Estimating Peak Load of Distribution Transformers

K. Halicka * A. Jurczuk * J. Nazarko* Z.A. Styczyński** 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. This paper presents possibilities of application of the statistical approach and the fuzzy set theory to electrical load estimation. The first of them is grounded on the application of diversity factor and conversion factor. In the second of them 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.

1. Introduction

Controlling the utility load curve shape is the most prior task of load management. The principal reason of considering these activities is to achieve a better match between the current and future capacity of supplement, transmission, distribution system and a customer demand. From the utility perspective, the magnitude and timing of load changes are both important. Utilities want the greatest load reduction to occur around the time of the system peak. In other words reducing of end-user demand should be agree with distribution system peak periods.

Demand Side Management (DSM) aims to control load growth (especially peak load) and alter shape of the load curve. These actions allow to reduce capital expenditures (avoided costs), improve load factor and system efficiency.

Controlling and management of electrical energy demand needs precise modelling and forecasting of peak loads. From the power utility it is seems the most important part of applying demand side activities.

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. The mathematical estimation of the loads at the system buses seems to be most realistically possible according to incomplete primary information. It demands earlier determination of the stable relations between bus loads and easier available data.

The statistical methods could be used to load estimation in distribution transformers. These methods require a great number of load data obtained from measurement experiment.

Traditional systems used to direct the power distribution system operations are not adapted for collecting and processing a vast amount of data and do not allow a dispatcher to utilize the information effectively. In this case the most renowned method for expressing the uncertainty in load models is fuzzy set theory. For the purpose of simplicity of mathematical operation the trapezoidal and triangular forms of fuzzy numbers are usually used. The fuzzy set theory could be used as an effective tool of load estimation.

This paper presents two approaches to load estimation - statistical analysis and fuzzy regression method. These approaches are illustrated by numerical examples.

---

* Bialystok Technical University
ul. Wiejska45A, 15-950 Bialystok, Poland. E-mail: jnazarko@cksr.ac.bialystok.pl
" 1EH University of Stuttgart
Pfaffenwaldring 47, D-70569 Stuttgart, Germany. E-mail: sty@iehsun.e-technik.uni-stuttgart.de

2. Statistical approach

2.1 Problem formulation

The most popular coefficient factors used by engineers to estimate the load on the distribution power system are the diversity factor *(D_f)* and KWh-to-Peak-kW or conversion factor (Q)[1].

The diversity factor is defined to be the sum of individual non-coincident customer peak demands divided by the coincident peak demand of customer as a group.

$$D_f = \frac{\sum Individual\ Peaks}{Group\ Peaks} \tag{1}$$

The KWh-to-Peak-KW is defined to be sum of all individual customers peaks divided by the total energy usage of the group over a given period.

$$C_f = \frac{\sum Individual\ Peaks}{Group\ Energy} \tag{2}$$

Load research data provide an important input to load estimation. With the wide availability of electronic demand recorders, it is now possible for utilities to automatically gather hourly test data for a large sample of diverse classes of customers. The load Research Department of Arkansas Power and Light Company maintains demand recorders on several hundred randomly selected customers throughout Arkansas. The devices store hourly KWHR measurements for each test customer. For each test customer a total of 8760 data points are taken annually [1].

Although D_f and C_f factors have been used for many years, they are still not statistically confirmed. This work presents the statistical properties of the two factors.

2.2 Statistical properties of diversity factor

The diversity factor was calculated several times for different samples. Two such procedures, useful for evaluation of statistical independence and underlying trends, are the run test and cumulative periodogram.

The run test is defined as a sequence of identical observations that is followed and preceded by different observations or no observations at all. The run test counts the number of runs that are completely above or below the median [2].

The computational results given in [3] indicate that the samples taken for calculating monthly diversity factor for different months are independent.

The diversity factor estimated from load research data was examined for normal distribution. The average diversity for each number of customers in a group represents a random variable. The calculations applied to 60 different sets of randomly selected 2, 5, 10,15, 20,25 and 30 customers in group.

One of the most convenient non-parametric test for normality is the chi-square goodness of fit test. Another widely used goodness of fit test is Kolmogorov-Smirnov test [2].

