# Cluster Analisis of Long-term Power Quality Data

Michał Jasiński, Tomasz Sikorski
Wroclaw University of Science and Technology
Faculty of Electrical Engineering
Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland

Jacek Karpiński, Marek Zenger
Tauron Dystrybucja S.A., Distribution System Operator,
Department Wroclaw, Poland

***Abstract*—Current methods for measurement and analysis of power quality is based on the selection of a representative period of time, usually a week, corresponding to normal operating conditions. However such conditions as variability of load and generation or changing the configuration of power system network may have a significant impact on the quality parameters of power system. In order to distinguish and perform comparative assessment of different states of the power system condition in terms of power quality parameters it is possible to use cluster analysis. This paper presents the results of the use of cluster analysis of long-term power data quality electricity recorded synchronously in selected fields of distribution substation 110/20/10kV.**

***Keywords: power quality, data mining, cluster analysis, partitioning.*** [1]

## I. Introduction

Nowadays power quality assessment uses parameters which are defined in literature and standards [1]÷[3] and include parameters of voltage variations and voltage events. The main voltage disturbances consist of: disturbances of power frequency (frequency variations), voltage changes (slow voltage changes, RMS voltage), voltage fluctuation (flicker severity), voltage unbalance (asymmetry), harmonics and interharmonics and in specifics conditions commutations notches or voltage signaling. The main voltage events are: rapid voltage changes, voltage dips and swells, interruptions, temporary overvoltage, transient overvoltage.

Measurement and assessment methods of that parameters are standardized among others in [1]÷[3]. The methods are based on choice of the representative period of time (commonly one week) which corresponds to the normal working condition of analyzed object (point of common coupling, nodes). This analysis are often not sufficient and does not represent the real state of the analyzed part of the power network because the **data** (measurement data) can depend on the analyzed **case** which for power quality can be change of load or generation as well as different configuration of the network. Due to this reason in this paper as the solution for comprehensive power quality analysis the cluster analysis was chosen. The cluster analysis is based on a division of the set of data (measurement data) into groups (clusters). The clustering, named also as data portioning, can be done in point of view of futures of the data or cases [4] ÷[6]. When the data futures are considered the clustering collects the data into groups according to their similarity.

The examples of clustering based on characteristic data are:

- Graphical clustering in which clusters are chosen during the identifying data that lie close together in the space of attributes. In other words the graphical clustering is based on the measurement of distance between data items.
- k-means method is based on the identification of k- clusters which are represented by its mean position.
- Self-organizing features map using competitive learning rule.

On the other hand it's possible to make a division of data into groups due to **cases**. i.e. by using external conditions, which can have influence on data futures. Furthermore it is possible to use combination of characteristic data and cases.

Clustering approach allows to distinguish similarity and differences between data selected in particular clusters. Due to that defining the clusters depends on futures of analyzed object it is important to choose correct characteristic date belong to the cluster or external condition which determined selection of the data for the cluster. Due to the mentioned rules of data division the process of clustering is very sensitive for data which are different from general cluster characteristic.

For the main reasons of using clustering it is mentioned:

- reduction of huge amount of data to few basic categories, which can be a base for next analysis,
- achievement of uniform group and their essential characteristic,
- achievement of classification of work condition of analyzing object,
- achievement of the division into groups as the element of analysis connected due to the characteristics data and cases.

In this paper it was used the clustering of long-term power quality data due to cases. The data were divided into clusters due to the external conditions for which the date are determined by different characteristics. The cases were associated with different level of load demand and cooperation with combined heat and power plant. The data was recorded synchronously in selected bays of the substation 110/20/10 kV

[1] This work was supported by Polish Ministry of Higher Education and Science, project 0401/0115/16



**978-1-5090-5518-0/16/$31.00 ©2016 IEEE**

including incoming feeder, outgoing feeder and the transformer. For the defined clusters it was performed analysis and assessment of power quality parameters in each cluster. It leaded to make conclusions about the impact of load and cooperation of heat and power plant on the power quality parameters.

