FACTA UNIVERSITATIS
Series: **Electronics and Energetics** Vol. 32, N° 3, September 2019, pp. 403-416
https://doi.org/10.2298/FUEE1903403S

ENERGY LOSSES ESTIMATION BY POLYNOMIAL FITTING AND K-MEANS CLUSTERING *

Lazar Sladojević, Aleksandar Janjić

University of Niš, Faculty of Electronic Engineering, Niš, Serbia

Abstract. *This paper represents an approach for the estimation and forecast of losses in a distribution power grid from data which are normally collected by the grid operator. The proposed approach utilizes the least squares optimization method in order to calculate the coefficients needed for estimation of losses. Besides optimization, a machine learning technique is introduced for clustering of coefficients into several seasons. The amount of data used in calculations is very large due to the fact that electrical energy injected in distribution grid is measured every fifteen minutes. Therefore, this approach is classified as the big data analysis. The used data sets are available in the Serbian distribution grid operator's report for the year 2017. Obtained results are fairly accurate and can be used for losses classification as well as future losses estimation.*

Key words: grid losses, least squares optimization, big data, clustering

1. INTRODUCTION

Big data analysis is rapidly becoming one of the most important tools in many aspects of engineering. Data are collected everywhere, and their numbers and collection rates are increasing each day. Therefore, various methods for processing of this data have been developed in recent years. These methods are efficient not only for extracting valuable information from a mass of data and their visualization, but also for developing predictive models for various applications.

Increasingly high amount of data can also be observed in a field of electrical power engineering. Electrical power grid is being modernized faster than ever, with large number of smart sensors being installed in many points of the grid. These sensors collect information about various electrical variables which are important for normal grid

Received February 26, 2019; received in revised form May 19, 2019
Corresponding author: Lazar Sladojević
University of Niš, Faculty of Electronic Engineering, Aleksandra Medvedeva 14, 18106 Niš, Serbia
(E-mail: lazar.sladojevic@elfak.rs)
* An earlier version of this paper was presented at the 4th Virtual International Conference on Science, Technology and Management in Energy, eNergetics 2018, October 25-26, 2018 [1].

operation. These data are used everywhere, from the power generation side management to the demand side management. A good overview of many different applications of big data in electrical energy management and most common methods for data processing can be found in [2].

One of the most important usage of big data is prediction of solar and wind power generation based on collected weather data. Weather has a major impact on production from renewable sources, and therefore it is very important to observe the relationship between the two.

Another, very interesting application of big data is detection of different consumption profiles based on measurements of different variables. Example would be [3], where the authors have used hourly electricity consumption readings and external temperature measurements to compute consumption profiles for residential customers.

Electrical faults in power grid can present a big problem, especially when the fault occurs on a geographically distant part of a network. Fault detection, identification and location [4] can also be obtained from the data collected in various measurement points in the grid.

Another big problem for electrical energy suppliers is energy theft. Theft of electrical energy can in some places reach astonishingly high values. Therefore, an approach for estimating the amount of stolen electrical energy based on smart meter data and least squares method for data processing has been developed and proposed in [5]. This topic is closely related to this paper, since the energy theft is observed as a non-technical loss which is also evaluated here.

Prediction of losses in distribution network has drawn more attention during the last years due to the deregulated energy market conditions, where distribution network operators are obliged to procure the energy for covering losses on the open electricity market [6, 7]. Increased market deregulation [7, 8] and shares of renewable and intermittent energy generation [9], makes line loss prediction difficult. Line losses themselves are also influenced by a multitude of factors and non-linear correlations which makes predictions model even more complicated.

Design of line loss prediction models have become a research priority in transmission networks as well and several different models have been proposed [10, 11, 12]. Losses are allocated using the quadratic expression in [10], formulated explicitly in terms of all the transactions in the system [11] with consideration of wind generation and varying loads in [12]. However, a majority of these models are mainly designed for line loss allocation issues for market applications as opposed to day-ahead predictions for TSO purposes.

In distribution networks, different methods for the calculation of losses are used, including heuristic algorithms [13] and neural networks [14]. The near quadratic relationship that exists between load and loss has been used to develop empirical relations for estimation of loss [15]. These relationships relate either the loss and load factors [16] or the loss and load [17, 18]. In these methods, using simplified feeder models for computation of the loss, the coefficients in the quadratic function are determined using a curve fitting approach.

Although the previous research and studies established satisfactory models for the energy losses calculation, they didn’t treat the process of the open market losses procurement. The contribution of this paper is therefore the seasonal classification and determination of curve fitting parameters for the purchase of energy losses. Least squares optimization method which is used in this paper is very similar to machine learning regression algorithms, but with certain restrictions attached to it. It is suitable for numeric data with linear or quadratic relationships of measured (input) and estimated (output) values. Least squares regression is a so called “supervised” learning algorithm, which will be described further in section 3. However, for experimental purposes, another machine learning algorithm was used for model improvement. This is the so-called clustering algorithm, which belongs to the “unsupervised” learning category. Even though these algorithms are used in various applications in electrical power engineering, to the authors knowledge, this is their first application in estimation of losses.

