Journal of Computations & Modelling, vol.4, no.1, 2014, 207-222
ISSN: 1792-7625 (print), 1792-8850 (online)
Scienpress Ltd, 2014

Load Estimation for War-Ships Based on Pattern Recognition Methods

George Tsamopoulos[1], Nikolas Giannitsas[2],

Fotis D. Kanellos[3], George J. Tsekouras[4] and Irene S. Karanasiou[5]

Abstract.

Load estimation is crucial information for every type of war-ships because it is the necessary base for a series of studies and operations, such as economic operation of electric generators, load shedding and power management systems. In this paper a pattern recognition methodology is presented, which consists of four steps: (a) data selection, (b) data pre-processing, in order to modify or exclude obviously incorrect values, (c) the application of pattern recognition algorithm, (d) load estimation. The clustering algorithm used in this paper is a modified k-means algorithm, which is properly calibrated by using the ratio of within cluster sum of squares to between cluster variation (WCBCR). This methodology is indicatively applied to a Hellenic Navy MEKO type frigate for the estimation of its total load demand while the usefulness of the obtained results for the power system design and operation is proved as well.

Keywords: Electric load profiles, pattern recognition, warship power system operation

---

[1] Lieutenant CDR, Hellenic Navy, Hellenic Naval Academy, Terma Hatzikiriaku, Piraeus, GR-18539, Greece
E-mail: tsamgeorge@ath.forthnet.gr

[2] Commander, Hellenic Navy, Hellenic Naval Academy, Terma Hatzikiriaku, Piraeus, GR-18539 Greece
E-mail: ngiannit@hotmail.com

[3] Production Engineering & Management Department, Technical University of Crete, Chania, GR-73100, Greece
E-mail: fkanellos@dpem.tuc.gr

[4] Hellenic Naval Academy, Terma Hatzikiriaku, Piraeus, GR-18539 Greece
E-mail: tsekouras@snd.edu.gr

[5] Department of Mathematics and Engineering Sciences, Hellenic Military Academy, GR-166 73 Vari Attikis, Greece
E-mail: ikaran@esd.ece.ntua.gr

1 Introduction

Warship safe seaway depends significantly on ship power system, with special note on aircraft carriers, new constructed frigates and submarines, comprising integrated fully electrified power system based on nuclear reactors, diesel machines, fuel cells and batteries. During last decade All Electric Ship (AES) offers a series of advantages, such as increased safety, survivability, manoeuvrability, precise and smooth speed control, low noise and lower “acoustic signature” than classical vessels. Combining different kinds of electric power plants one can minimize machinery space, pollutant emission levels, operational and maintenance cost after having resolved any technical problems emerged. Additionally, with energy storage systems (ESS) it is achieved to support voltage and frequency, manage peak loads, improve power quality and provide uninterruptible power supply for sensitive onboard loads. At the same time, systematic efforts have been done to connect a ship to a shore-side power supply in port with ship machinery shut-down provided that the port has the infrastructure and the green energy to serve ship load demand. This concept is known as *cold ironing*. Cold ironing is most effective when applied to ships remaining at the same port for large time periods, such as warships. In Greece, during the last 30 years the idea of cold ironing is partially applied on submarines and frigates, at Salamina Hellenic Naval Station.

Conclusively, warship power system design and operation under different conditions are complex processes, which require the total load demand as input. The usual solution of this problem is the estimation of total load demand for each operating condition based on multiplicative factors for each electric consumer, such as power efficiency, load factor, in order to calculate the average daily required load [2]-[3].

In this paper an alternative methodology for total load demand estimation in different operating conditions is presented based on pattern recognition methods. The main core of the methodology has already been implemented for warships in [4-5], but here emphasis is given to the selection of the representative chronological load curves which could last for different time periods (one hour, 2 hours, … 24 hours) based on classical forecasting criterions, such as mean absolute percentage error (MAPE). This methodology is applied on the electric power generation system of a MEKO-type frigate.

2 Load Estimation Based on Pattern Recognition Method

2.1 Previous Research in Pattern Recognition Methods

During the last years, a significant research effort has been devoted to load curve classification, in order to solve the short-term load forecasting of anomalous days [6], to cluster load profiles for ship electric consumptions [7], to cluster power systems customers [8-9] and to estimate the total load profile of power systems for the implementation of demand side management programs [10-11].