## II. Characteristic of Analysed Object and Measurement Data

The long-term data represents real measurement made in city distribution substation 110/20/10 kV. Distribution substation 110kV works in dual-capable system, 5 section in cooperation with two 3-coils transformers T1 and T2. Normally, the transformers are supplied from separated systems or 110kV section. The substation is connected with 110kV electric power distribution network by line bays and block bays with combined heat and power plant (CHP named as BC2). 20kV switching station consist of two section split by bus coupler bay. Normally, the substation is supplied from main transformers T-1 and T-2. 10kV switching station cooperate with synchronous water power plant. Substation have two auxiliary transformers with a group of winding connection ZNyn-11. System has good selectivity of cable earth fault protection and lower level of overvoltage caused by ground fault and lower effective tough voltage.

In the substation, there were installed power quality recorders in selected bays: incoming feeder 110kV from CHP block, transformer bay side 110kV and side 10kV. Measurements were made synchronously. Methods of the measurement and aggregation times were agreed with class A of standard [2]. The part of substation HV with location of power quality analyzers is presented in Fig. 1.

The measurement was made in spring time from 03.04.2012 to 24.04.2012. During the measurement there were noticed some dynamic voltage events. Table I presents parameters of registered events. The system of flagging was used in data analysis that means that selected aggregated data which includes events were omitted in analysis. Fig. 2 and Fig. 3 presents examples of the RMS voltage change and long term flicker severity for primary transformer side (110kV) and secondary (10kV). There are marked flags of voltage events in the figures.

TABLE 1 Characteristic of voltage event detected during the analysis

| Voltage level | Event | Event start | Duration time of event [ms] | Residual voltage [kV] |
|---|---|---|---|---|
| 10kV | 3-phase voltage dip | 03.04.2012 06:09:47 | 110.04 | 4.696 |
| 110 kV | 3-phase voltage dip | 03.04.2012 06:09:48 | 90.000 | 10.00 |
| 10 kV | 3-phase voltage dip | 07.04.2012 07:26:06 | 90.069 | 8.965 |
| 110 kV | 3-phase voltage dip | 07.04.2012 07:26:06 | 119.67 | 83.06 |
| 10 kV | 3-phase voltage dip | 13.04.2012 11:48:34 | 29.992 | 9.408 |
| 110 kV | 3-phase voltage dip | 13.04.2012 11:48:34 | 119.89 | 85.83 |

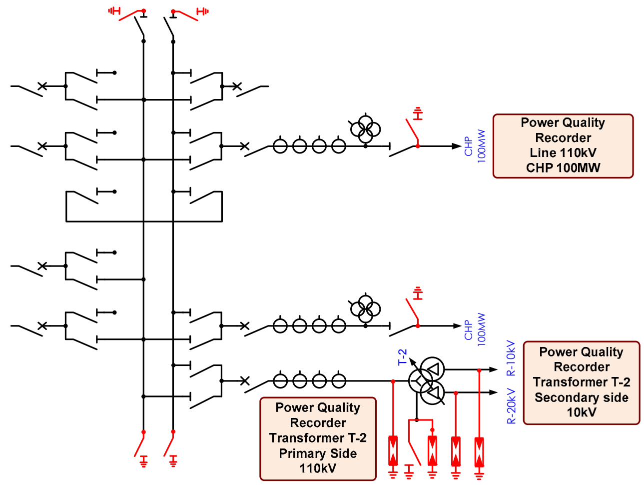


Fig. 1. Localization of measuring device in analyzed substation.