2. Problem Description

In this paper, an approach for estimation of losses in distribution grid based on the available data analysis is proposed. These losses comprise of two components, namely technical losses (TL) and non-technical losses (NTL). The proposed approach is based on analysis of losses data collected in year 2017. This data will be used to estimate parameters of a predictive model for future losses estimation.

2.1. Physical Interpretation

Technical losses can be split into two terms. The first term represents the constant losses. These mainly represent the losses in magnetic cores of distribution transformers, but other factors, such as losses due to corona, constantly operating measurement equipment, leakage currents and losses in dielectrics also contribute.

The other term is variable losses. They appear mainly in conductors but a small part of these losses can also be observed in other current carrying parts, such as switch contact resistances and busbars. These losses are proportional to the square of current or, equivalently, to the square of active power.

Non-technical or commercial losses appear due to infrequent or bad reading of measurement equipment and electrical power thefts. Therefore, these losses are proportional to the active power.

In distribution power grid of Serbia, electrical energy received from the transmission grid is measured every fifteen minutes throughout whole year. On the other hand, energy supplied to end users is measured once every month. Total losses represent the difference between total energy received from the transmission grid and total energy supplied to end users during one month. These data are collected and can be used for future losses estimation and losses classification. One method that allows this kind of estimation is described in the following section.

2.2. Mathematical Model

There are several ways to model the losses in the distribution grid, but they all need information about energy obtained from the transmission grid and distributed sources and energy delivered to end users. The model chosen here represents the losses in the following polynomial form [19, 20]:

$$\Delta W_{c,j} = \sum_i (a_j + b_j \cdot P_i + c_j \cdot P_i^2) \cdot \Delta t_i \tag{1}$$

where $\Delta W_{c,j}$ are the calculated (estimated) total losses for month *j*, *i* is the index of fifteen-minute interval Δt_i in month *j*, P_i is an average input power for the that interval, a_j represents the amount of constant losses in month *j*, b_j is the coefficient associated with the commercial losses and is proportional to input power P_i in month *j*, and finally c_j is the variable losses coefficient for month *j*, proportional to the square of power. For now, coefficients *a* and *c* are considered constant throughout the whole year, while the coefficient *b* varies by month. This assumption will be addressed in the following chapter.

On the other hand, measured losses, denoted as $\Delta W_{m,j}$ are already available as a difference between measured input and measured output energy.

For calculation of coefficients *a*, *b* and *c*, least squares method was used, which means that the sum of squared differences between the calculated and measured values of losses was minimized. Total number of variables is 36 (twelve for constant losses – coefficients a_j, twelve for commercial losses for each month – coefficients b_j and twelve for variable losses – coefficient c_j). Some of these variables are considered constant in some calculation, so the effective number of variables is lower. General objective function for minimization can now be written as:

$$\min F(a,b,c) = \sum_{j=1}^{12} (\Delta W_{c,j} - \Delta W_{m,j})^2 = \sum_{j=1}^{12} \left(\sum_i (a_j + b_j \cdot P_i + c_j \cdot P_i^2) \cdot \Delta t_i - \Delta W_{m,j} \right)^2 \tag{2}$$

Coefficient values are constrained to a certain range: for coefficient a_j: $a_{min} \le a_j \le a_{max}$, for coefficients b_j: $b_{min} \le b_j \le b_{max}$ and for coefficient c_j: $c_{min} \le c_j \le c_{max}$. Constraints for coefficients *a*, *b* and *c* have to be properly selected, based on their physical interpretation explained in the following chapter.

2.3. Restrictions of the Proposed Method

Proposed method has one drawback, it uses the monthly readings of energy consumption. This means that the value of measured losses is prone to errors due to bad or untimely readings. For example, in some rural areas, electricity consumption is read only every three months. This leads to slight under readings for certain months and slight over readings for others and, consecutively, to miscalculation of some parameters.

Better results would be obtained if the consumption was read with higher frequency, preferably the same as the input readings. This would require large number of smart meters installed at every point in the grid, which is not yet realized in practice. However, smart meters are being installed every day and, in the future, more reliable and accurate data will be available for analysis.

3. Mathematical Model Solving Method

As it was mentioned earlier the mathematical problem given with (2) is solved using the least square optimization algorithm. This problem is very similar to a linear regression problem which exists in the field of machine learning. In linear regression, output values are a linear combination of constant parameters and measured values of features (predictors). From (1) it can be observed that in this case, output values would be monthly measured values of losses, $\Delta W_{c,j}$, features would be Δt_i, $P_i \cdot \Delta t_i$ and $P_i^2 \cdot \Delta t_i$ and their coefficients would be *a*, *b* and *c*, respectively. However, since restrictions have been imposed on all parameters and some of them are also variable, this problem was reformulated into non-linear programming optimization problem, given with (2).

