Daily Load Curves in Distribution Networks - Analysis of Diversity and Outlier Detection

Grzegorz Dudek
Faculty of Electrical Engineering
Czestochowa University of Technology
42-200 Czestochowa, Al. Armii Krajowej 17, Poland
dudek@el.pcz.czest.pl

Jerzy Szkutnik
Faculty of Management
Czestochowa University of Technology
42-200 Czestochowa, Al. Armii Krajowej 19b, Poland
szkutnik@el.pcz.czest.pl

***Abstract*—This work is focused on diversity analysis of daily load curves and outlier curve detection in distribution networks. For diversity evaluation unified load curves are defined and the Euclidean distance between them and the average curve is calculated. Outlier detection is based on the distance between a load curve and an average curve representing curves of the same day of the week as the analyzed curve and from the same period of time. For detection of outliers in shape normalized daily curves are defined and the distance is measured between them and the average normalized curve. The analyzes are illustrated using real-world data from more than hundred substations.**

Keywords*—*daily load curve; outlier detection; diversity analysis

I. Introduction

This work focuses on distribution networks, which are crucial because they make up over 90% of the total electricity system network length [1]. Very large percentage of all electrical demand and renewable generation is connected to the distribution networks. This trends are expected to continue in the future. Some reports predict that the distribution network investments will have to make up between 65% and more than 80% of all the network investments to 2050, depending on the locations analyzed [2]. Such high investment costs are due to the resulting size and complexity of distribution networks.

Many research projects within the scope of power system loads are performed to get knowledge about the load variability in time (in daily, weekly and annual periods) and the factors which shape the load variation globally and also for specific groups of end users. This knowledge is useful or even necessary for making decisions about maintenance and development of distribution systems [3], [4]. Moreover, it is also crucial for short-, medium- and long-term forecasting models used in energy sector [5].

A load profile of electricity usage is very important to the efficiency and reliability of power transmission in an electricity distribution grid. Critical aspects of power distribution: sizing, modeling and optimization of power transformers and also batteries depends on the load profile. Load losses in the distribution network are dependent directly on the characteristics of the load profile such as average load factor, diversity factor, utilization factor, and demand factor. All these characteristics are calculated based on a given load profile.

In the era of rapid development of smart grid technologies the research area concerning system loads and load profiles is even more important. It can be applied in many smart grid projects such as [2]:

- advanced metering infrastructure (AMI) - revenue collection, reduction of electricity theft, outage notification, service and maintenance scheduling,
- customer-side systems - energy use awareness, support for demand response (DR), control of individual appliances, provision of smart grid-to-vehicle and vehicle-to-grid, management of heating and cooling devices,
- distributed energy resources - control, management and monitoring of variable and dispatchable distributed generation assets, system-impacts management,
- demand response - shift the load to decrease peak demand, increase customer awareness and choices, improve system flexibility, counteract system events, avail cheaper time-of-use tariff options
- storage - control, management and monitoring of storage assets,
- substation automation - optimize substation and support upstream and downstream use of assets,
- distribution automation - operation and management of the grid during normal, outage or maintenance conditions, minimizing impacts on customers,
- control centre systems - provide visual representation of system status and ability to manage operation and maintenance,
- cross-cutting: information and communication technology integration - collect, transmit and analyze electricity system data for specific purposes,
- asset management - manage and optimize asset utilization and maintenance.

In this work we describe a method of diversity analysis of daily load curves in distribution networks and methods of

978-1-5090-6406-9/17/$31.00 ©2017 IEEE

outlier daily curves detection. The analysis were performed using real-world data: 15-minute measurements recorded by the electricity meters installed in the primary substation 110/15 kV (one input meter, which measures total energy fed into the area and several output meters on feeders) and 125 electricity meters installed in the distribution substations 15/0.4 kV, which are fed through the primary substation 110/15 kV. Data are from the period spanning July 2013 - June 2014.