According to [12] the clustering methods based on pattern recognition concept are:

- the k-means (with random vectors choice or with specialized weight initialization),
- the fuzzy k-means,
- the adaptive vector quantization,
- the hierarchical methods (single link, complete link, unweighted pair group method average, weighted pair group method average, unweighted pair group method centroid, weighted pair group method centroid, Ward),
- the “modified follow the leader”,
- the self-organizing map (mono-dimensional and bi-dimensional).

Alternatively, classification problem can be solved by using data mining, wavelet packet transformation, frequency-domain data, stratified sampling etc.

The most commonly used respective adequacy measures are [12]:

- the mean index adequacy,
- the clustering dispersion indicator,
- the similarity matrix indicator,
- the Davies-Bouldin indicator,
- the modified Dunn index,
- the scatter index,
- he mean square error,

- the ratio of within cluster sum of squares to between cluster variation.

Especially, a pattern recognition methodology for the classification of ship power system load demand curves has already been presented by the authors in [4-5] giving emphasis on the determination of the necessary number of clusters based on the "knee" rule. In [4] a typical time period of the chronological load curves used, was 24 hours without the examination of other possible time periods. The last one has been studied in [5] without giving emphasis in forecasting results.

2.2 Proposed Methods

The proposed methodology can be used for warship load estimation based on pattern recognition techniques. Its flowchart is shown in Fig. 1. The main steps are the following:

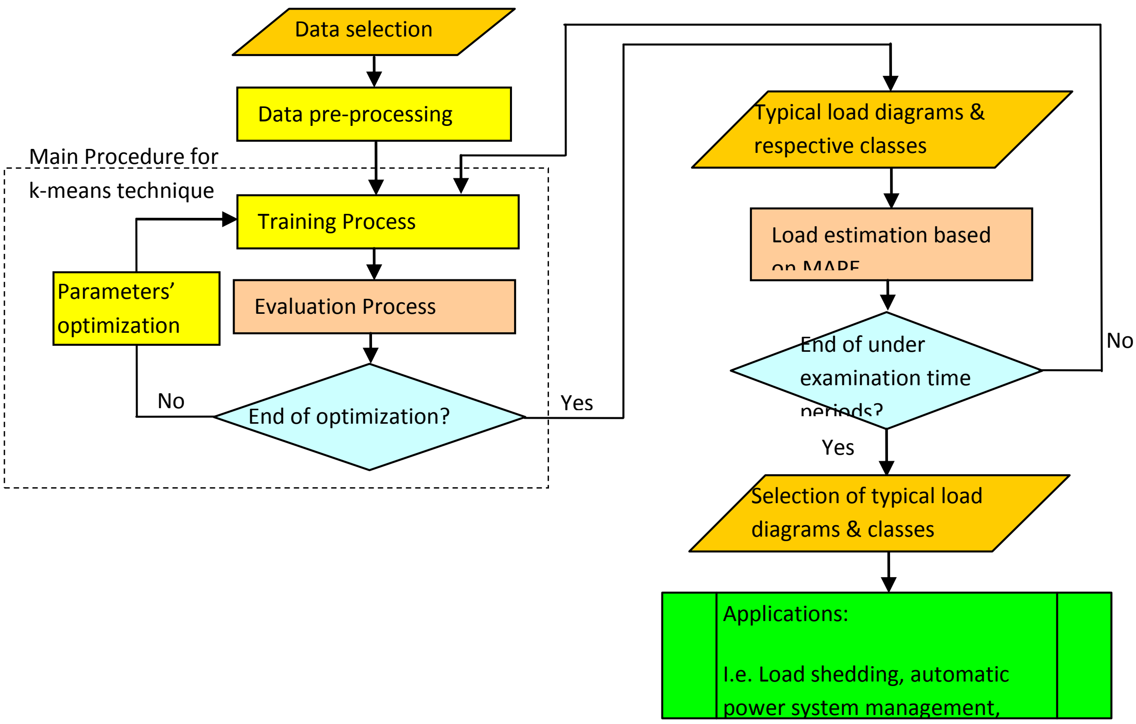

Figure 1. Flowchart of warship load estimation method based on pattern recognition classification of ship power system chronological total load demand curves.

i. *Data and features selection*: The chronological load curve of the ship power system defines the 15-minute average power demand over a specific time horizon,

T. The examined time period T is divided into M_D intervals, DT_j (=15 min) with j=1, 2,…, M_D. In each time interval DT_j ship load demand P_{Dj} is considered constant and it is calculated by:

$$\int_{t_{j-1}}^{t_j} p_D(t)\cdot dt = P_{Dj}\cdot DT_j \text{ [kWh]} \tag{1}$$

The respective transformation is presented in Fig. 2. The same process can be repeated for reactive power.

