Electrical Load Demand of Water Treatment Plant Profiled by Using K-Means Clustering Algorithm with Reliability Consideration

Soraphon Kigsirisin
Department of Waterworks Academic Development, MWA Waterworks Institute of Thailand (MWIT), Metropolitan Waterworks Authority, Thailand
Email: soraphon.kigs@gmail.com, soraphon.k@mwa.co.th

Onurai Noohawm
Department of Electrical Engineering, Kasetsart University, Thailand
Email: nonurai@hotmail.com

***Abstract*—Monitoring Electrical Load Demand (ELD) is very crucial for plants and manufacturers to observe the electrical consumption and to set the rate of a primary source, the source limit, for supplying electricity to their own electrical system. This study presents Calinski-Harabasz criterion method to find out K value in order to divide ELD of Bangkhen water treatment plant (BKP), the largest water treatment plant in Thailand and South-East Asia, as a case study into K groups. ELD of BKP recorded in twelve months is shown and analyzed by this method. Three seasons are done, too. Then, K-means clustering algorithm is applied to find the centroid value, the cluster center, of K group. The maximum cluster center is used to be the source limit to supply electricity to the plant. Moreover, this study evaluates reliability indices to find out the security of the plant from using the source limit. They are shown and discussed, when the peak load demand occurs.**

***Index Terms*—k-means, electrical load demand, reliability, water treatment plant**

I. INTRODUCTION

Managing Electrical Load Demand (ELD) is being campaigned in worldwide to observe electrical consumption and to make a plan for getting the efficient use of ELD. In general, ELD is recorded by SCADA system or data logger by current transformer and potential transformer for electrical suppliers and plants to monitor real-time and historical ELD [1]. It is often shown a measuring unit, volt, ampere, watt, var and else, on a station screen. ELD is mainly focused to reduce electricity cost by using many theories such as six concepts of Demand Side Management (DSM), etc. [2]. Furthermore, plants need the result from those theories for finding out the rate of a main source to supply electricity to their own electrical system as an islanding area without concerning any other electricity supplier outside [3]. A source is such as PV, hydropower, diesel generator. Islanding area has now been very popular for plants to generate, distribute and manage electricity by themselves and they get easy for ELD management. Nevertheless, since the cost of a source is too expensive, the most of the plants are trying to find the best approach to find out the optimal rate of a source, the source limit, by considering their actual ELD [4]. If the source limit supported generates electricity not to be enough to meet ELD, they must face electrical interruption and disruption of the treatment process. In other words, plants have no high reliability and people have no water to use in the period of the interruption.

Bangkhen water treatment plant (BKP), the largest water treatment plant of Metropolitan Waterworks Authority (MWA), Thailand, also in Thailand and South-East Asia. ELD is about 28 MW with maximum capacity of water treatment process around 4.4 million cubic meters per day [5]. BKP distributes quality water to 12 million people who live in Bangkok (The capital city of Thailand), Nonthaburi and Samutprakarn provinces. To supply electricity like an islanding area and gain the high electrical system reliability, MWA is finding and determining the source limit for BKP to supply quality water to those people to have quality water enough to use without any shortage from electrical interruption. Therefore, this study presents an approach to succeed in these issues by using Calinski-Harabasz Criterion (CH) and K-means clustering algorithm (KM) methods. They are simulated and applied to the actual recorded ELD of BKP. Then, those methods are coded in MATLAB, and the source limit is set. Moreover, this study evaluates the security of BKP from using the source limit through reliability indices, Expected Energy Not Supplied (EENS), Availability (A), Unavailability (UN) and Load Loss Factor (LLF). As a result, this study provides the approach and flowchart for BKP and other industries to follow, and discusses the reliability indices by using this approach.

Manuscript received January 10, 2018; revised June 20, 2018.