II. Diversity of Daily Load Curves

For analysis of the daily load diversity the daily load curves are unified. A daily curve can be expressed as a vector of n components: $\mathbf{L} = [L_1, L_2, ..., L_n]$, where $n = 24$ for hourly load time series, $n = 48$ for half-hourly load time series, and $n = 96$ for 15-minutes load time series. Unified daily curve $\mathbf{L}'' = [L''_1, L''_2, ..., L''_n]$ has components which are defined as the daily curve components divided by the mean daily load:

$$L''_t = \frac{L_t}{\overline{L}} \qquad (1)$$

where $t = 1, 2, ..., n$, $\overline{L}$ is a mean daily load.

Unified daily curves express shapes (profiles) of the daily load curves (curves of various daily average loads are reset to the same level but their shapes remain unchanged).

Fig. 1 demonstrates the unified curves for primary substation 110/15 kV. Selected curves for summer and winter are shown with thick lines, as well as the average curve. Usually different days of the week (Mondays, Tuesdays-Fridays, Saturdays, Sundays) express their own shapes. In our case there is no visible difference between Mondays and the other working days. Saturdays and Sundays have shifted night valley and deeper evening valley comparing to the working days. Moreover, the evening valley is deeper on Sundays than on Saturdays. Note differences between summer and winter curves. The evening peak for summer is about hour 22, while in winter it is several hours earlier. A midday peak is comparable with the evening peak for summer. In winter a daily curve for workdays is rather flat between hours 8-17, without a distinct peak.

From Fig. 1 one can conclude on:

- shape of the daily curve (the presence of valleys, peaks, load variation during the day),
- diversity in shapes in the analyzed period of time,
- degree of similarity of the daily curve shapes (higher when the gray curves are more focused around the average curve),
- differences in shapes between winter and summer (e.g. shifting of peaks),
- differences in shapes between different days of the week,
- outlier curves indicating failures of network elements, switching off customers etc.

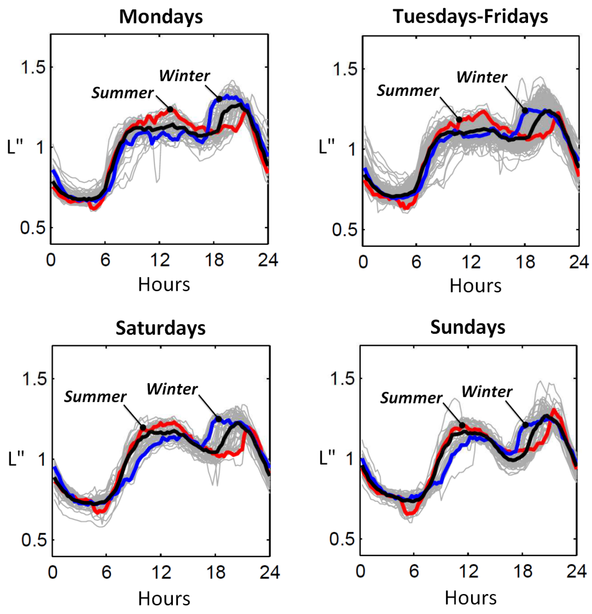

Fig. 1. Unified daily load curves for primary substation 110/15 kV (input meter M_1).

