1999 IEEE International Fuzzy Systems Conference Proceedings
August 22-25, 1999, Seoul, Korea

An Evaluation of an Accuracy of the Fuzzy Regression Analysis in the Electrical Load Estimation

J. Nazarko W. Zalewski
Bialystok Technical University
ul. Wiejska 45 A
15-950 Bialystok, Poland
(E-mail: jnazarko@cksr.ac.bialystok.pl
zalewski@cksr.ac.bialystok.pl)

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. In order to model system uncertainty, inexactness, and random nature of customers' demand, a fuzzy system approach is proposed. This paper presents possibilities of application of the fuzzy set theory to electrical load estimation and the effect of different factors on the result of calculations. Unreliable and inaccurate input data have been modeled by means of fuzzy numbers. A regression model, expressing the correlation between a substation peak load and a set of customer features (explanatory variables), existing in the substation population, is determined. 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.

1. Introduction

The knowledge of loads at system buses is one of the most important requirements for efficient operation of power distribution systems [3, 5]. Estimation of loads, particularly of peak 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 modeling of peak 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 [5].

The most renowned method for expressing the uncertainty in load models is fuzzy set theory [2, 4, 6, 7]. For the purpose of simplicity of mathematical operation the trapezoidal and triangular forms of fuzzy numbers are usually used (Fig. 1).

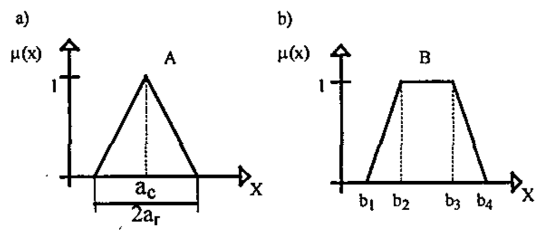


Fig. 1. Triangular and trapezoidal forms of fuzzy number

This paper presents a new scheme based on the fuzzy regression analysis for the estimation of peak load in distribution systems. Comparison of different factors influencing accuracy of the method has been done.

2. Fuzzy Regression Method

The general regression model is given by the following equation [1]:

$$\mathbf{Y} = \mathbf{ZA} + \mathbf{e} \tag{1}$$

where: **Y** - vector of output variables,
Z - matrix of independent variables,
A - vector of parameters,
e - vector of unobservable errors.

Two cases can be discriminated depending on the type of output variable. The first when the output variable y is a real number and the second when the output value is an interval $y \in \langle y_L, y_R \rangle$.

The first case is described in this section. It can be represented in the form:

$$\tilde{\mathbf{Y}} = \mathbf{Z}\tilde{\mathbf{A}} \tag{2}$$

where:

$$\tilde{y}_i(z_i) = \tilde{a}_0 + \tilde{a}_1 z_{i1} + \cdots + \tilde{a}_k z_{ik} \qquad i = 1, 2, ..., n \quad (3)$$

The linear fuzzy regression model (3) is represented using symmetric triangular fuzzy parameters $\tilde{a}_i = [a_{ic}, a_{ir}]$ (Fig. 1) as follows:

$$\tilde{y}_i(z_i) = [a_{0c}, a_{0r}] + [a_{1c}, a_{1r}] z_{i1} + \cdots + [a_{kc}, a_{kr}] z_{ik} \quad (4)$$

$$y_{ci}(z_i) = a_{0c} + a_{1c} z_{i1} + \cdots + a_{kc} z_{ik} \quad (5)$$

$$y_{ri}(z_i) = a_{0r} + a_{1r} z_{i1} + \cdots + a_{kr} z_{ik} \quad (6)$$

where: y_c, a_c - center parameters of fuzzy numbers (membership function $\mu = 1$),
y_r, a_r - spreads of fuzzy numbers (geometrically the spread is a half of the base of the triangle).