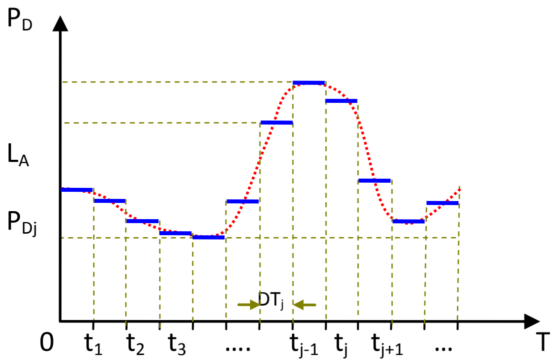

Figure 2. Chronological load demand curve during period T and its transformation in a chronological load curve of constant demand levels.

ii. *Data pre-processing:* The normality of the load diagrams is examined, in order to modify or delete the values that are obviously wrong. This process is known as *noise suppression*.

iii. *Main procedure*: The k-means clustering method is applied, which can be trained for the set of load diagrams, optimized and evaluated according to the ratio of within cluster sum of squares to between cluster variation (*WCBCR*), as it is expected to be the most descriptive adequacy measure [9]. More specifically, N defines the population of the input vectors, which are going to be clustered. Here, the input vector is the chronological total load demand curve of a war-ship power system. $\vec{x}_\ell = (x_{\ell 1},...x_{\ell i},...x_{\ell d})^T$ is the ℓ-th input vector and d its dimension. The corresponding set is given by $X = \{\vec{x}_\ell : \ell = 1,...,N\}$. Each classification process creates a partition of the initial N input vectors to M clusters. The j-th cluster has a representative load profile vector denoted with $\vec{w}_j = (w_{j1},...,w_{ji},...w_{jd})^T$. The vector $\vec{w}_j$ expresses the cluster centre. The corresponding set of clusters centres is the classes set defined by $W = \{\vec{w}_k, k = 1,...M\}$. The subset of input vectors $\vec{x}_\ell$, which belong to the j-th cluster, is denoted by Ω_j while the respective population of load diagrams by N_j . The Euclidean distance between input vectors ℓ_1, ℓ_2 of the set *X* is used for the study and evaluation of classification algorithms:

$$d\left(\vec{x}_{\ell_1},\vec{x}_{\ell_2}\right)=\sqrt{\frac{1}{d}\sum_{i=1}^{d}\left(x_{\ell_1 i}-x_{\ell_2 i}\right)^2} \tag{2}$$

The modified *k*-means clustering method groups the set of the *N* input vectors into *M* clusters using an iterative procedure. Initially the weights of the *M* clusters are determined as [9]:

$$w_{ji}^{(0)}=a+b\cdot(j-1)/(M-1) \tag{3}$$

where a and b are properly calibrated parameters. During epoch *t* Euclidean distances $d\left(\vec{x}_\ell,\vec{w}_j\right)$ for each training vector $\vec{x}_\ell$ are calculated for all centres. The ℓ-th input vector is assigned to the set $\Omega_j^{(t)}$, if the distance between $\vec{x}_\ell$ and the respective centre is the minimum obtained. When the entire training set is formed, the new weights of each centre are calculated as:

$$\vec{w}_j^{(t+1)}=\frac{1}{N_j^{(t)}}\sum_{\vec{x}_\ell\in\Omega_j^{(t)}}\vec{x}_\ell \tag{4}$$

where $N_j^{(t)}$ is the population of the respective set $\Omega_j^{(t)}$ during epoch *t*. This process is repeated until the maximum number of iterations is reached or the variation of the weights is not significant. The algorithm's final target is to minimize the error function, which expresses the unweighted mean value of the squared distance of each vector from its cluster centre.

In order to evaluate the performance of the clustering algorithm and to optimize its parameters, the ratio of within cluster sum of squares to between cluster variation *(WCBCR)* is used. It depends on the sum of the squared distances between each input vector and its cluster representative vector as well as the similarity of the clusters centres.

$$WCBCR=\sum_{k=1}^{M}\sum_{\vec{x}_\ell\in\Omega_k}d^2\left(\vec{w}_k,\vec{x}_\ell\right)\Bigg/\sum_{1\le q<p}^{M}d^2\left(\vec{w}_p,\vec{w}_q\right) \tag{5}$$

The proposed algorithm is repeated for different pairs of (*a*,*b*) and the pairs with the best *WCBCR* are recorded, leading to small values of *WCBCR* for the same final number of clusters. By increasing the number of clusters the respective measure decreases. It should be noted that in eq. (5) M is the number of the clusters without considering the blank ones. Instead of the modified k-means technique other pattern recognition techniques could be applied; however it has been used as it is an extremely rapid and robust method.