II. Method for Analyzing Data

A. K-Means Clustering Technique

Calinski-Harabasz Criterion (CH) or the variance ratio criterion (VRC) has been widely used for grouping objects into class, type or category introduced by T. Calinski and J. Harabasz in 1974 [6]. The core of CH is to solve the optimal number of cluster (k) (a positive integer number). CH evaluates the cluster validity based on the average between and within cluster sum of squares. Well-defined clusters have a large between cluster variance (SS_B) and a small within-cluster variance (SS_W). The larger the CH_K ratio is, the better the data partition is. To determine the optimal number of clusters, let's maximize CH_K with respect to k via equation (1), which should be in range between $2 \leq k \leq n$ [7].

$$CH_K = \frac{SS_B}{SS_W} \times \frac{(n-k)}{(k-1)} \quad (1)$$

where, k is the number of clusters, n is the number of ELD.

B. K-Means Clustering Technique

K-means clustering technique (KM) is applied to electrical load demand in this study. KM is about to group load demand equal to the number of k resulted by CH technique. KM considers the sum of squares of distance between load demand and the corresponding cluster centroid value for minimization [7]. In other words, the cluster centroid value is called the cluster center and the number of it is equal to k. When a cluster center is random and then set, the size of ELD nearby the cluster center will be chosen to be its member group.

In this study, ELD is split to a set of the n vectors $X_j, j = 1,2,3,\dots,n$ are partitioned into k groups $G_i, i = 1,2,3,\dots,k$ such that the cost function D is minimized. The cost function D is the sum of the Euclidian distance between each vector X_j and its corresponding cluster center C_i. It can be mathematically defined as equation (2) [8].

$$D = \sum_{i=1}^{k} \left(\sum_{X_j \in G_i} \left| X_j - C_i \right|^2 \right) \quad (2)$$

III. Reliability Evaluation for the Source Limit

A. Expected Energy Not Supplied

Expected Energy Not Supplied ($EENS$) indicates the expected number of megawatt hours that the plant must lack of electricity due to electrical inadequacy. This study lets maximum cluster center from k group and maximum ELD average to be the source limit. They are seen as a concept of this study. When peak of all considered ELD occurs which exceeds more than the source limit, the plant is not enough electricity to be supplied to all process. Remaining process must face a period of interruption time [9]. $EENS$ is defined by equation (3).

$$EENS = \sum \left(ELD_{peak} - S_{limit} \right) \times T_{down} \quad (3)$$

where, ELD_{peak} is the peak of considered ELD (MW), S_{limit} is the source limit from k group, or average (Avg) (MW), T_{down} is the period of interruption time (hr).

B. Availability and Unavailability

Availability (A) expresses the percentage of observed time a voltage source is uninterrupted. By the same token, the electrical system of the plant is energized and has electricity enough to supply ELD and all process without any interruption [10]. A is defined by equation (4).

$$A = \begin{cases} \dfrac{\sum T_{up}}{\sum T_{up} + \sum T_{down}} & ; \text{when, } ELD_{peak} < S_{limit} \\ 0 & ; \text{when, } ELD_{peak} \geq S_{limit} \end{cases} \quad (4)$$

where, T_{up} is the period of time that the plant and all process have electricity (hr).

In contrast, the percentage of observed time the plant and a process are interrupted or have no electricity enough to be supplied, so-called Unavailability (UN). UN is defined by equation (5).

$$UN = 1 - A \quad (5)$$

C. Load-Loss Factor

Load-Loss Factor (LLF) is the percentage factor that indicates how much the plant lacks of electricity or face interruption in a process due to electricity inadequacy. In this study, electricity lost when peak ELD incurs and exceeds more than the source limit. Normally, it is expected not to be more than 30 percent because the more LLF of the system has, the worse the system is [11]. It is based on the square of the load. LLF is defined by equation (6).

