

ARIMA Models in Load Modelling with Clustering Approach

J. Nazarko, *Member IEEE*, A. Jurczuk, W. Zalewski

Abstract – In distribution system, bus load estimation is complicated because system load is usually monitored at only a few points. As a rule receiving nodes are not equipped with stationary measuring instruments so measurements of loads are performed sporadically. In general, the only information commonly available regarding loads, other than major distribution substations and equipment installations, is billing cycle customer kWh consumption. The paper presents application possibilities of ARIMA modelling with fuzzy clustering approach to electrical load estimation. The clustering of the substations (customers) according to daily load curves is made on the base of the average alignment degrees. Simulation studies have been performed to demonstrate the efficiency of the proposed scheme and an effect of different parameters on its accuracy on the basis of actual data obtained at distribution system substations.

IndexTerms – Distribution System Planning, Load Modelling, Clustering, Fuzzy Set, ARIMA Models.

I. INTRODUCTION

The knowledge of loads at system buses is one of the most important requirements for efficient operation of power distribution systems [1]. Estimation of loads is the basis for the system state estimation and for technical and economic calculations. This makes possible improvement in operation and maintenance of electrical equipment and in planning of network operating configurations.

The main difficulties in the modelling of loads at receiving buses in distribution systems result from the random nature of loads, diversification of load shapes on different parts of the system, the deficiency of measured data and the fragmentary and uncertain character of information on loads and customers.

In the present stage of development of power distribution systems, the mathematical estimation of the loads at the system buses seems to be the most realistic strategy due to incomplete primary information on loads and customers. It demands earlier determination of the stable relations between bus loads and easier available data [1].

The probabilistic models are widely used to estimate system loads. In order to develop the relevant types and parameters of

probability distribution, large numbers of recorded consumption data are required. To obtain the above data a special measurement project has to be considered. The use of statistical methods is not always possible due to occurrence of a large deficit of measurements. The fuzzy set theory is a convenient mathematical tool that allows us partially to eliminate this problem and to limit the influence of deficit of measurements.

The paper presents application possibilities of ARIMA modelling with clustering approach to electrical load estimation. The clustering of the substations (customers) according to daily load curves is made on the base of the average alignment degrees. In the second step time series analysis was carried out. Simulation studies have been performed to demonstrate the efficiency of the proposed scheme on the basis of actual data obtained at distribution system substations.

II. APPLICATION OF ALIGNMENT DEGREES TO CLUSTERING

The loads on distribution transformers are the instantaneous summations of the individual demands of many customers. Since the pattern of electrical demand of each customer cannot be determined precisely, it is usually necessary to calculate system loadings on an estimation basis.

Planning engineers use load modeling to predict the loads on different parts and different time of distribution systems. The load estimation can be done on the basis of the fuzzy inference with clustering method of customers load profiles [1].

The 24-hour load curves of the different substation groups show the characteristic variation for each group. Example daily load profile for one substation is shown in Fig 1. There are three characteristic load values in the diagram:

- P_{dP} – the daily peak load,
- P_{dA} – the daily average load,
- P_{dB} – the daily base load.

On the basis of the analysis of the profiles from different days, the four characteristic intervals (columns) of the day are distinguished (Fig. 1):

- night – **n**,
- morning – **m**,
- afternoon – **a**,
- evening – **e**.

Authors are with Bialystok Technical University, Wiejska 45A St., 15-950 Bialystok, Poland (e-mail: nazarko@pb.bialystok.pl, ajurczuk@pb.bialystok.pl, zalewski@pb.bialystok.pl).

This paper reports work sponsored by the State Committee for Scientific Research (KBN) under contract 4T10B00225.

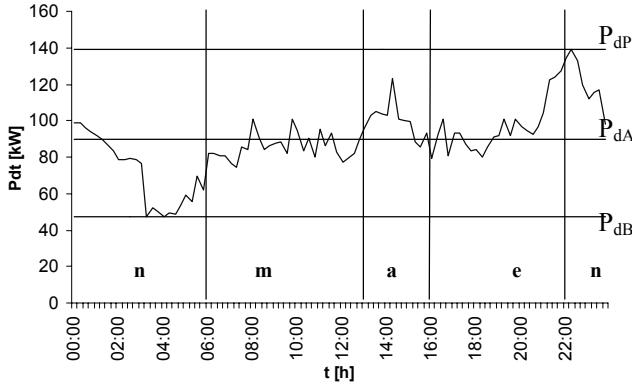


Fig. 1. Example of 24-hour load profile in substation

The clustering of the substations (customers) according to daily load curves is made on the base of the average alignment for each column. To avoid the influence of the instantaneous values on power changes consumed from substations, the average load of twenty four hours should be taken as reference quantity. The average alignment degree for each column is defined as ratio of the average load in the column to the daily average load:

$$I_c = \frac{P_{jA}}{P_{dA}} \quad (1)$$

where:

