Comparison of the Fuzzy Regression Analysis and the Least Squares Regression Method to the Electrical Load Estimation

W. ZALEWSKI

Bialystok Technical University
ul. Wiejska 45 A, 15-950 Bialystok, Poland
zalewski@cksr.ac.bialystok.pl

Abstract
An essential point in correct calculations and analysis of power distribution systems is a proper evaluation of loads. The acquisition of this data is complex because of a large number of nodes and their area distribution. As a rule receiving nodes are not equipped with stationary measuring instruments so measurements of loads are performed sporadically only. The theory which enables efficient description of unreliable and inaccurate data, and relationship between them, is the fuzzy set theory. The paper presents possibilities of application of the fuzzy set theory to power distribution system calculations. Unreliable and inaccurate input data were modeling 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. The fuzzy set approach and standard regression method are compared.

1. 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, 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.

The uncertainty appears at the input, at the output and in the nature of the system itself and it is caused by:
- the random character of the customers' demand for electrical power and energy;
- the influence of the random factor on the system,
- the fuzzy character of the control function, caused by the absence of the full formalization of aims of the system and by the presence of informal intuitive procedures in the decision process,
- the fuzzy character of the sets of the feasible system states, feasible controls and input information, caused by the uncertainty of corresponding a priori information.

The use of typical load shapes and additional information on customers (energy consumption in different periods, average level of utilization of rated power of transformers, single measurements of loads) in relation to the load modeling at distribution systems buses has been previously proposed [1]. Another approach consists in utilization of periodic energy consumed by customers who are divided into classes having different load shapes. A further approach is based on the employment of operational staff (experts) evaluations [1].

In the present stage of development of power distribution systems, the mathematical estimation of the loads at the system buses seems to be most realistically possible according 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 most renowned method for expressing the uncertainty in load models is fuzzy sets theory. Two approaches, fuzzy linear regression (FLR) model and least squares linear regression (LSLR) model are compared.

2. FUZZY MODELS
Many factors such as loads and voltages at system buses, energy consumption, and parameters of equivalent circuits are included in power distribution systems calculations. This data is loaded with different errors arising from inaccuracy of measuring instruments and deficiency of measurements. The theory which enables efficient description of unreliable and inaccurate data, and the relationship between them, is fuzzy set theory.

Fuzzy numbers which are elements of fuzzy sets may be presented in several forms and different shapes of
membership function can be used for description of the investigated quantity.

For the purpose of simplicity of mathematical operation the triangular and trapezoidal form of fuzzy numbers are usually used (Fig. 1).

![Fig. 1. Triangular a) and trapezoidal b) forms of fuzzy number](image)

The subject literature introduces several ways of formation of the membership function \( \mu_A(x) \) of fuzzy number [2, 3, 4, 5]. It is often possible, on the basis of expert experience, to define the parameters of trapezoidal fuzzy number by means of linguistic variables:

- the value of \( B \) is usually between \( b_2 \) and \( b_3 \),
- the value of \( B \) is never above \( b_3 \),
- the value of \( B \) is never below \( b_2 \).

Another method for description of a membership function is an application of data on the average value \( a_i \) and maximum error \( e \) of input quantity. In this case parameters of number \( A \) can be defined as follows:

\[
\begin{align*}
A_c &= a_0 \\
A_t &= e
\end{align*}
\]  

(1)

Fuzzy numbers may be also formatted by using a mean value of a few measurements (if the number of measurements does not allow us to define parameters of variable distribution) [2]. If we have a sample \( \{x_i\} \) where \( i = 1, 2, ..., n \) (\( n \leq 30 \)) then the parameters of the fuzzy number \( B \) are:

\[
\begin{align*}
b_1 &= \min \{x_i\} \\
b_2 &= E[x_i] - \left[ E[x_i] - \min \{x_i\} \right] d \\
b_3 &= E[x_i] + \left[ \max \{x_i\} - E[x_i] \right] d \\
b_4 &= \max \{x_i\}\n\end{align*}
\]  

(2)

where: \( E[x_i] \) - mean value of \( x \) calculated from the sample, 
\( d \) - factor of deviation, describing the degree of data uncertainty.

### 3. LEAST SQUARES REGRESSION METHOD

The general regression model is given by the following equations:

\[
Y = ZA + e
\]  

(3)

where: \( Y \) - vector of output variables, 
\( Z \) - vector of independent variables, 
\( A \) - dimension of parameters, 
\( e \) - vector of unobservable errors.

For linear regression the i-th row-vector of \( Y \) is presented in form [6]:

\[
y_i(z_i) = a_0 + a_1 z_{i1} + \cdots + a_k z_{ik} + e_i = 1, 2, ..., n
\]  

(4)

The parameters \( a_i \) of the vector \( A \) are determined by solution equation:

\[
A = \left( Z^T Z \right)^{-1} Z^T Y
\]  

(5)

To calculation dependence (5) the least squares method is applied. In this method confidence intervals for the expected value of the depended variable are define by:

\[
c_a = a \pm t_{a/2, n-1} S_a
\]  

(6)

where: \( t_{a/2, n-1} \) - critical value of t-Student test, determined for the significance level \( \alpha \) and \( n-1 \) degrees of freedom, 
\( S_a \) - standard deviation of parameter \( a \).

