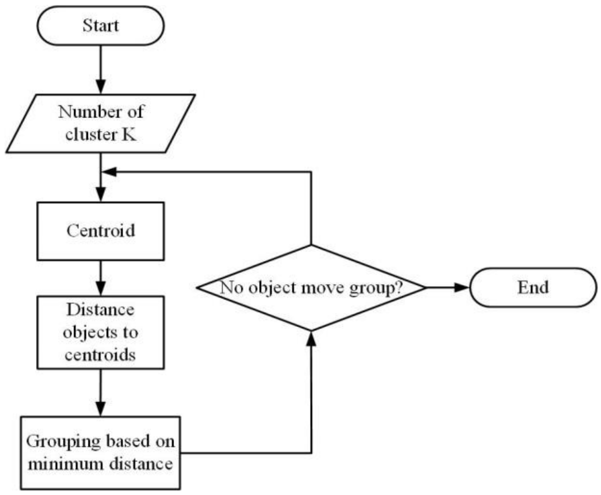

Fig.2. K-means Algorithm.

However, K-Means is still a simple algorithm; the number of *k* clusters is determined by humans. This leads to incorrect clustering. Therefore, when using this algorithm, a number of ways are used to see if *k* is the appropriate cluster number. One commonly used method is Elbow-method.

2.3 Elbow algorithm

The idea of this method is to run the clustering problem for the input data set within a range of values *k* (e.g *k* = 1:10). For each value *k*, the Sum of Squared Error (SSE) deviation of the data points belonging to a cluster versus centroid of that cluster will be calculated.

Then the line graph of SSE is plotted for each value of *k*. If the line chart looks like an arm then the "elbow" on the arm gives the *k* value the best. This algorithm wants to have a small SSE, but SSE will tend to decrease to 0 when k increases. Therefore, the author's goal is to choose a small value of k where SSE is low, and the elbow is where the error starts to decrease by increasing *k* (shown in Fig. 3).

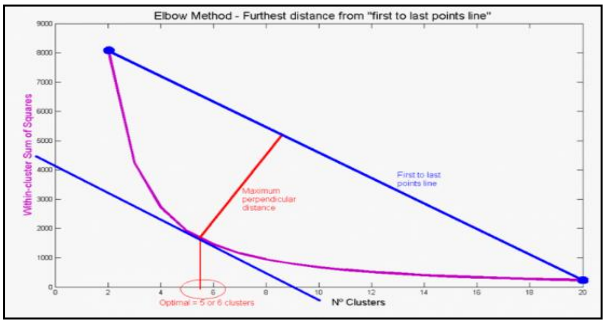

Fig. 3. Elbow Algorithm.

2.3. Creating typical load graphs

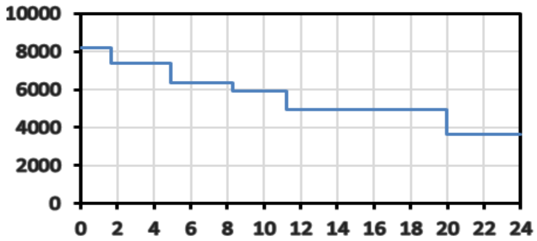

Fig.4. Extended graph of load.

Suppose the summer consists of n_1 days, winter consists of n_2 days. The power level P exists for time t_1 in the summer load graph and exists for time t_2 in the winter load graph. So, for a year, the P level exists for the time period:

$$T = t_1 n_1 + t_2 n_2 \tag{10}$$

Applying (10), we get the annual load graph shown as Fig. 4 which is taken into the form 24h.

3. TEST AND RESULT

Today, Quang Ngai electricity grid has a lot of outgoing routes such as the route 471 to supply electricity for the residential area, 477 to supply electricity to Quang Phu industrial park, and there are many other outbound routes. We use output 471 of E16.1 including one source and 174 nodes, the number of load: 58, 22kV voltage as shown in Fig. 5 to run the simulation.

Based on the load data series provided within one year 2017 (historical data), then we obtain a graph of some days drawn as shown in Fig. 6 because the authors select randomly 20 days to draw in Fig. 6 with the simplified purpose.

3.1. Clustering days

After running the proposed mode which is implemented in MATLAB environment [12], the load for one year of the outgoing routes 471 is divided into two groups as shown in Fig. 7 and the typical load graph for one year in Fig. 6 is the centroid representing these two load groups. In Fig. 7, the blue cells present the low load and the yellow cells means the high load.