- c – the column index (n, m, a, e),
- P_{jA} – the average load in the column j,
- P_{dA} – the daily average load.

The load diagrams are regarded as similar ones when their average alignment degrees for each column have similar values [1].

The right selection of the investigating objects to build the model is an important factor in load estimation process. The use of data from substations that belong to different classes can cause difficulties in working out a method to practical application. For this reason, it is necessary to compare all data characterizing substations especially attentively and to establish their membership to individual class of objects. In this work the clustering approach presented above was used in the ARIMA models to the electrical load modeling.

III. LOAD PROFILE MODELLING WITH ARIMA MODELS

Load modelling and forecasting has always been important for planning and operational decision conducted by utility companies. Effective analysis of load profiles and their adequate classification are getting more important because of market liberalisation. Knowledge about end-users' profiles is necessary to prepare the price strategy, reorganise a tariff system, to improve a network reliability and efficiency. Load modelling is also very important for balancing supply and demand side.

Most of modelling methods uses statistical techniques or artificial intelligence algorithms such as neural networks, fuzzy logic and expert systems. ARMA (Auto-Regressive Moving Average) and ARIMA (Auto-Regressive Integrated Moving Average), from among statistical methods are the most often used classical time series methods.

ARIMA is a class of stochastic process used to analyse time series. ARIMA technique is based on the assumption that the data have an internal structure, such as auto-correlation, trend or seasonal variation. The main idea of these methods is to identify and explore such a structure. The application of the ARIMA models for time series analysis is based on methodology of Box and Jenkins [3].

Contrary to standard regression models, ARMA and ARIMA models use the time and load as the only input parameters. These models forecast (estimate) a value in a response time series as a linear combination of its own past values and past forecast errors [4]. Equation (2) defines an example of $ARIMA(p,0,q)$ stationary process of order p, q :

$$y_t = c + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q} \quad (2)$$

where:

- $y_t, y_{t-1}, \dots, y_{t-p}$ - past values of time series,
- $e_t, e_{t-1}, \dots, e_{t-q}$ - past errors (values of white noise-series),
- ϕ_p, θ_q - parameters of model,
- p - order of non-seasonal auto-regressive model,
- q order of non-seasonal moving-average model.

In the presented approach, due to a specificity of analysed time series for load modelling a seasonal ARIMA model has been used. This model is defined as follows:

$$ARIMA(p,d,q)(P,D,Q)_s \quad (3)$$

where:

- d - number of non-seasonal differences,
- D - number of seasonal differences,
- P - order of seasonal auto-regressive model,
- Q - order of seasonal moving-average model,
- s - number of seasons.

Owing to its features ARIMA is a basic technique for analysis and modelling of time series. These models can be more complex to use than other statistical approaches.

ARIMA models are more often used for forecasting of load and electricity prices. But in the presented case, ARIMA models have been used for clustering load curves. The main idea was to compare classification, which has been performed with fuzzy sets, using similarity of classes of ARIMA appropriate models. It was assumed that it is possible to formulate clusters of load profiles through comparison of ARIMA models features. In order to examine the similarity of models for different load profiles it necessary to identify their classes and estimate appropriate parameters. For all parameters confidence intervals is estimated. A confidence

interval in this case is considered as a range of values that has a specified probability of containing the parameter being estimated. Procedure of model identification is consistent with Box and Jenkins methodology [3].

It has been assumed that parameters of model could be considered as points in a multi-dimensional ($\phi, \theta; \Phi, \Theta$) space. Each of these points has their own subspace specified by confidence interval. ARIMA models are considered as similar if their subspaces are convergent (Fig. 2).