iv. *Load estimation*: In order to use the chronological typical load curves as a forecasting tool, the mean absolute percentage error (*MAPE)* index between the measured and the estimated values of the load demand for each time step for all available vectors is used:

$$MAPE = 100\% \cdot \frac{1}{N \cdot d} \cdot \sum_{j=1}^{N} \sum_{i=1}^{d} \frac{\left| w(j,i) - x(j,i) \right|}{x(j,i)} \tag{6}$$

where $x(j,i)$ is the measured value of load demand for the i-th time step of j-th time period for the evaluation set and $w(j,i)$ the respective estimated value. This index is a practical measure and provides a way to approximate the actual value from the estimated one independently from the units and the quantity of the load demand.

This process is repeated for different time sub-periods varying from one hour to one day. Specifically, one hour, 2 hours, 3 hours, 4 hours, 6 hours, 8 hours, 12 hours and 24 hours can be examined, as these are practically sensible.

3 Application of the Proposed Method

3.1 Case Study

The developed methodology is applied to the total load demand of a Hellenic Navy MEKO type frigate comprising 4 generators of nominal active power of 750 kW and 0.8 inductive power factor. The total active energy produced by the ship electric generators was registered by the control and monitoring system “NAUTOS” every 15 minutes for a period of 20 days (8-31 January 1998) with the frigate at berth. Total electric power produced the ship generators equals the total load demand provided that no electric power is provided by the shore. The maximum load demand at anchor for the examined period was approximately 1200 kW and the average load demand approximately 500 kW. The equivalent load curve is given in Fig. 3.

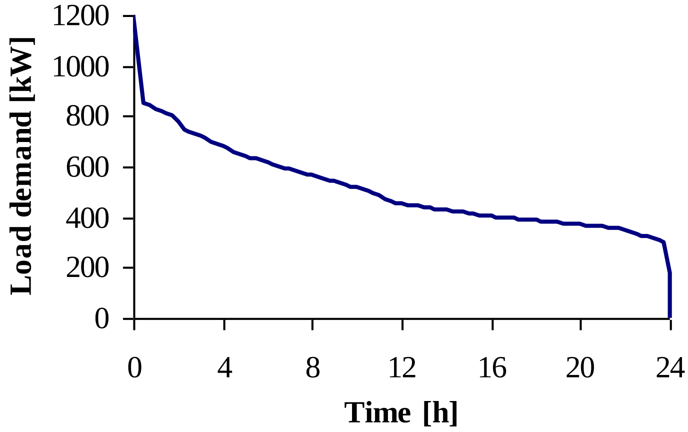

Figure 3. Total load demand duration curve for a period of 20 days (8-31 January 1998) with the ship at "anchor" condition.

3.2 Analytical Application of Modified K-Means for 6 Hours Period

The dimension of the 6-hour data vector is 24, as the respective time step is 15 minutes. The number of the formed vectors are 80 (=20 days * 4 vectors/ day). Following, the modified *k*-means method is executed for different pairs of parameters *a* and *b* defining the initialization weight centres and the number of the clusters ranging from 2 to 20, where $a=\{0.00, 0.01, \ldots, 0.45\}$ and $a+b=\{0.55, 0.56, \ldots, 1.00\}$. For each cluster 2116 different pairs (a,b) are examined and the pair with the smallest *WCBCR* is selected. In Fig. 4 the *WCBCR* measure for 2 up to 20 clusters is presented for the set of 80 training patterns. The number of the clusters, which provides satisfactory results, corresponds to the knee of the respective curve [8-9]. In this case, the respective number of clusters seems to be 3. However, *WCBCR* values for the clusters number 2 and 3 are large. If these values are not taken into account, Fig. 5 is obtained. In this case, the respective number of clusters is slightly greater than 7, so the proposed number of clusters is 8. Though it becomes apparent the respective solution is not unique while user's knowledge and experience is necessary.

The results for 8 clusters obtained with the modified k-means model and optimized *WCBCR* are presented in Table 1, where the calendar of the time period under study together with the cluster type and time sub-periods are included.