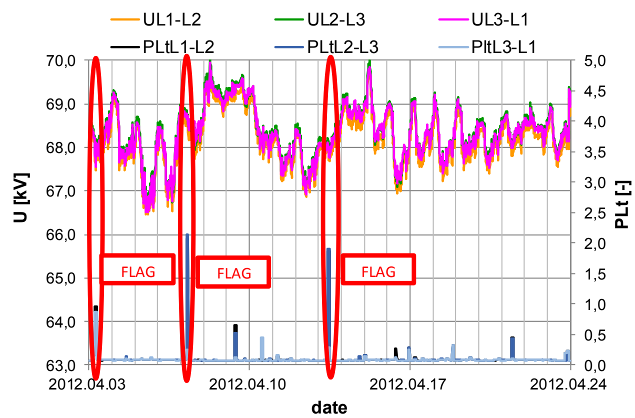


Fig. 2. Voltage (left axis of ordinates OY) and voltage fluctuations (right axis of ordinates OY) primary side of transformer, 110 kV

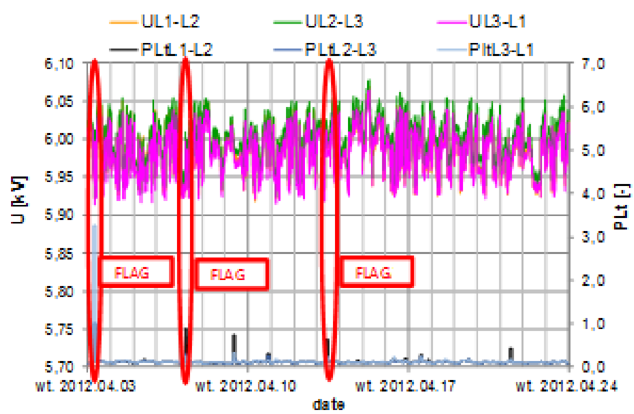


Fig. 3. Voltage(left axis of ordinates OY) and voltage fluctuations (right axis of ordinates OY) secondary side of transformer, 10 kV

## III. DATA CLUSTER ANALYSIS

The cluster chosen was made on the base of selected external conditions included load character in working day and no-working day and the cooperation of combined heat and power plant 100MW. Fig. 4 presents the trend of active and reactive power of transformer 110kV (mark as T2) and cooperated with the same system incoming feeder coupled with mentioned CHP (mark as BC2). On the top of the figure the clusters are marked to represents the rule for data division which includes cases related to:

- working day and non-working day,
- work of heat and power plant block and non-working of the CHP block.

Finally due to mentioned cases the data was divided into four clusters which represent particular case:

- I cluster: **working day** with working CHP block
- II cluster: **working day** without CHP block,
- III cluster: **non-working day** with working CHP block,
- IV cluster: **non-working day** without CHP block.

For such defined clusters the statistical analysis of power quality data was performed. Then the comparison of each characteristic was made. The results were presented as chart slides. The charts presents minimum, average, maximum of analyzed power quality parameters for each cluster. The flagged values (the aggregated values which includes mentioned voltage events) were not taken into consideration. Fig. 5÷Fig. 13 present comparative analysis of particular power quality parameters of 110kV and 10kV transformer side for each cluster I÷IV. Detailed statistics of the cluster analysis are presented in table II.

Using data analysis of particular clusters there is a possibility to extract influence of the selected conditions of power system which were the base of grouping. The grouping consist of: working days (I, III cluster) and non-working days (III, IV cluster) or work of heat and power plant block (I, III cluster) or it's non-work (II, IV cluster). Comparison of the results of the analysis allows to constitutes several quality conclusions.

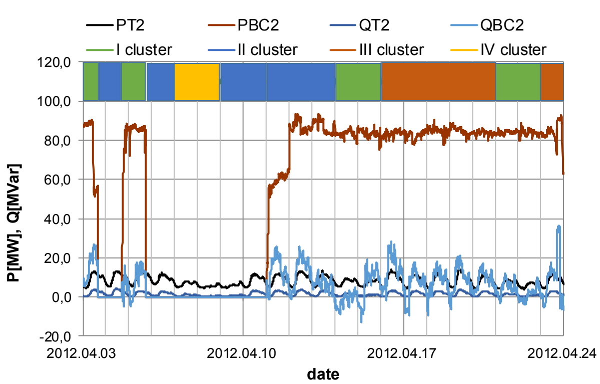


Fig. 4. Transformer load change(T2) and generation of heat and power block(BC2) with marked custers of data, 110kV