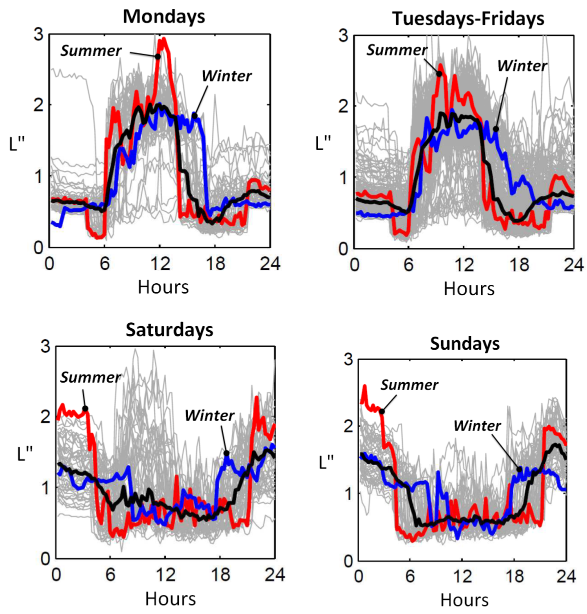

Fig. 2. Unified daily load curves for distribution substation 15/0.4 kV (meter M_9).

The shapes of curves representing large number of customers are generally more stable (less varied, similar to each other), as shown in Fig. 1. When a substation feeds a small group of specific customers, e.g. industrial ones, the daily load profiles are strictly dependent on the power consumption character of these customers. This is shown in Fig. 2. It can be seen from this figure the strong diversity of daily curve shapes, their unstable nature and the big differences between the shapes for workdays and shapes for Saturdays and Sundays. For workdays the daily period can be divided into two parts: the interval of increased load (hours 6-15) and the interval of reduced load (hours 1-6 and 15-24). On Saturdays and Sundays a higher load is observed at night and during the day there is a valley.

For diversity evaluation of the daily curve shapes Euclidean distance is used. The unified daily load curve in the case of 15-minute resolution can be expressed as a vector of 96 components: $\mathbf{L}'' = [L''_1, L''_2, ..., L''_{96}]$ or point in 96-dimensional space. Let us define the average unified curve |$\mathbf{M}'' = [M''_1, M''_2, ..., M''_{96}]$ with components defined as:

$$M''_t = \frac{1}{N}\sum_{i=1}^{N} L''_{i,t} \quad (2)$$

where N is the number of days in the period analyzed.

For similarity evaluation between unified curve $\mathbf{L}''$ and the mean curve $\mathbf{M}''$ we use a distance measure:

$$d = \sqrt{\sum_{t=1}^{96} (L''_t - M''_t)^2} \quad (3)$$

In Table I mean distances between unified daily curves and the average curves for selected substations were shown, where:

- M_k is an electricity meter symbol installed in a certain substation,
- L_{av} is a mean annual load measured by the electricity meter M_k,
- d_{av} is a mean distance for all days,
- d_{Mon} is a mean distance for Mondays (in this case $\mathbf{M}''$ is an average curve for Mondays),
- $d_{Tue\text{-}Fri}$ is a mean distance far Tuesdays-Fridays (in this case $\mathbf{M}''$ is an average curve for Tuesdays-Fridays),
- d_{Sat} is a mean distance far Saturdays (in this case $\mathbf{M}''$ is an average curve for Saturdays),
- d_{Sun} is a mean distance far Sundays (in this case $\mathbf{M}''$ is an average curve for Sundays),
- $d_{Work\text{-}Sat}$ is a mean distance between average curve for workdays and average curve for Saturdays,
- $d_{Work\text{-}Sun}$ is a mean distance between average curve for workdays and average curve for Sundays.

Note that diversity increases with decreasing in mean annual load, corresponding to the number of customers which are fed through the substation. The lowest diversity for substation 110/15 kV (input meter M_1) is observed, and the highest for the smallest substation (meter M_{10}). The distribution of the diversity expressed as d_{av} for all substations in Fig. 3 is shown.

III. Atypical Daily Curve Detection

Many outlier detection methods have been developed in recent years [6]. Taking into account the specificity of the problem of outlier detection in the daily load curves we propose new methods.