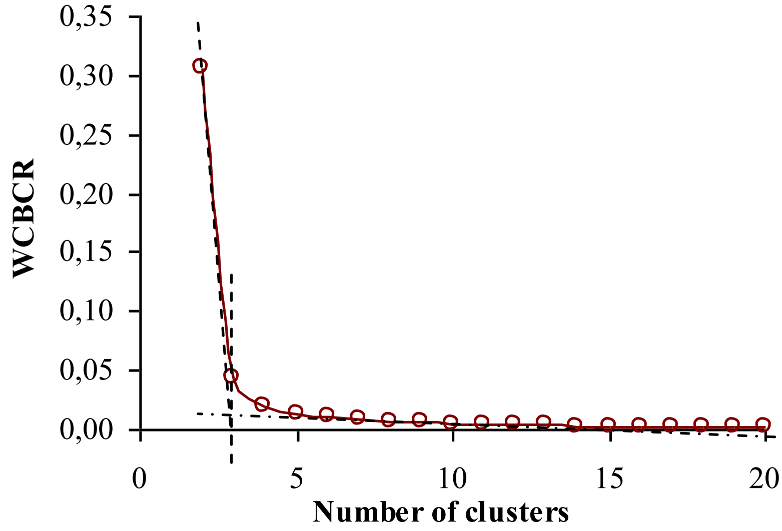

Figure 4. *WCBCR* measure of the k-mean for 2 to 20 clusters for the set of 80 training patterns of the total load demand for 6-hour time period & the use of tangents for the knee's estimation.

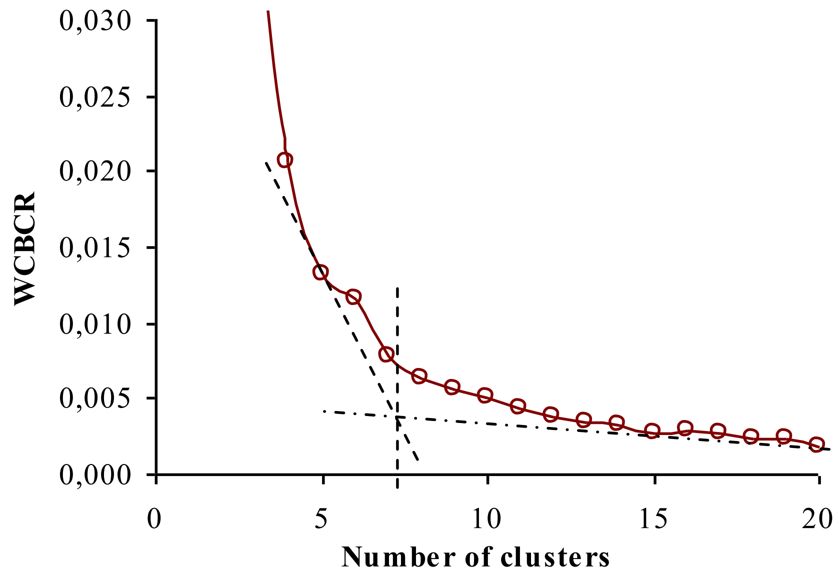

Figure 5. *WCBCR* measure of the k-mean for 4 to 20 clusters for the set of 80 training patterns of the total load demand for 6-hour time period & the use of tangents for the knee's estimation (zoom of Fig. 4).

In Table 2 the respective populations for 6-hour time sub-periods per cluster are registered. It is obvious that the most populated cluster is the 1st one for time sub-periods 0:15-6:00, 6:15-12:00, 12:15-18:00, while the 1st cluster and the 2nd one are equivalent for the last time sub-period 18:15-24:00. Finally, the most populated cluster for the whole study time period examined in this study, is the 1st one.

TABLE 1 Clusters Calendar with time sub-periods of the total load demand curves for HN MEKO type Frigate Power System.

Date January 1998	Time sub-period				Date January 1998	Time sub-period			
	0:15-6:00	6:15-12:00	12:15-18:00	18:15-24:00		0:15-6:00	6:15-12:00	12:15-18:00	18:15-24:00
8	1	1	1	1	22	2	2	2	2
9	1	1	2	2	23	2	2	2	2
11	1	1	1	1	24	4	2	1	1
12	1	1	2	4	25	4	2	1	1
14	6	3	2	6	26	1	1	6	3
16	1	1	1	1	27	1	1	1	1
17	1	1	1	3	28	1	1	1	2
18	1	1	2	4	29	8	7	1	1
20	4	1	1	5	30	1	1	2	2
21	2	2	4	2	31	2	2	2	2
			Most populated cluster			1	1	1	1 / 2

TABLE 2 Total number of time sub-periods per cluster of the total load demand curves for HN MEKO type Frigate Power System for 6 hours time sub-period.