The results of the tests confirmed that, on significance level 0.05, there was no reason to reject the hypothesis that distribution of D_f factor can be well approximated by normal distribution [3].

2.3 Statistical properties of KWh-to-Peak-kW factor

To exam if the average KWh-to-Peak-KW factor display any underlying trend, a test which analyse the residuals between the average of C_f values and the median was carried out. Figure 1 shows the relationship of the C_f factor in different number of customers with the months, which also confirm that no effect of the number of customers on the C_f factor.

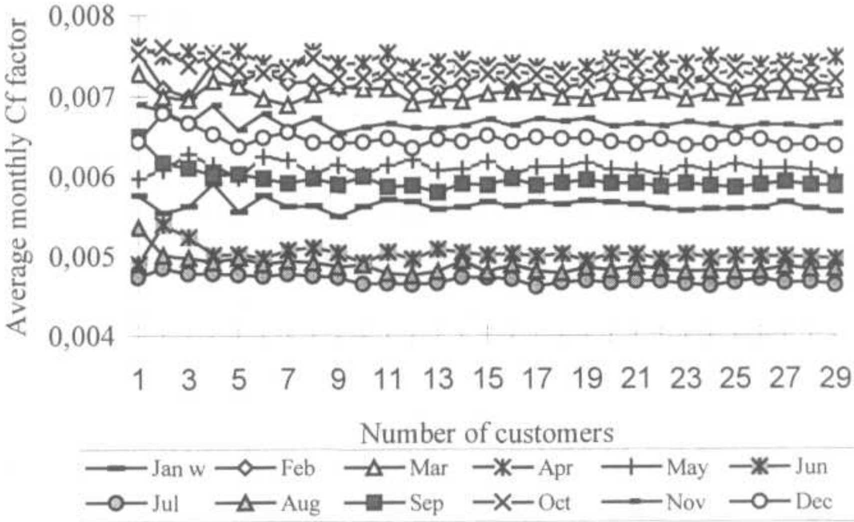

Fig. **1**. The average monthly KWh-to-Peak-KW factor versus number of customers for weekday period

The KWh-to-Peak-kW factor was calculated several times for different samples. It was assumed that the samples were independent of each other. This assumption was tested by the run test above. Computational results confirmed that the samples taken for calculating C_f factor were independent.

The shape of plots of frequency histograms of KWh-to-Peak-KW factors calculated for different groups of customers and also in monthly dimension clearly indicated that they had normal distribution (Fig. 2). This hypothesis was tested using Chi-square and Kolmogorov-Smirnov tests.

The results of the tests confirmed that, on the significant level 0.05, there was no reason to reject the hypothesis that distribution of C_f factor may be well approximated by normal distribution.

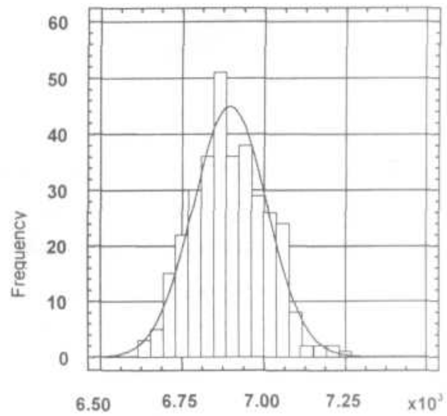

Fig. 2. Frequency histogram of the monthly diversity factor for February

2.4 Substation load estimation

In general, the only information commonly available regarding loads at locations, other than distribution substations and major equipment installation, is billing cycle customer kWh consumption. This energy data has been used to formulate estimates of loads on distribution circuits by applying diversity and KWh-to-Peak-KW conversion factors.

In order to estimate the demand of class of customer for the peak load day (either weekday or weekend) of any month it is necessary to the number of customers and the total kWh consumption of the class during the month. The peak demand may be obtained by using the following equation

$$P(m,n,k) = E \times \frac{C_f(m,n,k)}{D_f(N,m,n,k)} \tag{3}$$

where: P - peak demand,

E - kWh consumption for customer group,

C_f- KWh-to-Peak-KW conversion factor,

D_f- diversity factor,

m - month, n - type of day, k - class, N - number of customers.