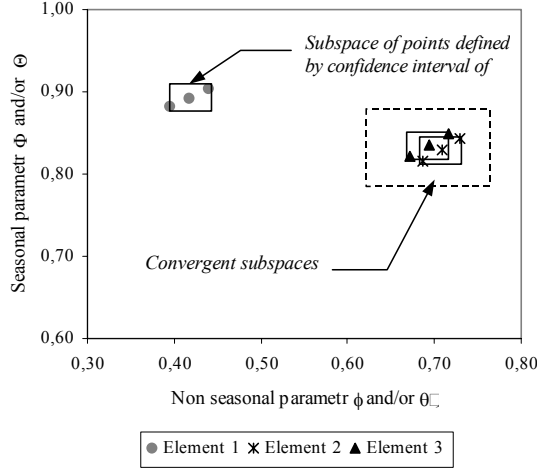


Fig. 2. Idea of clustering with ARIMA models

In the presented approach, load profiles are clustered according to following criteria:

- class and order of analysed models is identical,
- there is an intersection of parameters' confidence intervals (overlapping of confidence intervals).

According to authors' approach, it is possible to divide a load profiles into mutually exclusive clusters such that the elements of each group are similar. Results of comparisons are given in concentration plot. Data visualisation, on the one hand is aimed at identify a similarity of ARIMA models and on the other hand it indirectly allows to form clusters of analysed load profiles.

IV. NUMERICAL EXAMPLE

To verify the proposed method of peak load estimation the measurements of daily energy consumption A_d and daily peak load P_{dp} at selected distribution substations in Bialystok Power Distribution Utility Co. were made in June and July. Investigated objects are substations with transformers with 15/0.4 kV ratio of transformation and power ratio from 160 to 400 kVA. The frequency of measurement was 15 minutes. Description of substations and supplied customers are shown in Table I. Example daily load profiles for substation No. 35 and 734 are shown in Fig. 3 and 4.

TABLE I
DATA OF SUBSTATIONS AND SUPPLIED CUSTOMERS

Substation No.	Power of transformers	Number of customers	Type of customers
35	250 kV·A	50	commercial-services
54	250 kV·A	75	municipal-living + commercial-services
56	250 kV·A	88	individual houses + small church
638	400 kV·A	150	municipal-living
734	250 kV·A	56	individual houses + small services
902	400 kV·A	200	municipal-living + prison
1197	160 kV·A	67	music school
1643	250 kV·A	257	marketplace with small shops

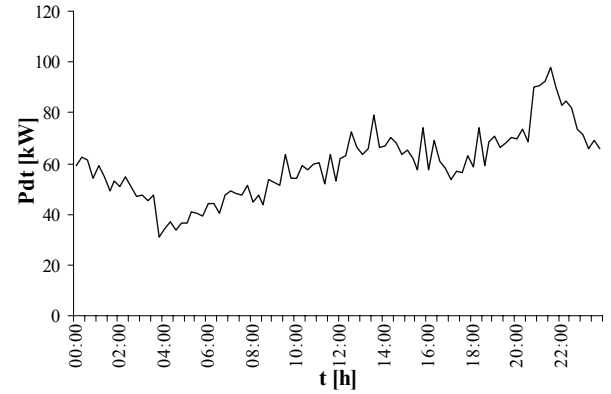


Fig. 3. Example of 24-hour load profiles in substation No. 734

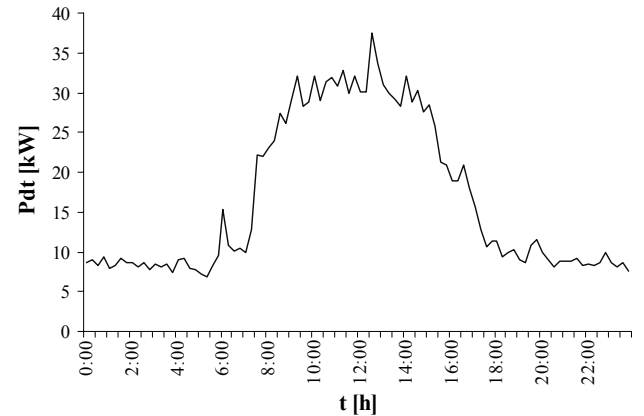


Fig. 4. Example of 24-hour load profiles in substation No. 35

On the basis of above considerations in the first step clustering of load in substation was made. On the ground of measurements of the profiles average alignment degrees for each column was calculated. The clustering of load profiles on

the plane l_e (the average alignment degree at the column **e** – evening) and l_m (the average alignment degree at the column **m** – morning) is shown in Fig.5a. The clustering of load profiles on the plane l_n (the average alignment degree at the column **n** – night) and l_a (the average alignment degree at the column **a** – afternoon) is shown in Fig. 5b. To the comparing process fuzzy set approach was applied [2].