For solving of this problem, standard mathematical methods were used, in this case the interior point algorithm. Input parameters are the fifteen-minute readings of electrical energy injected from the transmission grid and distributed sources into distribution grid and the monthly measured values of losses in the distribution system. Output consists of the values of coefficients *a*, *b* and *c*.

For easy programing and formulation of problem, an open source optimization platform Yalmip [21] was used. Yalmip's syntax allows easy and intuitive definition of variables, objective function, constraints and other options. Yalmip was used with one of Matlab's integrated solvers for performing computations. It can select solver for a problem automatically, based on its structure, but also permits users to select the solver they think it fits best. This allows all kinds of problems to be defined in the same way, unlike the case of using each solver individually, where user would have to define the problem in a form specific to that particular solver.

The solver used for calculation of coefficients is Matlab's *fmincon* nonlinear programming solver. This solver utilizes several different algorithms for objective function minimization, but the one used here was the interior point algorithm [22].

Variables involved in calculations are already denoted a_j, b_j and c_j, with j = 1...12. Objective function is given with (2).

Constraints are chosen based on real data, and the realistic values of coefficients. The total nominal iron core losses power of all transformers in the distribution system of Serbia is approximately 32 MW. Therefore, the parameter *a* constriction adopted is $30 \leq a \leq 40$. Commercial losses always exist, but they do not exceed 10 % in Serbian distribution grid. Thus, adopted constraint for parameter *b* is $0.01 \leq b \leq 0.1$. Unlike the previous parameters whose extreme values are relatively easy to estimate, parameter *c* cannot be constrained in such a straight-forward manner. Since it is multiplied by a square of power, its value is undoubtedly very small. Based on author's previous experience, adopted constriction for this parameter is $0.00002 \leq c \leq 0.00006$. As it was mentioned in the previous section, for this calculation the coefficients *a* and *c* are considered constant, while the coefficient *b* varies by month. This is not to be confused with constraints which are simply the minimum and maximum sensible values that can be obtained as calculation results. Since the grid topology and number of transformers remains very much the same throughout the whole year, it makes sense to keep coefficients *a* and *c* constant. On the other hand, coefficient *b* is affected by many external factors and therefore it is considered variable. All of these assumptions are used for the first calculation.

4. Numerical Results

Obtained results are presented in Table 1 and Fig. 1. Table 1 contains the real values of coefficients, while the values of coefficients b_j and *c* in Fig. 1 are scaled. The scaling is used only to make all values visible on a chart. After the scaling, coefficients b_j are shown in percent, while the coefficient *c* is now dimensionally equal to W^{-1}. Fig. 2 represents the comparison between calculated and measured values of losses.

Table 1 Calculated values for coefficients

Month	Coefficient *a* (MW)	Coefficient *b* (pu)	Coefficient *c* (MW^{-1})
January	32.739	0.091499	0.000020609
February	32.739	0.072052	0.000020609
March	32.739	0.061105	0.000020609
April	32.739	0.039002	0.000020609
May	32.739	0.057862	0.000020609
June	32.739	0.013125	0.000020609
July	32.739	0.019671	0.000020609
August	32.739	0.014427	0.000020609
September	32.739	0.018085	0.000020609
October	32.739	0.043728	0.000020609
November	32.739	0.051645	0.000020609
December	32.739	0.065777	0.000020609

It can be observed from Fig. 2 that the computations were done successfully. Calculated and measured losses are equal which means that the coefficients are well estimated. Statistically speaking, it can be said that the input data, defined in the previous chapter are the training data for the model (1).