Four substations were considered with the developed algorithm [1]. Percentage errors between actual and estimated peak values for all months are presented in table 1.

Table 1. Results of daily peak load estimation

Month	P_{mest}	P_{estm}	ΔP	$\lvert\Delta P_{\%}\rvert$
Jan	5.528	5.716	-0.19	3.40
Feb	5.099	6.017	-0.92	18.00
Mar	4.379	5.044	-0.67	15.19
Apr	3.176	3.165	0.01	0.35
May	4.102	4.408	-0.31	7.45
Jun	4.886	4.884	0.00	0.00
Jul	5.127	5.329	-0.20	3.94
Aug	5.300	5.976	-0.68	12.77
Sep	4.778	5.326	-0.55	11.48
Oct	3.117	3.417	-0.30	9.62
Nov	5.445	6.068	-0.62	11.44
Dec	5.724	4.957	0.77	13.40

Note: P_{mest}, - measured peak load of the peak day of month, P_{estm} - estimated peak load of the peak day of month, ΔP - estimation error, $|\Delta P_{\%}|$ - percentage errors.

3. Fuzzy regression approach

3.1 Initiation in the fuzzy regression

The general regression model is given by the following equation [4]:

$$Y = ZA + \mathbf{e} \tag{4}$$

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 (daily peak load -P_{dP}) is a real number and the second when the output value is an interval $P_{dP} \in <P_{dP}^{L}, P_{dP}^{R}>$-

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

$$\tilde{P}_{dP} = Z\,\tilde{A} \tag{5}$$

where: $$\tilde{P}_{dP}(Z_i) = \tilde{a}_0 + \tilde{a}_1 z_{i1} + ... + \tilde{a}_k z_{ik} \qquad i=1,2,\ ...,\ \mathbf{n} \tag{6}$$

The linear fuzzy regression model (6) is represented using symmetric triangular fuzzy parameters $aj=[a_{ic}, a_{ir}]$ as follows:

$$\tilde{P}_{dPi}(z_i) = [a_{0c}, a_{0r}] + [a_{1c}, a_{1r}]z_{i1} + ... + [a_{kc}, a_{kr}]z_{ik}$$
$$P_{dPic}(z_i) = a_{0c} + a_{1c}z_{i1} + ... + a_{kc}z_{ik}$$
$$P_{dPir}(z_i) = a_{0r} + a_{1r}z_{i1} + ... + a_{kr}z_{ik}$$

where: P_{dPc}, a_c - center parameters of fuzzy numbers (membership function μ = 1),
P_{dPr}, a_r - spreads of fuzzy numbers (geometrically the spread is a half of the base of the triangle).

The parameters aj of the vector $\tilde{A}$ of the linear fuzzy regression model are determined by a solution of a linear programming (LP) problem which is to minimize the sum of spreads $P_{dpr}(z_i)$ of elements of vector $\tilde{P}_{dP}$ [5].

Therefore the following LP problem is formulated.

$$c = p_{dP1r}(z_1) + p_{dP2r}(z_2) + \cdots + p_{dPnr}(z_n) \rightarrow \text{Minimum} \tag{10}$$

subject to $$p_{dPi} \in \tilde{\mathbf{P}}_{dP}(z_i), \qquad i = 1, 2, ..., n \tag{11}$$

$$a_{ir} \geq 0, \qquad i = 0, 1, 2, ..., k \tag{12}$$

From (7) - (9), the LP problem (10) - (12) can be written as follows:

$$\sum_{i=1}^{n}\left(a_{0r}+a_{1r}\left|z_{i1}\right|+\cdots+a_{kr}\left|z_{ik}\right|\right) \rightarrow \text{Minimum} \tag{13}$$

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

$$a_{0c}+\sum_{j=1}^{k}\left(a_{jc}z_{ij}\right)+a_{0r}+\sum_{j=1}^{k}\left(a_{jr}\left|z_{ij}\right|\right) \geq p_{dPi}, \quad i = 1, 2, ..., n \tag{15}$$

The parameters $a_j = [a_{jC}, a_{jr}]$ of vector $\tilde{A}$ are determined as the optimal solution of the LP problem (13) - (15). Since the LP problem always has feasible solutions, the fuzzy parameters are obtained from the LP problem for any data.