The coefficients of polynomial (3) should be optimally determined in order that value of membership function of estimator is bigger than imposed level *h*. It means that [10]:

$$\mu_{\tilde{Y}_i}(y_i) \geq h \qquad i=1,2,...,n \quad (7)$$

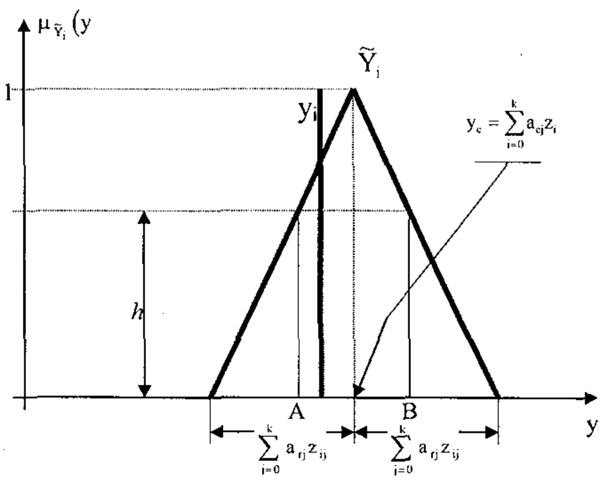


Fig. 2. The membership function of estimator $\mu_{\tilde{Y}_i}(y_i)$ and estimating value y_i

The Fig. 2 presents graphic interpretation of this way. From equation (7) appears that output value y_i should be between points *A* and *B* (Fig. 2). In the other words value y_i should belong to support of *h*-cut set:

$$y_i \in \text{supp}(Y_h) \qquad h \in [0,1] \quad (8)$$

Author of model determines the value of level *h*. It is the threshold value representing the degree to which the model should satisfy all the experimental data points. The dependence between *h* level and spread of the fuzzy set is proportional.

The method uses the criterion of minimizing the total vagueness $J = f(y_r(z_i))$, defined as the sum of individual spreads of elements of vector $\tilde{\mathbf{Y}}$ [9, 10].

$$J = y_{1r}(z_1) + y_{2r}(z_2) + \cdots + y_{nr}(z_n) \rightarrow \text{Minimum} \quad (9)$$

$$\text{subject to} \quad y_i \in \tilde{\mathbf{Y}}(z_i), \qquad i = 1, 2, ..., n \quad (10)$$

$$a_{jr} \geq 0, \qquad j = 0, 1, 2, ..., k \quad (11)$$

Using equations (4) – (6) the problem can be written as follows:

$$J = \sum_{i=1}^{n} \left(a_{0r} + a_{1r}|z_{i1}| + \cdots + a_{kr}|z_{ik}| \right) \rightarrow \text{Minimum} \quad (12)$$

$$a_{0c} + \sum_{j=1}^{k} \left(a_{jc} z_{ij} \right) - (1-h) \left(a_{0r} + \sum_{j=1}^{k} \left(a_{jr} |z_{ij}| \right) \right) \leq y_i \quad (13)$$
$$i = 1, 2, ..., n$$

$$a_{0c} + \sum_{j=1}^{k} \left(a_{jc} z_{ij} \right) + (1-h) \left(a_{0r} + \sum_{j=1}^{k} \left(a_{jr} |z_{ij}| \right) \right) \geq y_i \quad (14)$$
$$i = 1, 2, ..., n$$

Dependence (12) presents a linear programming (LP) task, which can be easily solved by conventional techniques. The parameters $a_i = [a_{ic}, a_{ir}]$ of vector $\tilde{\mathbf{A}}$ are determined as the optimal solution of the LP problem (12) – (14).

3. Application of the Fuzzy Regression Analysis to the Electrical Load Estimation

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 estimation to predict the loads on different parts of distribution systems.

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 values are required. To obtain the above data a special measurement project has to be considered.

In practice the only information commonly available regarding loads, other a major distribution substations is kWh consumption. In modeling process also can be used other input quantities (number of customers, rated power of transformers, installed capacity).

Many relationships between output quantities (peak load, load flow, losses of power and energy) and describing values coming from measurements can be represented by a regression model. 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 to partially eliminate unreliability from input information and to limit the influence of deficit of measurements.

The daily 15-minutes peak power consumption for a given substation may be found on the basis of input quantities using fuzzy regression models (3) and (12) - (14).

4. Factors Influencing Accuracy of the Method

The accuracy of electrical load estimation method (presented in section II) depends on following factors: significance of explanatory variables used to building of the model, value of the h level, right selection of the objects, which were a basis of the modeling process.