The variability of **frequency** is noticeable from the character of a day. In Fig. 5 there is observable larger changes of frequency in working day than in non-working day (clusters I, II in comparison to clusters III, IV). The impact on frequency variation is less noticeable when the state of working or not heat and power plant block is considered (cluster I relative to cluster II and cluster II relative to cluster IV). In each clusters (I, II, III, IV) all values are focused around the mean value. The confirmation of frequency concentration around mean value are detailed statistics in table II expressed by small value of standard deviation.

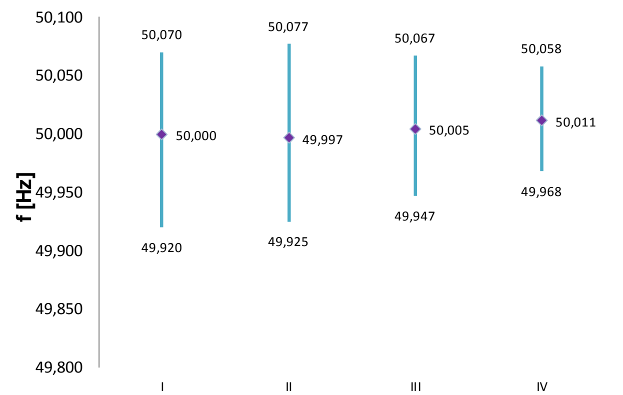


Fig. 5. Comparative voltage frequency for 110kV transformer side for each cluster I÷IV

Primary **voltage** of the transformer 110 kV depends on both character of a day as well as contribution of the power produced by CHP. Fig. 6 presents higher values and smaller variety of voltage in non-working days (clusters III, IV) than in working days (clusters I, II). Less noticeable in comparison to character of the day is an impact of CHP contribution (cluster I relative to cluster II and cluster III relative to cluster IV). Due to the automatic voltage control using tap changer secondary voltage of the transformer 10kV does not express so clear relations to load condition and CHP contribution as in case of primary voltage. In each cluster(I, II, III, IV) all values are focused around the mean value.

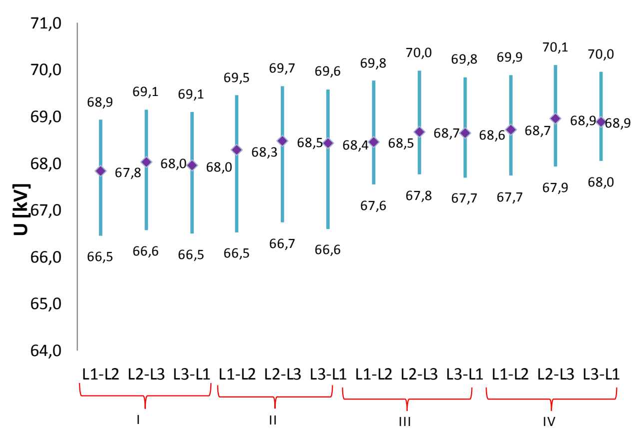


Fig. 6. Comparative line-to-line voltage for 110kV transformer side for each cluster I÷IV

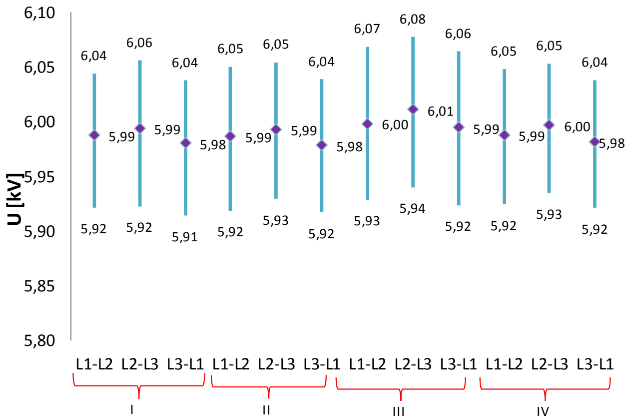


Fig. 7. Comparative line-to-line voltage for 10kV transformer side for each cluster I÷IV