The least squares regression method is widely used in a lot of estimation processes.

### 4. FUZZY REGRESSION METHOD

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.

Two cases can be discriminated in dependence on kind of output variable. The first when the output variable \( y \) is the real number and the second when the output value is an interval \( y \in \langle y_L, y_R \rangle \).

The first case will be presented in this section. On the basis dependence (3) it can be represented in the form:

\[
\bar{Y} = Z\bar{A}
\]  

(7)

where:

\[
\bar{y}_i(z_i) = \bar{a}_0 + \bar{a}_1 z_{i1} + \bar{a}_2 z_{i2} + \bar{a}_k z_{ik} \quad i = 1, 2, ..., n
\]  

(8)

The fuzzy linear regression model (8) is represented using symmetric triangular fuzzy parameters \( \bar{a}_i = [a_{ic}, a_{is}] \) (Fig. 1a) as follows:

\[
\bar{y}_i(z_i) = a_{ic} + a_{is} z_{i1} + \cdots + a_{is} z_{ik}
\]  

(9)

\[
y_{ic}(z_i) = a_{ic} + a_{is} z_{i1} + \cdots + a_{is} z_{ik}
\]  

(10)

\[
y_{is}(z_i) = a_{ic} + a_{is} z_{i1} + \cdots + a_{is} z_{ik}
\]  

(11)

where: \( y_c, a_c \) - center parameters of fuzzy numbers (membership function \( \mu = 1 \)), 
\( y_s, a_s \) - spreads of fuzzy numbers (geometrically the spread is half of the base of the triangular).
The parameters $\bar{\alpha}$ of the vector $\bar{A}$ of the FLR model are determined by a solution of a linear programming (LP) problem which is to minimize the sum of spreads $y_i(\alpha)$ of elements of vector $\bar{Y}$ [7, 8].

Therefore the following LP problem is formulated:

\[ c = y_1(z_1) + y_2(z_2) + \ldots + y_n(z_n) \rightarrow \text{Minimum} \quad (12) \]

subject to \[ y_i \in \bar{Y}(z_i), \quad i = 1, 2, \ldots, n \quad (13) \]

\[ a_{t\sigma} \geq 0, \quad i = 0, 1, 2, \ldots, k \quad (14) \]

From (9) - (11), the LP problem (12) - (14) can be written as follows:

\[ \sum_{i=1}^{n} \left( a_{t\sigma} + a_{t\sigma} |z_{i\sigma}| + \ldots + a_{k\sigma} |z_{k\sigma}| \right) \rightarrow \text{Minimum} \quad (15) \]

\[ a_{t\sigma} + \sum_{j=1}^{n} (a_{t\sigma} z_j) - a_{t\sigma} - \sum_{j=1}^{n} (a_{t\sigma} z_j) \leq y_1, \quad i = 1, 2, \ldots, n \quad (16) \]

\[ a_{t\sigma} + \sum_{j=1}^{n} (a_{t\sigma} z_j) + a_{t\sigma} + \sum_{j=1}^{n} (a_{t\sigma} z_j) \geq y_1, \quad i = 1, 2, \ldots, n \quad (17) \]

The parameters $\alpha = [a_{0\sigma}, a_{t\sigma}]$ of vector $\bar{A}$ are determined as the optimal solution of the LP problem (15) - (17). Since the LP problem always has feasible solutions, the fuzzy parameters are obtained from the LP problem for any data.

5. ELECTRICAL LOAD ESTIMATION

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

Planning engineers use load estimation to predict what the loads looks like 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 the special measurement project has to be considered.

In practice the only information commonly available regarding loads, other then major distribution substation is kWh consumption.

Many relationships between output quantities (peak load, load flow, losses of power and energy) and describing values coming from measurements can be represented by 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 [1].

The daily 15-minutes peak power consumption for a given substation may be found on the basis of kWh consumption using FLR model (8) or LSLR model (4) - (6).

6. NUMERICAL EXAMPLE

The measurements of daily energy consumption $A_d$ and daily peak load $P_{dp}$ at selected distribution substation in Białystok Power Distribution Utility Co. were made in April. This substation supplies municipal-living and commercial customers [9]. The sample 24-hour load profiles for one are shown on Fig. 2 and Fig. 3.
* for work days (sample size = 19)
  \[ P_{d_1} = 24.885 + 0.0461 \cdot A_{d_1} \]
  (21)
  Confidence intervals for \( \alpha \) level = 0.10
  \[ c = a_0 \pm 19.949 + (a_1 \pm 0.0212) \cdot A_{d_1} \]
  (22)
* for weekend days (sample size = 10)
  \[ P_{d_1} = -2.337 + 0.0728 \cdot A_{d_1} \]
  (23)
  Confidence intervals for \( \alpha \) level = 0.10
  \[ c = a_0 \pm 13.097 + (a_1 \pm 0.0166) \cdot A_{d_1} \]
  (24)

On the basis of models (19) - (24) daily peak loads at the same substation in May were estimated. The results are shown in Tables I and II and Fig. 4 - 9 together with the corresponding measurements data.