In the first case, verification of significance of input quantities is made. This is a basis for the eventual elimination from the qualitative model those input quantities that may be recognized as insignificant in assumed variation interval of all input quantities. The significant input quantities are placed in the sequence from the most significant to the least one.

On initial stage of analysis and optimization of distribution system operating conditions, possible large number of quantities that characterize system and have influence investigated output characteristic should be taken into account. Elimination of same quantities without sufficient justification but only for fear that a model might be too large is a mistake that may shatter the sense of all work when omitted quantities have a significant influence investigated system characteristic. In order to verify the significance of investigated factor on analyzed output quantity, the theory of experimental design is applied [5].

The design of experiment is a procedure of selection of the number and conditions of realization of experiments that are necessary and sufficient for solution of a stated problem with required accuracy as well as the procedure of selection of a mathematical method of analysis of experiment results and a decision criterion [4]. The following sets are defined in this method:

- set of input quantities

$$\mathbf{X}_C,\ \{\mathbf{X}_i\colon i = 1, 2, \ldots, k\}, \tag{15}$$

where k is a number of input quantities;

- set of output quantities

$$\mathbf{W},\ \{W_p\colon p = 1, 2, \ldots, m\}, \tag{16}$$

where m is a number of output quantities;

- set of constants

$$\mathbf{C},\ \{C_s\colon s = 1, 2, \ldots, r\}, \tag{17}$$

where r is a number of constants.

A qualitative mathematical model of an investigated object is introduced in the following form:

$$\mathbf{F}(X_1, X_2, \ldots, X_i, W_1, W_2, \ldots, W_m) = 0. \tag{18}$$

Then the investigated object is decomposed on m formal objects and each of them is described by only one output quantity. In presented method of load estimation only one output quantity (daily peak load) and three input quantities (daily energy consumption, number of customers, rated power of transformers) were determinate. This leads to the succeeding relation:

$$F_1(X_1, X_2, X_3, W_1) = 0. \tag{19}$$

The above decomposition is used to define the response function of the investigated object:

$$W_1 = F_1(X_1, X_2, X_3). \tag{20}$$

This function is approximated by linear model:

$$W = b_0 + b_1X_1 + b_2X_2 + \ldots + b_kX_k\,, \tag{21}$$

where $b_0, b_1, b_2, \ldots, b_k$ are polynomial coefficients.

Unknown polynomial coefficients are calculated on the basis of the solution of the general regression equation [1].

Verification of the significance of polynomial coefficients is usually made by means of the t-Student test [1]. In order to attain this it is necessary to calculate t statistics for each coefficient b_i of polynomial (21).

$$t(b_i) = \frac{|b_i|}{S(b_i)}, \qquad i = 0, 1, \ldots, k \tag{22}$$

where $S(b_i)$ is a standard deviation of b_i coefficient.

Verification of significance is performed by comparing the each statistic $t(b_i)$ to the assumed critical statistic t_α for significance level α.

It allows us to order investigated quantities in accordance with increasing values of their significance levels. Because values of statistics (significance level) are a quantitative measure of significance of their influence investigated output quantity it has a great importance in further merit analysis.

The assumption of a given value of significance level α allows us to distinguish investigated input quantities as:

- significant, if $t(b_i) \geq t_\alpha$ (23)
- insignificant, if $t(b_i) < t_\alpha$ (24)

The value of h-level is the next factor, which has an effect on the accuracy of the method. This criterion expresses a fact that the fuzzy output of the model should include all of the data points to certain degree h making the model as specific as possible. A choice of the value of h influences widths of coefficient a_r of the fuzzy parameters.

On the basis equations (13) – (14) variation of width of spreads of fuzzy numbers is determined as:

$$a_r^h = \frac{a_r}{1-h} \tag{25}$$

where a_r is the spread of the fuzzy parameter for h equal 0.

Dependence of parameter a_r on the value h is presented on Fig. 3.