	Load cluster							
6 hours period	1	2	3	4	5	6	7	8
0:15-6:00	**11**	4	0	3	0	1	0	1
6:15-12:00	**12**	6	1	0	0	0	1	0
12:15-18:00	**10**	8	0	1	0	1	0	0
18:15-24:00	**7**	**7**	2	2	1	1	0	0
Time periods per cluster	40	25	3	6	1	3	1	1

In Fig. 6 the representative load curves per cluster are presented in case of using 8 clusters. Additionally, the confidence limits of the variations (mean value ± standard deviation) are presented. The intermediate area between the confidence limits has a probability of occurrence equal to 68.27% if normal distribution is assumed.

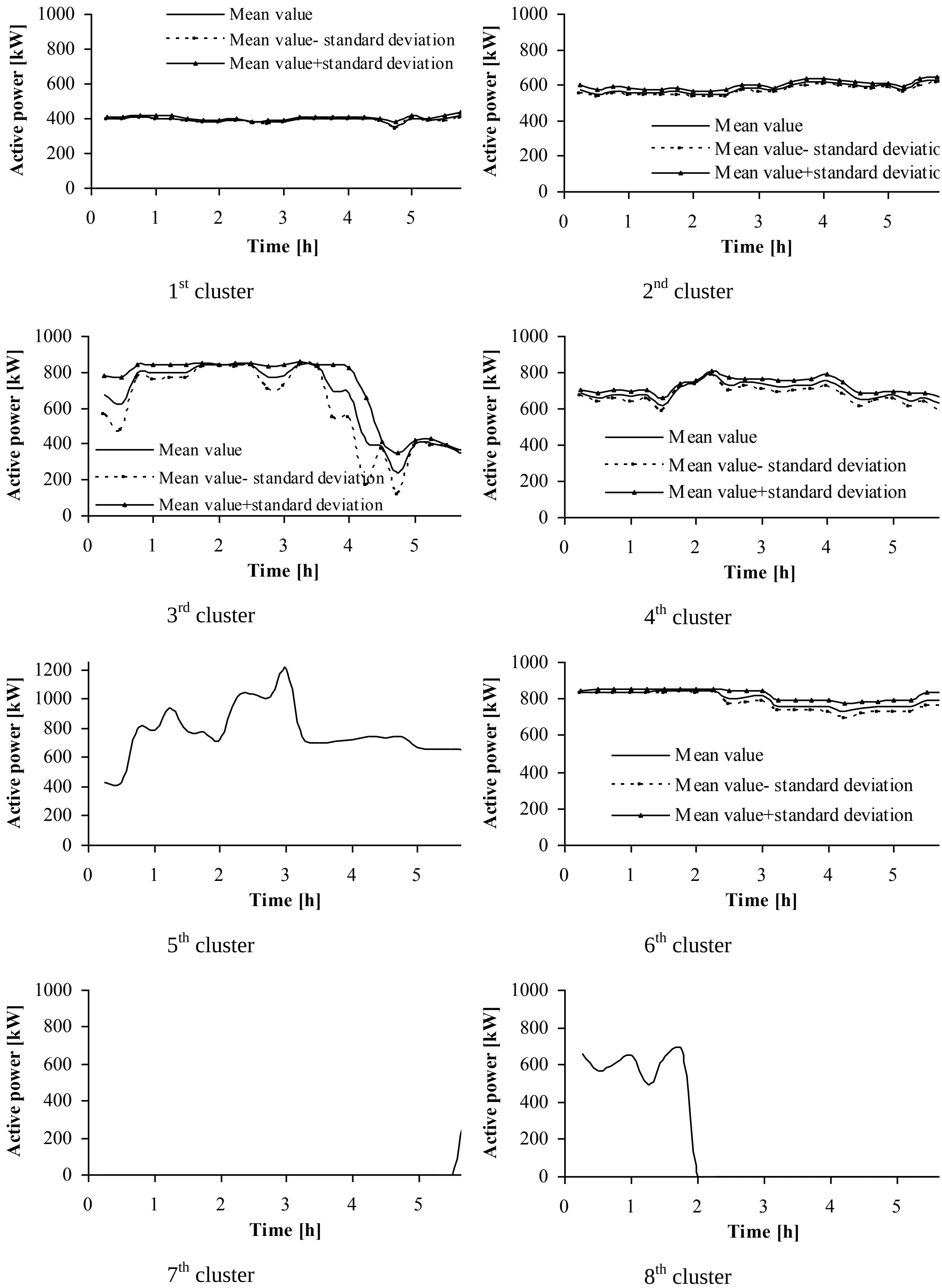

Figure 6. 6 hours chronological load curves for total load demand with the respective confidence intervals using the modified k-means technique with *WCBCR* as evaluation measure and assuming 8 clusters.