$$LLF = \begin{cases} \dfrac{\sum \left(ELD_{peak} - S_{limit} \right)^2}{ELD_{peak}^2} & ; \text{when, } ELD_{peak} > S_{limit} \\ 0 & ; \text{when, } ELD_{peak} \leq S_{limit} \end{cases} \quad (6)$$

IV. Data and the Result Discussion

Electricity of BKP is supported by 69 kV from an electrical supplier to supply electricity to the main incoming switchgear and transformer. Current and voltage transformers are installed at up-rising of incoming switchgear. Current and voltage values are measured and automatically recorded in every 30 minutes by SCADA system. They are multiplied together for megawatt unit used in this study. Because of the pattern of ELD of each month is very different from another, it results from water consumption of its customers which BKP has to run the water treatment process to respond to their need.

In this study, actual ELD of each month between March of 2014 and Feb of 2015 is evaluated. This study does not regard any activity affecting load demand, mainly such as preventive maintenance. Moreover, ELD is separated into three seasons in Thailand to observe the result as follows:

1) Summer: March, April and May

2) Rainy: June, July, August, September and October
3) Winter: November, December, January and February

To illustrate this study, the procedure of clustering ELD by using CH and KM methods, and reliability indices evaluation is shown in a flow chart in Fig. 1.

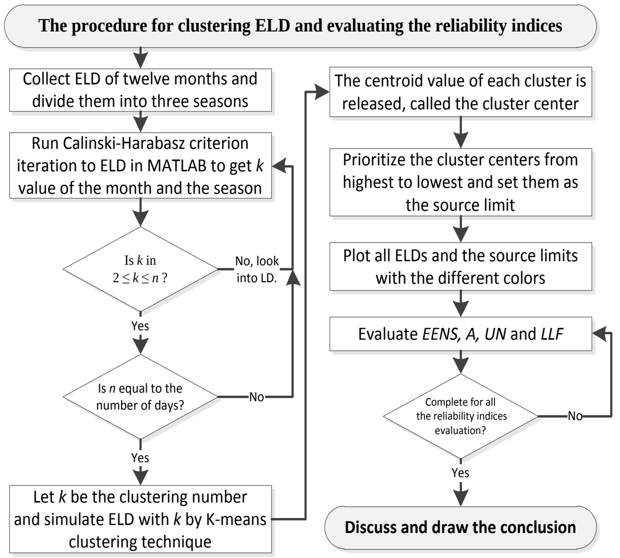


Figure 1. The flow chart of the procedure for clustering ELD and evaluating the reliability indices

After ELD of all days in each month and season is coded and computerized by CH and KM methods in MATLAB. k value and the source limit of each k group are released and shown in Table I. This study focusses on both the source limit from the cluster center and the source limit of maximum ELD average (Avg), to look into the result then. From Fig. 2, 30-min-recorded time from 0.00 a.m. to 23.30 p.m. is depicted on horizontal axis from 1 to 48, respectively. ELD of all considered days is shown on vertical axis. The definition of line color is described as follows,

- Grey line: load demand
- Red line: maximum load demand
- Yellow line: average load demand (nonlinear), the source limit of load demand average (straight line)
- Cyan line: the lowest load demand
- Purple line: the cluster center of k group (nonlinear), the source limit of k group (straight line)

The reliability indices in each month and season are evaluated and shown in Table II. As the result, using the source limit results in interruption differently to each month and season when the peak load demand occurs, the plant must encounter inadequate electricity. The result can be concluded into four cases as follows:

- Inadequate electricity all time long: it occurs in only C3 of April, C3 and Avg of July, C2 of Summer, and C2 of Rainy. A is equal to 0%, UN is equal to 100% and LLF is close and much more than 1000%.
- Inadequate electricity in period of time almost similar to each other: it occurs in C1, C2, C3 and Avg of September.
- Inadequate electricity in range between cluster centers: it occurs in all of months and seasons.
- Inadequate electricity between months and seasons: all the reliability indices values of all the seasons are more than the value of its month's season, excepting all Avg and C2 of winter.

