ABSTRACT – The paper deals with load profiles of electrical consumers of distribution networks (low voltage busbars of distribution transformers). These profiles are the most important data not only for performing off-line network analyses (what-if analysis, operation and development planning, etc.), but also for providing satisfactory results of running of all real-time functions (State Estimation, Power Flow, Fault Management, etc.). The load profiles are treated as random variables. The following two proofs are the main objective of the paper: 1) The load profiles with the necessary accuracy can be derived on the basis of a relatively small number of experiments (in-field measurements) and 2) The normal distribution is the best guess for the statistical behaviour of the most number of loads.

1. INTRODUCTION

By restructuring the electric power industry, distribution power utilities (DPU) have become independent utilities, faced with the laws of the electricity market. This forces them to increase the efficiency of their business. Already established SCADA systems, which usually cover the supply substations and small parts of medium voltage (MV) network, are not sufficient to achieve the desired efficiency in business from the technical point of view. Therefore, SCADA systems are being more and more integrated into distribution management systems (DMS) during the last ten years. The quality of their application depends on the quality and the quantity of both real-time data provided by SCADA system and historical data. Among the historical data, the most necessary but also the least reliable are certainly the data about electrical loads – daily load profiles (DLPs). Considering the fact that a DPU usually has from several to several tens (even hundreds) thousand of loads, it is clear that the measurement of DLPs of all loads is practically undosable, that is, economically unjustifiable. Therefore, in order to provide the image about loads with sufficient quality, the measurements of a relatively small number of characteristic representatives of loads are used (e.g. the loads of distribution transformers). Since the DLP is treated as random variable, for every measurement it is necessary to define not only the load type but also the statistical distribution which best describes it. The references [1,2] stress that the statistical distribution of loads is certainly not of Normal type. In accordance with this, Normal, Lognormal, Beta and other standard distributions are considered. Their qualities, goodness of fit, are quantified on the basis of the standard statistical tests: Kolmogorov-Smirnov, Anderson-Darling and Chi-square ($\chi^2$) [3].

The basic ideas and used statistical tests are presented in Section 2. In Section 3, the results of numerical studies on the example of the DPU of ELEKTROKRAJINA – Banja Luka, Republic of Srpska, Bosnia and Herzegovina, are presented. The conclusion and references are given in Sections 4 and 5, respectively.

2. DAILY LOAD PROFILES

The fact is that the simultaneous and long-term measurements of all loads [e.g. the loads of all distribution transformers low voltage (LV) sides] is the most quality way to achieve their DLPs. But, at the same time, it is clear that in classic DPU the measurement of DLPs of all loads is practically undosable, that is, economically unjustified. Yet, any DPU, even the one without applied DMS, can provide the necessary DLPs for all characteristic days, for all characteristic days and periods (seasons), with a relatively small number of measurements, in a really short time.

2.1 Characteristic load

A group of loads with DLPs which are mutually "proportional", i.e. the DLP of any load can be derived from the DLP of another load from the group by multiplication with a constant, are termed similar loads. The normalized DLP (NDLP) of a load equals to its DLP divided with a selected quantity indicator (e.g. peak value, or the supplied daily, monthly or annual electric energy to the considered load [1]). Similar loads have only four NDLPs: one for the active powers, one for the reactive powers, one for the current magnitudes and one for their power factors (equal to DLP).

Thus, all similar loads can be described with a single quadruplet of NDLPs – active and reactive power, current and power factor. Each of the following two pairs of NDLPs represents sufficiently well the considered similar loads: (active and reactive power) or (current and power factor). These NDLPs, associated with the unity quantitative indicators, represent the characteristic load. Characteristic loads are usually associated with household, commercial, industry, etc. For better presentation of loads, each class can be divided in sub-classes (e.g. households with locally or remotely provided heating, etc.).

2.2 Characteristic day and season

It is assumed that the same DLP of a load appears in every day of the "same type" – e.g. for all weekdays of a season. As noted above, the characteristic load can be represented sufficiently well with one of two noted pairs of NDLPs: (active and reactive power) or (current and power factor). The day type associated with the selected pairs of NDLPs is termed characteristic day. The set of characteristic days usually consists of weekday, the first and the second day of the weekend (Saturday and Sunday) and holidays. The period in which the quadruplets of NDLPs, for each characteristic day sufficiently well represent the loads is termed season or characteristic period. The set of characteristic periods usually consists of spring, summer, autumn and winter.

The load model based on NDLPs with corresponding quantitative indicators associated with all network loads, for all characteristic days and periods, has significant advantages over the load model based on individual DLP of all loads. These advantages consist of the significant decrease in the amount of necessary data for network
load representation and the speed of their processing, all without significant decrease in the quality of DMS power calculations.