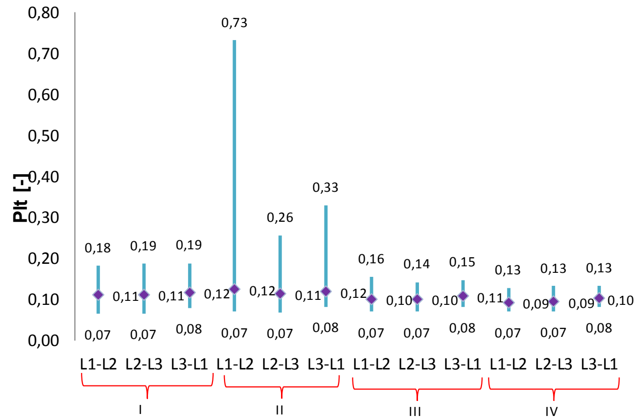


Fig. 9. Comparative analysis voltage fluctuations for 10kV transformer side for each cluster I÷IV

The significant **voltage fluctuations** were noticed in working day with without CHP (cluster II). Furthermore it is noticeable similar relation as in frequency and voltage which reveals by larger voltage fluctuation in working days than in non-working days (clusters I, II relative to clusters III, IV). It is showed in Fig. 8. Similar relations are noticed on 10kV transformer side, Fig. 9. On the other hand, working days (cluster III, IV) are expressed by smaller voltage fluctuation but they have bigger level of dispersion around mean value. The confirmation of the statement is confirmed in table II by higher value of standard deviation.

Asymmetry designated by 110kV voltage is slightly higher in non-working days without block (cluster IV) but the bigger variety is notice in working day (cluster I) – Fig. 10. Relations for 10kV voltage are similar i.e. higher values in days without block that depicts Fig. 11. Differences are however slight.

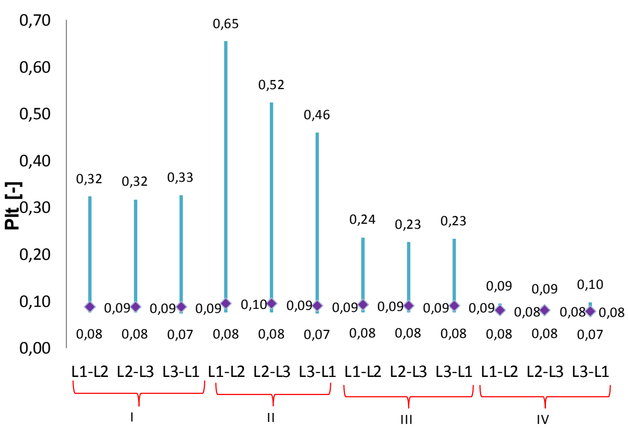


Fig. 8. Comparative analysis voltage fluctuations for 110kV transformer side for each cluster I÷IV

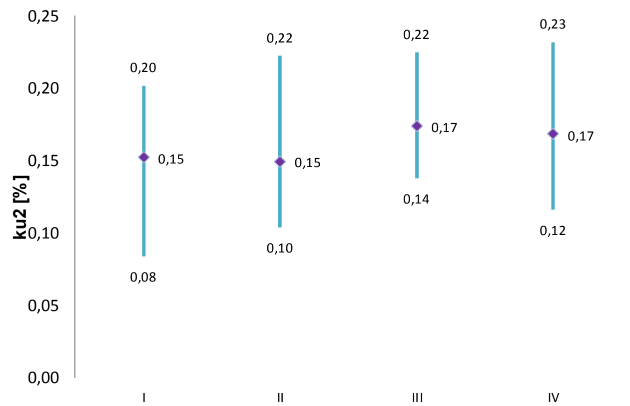


Fig. 10. Comparative analysis voltage unbalance for 110kV transformer side for each cluster I÷IV