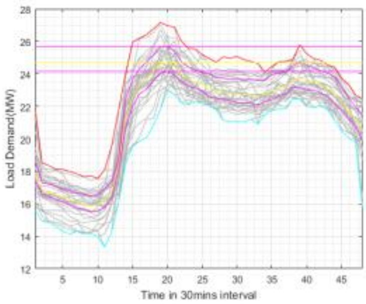
TABLE I. THE NUMBER OF DAYS, K VALUE, MAXIMUM ELD AND THE SOURCE LIMITE OF MONTH AND SEASON

Month	Day	k	ELD_{peak}	S_{limit}			
				C1	C2	C3	Avg
Mar	31	2	27.17	25.69	24.17	N/A	24.70
Apr	30	3	27.28	24.94	20.91	11.59	22.18
May	31	2	26.86	25.29	23.81	N/A	24.67
Jun	30	2	26.56	25.28	24.44	N/A	24.92
Jul	31	3	26.60	25.54	25.01	9.93	12.92
Aug	31	2	10.84	9.96	9.70	N/A	9.84
Sep	30	3	11.78	9.99	9.93	9.89	9.93
Oct	31	2	26.56	24.88	22.32	N/A	24.05
Nov	30	2	26.40	25.30	24.04	N/A	24.50
Dec	31	2	27.30	24.68	21.27	N/A	24.34
Summer	92	2	27.17	24.56	13.58	N/A	23.84
Rainy	153	2	26.60	24.55	9.90	N/A	16.27
Winter	120	2	27.30	24.73	22.25	N/A	23.84
Jan	31	2	23.76	22.62	21.29	N/A	22.14
Feb	28	2	26.31	25.10	23.74	N/A	24.47
Note: C1: Cluster center 1, C2: Cluster center 2, C3: Cluster center 3							

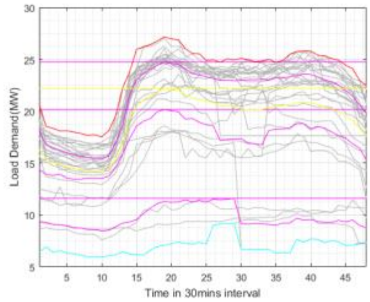
TABLE II. THE RESULT OF RELIABILITY INDICES

Month	$EENS$ (MWh)				A			
	C1	C2	C3	Avg	C1	C2	C3	Avg
Mar	3.46	17.79	N/A	10.40	0.81	0.40	N/A	0.48
Apr	11.96	90.31	284.84	53.71	0.42	0.23	0.00	0.25
May	8.99	31.42	N/A	17.40	0.56	0.31	N/A	0.33
Jun	4.23	12.62	N/A	7.07	0.75	0.44	N/A	0.63