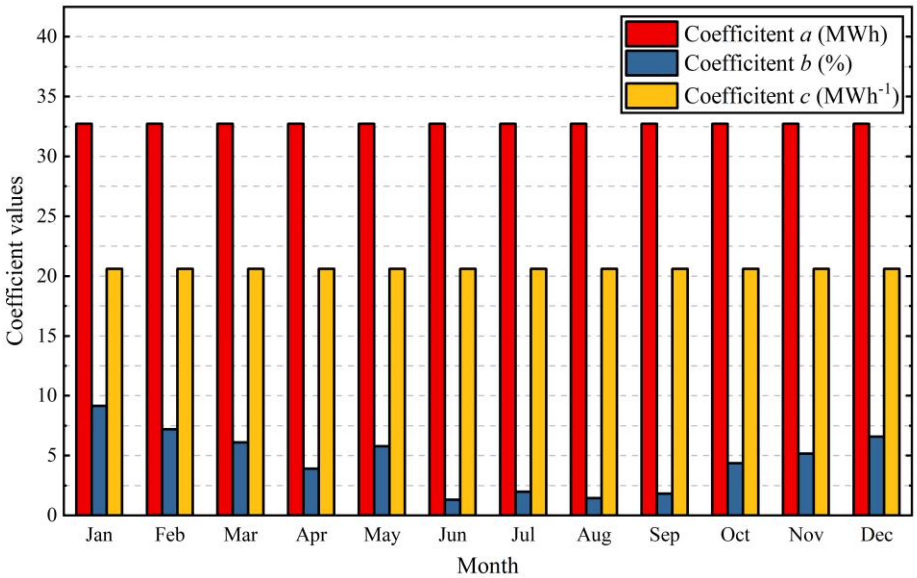

Fig. 1 Scaled values of the calculated coefficients

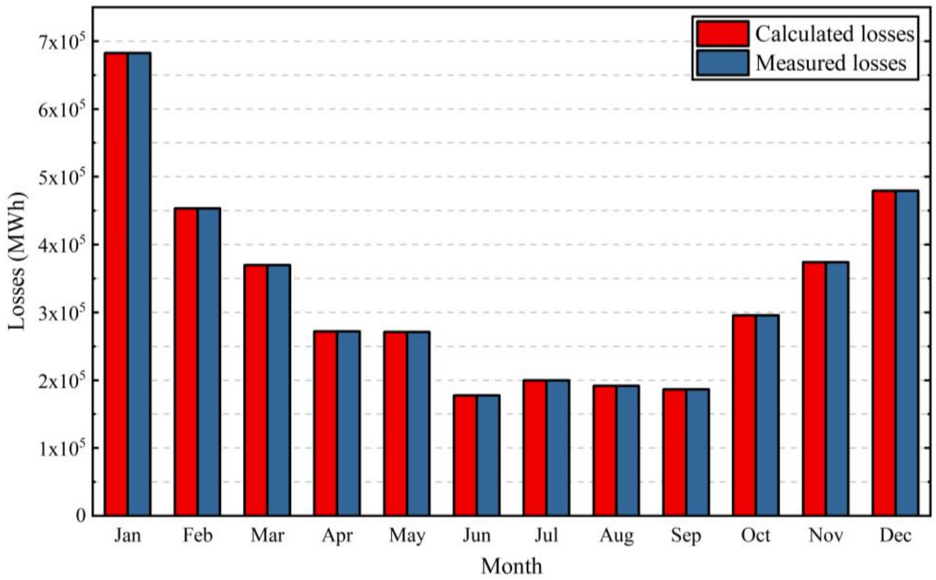

Fig. 2 Comparison between calculated and measured losses

From Fig. 2 it is obvious that the model fits the training data well. Now these coefficients can be used for future estimation of losses under the previously introduced assumptions.

5. Model Improvement

Distribution companies are procuring the energy for covering losses on the open electricity market. In order to simplify the procurement procedure, the whole concept can be extended by grouping months into several distinct seasons. The advantage of distinction is grouping of coefficients by season instead of having different coefficients for each month and potentially better prediction accuracy in the certain season. The accuracy however depends on quality of the collected data. Grouped coefficients are less prone to various stochastic errors than individual coefficients. Besides that, there are noticeable differences of the climate factors, such as external temperature and humidity in the same months during the years. Even the reading of electrical energy consumption is not as frequent in winter as it is at summer. On the other hand, adopting one set of coefficients for the whole year would cause too high error values in future calculation of losses. Therefore, a three-season model was adopted, and months were clustered into winter, summer and “transient” seasons.

Clustering itself is an “unsupervised” machine learning algorithm, which means that, unlike regression, it doesn’t have the output data to compare inputs to. Instead, it seeks to find similarities among the input measurements. In this case, there are twelve input points - twelve months and three features – coefficients a, b and c. This means that there is a total of 36 input values among which a clustering algorithm should find similarities. A twelve by three matrix was used for convenient storage of these data. In this particular case, that matrix is given in tabular form with Table 2, where rows represent input measurements for each month and columns represent features.

The number of clusters is another input to the clustering algorithm. As explained earlier, the number of clusters was chosen to be three, since it was utmost logical to divide year into three seasons according to weather conditions. Nevertheless, three different number of clusters were examined, ranging from two to four and the clustering results were observed for each calculation. Since the results depend on both number of clusters and initial (random) clustering [23], there have been several possible solutions, among which the one with most sense was chosen. Generally, there is no single “best” way of choosing the number of clusters. Rather, a certain expertise in the field that the data belong to is required in order to choose the appropriate number [23].

Another important issue with clustering is different scaling of data features. Features’ scales can be different from each other by several orders of magnitude, such as in Table 2. To address this issue, a normalization is required in order to obtain meaningful clustering results.