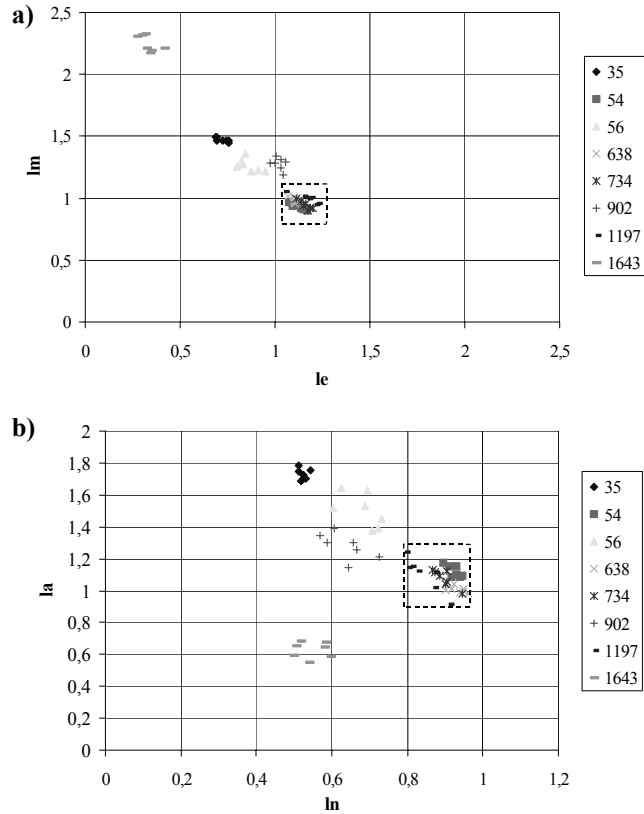


Fig 5. The example of clustering of seven daily profiles on the plane l_e and l_m (a) and l_n and l_a (b) for eight substations

Each average value of alignment degrees for seven work days is represented by symmetric triangular fuzzy number as follows: $\tilde{l}_i = [l_{ic}, l_{ir}]$

where: l_{ic} - center parameters of fuzzy numbers

(membership function $\mu = 1$),

l_r - spreads of fuzzy numbers (geometrically the spread is a half of the base of the triangle)

Parameters l_{ic} were calculate as average values of seven points for each substation. The spread of fuzzy alignment degrees is difference between max and min value for each substations. In accordance with above approach to load profiles clustering, substation no. 54, 638, 734 and 1197 could be considered as similar.

In the second step time series analysis was carried out. ARIMA models of examined load profiles have been built in consistent with Box-Jenkins methodology [4]. According to this procedure, for each variable representing load data from considered substations, a class of ARIMA model was described.

To identify a class of ARIMA model, according to Box-Jenkins methodology, have been used plots of auto-correlation

and partial auto-correlation functions. [4]. In the analysed case, load series have been non-stationary. The most common cause of non-stationarity of examined variables has been seasonality (typical for load data). An example of ACF and PACF plots for one of analysed load profiles is given in the Fig. 6 (daily variation, number of seasons $s=96$).