TABLE I. Diversity of the Daily Load Curves

	L_{av} kW	d_{av}	d_{Mon}	$d_{Tue\text{-}Fri}$	d_{Sat}	d_{Sun}	$d_{Work\text{-}Sat}$	$d_{Work\text{-}Sun}$
M_1	10521	0.61	0.58	0.54	0.56	0.59	0.42	0.70
M_2	2348	0.67	0.78	0.61	0.52	0.61	0.42	0.56
M_3	1679	0.70	0.64	0.63	0.64	0.63	0.53	0.70
M_4	709	0.91	0.80	0.73	0.80	0.68	0.95	1.33
M_5	278	0.81	0.78	0.77	0.76	0.73	0.36	0.76
M_6	104	0.97	0.90	0.88	0.89	0.87	0.81	0.98
M_7	51	1.34	1.27	1.27	1.34	1.45	0.63	0.72
M_8	30	1.55	1.35	1.36	1.07	1.08	1.91	2.15
M_9	14	4.56	3.51	3.49	4.00	3.08	6.47	7.94
M_{10}	8	5.53	4.06	4.12	5.00	3.37	5.29	10.46

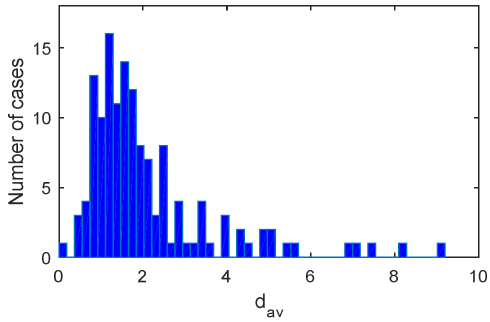

Fig. 3. Histogram of the mean distance between unified curves $\mathbf{L}''$ and the mean curves $\mathbf{M}''$.

Atypical daily curve (called outlier) differs in some way from typical daily curves from the same time period (e.g. a few neighboring weeks) and representing the same day of the week. The difference between daily curves may relate to:

- average daily load (outlier is located above or below the typical curves),
- range of variability of daily curve (deeper/shallower valleys or peaks in relation to the typical curves) or
- shape of the curve (e.g. peaks shifted in time comparing to typical curves).

Atypical curves which differ from typical ones in the average daily load or in the range of variability can be detected using some measure of distance between curves. An average daily curve $\mathbf{M}$ is defined representing n curves of the same day of the week as the analyzed curve and from the same period of time $\mathbf{\Phi}$. It is assumed that $\mathbf{\Phi}$ includes five weeks before and five weeks after the analyzed day. The daily load curve for 15-minute resolution can be expressed by the vector of 96 components: $\mathbf{L} = [L_1, L_2, ..., L_{96}]$. The Euclidean distance between analyzed load curve of the day i, $\mathbf{L}_i$ and the average curve $\mathbf{M}_i = [M_{i,1}, M_{i,2}, ..., M_{i,96}]$ is defined as follows:

$$d = \sqrt{\sum_{t=1}^{96} (L_{i,t} - M_{i,t})^2} \quad (4)$$

where the components of $\mathbf{M}_i$ are medians of the corresponding components of n curves from $\mathbf{\Phi}$ representing the same day of the week as the analyzed curve $\mathbf{L}_i$:

$$M_{i,t} = median(L_{i-5\cdot 7,t}, L_{i-4\cdot 7,t}, \ldots, L_{i+5\cdot 7,t}) \quad (5)$$

For detection of outliers in shape normalized daily curves are defined $\mathbf{L}' = [L'_1, L'_2, \ldots, L'_{96}]$, which components are calculated as follows:

$$L'_t = \frac{L_t - \bar{L}}{\sqrt{\sum_{t=1}^{96} (L_t - \bar{L})^2}} \quad (6)$$

where $\bar{L}$ is a mean load of the day and the square root in the denominator is a measure of dispersion of the daily curve.