2.3 Load as a random variable

The results of fairly detailed researches presented in [1], based on the bottom-up approach, prove that individual loads are not ideally Gaussian normally distributed (GND). The analysis of loads showed a left skewness of histograms in particular hours. Thus, it was concluded there, that customers load is better fitted with LND than GND and even better with modified LND. In addition, reference [2] stresses that GND has poor goodness of fit and proclaimed the Beta distribution as the distribution of load in particular hours. On the other side, on the basis of the consideration about the application of confidence intervals and pseudo-Gaussian robustness of data, reference [4] stresses that there is a very small number of situations when one has to abandon the GND confidence intervals. One of these situations refers to the loads which behave in mixed discrete-continuous way. Then, the aggregate of the loads has a multi-modal character and tends to a GND, only slowly. For example, if there is a sufficiently large probability of a load to be switched-off and (1-\(p\)) to be switched-on, the uniform distribution describes the load in the best way.

The disagreement in the description of the load as a random variable represents the first reason why, in this research, the measurement data from LV sides of distribution transformers (DTs) are submitted to statistical tests to check the type of the distribution of the DTs loads. The second reason is the well known Central Limit Theorem which claims that for a sufficiently large sample size the probability distribution for the sum of independent variables, \(X\) with standard deviations \(\sigma\) approximates a GND [4]. Since the DT’s load present the sum of many individual loads, it could be expected that the load as random variable would be best described with GND.

2.2 Statistical tests

Standard statistical (non-parametric) tests used in this paper are: Pearson’s chart, Kolmogorov-Smirnov, Anderson-Darling and \(\chi^2\).

**Pearson’s chart (PC)** [5]: The experience has shown that the shapes of unimodal distributions for a sample \((x_1, x_2, \ldots, x_n)\) can be described by parameters \(\beta_1\) (called skewness) and \(\beta_2\) (curtosis):

\[
\beta_1 = \mu_3 / \mu_2^\frac{3}{2}, \quad \beta_2 = \mu_4 / \mu_2^2, \quad (1a)
\]

\[
m_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k, \quad \mu_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^k. \quad (1b)
\]

The point defined with parameters \(\beta_1\) and \(\beta_2\) is drawn at PC and one observes its position in reference to the zones (points, lines, regions) which theoretically correspond to known distributions. This is a very simple test that gives directions about the type of random variable.

**Kolmogorov-Smirnov (KS)** [6]: This test uses null hypothesis that distribution function of a sample is \(F(x)\). The used statistics says:

\[
D_n = \max \{D_n^+, D_n^-, \}, \quad (2)
\]

\[
D_n^+ = \sup_{x \in R} \{S_n(x) - F_0(x)\}, \quad D_n^- = \sup_{x \in R} \{F_0(x) - S_n(x)\}. \quad (3)
\]

The KS test is given for distributions with all specified parameters.

But, using the table from [7], with unknown parameters calculated from a sample, critical values for the statistics when \(F_0(x)\) is GND are obtained. In order to use the same test for LND, the logarithm of the sample is being derived first and then proceed as for the GND.

**Anderson-Darling (AD)** [8]: This test is one of the most powerful and important goodness of fit tests in the statistical literature. It is more sensitive to deviations in the tails of the distribution than KS test. The null hypothesis is that the empirical distribution fits to GND. The AD statistics is [3]:

\[
A^2 = -\frac{n}{n-0.5} \left(\sum_{i=1}^{n} \frac{2i-1}{n} \ln z_i + \ln(1-n_{n+1-i})\right) - n. \quad (4)
\]

where \(n\) denotes the sample size and \(z_i\) is the \(i\)-th value of sorted sample. The critical values for the modified AD statistics [3]:

\[
A_m^2 = A^2 \left(1 + \frac{0.75}{n} + \frac{2.25}{n^2}\right). \quad (5)
\]

amounts to 0.752 for significance level \(\alpha=0.05\) and 1.035 for \(\alpha=0.01\). The tests are made for normality. As the logarithm of a LND gives a GND, AD test can be applied for testing both normality and log normality of data.

**Chi square (\(\chi^2\))** [3]: To test the hypothesis that a random sample \(X_1, \ldots, X_n\) has the distribution function \(F_0(x)\), the range of \(X_j\) is partitioned into \(r\) cells, say \(E_1, \ldots, E_r\). Let \(n_1, \ldots, n_r\) be the observed number of \(X_i\) in these cells and \(p_1, \ldots, p_r\) be the expected number of \(X_i\) in these cells. The statistic is:

\[
\chi^2 = \sum_{i=1}^{r} \frac{(n_i - n p_i)^2}{n p_i}. \quad (6)
\]

The value of the \(\chi^2\) statistics is compared to the critical values which are obtained from tables as \(\chi^2_{r-s-1,\alpha}\), with \(r-s-1\) called degree of freedom, where \(s\) is the number of parameters calculated from the sample (e.g. 2 for normal and log-normal, 3 for beta distribution, etc). This test is very general. It does not depend on whether \(F\) is univariate or multivariate, discrete or continuous. However, it is not as powerful as AD and KS tests. If the number of cells is fixed for testing different distributions, the values for \(\chi^2\) statistics can tell which distribution fits "better".