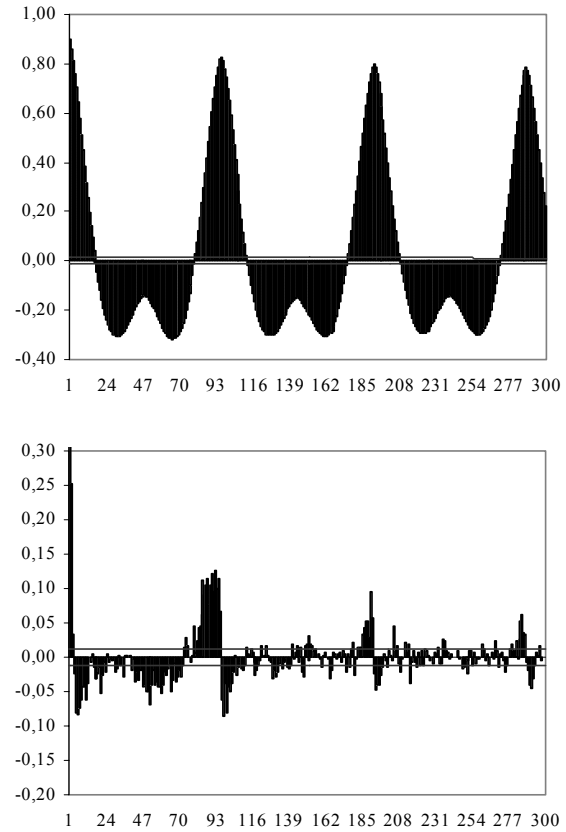
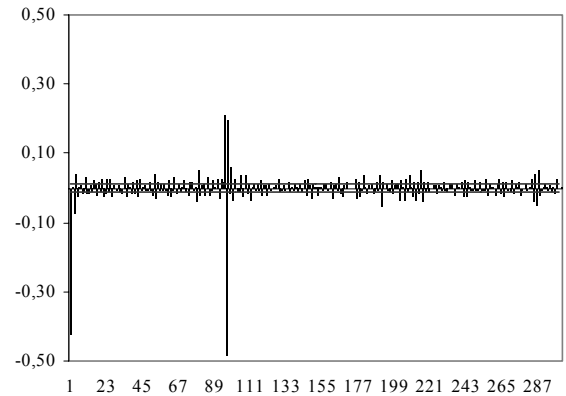


Fig. 6. ACF and PACF plots of time series

Therefore, analysed time-series have been de-seasonalized. Owing to distinctive features of these ACF and PACF plots (for stationary time series) it has been possible to identify a tentative class of ARIMA models (Fig. 7).



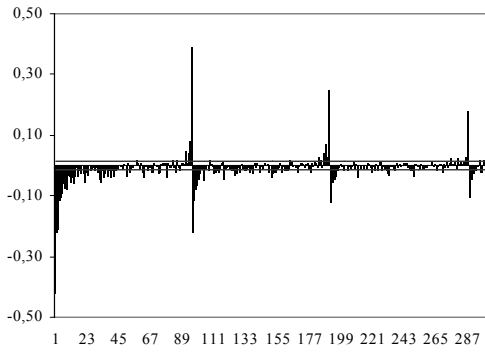


Fig. 7. ACF and PACF plots of differenced time series

Due to the main purpose of research, one adequate class of ARIMA model has been chosen for all load profiles. Examining correlograms it has been observed one significant spike in ACF plot at lag 1 and 96. Considering PACF plot, exponential decay of function values is observed (Fig. 7). These two characteristics are typical for seasonal moving-average model of order 1, which takes the following form:

$$ARIMA(0,1,1) \times (0,1,1)_{96}. \quad (4)$$

For all analysed time-series representing different profiles of end-user ARIMA models (3) have been identified and verified. Parameters in all examining models are given in the Table II.

TABLE II
PARAMETERS OF ANALYSED MODELS $ARIMA(0,1,1) \times (0,1,1)_{96}$

Substation	35	54	56	638	734	902	1197	1643
ϕ	0,416	0,663	0,681	0,735	0,708	0,623	0,694	0,575
θ	0,893	0,811	0,873	0,834	0,830	0,859	0,835	0,863

Achieved results have been used to verify the first attempt to clustering load curves. As it was mentioned above, the concentration of models' parameters has been chosen as main criteria of load profiles clustering. Results of this step of analysis are presented in the Fig. 8.

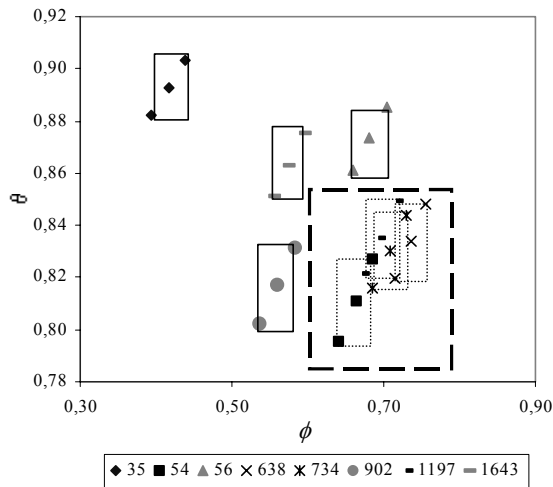


Fig. 8. Concentration plot of ARIMA models' parameters

In accordance with assumptions of ARIMA approach to load profiles clustering, substation no. 54, 638, 734 and 1197 could be considered as similar (Fig. 8). Confidence intervals in these cases are overlapped. These four substations may form a main cluster represented by one model of time series.