Based on the Fig. 7, it can be predicted that: from June to September every year in the hot and dry season, the residential load is high. From October to May next year, the load is lower.

In order to simplify the data to serve the purpose of calculation more convenient, we perform the linearization of the graph in Fig. 8, we have the graph as shown in Fig. 9 as the graph after being linearized. Here, the blue and red line represent for two different groups of grahp of load, the x-axist is for period and y-axist is for active power.

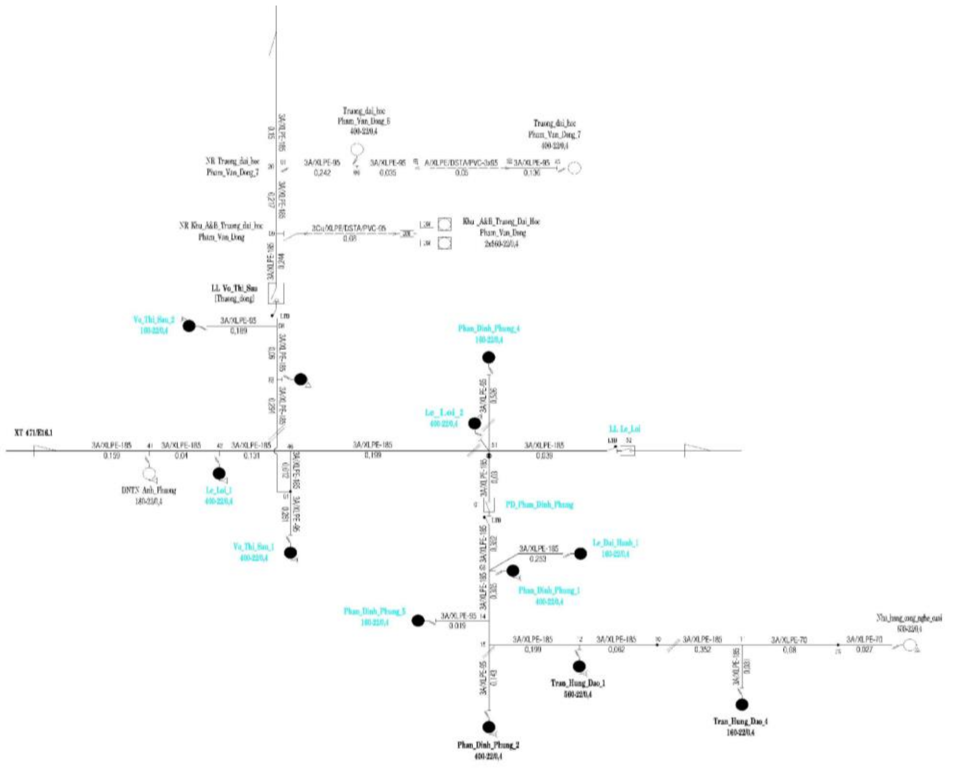

Fig. 5. Outgoing routes 471-E16.1 Quang Ngai residential area.

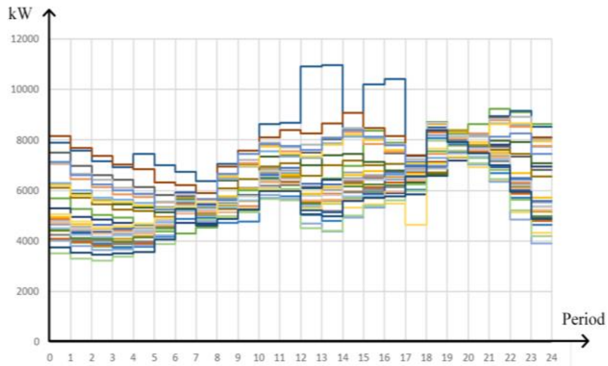

Fig.6. Load graphs of days in 2017.

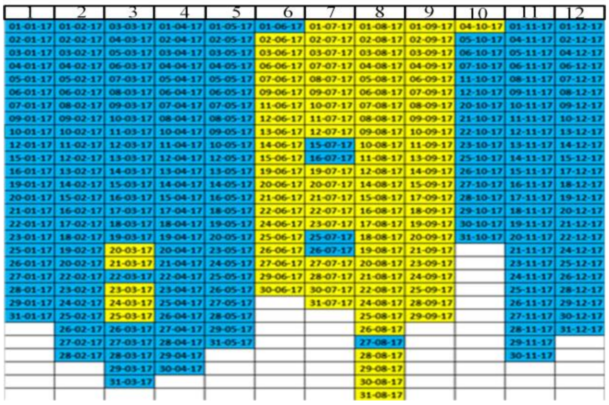

Fig. 7. Results of subgroup load.