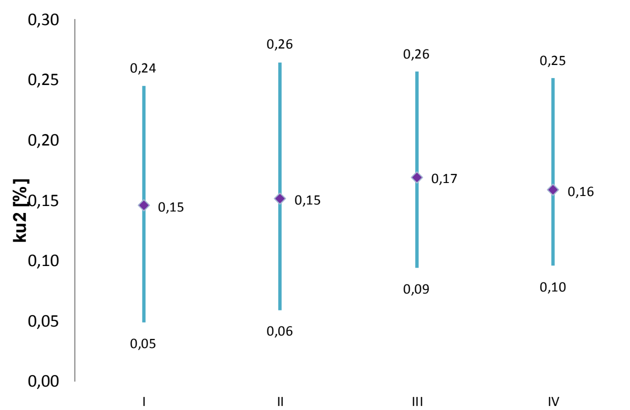


Fig. 11. Comparative analysis voltage unbalance for 10kV transformer side for each cluster I÷IV

**Total harmonic distortion** factor of 110kV voltage is higher in non-working days and days without CHP support , Fig. 12. This relation is not clearly noticeable on 10 kV side, however noticeable is higher level of relative harmonic contribution on 10kV than 110kV side, Fig. 13.

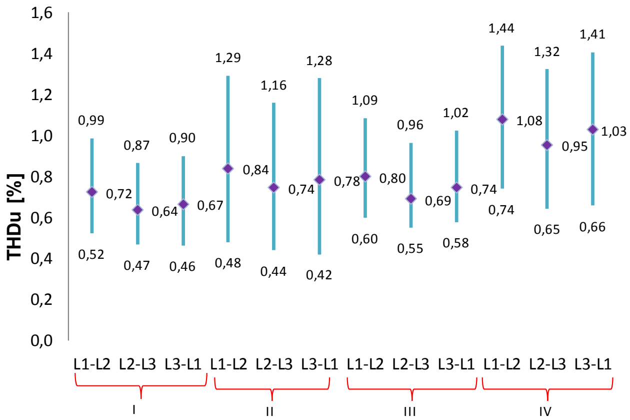


Fig. 12. Comparative analysis voltage total harmonic distortion (THDu) for 110kV transformer side for each cluster I÷IV

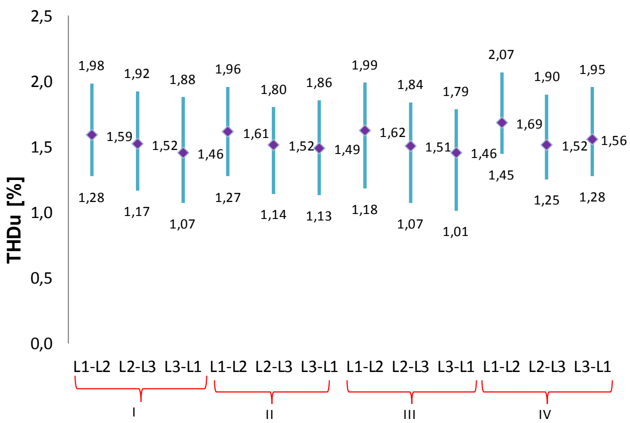


Fig. 13. Comparative analysis voltage total harmonic distortion (THDu) for 10kV transformer side for each cluster I÷IV

## IV. Conclusions

Presented results of analysis of three weeks power quality data using clustering on the basis of selected cases makes possible to define the influence of working conditions of system on power quality. Comparison the statistic characterization of the grouped data set of particular clusters allowed to assess the influence of load changing and support of combined heat and power plant. Authors see the need of continuous analysis for using clustering, together with correlation tools in process of decision-making [7]. Whereas it's important to recognize that cluster analysis can serves as a supplement to the classical power quality assessment which assure compliance with requirements for power quality described in regulations [1]÷[3].