For this particular case of clustering months into seasons, a somewhat different approach than before was used. The previously introduced constrictions still apply, but now only coefficient *a* is considered constant and its value fixed to 32 MW. All the other coefficients are considered variable for every month. This way, coefficients are optimized so that calculated losses are equal or almost equal to measured losses, similar to the previous case, depicted in Fig. 2. These values of coefficients serve as an input data set for the process of clustering. Therefore, they will be referred to as the initial coefficients. Obtained values of these coefficients are shown in Table 2, while the Fig. 3 shows their scaled versions (same scaling as Fig. 1). Comparison of calculated and measured losses is shown in Fig. 4. Values of errors i.e. differences from Fig. 4, both in absolute and relative units are given in Table 3.

Table 2 Calculated values for coefficients

Month	Coefficient *a* (MW)	Coefficient *b* (pu)	Coefficient *c* (MW^{-1})
January	32	0.0353882	0.00003250295
February	32	0.0380628	0.00002887427
March	32	0.0198292	0.00003235119
April	32	0.0360867	0.00002157678
May	32	0.0337267	0.00002904083
June	32	0.0125835	0.00002087766
July	32	0.0158188	0.00002198364
August	32	0.0132383	0.00002108569
September	32	0.0150446	0.00002172672
October	32	0.039316	0.00002210148
November	32	0.0455431	0.00002230626
December	32	0.0357529	0.00002795988

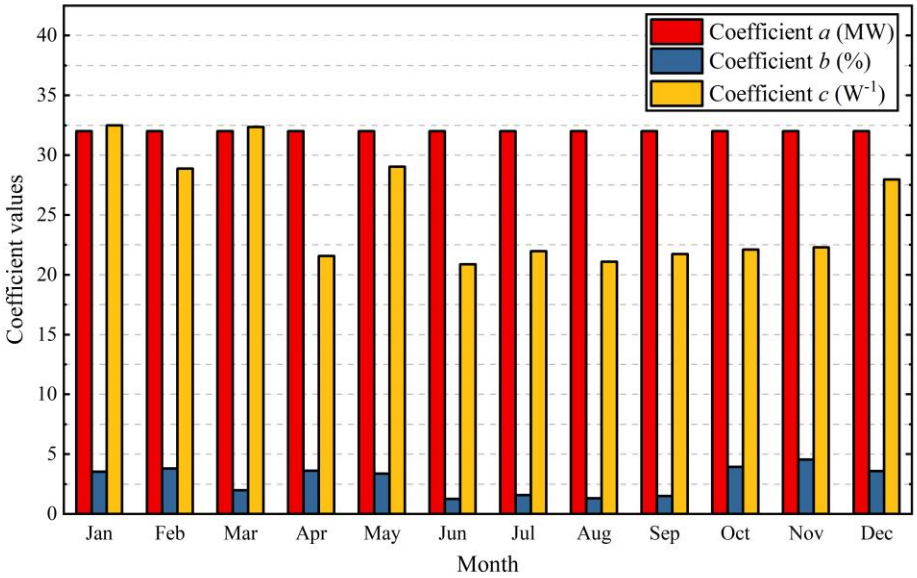

Fig. 3 Scaled values of the calculated coefficients after initial optimization

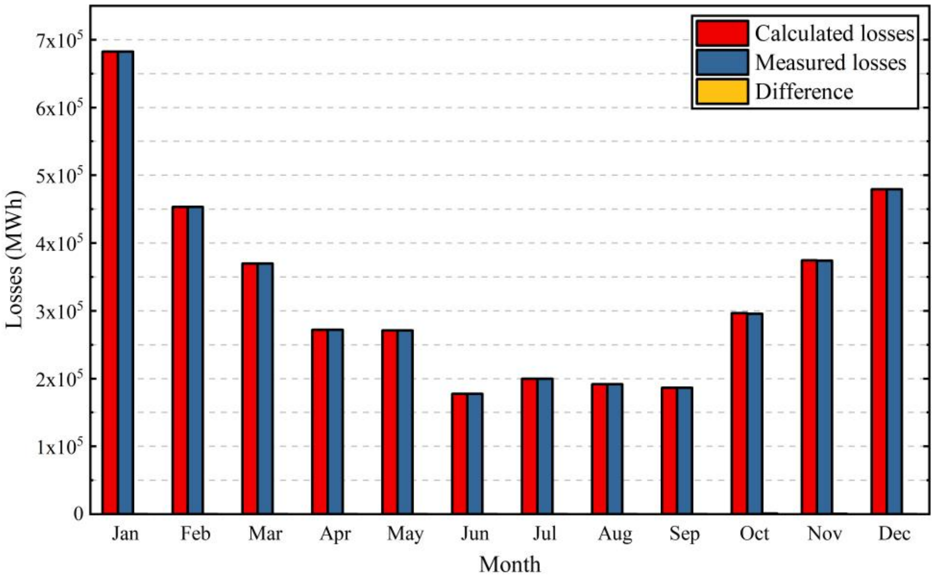

Fig. 4 Comparison between calculated and measured losses after initial optimization

Table 3 Absolute and relative differences in calculated and measured losses after initial optimization