3.2 Applications 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.

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.

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. 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 [6]. 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 kWh consumption using fuzzy regression models

3.3. 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 five 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 kVA. On the ground of measurements the fuzzy linear regression models (6) were determined.

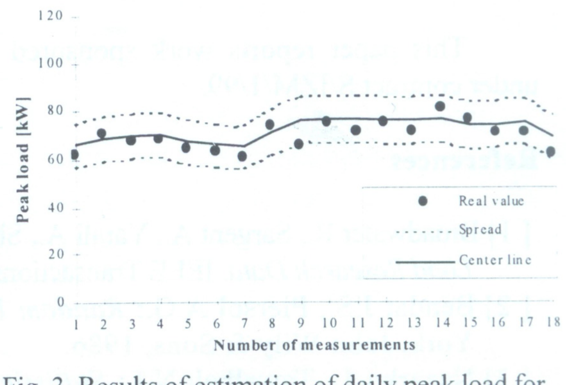

Fig. 3. Results of estimation of daily peak load for substation February on the basis of model (17)

For the general fuzzy model presented in form:

$$P_{dP} = [a_{oc}, a_{Or}] + [a_{1c}, a_{1r}]A_d$$

where: Pd_P - the daily 15-minutes peak power consumption,

A_d - the daily energy consumption,

the LP problem corresponding to the given data was formulated from (13) - (15). By solving this one, the following model was obtained:

$$P_{dP} = [18.90, 4.5336] + [0.0529, 0.019]\text{-}A_d \tag{17}$$

On the basis of model (17) daily peak loads at the tested substation in February were estimated. The results are shown on Fig. 3 together with the corresponding measurement data.

4. Conclusions

In this paper possibilities of application of the statistical approach and the fuzzy set theory to electrical load estimation was presented.

The diversity and KWh-to-Peak-KW factors have stable statistical results. The averages are normally distributed. The means of diversity and conversion factors for all calculated months have narrow confidence intervals.

Calculations and statistical analysis indicate that the value of diversity and KWh-to-Peak-KW factors depend on months, type of day and class of customer, while the number of customer in a group does not have any effective influence on evaluation of the Cj factor.

The results of investigations and calculations allow us to conclude that the diversity and conversion factors are a useful tool used in practice by electrical engineers. To obtain more general results it is preferable to collect a substantial amount of data for more wide variety of classes and large number of customers.

In case when realization of extensive measurement experiment is impossible, the use of the fuzzy approach allows to perform estimation process. The proposed fuzzy regression analysis allows us to estimate daily 15-minutes peak power demand at distribution transformers during normal state conditions, on the basis of kWh consumption. The method does not handle the instantaneous loads changes caused for example by switching operations or line outages.

The spread of the fuzzy models depends on maximum and minimum value of a given data. It does not depend on sample size. In standard regression models the width of confidence intervals depends on sample size, standard deviation and significance level. In case of small sample size, standard deviation was become excessive. The application of the fuzzy approach eliminates this difficulty.

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 authors see usefulness of applying of presented methods to problems of load forecasting and load estimation in power distribution systems. The presented approaches may be a useful tool supporting planning distribution engineers in planning and design to estimate the load of distribution networks.

Acknowledgement

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

References

[1] Broadwater R., Sargent A., Yarali A., Shaalan H., Nazarko J.: *Estimating Substation Peaks from Load Research Data.* IEEE Transactions on Power Delivery, Vol. 12, No. 1, January 1997.

[2] Bendat J.S., Piersol A.G.: *Random Data Analysis and Measurements Procedures.* New York, John Wily & Sons, 1986.

[3] Nazarko J., Tawalbeh N.I.: *Statistical Properties of Diversity and Conversion Factors and their Use to Load Estimation.* In "Distributed Energy storage for Power Systems" (eds. K. Feser, Z.A. Styczynski) Verlag Mainz, Aachen, 1998.

[4] Vaderman S.B.: *Statistics for Engineering Problem Solving.* PWS Publishing Company. Boston, 1994.

[5] 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.

[6] 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.* (To be published).