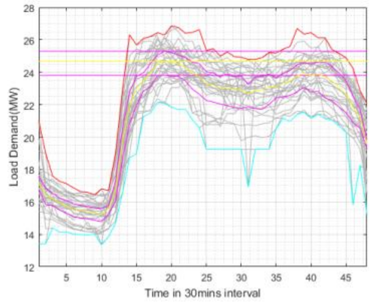
Jul	2.31	6.73	318.10	246.50	0.79	0.50	0.00	0.00
Aug	4.20	6.71	N/A	5.33	0.63	0.58	N/A	0.60
Sep	1.99	3.52	4.45	3.91	0.81	0.69	0.63	0.67
Oct	6.03	43.24	N/A	14.81	0.67	0.25	N/A	0.46
Nov	2.82	15.76	N/A	9.40	0.81	0.38	N/A	0.54
Dec	7.67	63.15	N/A	11.34	0.58	0.25	N/A	0.48
Jan	3.17	19.90	N/A	7.87	0.71	0.38	N/A	0.44
Feb	4.11	20.89	N/A	10.09	0.69	0.33	N/A	0.52
Summer	20.34	244.64	N/A	32.00	0.33	0.00	N/A	0.31
Rainy	15.05	326.95	N/A	174.01	0.42	0.00	N/A	N/A
Winter	8.51	48.17	N/A	20.81	0.58	0.25	N/A	0.33
Month	***UN***				***LLF* (%)**			
Mar	0.19	0.60	N/A	0.52	1.01	9.02	N/A	4.61
Apr	0.58	0.77	1.00	0.75	4.92	144.86	1207.9	54.93
May	0.44	0.69	N/A	0.67	2.85	19.91	N/A	7.50
Jun	0.25	0.56	N/A	0.38	1.08	5.02	N/A	2.27
Jul	0.21	0.50	1.00	1.00	0.42	1.75	1539.1	928.87
Aug	0.38	0.42	N/A	0.40	5.12	10.17	N/A	7.15
Sep	0.19	0.31	0.38	0.33	2.62	4.82	6.36	5.46
Oct	0.33	0.75	N/A	0.54	1.87	37.96	N/A	6.92
Nov	0.19	0.63	N/A	0.46	0.64	6.92	N/A	3.41
Dec	0.42	0.75	N/A	0.52	2.92	75.08	N/A	4.86
Jan	0.29	0.63	N/A	0.56	0.79	11.66	N/A	2.68
Feb	0.31	0.67	N/A	0.48	1.04	10.50	N/A	3.66
Summer	0.67	1.00	N/A	0.69	9.76	872.76	N/A	20.73
Rainy	0.58	1.00	N/A	1.00	5.76	1582.65	N/A	471.91
Winter	0.42	0.75	N/A	0.67	3.11	44.64	N/A	10.71