Month	Absolute difference (MWh)	Relative difference (%)
Jan	0.0000	0.0000 %
Feb	0.0000	0.0000 %
Mar	0.0000	0.0000 %
Apr	117.5637	0.0432 %
May	0.0000	0.0000 %
Jun	0.0000	0.0000 %
Jul	0.0000	0.0000 %
Aug	0.0003	0.0000 %
Sep	0.0000	0.0000 %
Oct	784.1185	0.2644 %
Nov	407.3877	0.1088 %
Dec	0.0000	0.0000 %

From Table 2, it can be noted that the values of coefficients b_j are different than those in Table 1. The reason for this lies in the fact that now all the other coefficients are different too (although coefficient *a* is only slightly different). This means that the coefficients b_j also had to change in order to achieve the best possible fit to the input data.

From Fig. 4 and Table 3 it is obvious that coefficients fit the input data almost perfectly, since the yellow bars are almost invisible for every month.

Coefficients *b* and *c* are now variable throughout months, and discovering similarities among them is the key for clustering of months. This approach theoretically gives better results than the approach with two constant parameters because in the former case, the clustering is done on the basis of two features (parameters *b* and *c*), while in the latter case, clustering would have been done on the basis of one feature only (parameter *b*). This theory has also been proven experimentally. Based on these values, months are clustered into seasons and clustering results are shown in Table 4.

Table 4 Clustered months

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Cluster index	1	1	1	3	1	2	2	2	2	3	3	1
Season	Win	Win	Win	Tra	Win	Sum	Sum	Sum	Sum	Tra	Tra	Win

Season labels in Table 4 stand for Winter season (cluster 1, label “Win”), Summer season (cluster 2, label “Sum”) and Transient season (cluster 3, label “Tra”). From the Table 4 it can be observed that months are clustered almost completely as expected. The only exception is May, which has been clustered as a winter month. Looking back to the previous calculations, one can notice that May does not follow the usual pattern like other months. In Fig. 1, all months follow general pattern that coefficient *b* values get lower during summer and higher during winter in characteristic “elbow” shape. However, value for May presents an outlier from that pattern since its value is higher than values for the surrounding months. There is also a visible difference of parameter *c* for May in Fig. 3

compared to surrounding months. This value corresponds more to the winter months than to other months. The reason for this could be a significant under reading of electrical energy consumption in May, untimely data collection and report creation, error in statistical processing of data or simply higher energy theft rate in that particular month.

Final coefficients were calculated as centroids of each cluster from Table 4. Centroid of a cluster is a vector of mean values of all features for all points in that cluster. These coefficients are shown in Table 5 and their scaled values depicted in Fig. 5 while the comparison of calculated and measured losses for each month is shown in Fig. 6 and absolute and relative differences given in Table 6.

Table 5 Grouped coefficients for three seasons

Cluster index	Season	Coefficient *a* (MW)	Coefficient *b* (pu)	Coefficient *c* (MW^{-1})
1	Winter	32	0.0350917	0.00002942712
2	Summer	32	0.0141719	0.00002141825
3	Transient	32	0.0403153	0.00002199484