As it is shown in the Fig. 8, load profile representing substation no. 35 is not concentrated with the rest of them. Confidence interval in this case has not any intersection with other one. ARIMA models and alignment degrees in analysed case generate similar results in the clustering process.

V. CONCLUSIONS

The paper presents possibilities of ARIMA modelling with clustering approach to electrical load estimation. The study was aimed to demonstrate the efficiency of the proposed method.

On the basis of the results it is possible to combine the fuzzy clustering and ARIMA models for load profile clustering. Researches in time series clustering have revealed that these models give the similar results as fuzzy coefficient approach. But in practise, estimation of ARIMA models demands in many cases statistical experience and sophisticated tools. Therefore the first of examined method seems to be more advantageous. But considering supply-side planning determined classes of ARIMA models could be useful in short-term load forecasting.

Achieved results show usefulness of these kinds of methods in two following cases:

- formulation of the substation classes,
- load estimation for substation clusters.

Taking into account changes of energy policy, effective classification and modelling tools could helpful in re-organising of tariff systems, formulating price strategies and capacity planning.

The authors see usefulness of ARIMA modelling with fuzzy clustering approach to solve problems of load forecasting and load estimation in power distribution systems. The presented method may be also a useful tool supporting planning distribution engineers.

VI. ACKNOWLEDGMENT

The authors are grateful to the Board of Zakład Energetyczny Białystok SA for a technical support in the field load research.

VII. REFERENCES

- [1] J. Nazarko, *Modeling of Electrical Power Distribution Systems*, Białystok Technical University Publisher, Białystok, 1993.
- [2] Zalewski W.: "Application of fuzzy inference to electric peak load modelling", in *Proc. IEEE International Conference on Power Systems, Kathmandu, Nepal, 3-5 November, 2004*.
- [3] Box G.E.P., Jenkins G.M., *Time series analysis: forecasting and control*, Englewood Cliffs, NJ, Prentice Hall, 1994.
- [4] J. H. Wilson, B. Keating, *Business Forecasting*, Irwin McGraw-Hill, Boston, 1998, pp. 299-369.

VIII. BIOGRAPHIES



Joanicjusz Nazarko received his Ph.D. and D.Sc. degrees from the Warsaw University of Technology in 1983 and 1992, respectively. He is currently a Full Professor at the Bialystok Technical University, Poland. He heads the Chair of Business Informatics and Logistics. He has been a Visiting Professor at the Virginia Polytechnic and State University at Blacksburg, USA, the University of Ulster at Jordanstown, UK, and the University of Stuttgart, Germany. He has participated in several international research projects. He has also been a speaker at numerous international conferences. He has produced over 100 papers and reports, receiving several awards for his research, publication, and teaching activities. He is a member of IEEE, IIF and SEP.



Arkadiusz Jurczuk received MSc in electrical engineering from Bialystok Technical University, Ph.D. degree in management science from Management Department of Lodz University, respectively in 1995 and 2001. Since 1995 he had worked for Bialystok Power Utility as a specialist of electricity trade. His utility experience includes topics from customer service area, load management and electricity markets. In the 1998 he joined Bialystok Technical University. He has been working for Faculty of Management, where he is an assistant professor. The main interest's areas are demand-side management, local energy markets, modelling of a decision process and risk management. He is author of over than twenty papers connected with his fields of interests. He is a member of Technical Committee in Podlaska Agency For Energy.



Wojciech Zalewski was born in Bialystok, Poland. He graduated from the Bialystok Technical University with M.Sc. degree in Electrical Engineering in 1988. He received his Ph.D. degree in Electrical Engineering from the Warsaw University of Technology, Poland in 1997. He is presently an Assistant Professor at the Chair of Business Informatics and Logistics at the Bialystok Technical University. His research interest areas are the application of expert systems, and probabilistic and fuzzy concepts to power distribution systems modelling and analysis with emphasis on load estimation. He is an author of 40 publications and member of Polish Society of Electrical Engineers (SEP).