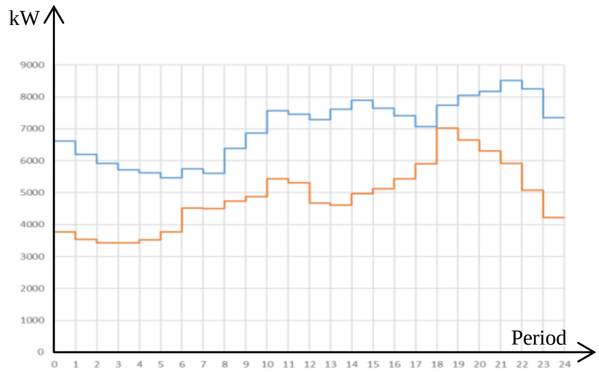

Fig. 8. Clustering result of two groups of selected days.

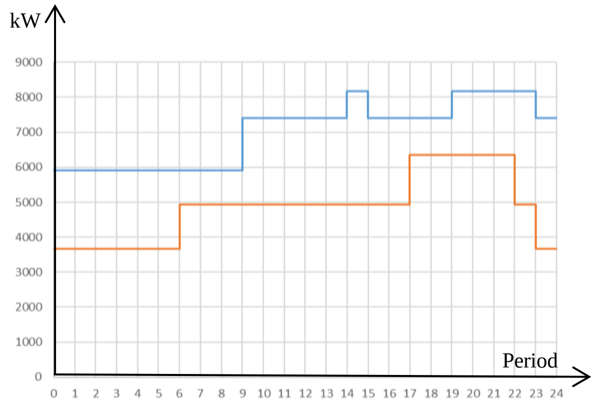

Fig. 9. Graph of clustering results after linearity.

3.2 Apply the proposed method to optimize number of capacitors in distributed network

Applying the proposed model to the distribution grid in Quang Ngai province on the route 471 belongs to E16.1 includes one source and 174 nodes, 22kV voltage as shown in Fig. 5.

With the above outgoing route, the author will perform the distribution grid optimization through the CAPO problem in PSS/ADEPT software [11] so that there can be comparisons, assessments and proofs for this new method. The input data of PSS/ADEPT is set following:

- Price of electrical energy: 0.06 $/KWh;
- Discount rate: 0.06 pu/year;
- Inflation rate: 0.03 pu/year;
- Evaluation period: 5 years;
- Installation cost for fixed capacitor banks: 5 $/Kvar;
- Installation cost of swiched capacitor banks: 8 $/kVar;
- Maintenance rate for switched capacitor banks: 0.5 $/kVar for one year;
- Size of fixed capacitor banks: 300 KVar;
- Size of switched capacitor banks: 100 Kvar.

The load graphs are shown for both cases: (i) maximum load in one year denoted maximum load and (ii) typical day after applying (10), denoted proposed method. The maximum load is selected by Quang Ngai power company, meanwhile the load graph of typical day is obtained after running the proposed model.

In order to simplify the computation, we conduct linearization of the load graph of the typical day into three levels corresponding to three different periods of one day in Table 1.

Table 1. Post-linearization capacity levels

Level	Active power (kW)	Period
1	8371.41	19h – 24h
2	7300.28	9h – 19h
3	5918.69	0h – 9h

After performing the CAPO problem in PSS/ADEPT of two cases for the actual load data corresponding to the power levels in Table 1. We have a comparison table of the two above cases in Table 2. Here, the first column shows the methods of typical load, the second column present the type of capacitors and their placements, while the last column shows the active power of two methods after compensation.

Based on the comparison of Table 2 of two cases, the number of installed capacitors of the maximum load case is more than that of the proposed method but the efficiency on reducing losses is negligible. Here, the different number of capacitors installed in the two methods is 2 capacitors, e.g. 9 capacitors for the maximum load case and 7 capacitors for the proposed method case. Meanwhile, the amount of energy losses before and after compensates 606.615 KWh for the maximum load case and 615.025 KWh for the proposed method case, thus the difference of power losses of two cases is only 9 KWh.