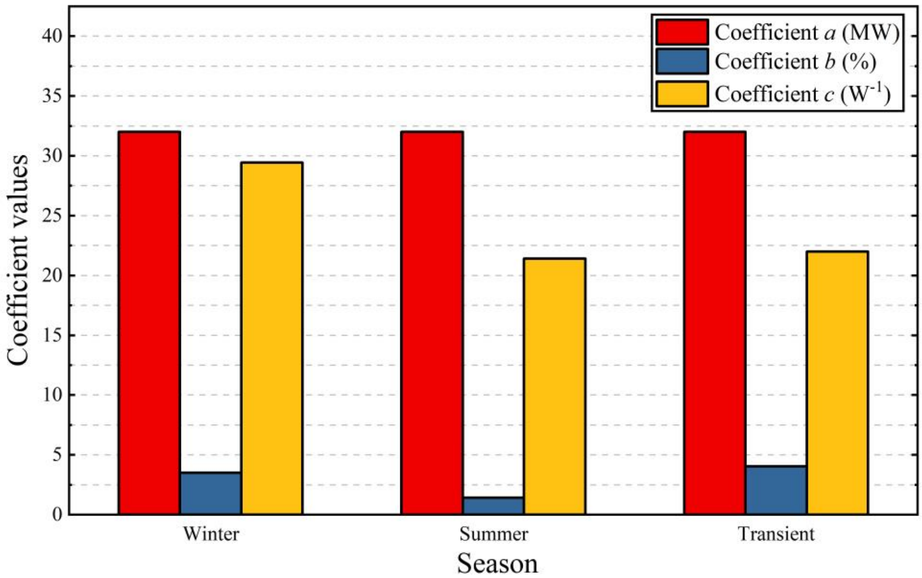

Fig. 5 Grouped coefficients for the three seasons

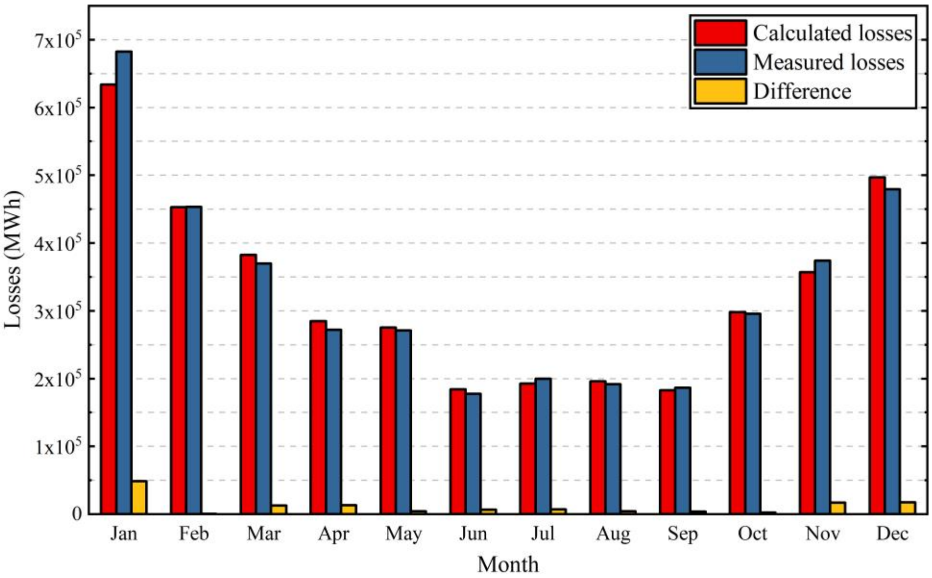

Fig. 6 Comparison between calculated and measured losses after clustering

Table 6 Absolute and relative differences in calculated and measured losses after initial optimization

Month	Absolute difference (MWh)	Relative difference (%)
Jan	48709.5891	7.6852 %
Feb	697.9128	0.1541 %
Mar	12719.8945	3.3250 %
Apr	13128.5232	4.6069 %
May	4248.3817	1.5425 %
Jun	6624.1044	3.5945 %
Jul	7211.0681	3.7413 %
Aug	4227.7345	2.1589 %
Sep	3686.2140	2.0147 %
Oct	2343.3930	0.7859 %
Nov	16880.7436	4.7272 %
Dec	17489.1562	3.5188 %

It can be seen that small error appears in the calculation. This error is a consequence of fact that one set of coefficients cannot perfectly fit all months, but seeks to reduce the overall error instead. Higher error can only be observed for winter months, due to the fact that May once again influenced this calculation. This also reflected to the somewhat higher value of coefficient *c* for winter season. Nevertheless, the overall error value is very low compared to values of losses.

6. CONCLUSIONS

In this paper, a new approach for calculation of losses in the electrical distribution grid was presented. There are two general classes of losses: technical and non-technical losses. Both types are unavoidable, but it is important to know how each type affects the total amount of losses, i.e. they have to be classified. This is done by analyzing the data available from the distribution grid operator. The data contain the distribution grid input energy measurement for every fifteen-minute time interval and the monthly measurements of energy delivered to end users. Difference between these two are the real total losses for a certain month. On the other hand, a polynomial equation is introduced to calculate that same losses based on the grid input power. Coefficients for this equation are computed using the least squares method, by minimizing the squared differences between the calculated and measured losses. These coefficient values are constrained based on their physical interpretation and authors experience. Results show that the minimization was successful and that the losses can clearly be classified this way. Additionally, calculated coefficients can be used for future estimation of losses. This concept was further expanded by introducing clustering of months into seasons. The obtained results show expected distinction of months, with the exception of May, which was classified as a winter month. This, along with the results of previous calculations, lead to a conclusion that there has probably been an error in energy consumption readings in this particular month. However, the overall results are very good, and the whole concept can be further improved. Future research will be focused on processing the data for the previous few years. This could allow the researchers to discover some specific trends and obtain better clustering accuracy since some seasons may begin in one year and end in another.

Acknowledgement: *This research is partly supported by project grant III44006 financed by the Ministry of education, science and technology development of the Republic of Serbia.*

REFERENCES

[1] L. Sladojević, A. Janjić, M. Ćirković „Calculation of Losses in the Distribution Grid Based on Big Data", In Proceedings of the 4th Virtual International Conference on Science, Technology and Management in Energy, eNergetics 2018, October 25-26, 2018, pp. 19-22

[2] K. Zhou, C. Fu and S. Yang, "Big data driven smart energy management: From big data to big insights", *Renewable and Sustainable Energy Reviews*, vol. 56, pp. 215–225, 2016.