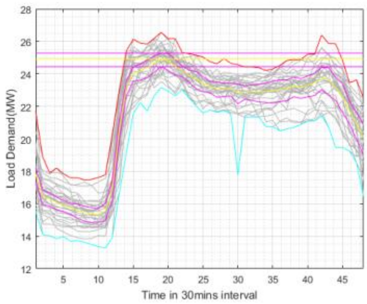
March of 2014



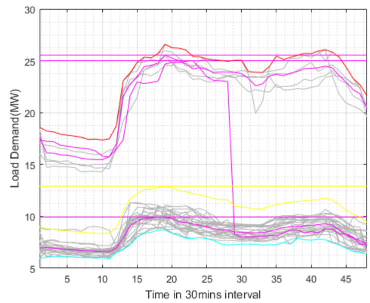
April of 2014



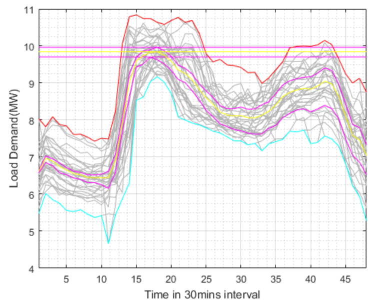
May of 2014



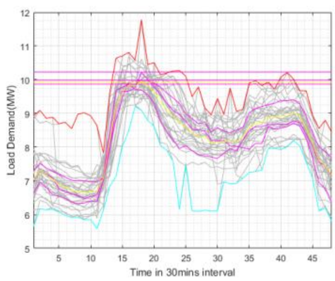
June of 2014



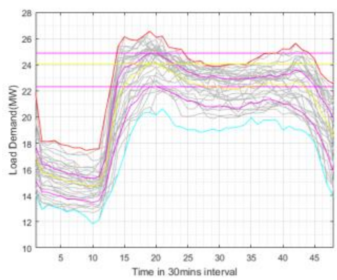
July of 2014



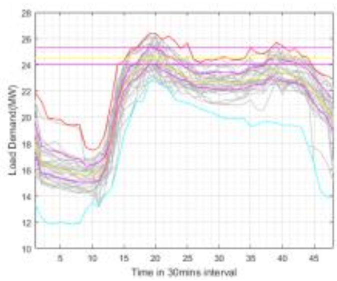
August of 2014



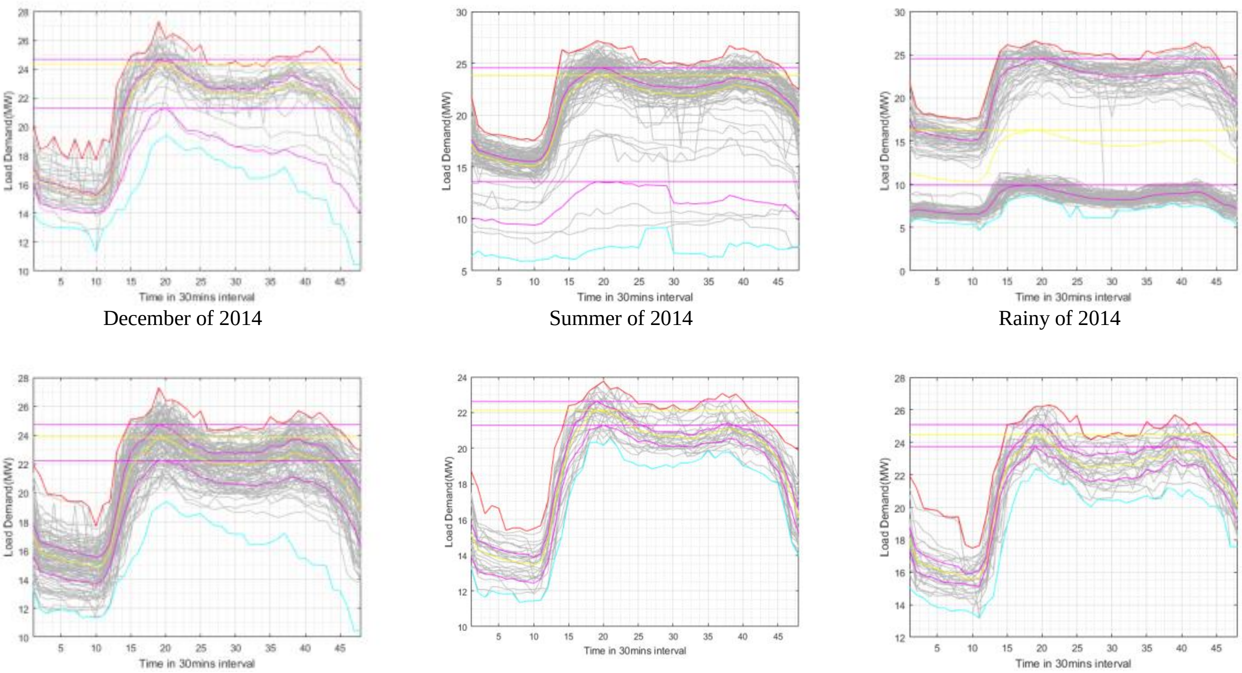
September of 2014



October of 2014



November of 2014



December of 2014 | Summer of 2014 | Rainy of 2014

Winter of between 2014 and 2015 | January of 2015 | February of 2015

Figure 2. ELD, ELD_{peak} and S_{limit}

V. Discussion and Conclusion

This study uses Calinski-Harabasz Criterion to solve K value for K-means clustering technique. The actual load demand of Bangkhen water treatment plant from Mar. 2014 to Feb. 2015 is proposed and clustered by this technique to find out the centroid value of each K group, called the cluster center. This study also separates those months into three seasons in Thailand for load demand and result observations. After using those methods to load demands of a month and a season, they are divided into two groups or three groups, K value. The maximum cluster center is set to be the source limit for electrical supply. The source limit of load demand average is established alike. By using the source limit, EENS, A, UN and LLF are evaluated to observe the security of the plant as reliability indices. The result shows that if EENS is less, UN and LLF will be less. On the other hand, A will be much. And, the reliability indices of the source limit of load demand average are in the range between the highest source limit and the lowest source limit. However, in some case, the plant has no electricity enough all time long when the peak load demand occurs because the source limit is less than all load demand. Moreover, the reliability indices of some season are more than its month's season. For this reason, the plant and others should look at a source limit, as a main source with evaluating these reliability indices, before ordering a main source to supply electricity to their electrical system as Islanding area. Otherwise, they must face all interruption and disrupt the treatment process when the peak of load demand occurs.