More specifically, the first cluster is the most populated one with the half population of periods belonging in this cluster. The respective mean total load demand is about 393 kW with mean standard deviation of about 11 kW. The second cluster includes the 31.25% of the total population and it corresponds to mean total load demand of about 583 kW with mean standard deviation of about 20 kW. The third cluster includes the 3.75% of the total population. It contains the load curve of 14 January 1998, which is a special day as it corresponds to a power system failure occurred at 10:45 p.m. and it could be excluded during data pre-processing. The mean total load demand is about 644 kW and with mean standard deviation of about 57 kW. The forth cluster includes the 7.5% of the total population and it corresponds to mean total load demand of about 694 kW and with mean standard deviation of about 29kW. The fifth cluster has only one member representing the 6-hours time sub-period with the peak load. The sixth cluster includes the 3.75% of the total population and it corresponds to large mean total load demand of about 799 kW with mean standard deviation of about 26 kW. The seventh and eighth clusters have one member each one, which represents the load curve of 29 January 1998, when the frigate was supplied by the shore connection for 8 hours. From the analysis of the examined period, it seems the frigate power system presented the same behaviour for all days, working days or not.

The obtained 6-hours chronological load curves can be used for the load estimation of the next day. The 5th, 7th and the 8th clusters are "special" curves, so they can not be used for load forecasting. Because lack of data (only 18 days, no available temperatures etc) a separation rule between the rest clusters cannot be found. So, for each 6-hour time sub-period (one from 0:15-6:00, 6:15-12:00, 12:15-18:00, 18:15-24:00) the 6-hour chronological load demand curve of the most populated cluster can be easily used as the prospective load demand curve of the respective 6-hour time period of the next day. In case of having two clusters with the same population, the respective mean value is used for prediction purposes. Here, the 1st representative load demand curve is selected for time sub-periods 0:15-6:00, 6:15-12:00, 12:15-18:00, while the mean value load curve of the 1st and the 2nd curves is calculated for the sub-period 18:15-24:00. The respective mean absolute percentage error (*MAPE*) is

24.46% for all sub-periods. If the connection between sub-periods and clusters was known a priori, the respective *MAPE* would be 11.86%, which is the lower limit that can be achieved.

3.3 Application of Proposed Method

The aforementioned process has been repeated for different time periods. Specifically, the examined time periods are one hour, 2 hours, 3 hours, 4 hours, 6 hours (from the previous paragraph), 8 hours, 12 hours and 24 hours [5], which lead to different size of vectors and sets, as registered in Table 3. Following, the modified *k*-means technique is executed according to the optimization process and the number of the clusters, which corresponds to the knee of the respective curve, is also registered in Table 3.

If load forecasting for the next day is performed based on the most populated cluster per time period, the respective mean absolute percentage error ($MAPE_{\text{all}}$) for all days for the different time sub-periods is registered in Table 4. If the connection between time periods and clusters was known a priori, the respective mean absolute percentage error ($MAPE_{per}$) is obtained as registered in Table 4. These results are compared with the classical load estimation which is based on the mean values of the respective time sub-periods for all days using the criterion of mean absolute percentage error (*MAPE*).

TABLE 3 Size of vector, number of vectors and number of selected clusters from pattern recognition method execution for HN MEKO type Frigate Power System for different time periods [5].

Time period [h]	1	2	3	4	6	8	12	24
Size of vector	4	8	12	16	24	32	48	96
Number of vectors	480	240	160	120	80	60	40	20
Number of clusters	10	8	9	7	8	7	7	4

TABLE 4 Mean absolute percentage error ($MAPE_{all}$) for 20 days based on different time sub-periods.

Time period [h]	1	2	3	4	6	8	12	24
$MAPE_{all}$ [%]	25.30	24.83	25.13	25.91	**24.46**	27.49	25.07	30.01
$MAPE_{per}$ [%]	**8.97**	10.35	10.21	11.88	11.86	12.50	11.32	18.63
$MAPE$ [%]	28.10	28.10	28.11	28.13	28.14	25.62	**25.52**	27.14

The best results for mean absolute percentage error based on the most populated time sub-period appear for the 6-hour time sub-period. It is noted that the proposed load estimation methodology is proved superior to the classical one for all cases, except from 8 and 24-hour time periods. The different behaviour of the last ones ought to the small available number of vectors.