[3] O. Ardakanian, N. Koochakzadeh, R. P. Singh, L. Golab, and S. Keshav, "Computing Electricity Consumption Profiles from Household Smart Meter Data," In Proceedings of the EDBT Workshop on Energy Data Management, 2014, pp. 140–147.

[4] H. Jiang, J. Zhang, W. Gao, and Z. Wu, "Fault detection, identification, and location in smart grid based on data-driven computational methods," *IEEE Transaction on Smart Grid*, vol. 5, no. 6, pp. 2947–2956, 2015.

[5] D. N. Nikovski, Z. Wang, A. Esenther, H. Sun, K. Sugiura, T. Muso and K. Tsuru, "Smart meter data analysis for power theft detection", In Proceedings of the Int. Workshop Machine Learning and Data Mining in Pattern Recognition, ser. LNCS, vol. 7988. Springer, 2013, pp. 379–389.

[6] P. Nallagownden and T. P. Hong, "Development of a new loss prediction method in a deregulated power market using proportional sharing," In Proceddings of Power Engineering and Optimization Conference (PEOCO), 2011 5th International. Shah Alam, Selangor, Malaysia: IEEE, June 2011, pp. 48–53.

[7] A. Petrušić, A. Janjić, "Economic Regulation of Electric Power Distribution Network", In Proceedings of 2nd Virtual International Conference on Science, Technology and Management in Energy, Niš, Serbia, 2016, pp. 25-32
[8] W. D. Liu, T. J. Zeng, T. Jun, L. L. Shang Guan, and B. J. Li, "Research on Electric Power with Development and Application of Line Loss Rate Forecasting Software Based on MLRM-GM," *Advanced Materials Research*, vol. 977, pp. 182–185, June 2014.
[9] K. Li, Z. Q. Sun, and M. Wang, "Theoretical Line Loss Calculation of Distribution Network Considering Wind Turbine Power Constraint," *Advanced Materials Research*, vol. 986-987, pp. 630–634, July 2014.
[10] Q. Ding and A. Abur, "Transmission Loss Allocation Based on a New Quadratic Loss Expression," *IEEE Transactions on Power Systems*, vol. 21, pp. 1227–1233, August 2006.
[11] G. Gross and S. Tao, "A Physical-Flow-Based Approach to Allocating Transmission Losses in a Transaction Framework," *IEEE Transactions on Power Systems*, vol. 15, pp. 631–637, May 2000.
[12] Elmitwally, A. Eladl, and S. M. Abdelkader, "Efficient algorithm for transmission system energy loss allocation considering multilateral contracts and load variation," *IET Generation, Transmission & Distribution*, vol. 9, pp. 2653–2663, August 2015.
[13] L. Tian, Q. Q. Wang, and A. Z. Cao, "Research on SVM Line Loss Rate Prediction Based on Heuristic Algorithm," *Applied Mechanics and Materials*, vol. 291-294, pp. 2164–2168, February 2013.
[14] Y. Ren, X. G. Zhang, and X. C. Huang, "Study on the Prediction of Line Loss Rate Based on the Improved RBF Neural Network," *Advanced Materials Research*, vol. 915-916, pp. 1292–1295, April 2014.
[15] P. S. Nagendra Rao and R. Deekshit Energy Loss Estimation in Distribution Feeders IEEE Transactions on Power Delivery, vol. 21, no. 3, July 2006.
[16] M.W. Gustafson, J. S. Baylor, and S. S. Mulnix, "Equivalent hours loss factor revisited," *IEEE Trans. Power Syst.*, vol. 3, no. 4, pp. 1502–1507, Nov. 1988.
[17] M. W. Gustafson, "Demand, energy and marginal electric system losses," *IEEE Trans. Power App. Syst.*, vol. PAS-102, no. 9, pp. 3189–3195, Sep. 1983.
[18] W. Gustafson and J. S. Baylor, "Approximating the system losses equation," *IEEE Trans. Power Syst.*, vol. 4, no. 3, pp. 850–855, Aug. 1989
[19] M. Järvinen, "Developing network loss forecasting for Distribution System Operator", Master of Science Thesis, Tampere University of Technology, 2013
[20] https://www.ofgem.gov.uk/ofgem-publications/43519/sohn-overview-losses-final-internet-version.pdf, accessed on 26.2.2019.
[21] J. Löfberg, "YALMIP: A Toolbox for Modeling and Optimization in MATLAB", In Proceedings of the CACSD Conference, Taipei, Taiwan, 2–4 September 2004.
[22] Waltz, R. A., J. L. Morales, J. Nocedal, and D. Orban, "An interior algorithm for nonlinear optimization that combines line search and trust region steps," *Mathematical Programming*, vol. 107, no. 3, pp. 391–408, 2006.
[23] G. James, D. Witten, T. Hastie, R. Tibshirani, " An Introduction to Statistical Learning with Applications in R", Springer Texts in Statistics, Springer Science + Business Media New York, 2013 (Corrected at 6th printing 2015).