Table 2. Comparison of two cases

Method	Placement		Losses (KWh)
Maximum load	Fixed Capacitor	2354 2373 2281 2278 2310 22	606.615
	Switched Capacitor	2283 2293 2315	
Proposed moethod	Fixed Capacitor	2354 2373 2278 2270	615,025
	Switched Capacitor	2315 2293 2364	

- 2354: Before Nguyen_Cong_Phuong_2 Transformer
- 2364: After Nguyen_Cong_Phương_2 Transformer.
- 2281: Before Phan_Dinh_Phung_2 Transformer.
- 2310: Before Quang_Trung_3 Transformer.
- 2315: After Quang_Trung_3 Transformer.
- 22: A&B zone of Pham_Van_Dong University.
- 2373: Nguyen_Cong_Phuong.
- 2278: Before Le_Trung_Dinh Transformer.
- 2293: After Le_Trung_Dinh Transformer.
- 2283: After Phan_Dinh_Phung_1 Transformer.
- 2270: Before Phan_Dinh_Phung_1 Transformer.

4. CONCLUSION

This paper has introduced an advanced method of creating a typical load graph to optimize the distribution grid in term of power losses and the number of capacitors. The calculation tool is integrated with the proposed method when applying the model to the actual grid in Quang Ngai province has proved the correctness. Here, the simulation shows that the proposed method has less number of capacitors than that of the maximum load graph, but the deviation of power losses of both methods is small. Therefore, the result can help design engineers, investors get an overview of the optimal plan when planning as well as choosing the suitable plan with characteristics and the actual situation. Besides, the accuracy of the proposed method still has not reached the maximum reliability. However, based on this algorithm, authors can expand the plan to continue to build a program to optimize multi-objective economic compensation for power distribution network taking into account the installation of the static capacitor and dynamic compensation capacitor with load is uncertain and changes over time.

ACKNOWLEDGMENT

The authors wish to express their appreciation to the Ministry of Education and Training for supporting this research project as part of the Ministerial Project of Science and Technology CTB2018_DNA.04 "Establishing models and basic parameters of hybrid biogas-solar energy system for normal usage in rural Vietnam".

Gratefully thanks Quang Ngai electric power company provided data for testing the proposed method.

REFERENCES

[1] "Circular 428/qd-ttg vietnam government," http://vanban.chinhphu.vn/portal/page/portal/chinhphu/hethongvanban?Classid=2&mode=detail&document id=183986.

[2] Hieu, N.H., Lam, L.H., Luu, C.T. and Tuan, T.Q., 2015, June. Effects of DFIG wind power generation on Vietnam power system operation. In 2015 IEEE Eindhoven PowerTech (pp. 1-4). IEEE.

[3] Duong, M.Q., Dinh, T.V., Van Tan Nguyen, H.V.P., Nguyen, N.T.N.T. and Le, T.T.M., 2018. Effects of FSIG and DFIG Wind Power Plants on Ninh Thuan Power Grid Vietnam. GMSARNInternational Journal, 12(3), pp.133-138.

[4] Duong, M.Q., Tran, N.T.N., Sava, G.N., Leva, S. and Mussetta, M., 2018, October. The Impact of 150MWp PhoAn Solar Photovoltaic Project into Vietnamese QuangNgai-Grid. In 2018 International Conference and Exposition on Electrical And Power Engineering (EPE) (pp. 0498-0502). IEEE.

[5] Hieu, N.H. and Lam, L.H., 2016, November. Using double Fed induction generator to enhance voltage stability and solving economic issue. In 2016 IEEE International Conference on Sustainable Energy Technologies (ICSET) (pp. 374-378). IEEE.

[6] Vu Van Thang, Bach Quoc Khanh, Location and optimal capacity of capacitors in planning and improving distribution systems, Journal of Science and Technology - The University of Danang, No. 3(112).2017.

[7] Nguyen Van Minh, Bach Quoc Khanh, Optimal Sizing and Placement of Capacitors in the 16 Buses Test System Using GA for Loss Reduction, Journal of Science and Technology - Hanoi University of Industry, No 42, Pp. 20-24, 2017.

[8] Nguyen Huu Hieu, Optimizing reactive power compensation in a distribution system in case taking into account actual capacitors, Journal of Science and Technology - The University of Danang, No. 6(115).2017, pp.14-18.

[9] S.Saraswathi, Dr. Mary Immaculate Sheela, A comparative Study of Various Clustering Algorithms in Data Mining

[10] Machine learning, website: https://machinelearningcoban.com/

[11] PSS/ADEPT, website: https://pss-store.siemens.com/

[12] Matlab, website: https://www.mathworks.com/products/matlab.html