Normalized curve $\mathbf{L}'$ carries information about the shape of the daily curve. The mean value of the load curve and its dispersion are filtered out. This leads to unification of daily curves. Now, they can be compared only in terms of the shape similarity. Similarity is based on the Euclidean distance between the analyzed curve $\mathbf{L}'$ and the average normalized curve $\mathbf{M}'$, which components are medians of the corresponding components of n normalized curves from $\mathbf{\Phi}$ representing the same day of the week as the analyzed curve. The distance and medians are calculated form equations similar to (4) and (5), respectively.

Distance d can be a measure of atypicality of the daily curve. If $d = 0$ a daily curve is identical to the average curve $\mathbf{M}$ which represents typical curve for the period $\mathbf{\Phi}$ covering the analyzed curve. High value of d informs about atypicality of the curve. A source of atypicality may be national or religious holidays, measurement system errors, network failures, switching of customers, weather anomalies etc.

Fig. 4 shows distances between a given daily curve $\mathbf{L}_i$ from the analyzed period of time and average curve $\mathbf{M}_i$. As a threshold for outliers it was assumed $d = 5000$. Examples of the detected curves which are atypical because of the average daily load or the range of variability in Fig. 5 are shown. In this figure ten curves of the same day of the week as the analyzed curve and from period $\mathbf{\Phi}$ are drawn with thin lines. Top charts demonstrate holidays (Christmas Day and New Year's Day) for which load curves lie below the average curves. Left bottom panel shows the day of increased demand. This day the substation supplied energy also to the adjacent area. Right bottom panel shows a temporary interruption in the supply of energy.

Distances between normalized daily curves $\mathbf{L}'_i$ and the average curves $\mathbf{M}'_i$ in Fig. 6 are shown. The threshold for outliers is set to $d = 0.4$. Examples of outliers which differ in shape form the typical shape in Fig. 7 are shown. Left top panel demonstrate the day before Christmas which has characteristic evening pick. It corresponds to the celebratory dinner which is traditionally served in Poland on Christmas

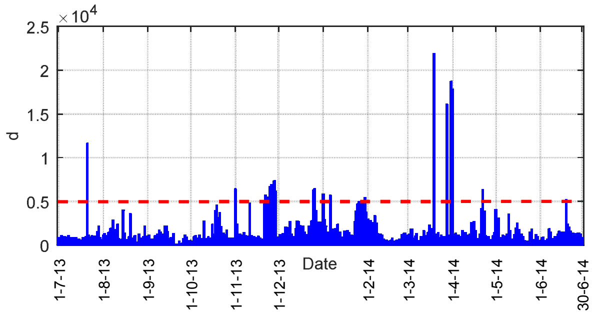

Fig. 4. Distances between daily curves $\mathbf{L}_i$ and average curves $\mathbf{M}_i$.

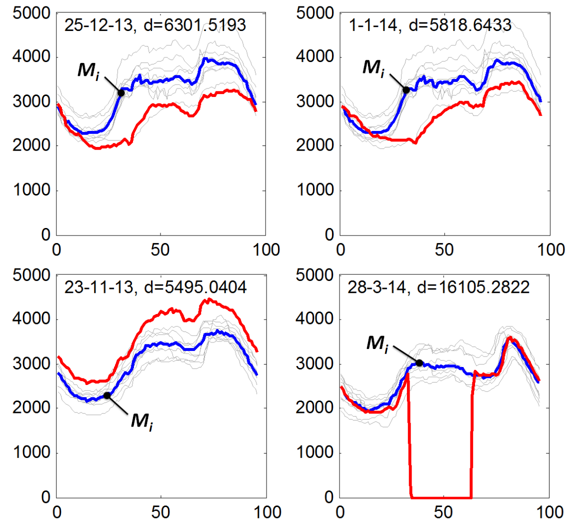

Fig. 5. Outliers detected using distance between daily curves $\mathbf{L}_i$ and average curves $\mathbf{M}_i$.