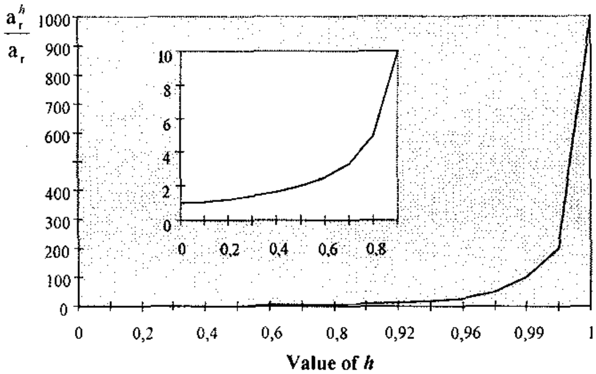


Fig. 3. The effect of value *h* on the width of fuzziness

In case of the large value of *h* level the width of spread of fuzzy models was become excessive. On account of the big width of fuzzy intervals determined in the fuzzy regression method the use of this method can disable the proper estimation process.

The right selection of the investigating object to build of the fuzzy regression model is an important factor in load estimation process. The use of data from substations belong to different classes can cause unsuitable of working out method to practical application. For that reason it is necessary to especially attentively compare all data characterizing of substations and to establish of their membership to individual class of objects. The number and type of customers supplied from substation, level of load of transformer, type of tariff of energy price can be use as helpful indexes of evaluation each object.

5. Numerical Example

To verify the proposed method of peak load estimation and description the effect different factors on its accuracy 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 January. Investigating objects are substations with transformers with 15/0.4 kV ratio of transformation and power ratio from 160 to 400 kV·A. The frequency of measurement was 15 minutes. Data of substations and supplied customers are shown in table I. On the ground of measurements the fuzzy linear regression models (3) were determined. Each model was calculated for work days. To building model of daily peak load all substations were used except substation No. 191, this was a test object. For the general fuzzy model presented in form:

$$P_{dP} = [a_{0c}, a_{0r}] + [a_{1c}, a_{1r}]\cdot A_d + [a_{2c}, a_{2r}]\cdot S_T + [a_{3c}, a_{3r}]\cdot n_o \quad (26)$$

where: P_{dp} - the daily 15-minutes peak power consumption,
A_d - the daily energy consumption,
S_T – rated power of transformer in substation,
n_o – number of customers

Table I
Data of substations and supplied customers

Substation No.	Power of transformers	Number of customers	Type of customers
191	160 kV·A	200	municipal-living + commercial-services
112	400 kV·A	45	municipal-living + commercial-services
483	250 kV·A	113	municipal-living + small hotel
853	250 kV·A	40	commercial-services
512	400 kV·A	205	municipal-living
1059	630 kV·A	2	Bank and small factory

the LP problem corresponding to the given data was formulated from (12) - (14). By solving this one, the following model was obtained:

$$P^I_{dP} = [-18.166, 0.0] + [0.0945, 0.0]\cdot A_d + [0.0196, 0.0487]\cdot S_T + [0.028, 0.0]\cdot n_o \quad (27)$$

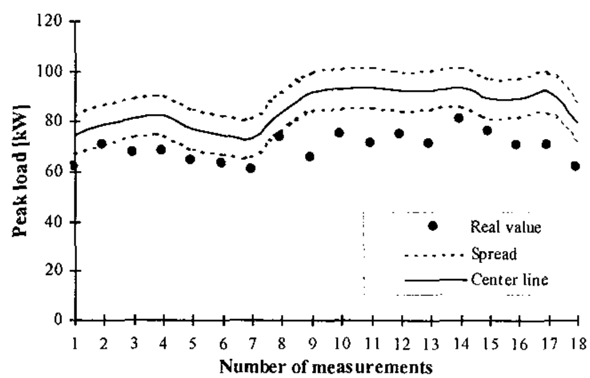


Fig. 4. Results of estimation of daily peak load for substation No. 191 on the basis model (27) $P^I_{dP} = f(A_d, S_T, n_o)$

Because substation No. 1059 does not belong to the same class of the objects, it was eliminated from the model. The model (27) was improved and assumed following form:

$$P^{II}_{dP} = [19.733, 0.0] + [0.0497, 0.0]\cdot A_d + [0.0038, 0.0217]\cdot S_T + [0.011, 0.068]\cdot n_o \quad (28)$$

On the basis of model (27) and (28) daily peak loads at the test substation No. 191 in February were estimated. The results are shown on Fig. 4 - 5 together with the corresponding measurement data.