<table>
<thead>
<tr>
<th>Real values</th>
<th>FLR model</th>
<th>LSLR model</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P_{SP} )</td>
<td>( P_{SP}^{cont} )</td>
<td>Spread</td>
</tr>
<tr>
<td>[kW]</td>
<td>[kW]</td>
<td>[%]</td>
</tr>
<tr>
<td>80.06</td>
<td>70.53</td>
<td>11.56</td>
</tr>
<tr>
<td>72.07</td>
<td>69.80</td>
<td>11.28</td>
</tr>
<tr>
<td>68.64</td>
<td>69.07</td>
<td>11.31</td>
</tr>
<tr>
<td>67.01</td>
<td>67.51</td>
<td>11.05</td>
</tr>
<tr>
<td>76.47</td>
<td>69.98</td>
<td>11.46</td>
</tr>
<tr>
<td>70.51</td>
<td>73.89</td>
<td>12.13</td>
</tr>
<tr>
<td>66.48</td>
<td>67.63</td>
<td>11.07</td>
</tr>
<tr>
<td>62.79</td>
<td>63.05</td>
<td>10.29</td>
</tr>
<tr>
<td>59.14</td>
<td>65.55</td>
<td>10.71</td>
</tr>
<tr>
<td>71.46</td>
<td>66.95</td>
<td>10.95</td>
</tr>
<tr>
<td>62.18</td>
<td>67.17</td>
<td>10.99</td>
</tr>
<tr>
<td>74.82</td>
<td>71.74</td>
<td>11.76</td>
</tr>
<tr>
<td>75.85</td>
<td>67.52</td>
<td>11.05</td>
</tr>
<tr>
<td>74.38</td>
<td>67.12</td>
<td>10.98</td>
</tr>
</tbody>
</table>

Fig. 4. FLR model \( P_{d_1} = f(A_{d_1}) \) for work days

Fig. 5. LSLR model \( P_{d_1} = f(A_{d_1}) \) for work days

Fig. 6. Errors of estimation for the both models for work days

Table I

RESULTS OF PEAK LOAD ESTIMATION FOR WORK DAYS
7. 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. Owing to adequacy and simplicity it can be easily adapted to practical algorithms based on available data which may contain different types of errors.

The proposed method of application of the FLR models allows us to estimate daily 15-minutes peak power demand at distribution transformers during normal state conditions, on the basis of kWh consumption.

Average errors in fuzzy regression analysis and standard linear regression are similar. In case of small sample size standard deviation were become excessive. The spread of the fuzzy models depends on max and min value of a given data. It does not depend on sample size. In LSLR models the width of confidence intervals depends on sample size, standard deviation and significance level.

On account of large width of confidence intervals determined in standard regression method the use of this method often disable the proper estimation process. The application of the fuzzy approach eliminates this difficulty.

TABLE II

RESULTS OF PEAK LOAD ESTIMATION FOR WEEKEND DAYS

<table>
<thead>
<tr>
<th>Real values</th>
<th>FLR model</th>
<th>LSLR model</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_{0p} [kW]</td>
<td>P_{0p,sim} [kW]</td>
<td>Spread [%]</td>
</tr>
<tr>
<td>P_{0p} [kW]</td>
<td>Con. interv. [kW]</td>
<td></td>
</tr>
<tr>
<td>47.23</td>
<td>46.34</td>
<td>2.70</td>
</tr>
<tr>
<td>53.58</td>
<td>56.94</td>
<td>3.28</td>
</tr>
<tr>
<td>71.38</td>
<td>73.70</td>
<td>4.19</td>
</tr>
<tr>
<td>51.75</td>
<td>55.86</td>
<td>3.22</td>
</tr>
<tr>
<td>50.22</td>
<td>48.63</td>
<td>2.83</td>
</tr>
<tr>
<td>63.6</td>
<td>61.22</td>
<td>3.51</td>
</tr>
<tr>
<td>50.25</td>
<td>47.64</td>
<td>2.77</td>
</tr>
</tbody>
</table>

Author has seen usefulness of applying of fuzzy regression method to problems of load forecasting and load estimation in power distribution systems.

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

Acknowledgments

This paper reports work sponsored by State Committee for Scientific Research (KBN) under contract W/WE/98. I wish to thank prof. J. Nazarko for valuable comments on this paper.

References