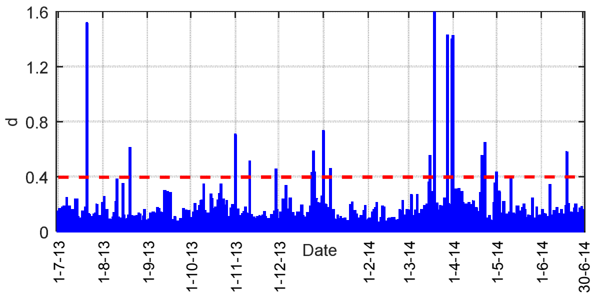

Fig. 6. Distance between normalized daily curves $\mathbf{L}'_i$ and average curves $\mathbf{M}'_i$.

Eve. Right top panel shows Christmas day with shifted night valley. Similar shift of the night valley is observed in right bottom panel, where the day after Easter is shown (this day is also a holiday in Poland). This day is characterized also by the deeper evening valley. Left bottom panel shows completely different shape from typical ones, which is caused by the interruption in the energy supply.

IV. CONCLUSION

In this work a method for diversity evaluation of daily load curves in distribution networks is proposed as well as methods for outlier daily curves detection. The experimental study is performed using real-world data from more than hundred substations. The analysis of daily curves is extremely important for efficiency and reliability of distribution systems and enables the distribution system owner to make decisions about its maintenance and development.

The proposed methods are based on Euclidean distance between daily load curves. The distance is an indicator of diversity or outliers. In the case of diversity we calculate the distance between unified curves. This allow us to compare diversities for different substations (they are comparable). As expected, the diversities for bigger substations, which feeds larger number of customers, are smaller.

Outlier detection is carried out in two ways. Outliers in the average daily load or in the range of daily curve variability are detected using distances between daily curves of the same day of the week and from the same period of time. While outliers in shape are detected using normalized daily curves. Both ways lead to effective detection of outliers of different type.

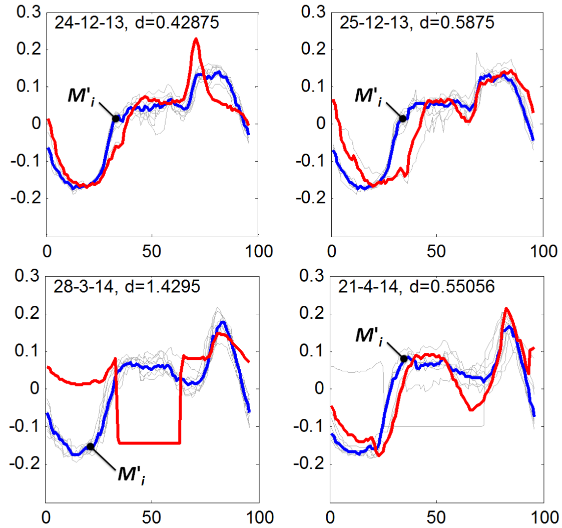

Fig. 7. Outliers detected using distance between normalized daily curves $\mathbf{L}'_i$ and average curves $\mathbf{M}'_i$.

REFERENCES

[1] "Global Transmission and Distribution Report 2010," ABS Energy Research, London, 2010.

[2] "Technology Roadmap: How2Guide for Smart Grids in Distribution Networks," International Energy Agency Report, 2015.

[3] J. Szkutnik, "The energy efficiency as the necessary element of the planning in the sector of the electrical energy," Proc. 9th Inter. Conf. Electric Power Engineering 2008, pp. 397-400, 2008.

[4] A. Gawlak, and L. Poniatowski, "Main directions of investment in the sector of electrical energy distribution," Proc. 14th Inter. Conf. Electrical Power Engineering 2014, pp. 39-43, 2014.

[5] G. Dudek, "Pattern Similarity-based Methods for Short-term Load Forecasting – Part 2: Models," Applied Soft Computing, vol. 36, pp. 422-441, 2015.

[6] V. Barnett, and T. Lewis, "Outliers in Statistical Data." Wiley, 1994.