Taking advantage of consideration presented in section IV the significance of input quantities was investigated. On the basis of the theory of experimental

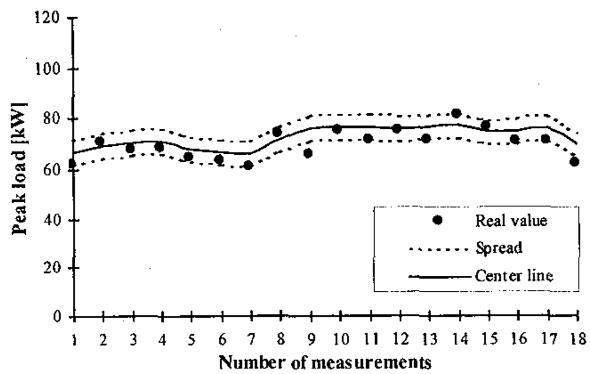


Fig. 5. Results of estimation of daily peak load for substation No. 191 on the basis model (28) $P^{II}_{dP} = f(A_d, S_T, n_o)$

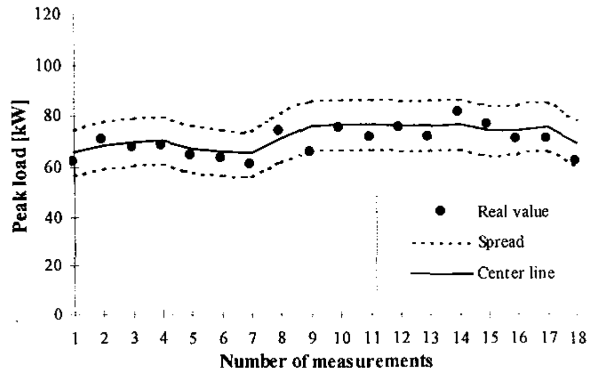


Fig. 6. Results of estimation of daily peak load for substation No. 191 on the basis model (29) $P^{III}_{dP} = f(A_d)$

design (15) - (24) significance three input quantities was carried out. The daily energy consumption - A_d, rated power of transformer in substation - S_T and number of customers - n_o was tested. Verification of significance was performed for significance level $\alpha = 0{,}05$. Results of significance test are shown in table II.

Table II
The results of significance test of input quantities

Input quantities	Polynomial (21) coefficients	Value of coefficients	Statistic t - Student test	$\alpha = 0{,}05$ t_α
	b_0	-10,9107	2,45954	2,00
A_d	b_1	0,0808	9,11275	2,00
S_T	b_2	0,0619	2,72541	2,00
n_o	b_3	-0,0576	2,15351	2,00

Consequently above test fuzzy regression model with the most important quantity (A_d) was prepared. The following form was obtained:

$$P^{III}_{dP} = [18.90, 4.5336] + [0.0529, 0.019] \cdot A_d \qquad (29)$$

Fuzzy load model was verified on the basis of measurements in the substation No. 191 in February. The use only one, the most significant input quantity (daily energy consumption) does not make results of calculation worse (Fig. 6). All models (27 – 29) presented above were solved for value of h level equals 0.

In numerical example the effect of value of h level on the width of spread of fuzzy model was compared. In test three values of h were used (h = 0, = 0.25 and = 0.5). The results of calculations for model (29) are shown on Fig. 7.

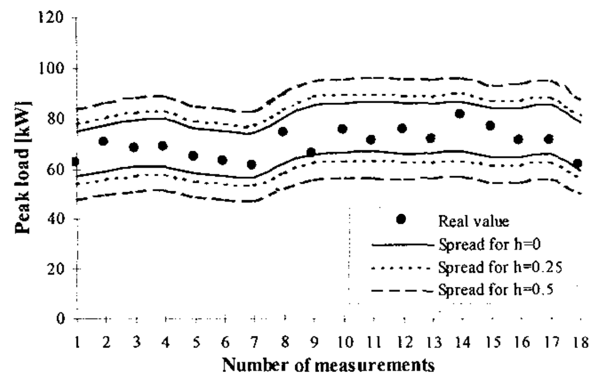


Fig. 7. Results of estimation of daily peak load for substation No. 191 on the basis model (29) for different h level

6. Conclusions

It is seen from the considerations and relationships described above that the fuzzy set approach to electrical load estimation puts a new quality into the system analysis in uncertain conditions.