5 Conclusions

In this paper, a pattern recognition methodology is presented, which is used for ship load estimation. The applied method consists of four steps: (a) data selection, (b) data preprocessing, (c) pattern recognition algorithm application and (d) load estimation. The clustering algorithm used is a modified k-means algorithm properly calibrated regarding the ratio of within cluster sum of squares to between cluster variation. The application example of this methodology was a Hellenic Navy MEKO type frigate with its total load demand at anchor for 20 days was available. From the examination of different time sub-periods of chronological load curves it appears that 6-hour time sub-period is the best, as it gives the smallest mean absolute percentage error (*MAPE*). The results obtained by the proposed method are compared with the classical load estimation based on the mean values of the respective time sub-periods for all days. A significant improvement of *MAPE* by 3.68% (from 28.14% to 24.46%) is obtained with the proposed method.

6 Acknowledgements

The authors would like to express their appreciation to Dr. I.K. Hatzilau, Emeritus Professor at Hellenic Naval Academy, for his overall contribution on the development of the Electric Circuits Laboratory of Hellenic Naval Academy.

References

[1] J. M. Prousalidis, G. J. Tsekouras, F.D. Kanellos: *New challenges emerged from the development of more efficient electric energy generation units*, in IEEE Electric Ship Technologies Symposium (IEEE ESTS '11), Westin Alexandria, Alexandria, Virginia, USA, 2011

[2] US Navy: *Electrical system load and power analysis for surface ships*, DOS 310-1, 1980

[3] I.K. Hatzilau, G. J. Tsekouras, J. Prousalidis, I.K. Gyparis: *On Electric Load Characterization and Categorization in Ship Electric Installations*, INEC-2008, 1-3 April 2008, Hamburg, Germany

[4] G. J. Tsekouras, I.S. Karanasiou, F.D. Kanellos: *Classification of total load demand profiles for war-ships based on pattern recognition methods*, in 15th WSEAS International Conference on Communications (CSCC '11), Corfu, Greece, 2011

[5] A.A. Gialketsi, G. J. Tsekouras, F.D. Kanellos, J.M. Prousalidis: *Ship electric load estimation using pattern recognition methods for the classification of the total electric load demand curves*, 1st MARINELIVE Conference on "All-Electric Ship", June 4-5, 2012, Athens, Greece

[6] G. Chicco, R. Napoli, F. Piglione: *Load pattern clustering for short-term load forecasting of anomalous days*, in 2001 IEEE Porto Power Tech Conference 10th-13th, September, Porto, Portugal p.6.

[7] G. J. Tsekouras, I.K. Hatzilau, J. Prousalidis: *A new pattern recognition methodology for classification of load profiles for ships electric consumers*, Journal of Marine Engineering and Technology, No. A14 (2009), pages 45-58,

[8] G. J. Tsekouras, P. B. Kotoulas, C.D. Tsirekis, E. N. Dialynas, N. D. Hatziargyriou: *A Pattern Recognition Methodology for Evaluation of Load*

Profiles and Typical Days of Large Electricity Customers, Electrical Power Systems Research, Vol. 78 (2008), pages 1494-1510

[9] G. J. Tsekouras, N.D. Hatziargyriou, E. N. Dialynas: *Two-Stage Pattern Recognition of Load Curves for Classification of Electricity Customers*, IEEE Transactions on Power Systems, Vol. 22, No. 3,(2007), pages 1120-1128

[10] G. J. Tsekouras, C.A. Anastasopoulos, F.D. Kanellos, V.T. Kontargyri, I.S. Karanasiou, A.D. Salis, N.E. Mastorakis: *A Demand Side Management Program of Vanadium Redox Energy Storage System for an Interconnected Power System*, WSEAS International Conference on Energy Planning, Energy Saving, Environmental Education, (EPESE '08), Corfu Island, Greece, 2008

[11] G. J. Tsekouras, F.D. Kanellos, V.T. Kontargyri, I.S. Karanasiou, A.D. Salis, N.E. Mastorakis: *A New Classification Pattern Recognition Methodology for Power System Typical Load Profiles*, WSEAS Transactions on Circuits & Systems, Vol. 12, Issue 7, 2008, pp. 1090-1104

[12] G. J. Tsekouras, A. D. Salis, M. A. Tsaroucha, I. S. Karanasiou: *Load time-series classification based on pattern recognition methods*, 16th chapter in the book of Peng-Yeng Yin. Pattern Recognition; Techniques, Technology and Applications, 1st Edition, I-Tech Press, Vienna, 2008.