3. NUMERICAL STUDIES

The considered part of the DN of the PDU ZP "Elektrokrainja" consists of the area of the supply substation "Banja Luka 3", 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA. The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA. The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA.

The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA. The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA. The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA. The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA. The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA. The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA. The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA. The transformers supply five 20 kV and eleven 10 kV feeders with 110/20/10 kV/kV, with two supply transformers, 40 and 20 MVA.
13), 2) individual households with local heating (5 out of 13) and 3) collective residence with remote heating (4 out of 13). The considered period involves three characteristic seasons: winter, spring and summer. Each season consists of three characteristic days: weekday, weekend and holiday.

2) Data filtering: This process refers to rejecting or correction of data which result from network topology changes, switching on/off large loads, measurement instruments malfunction, exceptional events etc.

3) NDLPs establishment: Each of the measured DLP is divided with the corresponding (daily) energy. This step provides mutual comparison of load profiles, their simultaneous processing and establishment of NDLP representatives.

4) NDLP representatives establishment: These representatives are established for each measurement location and each characteristic load. The procedure is repeated for all characteristic days and seasons. These representatives are associated with their expected values and dispersions.

5) Data verification: By comparison of representatives of measurements which belong to a group of similar loads, with a representative of its characteristic load, this procedure enables the mistakes made in the previous steps to be stressed. Fig. 1 presents the mean values of the similar loads representatives for summer weekday. These values are compared mutually and also with the mean values of characteristic loads representatives.

This figure proves that fair quality load representatives can be established with a relatively small number of measurements, but with the right selection of locations of these measurements. For the considered example, the average sums of square distances of representatives of similar to characteristic loads are presented in Table 2. E.g., for weekday/summer, similar loads of type 2 have better fit with characteristic load (0.147) than type 3 (0.228), and especially type 1 (3.620).

Table 2 – Average distance of similar to characteristic loads

<table>
<thead>
<tr>
<th>type</th>
<th>winter</th>
<th>spring</th>
<th>summer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.655</td>
<td>0.274</td>
<td>0.235</td>
</tr>
<tr>
<td>2</td>
<td>0.153</td>
<td>0.201</td>
<td>0.147</td>
</tr>
<tr>
<td>3</td>
<td>0.287</td>
<td>0.211</td>
<td>0.228</td>
</tr>
</tbody>
</table>

Statistical tests: The testing process of the load as random variable is presented for a load of type 2, for a summer weekday. Mean values (blue line) are shown in Fig. 2. The range of 99.87% of the sample (3σ) is emphasised in both ways – beside the mean value and separately for the purpose of stressing the dynamics of its change (red line). The values of KS, AD and χ² statistics are shown in Fig. 3, 4 and 5, respectively. Tests are performed for GND and LND, for critical value (cv) 1% and 5%. Hourly load's skewness and curtosis are plotted in the PC, Fig. 6. The colour of a circle defines its 15-minute time as seen on the 12-o'clock legend on the right.
The presented results imply:

- The conclusions from AD test are almost the same as the results from KS test.
- The tests claim that in certain moments of the day the sample is neither GND nor LND (moments when the blue and green curves cross red lines that represent the critical values).
- In tests where null hypothesis for GND fails, also fails the null hypothesis for LND (since the logarithm function for values near the unity are similar to the linear function: $\ln x \approx 1 - x$, $x \to 1$, and linear functions keep normality). Guided by this fact and the fact that logarithmed original data almost never conforms to the normal distribution, it can be concluded that LND is worse guess than GND.
- Moments during the day when data does not conform to the GND are in the periods of day when load changes rapidly. In those periods, all three KS, AD and $\chi^2$ tests have the highest values. In those periods, the standard deviation of the sample has the highest values, as seen in Figure 3 (bottom red line).

4. CONCLUSION

The paper proves that the necessary and satisfactory DLPs for all characteristic loads, for all characteristic days and periods (seasons) can be derived from a relatively small number of measurements. Also, on the basis of Anderson Darling’s and other statistical tests, the paper confirms that the randomness of the obtained load profile is not distributed ideally in normal or in lognormal way, but generally, the normal distribution is the best guess for the most number of cases.

5. REFERENCES


Acknowledgments

The results of the researches presented in this paper are derived from the frames of the project "DMS system", which is financed by DPU Elektrokrajina – Banja Luka, Republic of Srpska, Bosnia and Herzegovina, for its own purposes. It was implemented in cooperation with DMS GROUP LLC, for Power Engineering Novi Sad and Faculty of Technical Sciences Novi Sad, Serbia, in accordance with the Protocol No. 01-12-PP/08, dated March 19, 2008.