Acknowledgement

The authors wish to express kind thanks to Metropolitan Waterworks Authority, Thailand, for all support to do this research successfully.

References

[1] N. Baranovic, P. Andersson, I. Ivankovic, K. Zubrinic-Kostovic, D. Peharda, and J. E. Larsson, "Experiences from intelligent alarm processing and decision support tools in smart grid transmission control centers," *Cigre Session*, D2-112, 2016

[2] UNIDO, Demand-Side Management. Sustainable Energy Regulation and Policymaking for Africa, 2015.

[3] N. K. Paliwal and R. K. Rai, "Micro-grid operation during grid connected and islanding mode using conventional and inverter interfaced source," in *Proc. International Conference on Smart Electric Grid*, 2014.

[4] D. Colley, N. Mahnoudi, D. Eghbal, and T. K. Saha, "Queensland load profiling by using clustering techniques," in *Proc. AUPEC*, 2014, pp. 1-6.

[5] Department of Water Transmission and Distribution, Yearly Energy Consumption Report, Metropolitan Waterworks Authority, Thailand, 2016.

[6] S. Lukasik, P. A. Kowalski, M. Charytanowicz, and P. Kulczycki, "Clustering using flower pollination algorithm and Calinski-Harabasz index," in *Proc. IEEE Congress on Evolutionary Computation*, 2016.

[7] B. Mamum and I. Reaz, "Assessment of electrical load in water distribution systems using representative load profiles-based method," *Advances in Electrical Engineering*, vol. 2014, articale ID 865621, pp. 1-10, 2014.

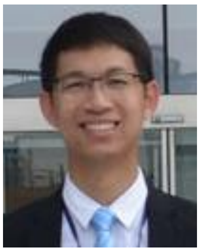
[8] R. Damayanti, A. G. Abdullah, W. Purnama, and A. B. D. Nandiyanto, "Electrical load profile analysis using clustering techniques," in *Proc. IOP Conference Series: Materials Science and Engineering*, 2017.

[9] J. Hokierti, *Reliability and Maintenance for Distribution System*, Charun, Bangkok, 2011.

[10] H. Sabouhi, A. Abbaspour, M. F. Firuzabad, and P. Dehghanian, "Reliability modeling and availability analysis of combined cycle

power plants," *International Journal of Electrical Power & Energy Systems*, vol. 79, pp. 108-119, July 2016.

[11] Top Energy, TePuna Hihiko, "Network loss factor methodology," Jan. 2015.



Soraphon Kigsirisin was born in Bangkok, Thailand, in 1988. He received his M.Eng and B.Eng degrees in Electrical Engineering with Outstanding Acedemic Performance Award from Kasetsart University, Bangkok, Thailand in 2011 and 2013 respectively.

He worked as a researcher at the Department of Electrical Engineering, Kasetsart University. He is now working as an Engineer at MWA Waterworks Institute of Thailand, Metropolitan Waterworks Authority (MWA), Thailand. His research interests are electrical power system, electrical system reliability, power quality, sustainable energy and water productivity.

Onurai Noohawm was born in Nakornpatom, Thailand, in 1975. She received her M.Eng and B.Eng degrees in Electrical Engineering from King Mongkut's University of Technology Thonburi, Thailand, in 1996 and 2002, respectively. Currently, she is working as a researcher at the Department of Electrical Engineering, Kasetsart University, Thailand.