The proposed method allows us to estimate daily peak power demand at distribution transformers during normal state conditions, on the basis of different input information. Results of investigations made on the basis of experimental design show that the energy consumption is the most correlated factor with the peak load demand.

The presented results (also in [8, 11]) show that right selection of investigated objects and choice of value h have a large influence accuracy of proposed method. The spread of the fuzzy models depends on maximum and minimum value of a given data and value of h coefficient.

The authors see usefulness of applying of fuzzy regression analysis to problem of load forecasting and load estimation in power distribution systems. The presented method may be a useful tool supporting planning distribution engineers.

The example is used here to illustrate the output from calculations where the fuzzy regression analysis is applied.

7. Acknowledgement

This paper reports work sponsored by State Committee for Scientific Research (KBN) under contract W/IZM/1/99.

8. References

[1] Brandt S.: *Statistical and Computational Methods in Data Analysis.* Springer Verlag, New York 1997.

[2] El-Hawary M.E.: *"Electric Power Application of Fuzzy Systems".* IEEE Press, 1998.

[3] Kujszczyk Sz.: *„Elektroenergetyczne sieci rozdzielcze".* PWN, Warszawa 1994. (In Polish)

[4] Miranda V., Pereira J., Saraiva J. T.: „Experience in State Estimation Models for Distribution Systems Including Fuzzy Measures". *Proceedings of International Symposium on Electric Power Engineering „Stockholm Power Tech",* IEEE/RIT, Stockholm, Sweden, June18-22, 1995, pp. 288-293.

[5] Nazarko J.: *„Modeling of Electrical Power Distribution Systems".* Bialystok Technical University Publisher, Bialystok, 1993.

[6] Nazarko J., Zalewski W.: „An Application of the Fuzzy Regression Analysis to the Electrical Load Estimation". *Proceedings of IEEE 8th Mediterranean Electrotechnical Conference MELECON '96 „Industrial Applications in Power Systems, Computer Science and Telecommunications",* Bari, Italy, May 13-16, 1996, pp. 1563-1566.

[7] Nazarko J., Zalewski W.: „The Fuzzy Set Approach to Substation Peak Load Estimation". *Proceedings of 31st Universities Power Engineering Conference UPEC'96,* Iraklio, Greece, September 18-20, 1996, pp. 1291-1294.

[8] Nazarko J., Zalewski W.: „The Fuzzy Regression Approach to Peak Load Estimation in Power Distribution Systems". *IEEE Transactions on Power Systems, PE-036-PWRS-0-08-1998.* (Will be published).

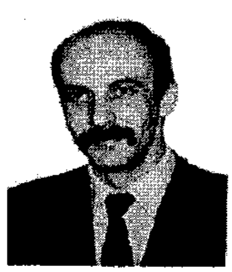
[9] Tanaka H., Uejima S., Asai K.: „Linear Regression Analysis with Fuzzy Model". *IEEE Transactions on Systems, Man and Cybernetics,* Vol. 12, No. 6, December 1982, pp. 903-906.

[10] Tanaka H., Ischibuchi H.: *„Possibilistic Regression Analysis Based on Linear Programming"* in „Fuzzy Regression Analysis" (eds. J. Kacprzyk, M. Fedrizzi), Physica-Verlag, Heidelberg, 1992, pp.47-60.

[11] Zalewski W.: „Comparison of the Fuzzy Regression Analysis and Least Squares Regression Method to the Electrical Load Estimation". *Proceedings of the IEEE 9th Mediterranean Electrotechnical Conference MELECON'98 „Regional Cooperation in Technology",* Tel-Aviv, Israel, May 18-20, 1998, pp. 207 - 211.

Biographies

Joanicjusz Nazarko was born in Michalowo, Poland. He received his Ph.D. and D.Sc. degrees in Electrical Engineering from the Warsaw University of Technology in 1983 and 1992, respectively. He is currently the Professor at the Bialystok Technical University, Poland and he holds a Chair of Business Informatics and Logistics. He is also a visiting lecturer at the Warsaw University of Technology. His research activity is centered on automation of power distribution systems with emphasis on modelling and analysis of distribution systems in uncertain conditions. He is the author over 80 papers. He is a member of IEEE, IEE and CIGRE.



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.