## References

[1] EN 50160Voltage characteristics of electricity supplied by public distribution network.

[2] IEC 61000 4-30 Electromagnetic compatibility (EMC) – Part 4-30: Testing and measurement techniques – Power quality measurement methods.

[3] Rozporządzenie Ministra Gospodarki z dnia 4 maja 2007 r. w sprawie szczegółowych warunków funkcjonowania systemu elektroenergetycznego, ze zmianami 21 sierpnia 2008 r. (Dz.U. 2008 nr 162 poz. 1005), in Polish.

[4] S. Wierzchoń, M. Kłopotek, "Alorithms for cluster analysis", Polish Academy of Science, 2015

[5] A. Paderewski, „Analiza Skupień", course materials SGGW, http://agrobiol.sggw.waw.pl/biometria/media//paderewski/Statystyka/Analiza%20skupie%C5%84.pdf (accessable 10.09.2016)

[6] CIGRE Broshure 292, „Data mining techniques and applications in the power transmission field, CIGRE, 2006.

[7] M. Jasiński, „ Proces decyzyjny w wnioskowaniu o stanie jakości energii elektycznej w sieciach elektroenergetycznych zasilających zakłady przeysłowe", Conferrence „Innowacyjne pomysły młodych naukowców: Nauka – Startup – Przemysł", Kraków 30-31.05.2016, in Polish.

TABEL II. STATISTIC POWER QUALITY ANALISIS IN EACH CLUSTER I÷IV.

| 110kV (primary side of the transformer) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Aggregated values | I cluster | | | II cluster | | | III cluster | | | IV cluster | | |
| U | Average value | 67,8 | 68,0 | 68,0 | 68,3 | 68,5 | 68,4 | 68,5 | 68,7 | 68,6 | 68,7 | 68,9 | 68,9 |
| | Variance | 0,21 | 0,22 | 0,22 | 0,56 | 0,56 | 0,56 | 0,19 | 0,20 | 0,18 | 0,22 | 0,23 | 0,17 |
| | Standard deviation | 0,45 | 0,47 | 0,47 | 0,75 | 0,75 | 0,75 | 0,44 | 0,45 | 0,43 | 0,47 | 0,48 | 0,41 |
| | **Relative standard deviation** | **0,01** | **0,01** | **0,01** | **0,01** | **0,01** | **0,01** | **0,01** | **0,01** | **0,01** | **0,01** | **0,01** | **0,01** |
| THD | Average value | 0,72 | 0,64 | 0,67 | 0,84 | 0,74 | 0,78 | 0,80 | 0,69 | 0,74 | 1,08 | 0,95 | 1,03 |
| | Variance | 0,01 | 0,01 | 0,01 | 0,04 | 0,03 | 0,04 | 0,01 | 0,01 | 0,01 | 0,03 | 0,03 | 0,04 |
| | Standard deviation | 0,08 | 0,07 | 0,08 | 0,19 | 0,16 | 0,20 | 0,09 | 0,08 | 0,09 | 0,17 | 0,16 | 0,19 |
| | **Relative standard deviation** | **0,12** | **0,11** | **0,13** | **0,22** | **0,21** | **0,25** | **0,11** | **0,11** | **0,12** | **0,15** | **0,17** | **0,19** |
| $P_{lt}$ | Average value | 0,09 | 0,09 | 0,09 | 0,10 | 0,09 | 0,09 | 0,09 | 0,09 | 0,09 | 0,08 | 0,08 | 0,08 |
| | Variance | 0,00 | 0,00 | 0,00 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| | Standard deviation | 0,03 | 0,03 | 0,03 | 0,08 | 0,06 | 0,05 | 0,03 | 0,03 | 0,03 | 1,17 | 1,19 | 1,21 |
| | **Relative standard deviation** | **0,33** | **0,32** | **0,35** | **0,82** | **0,65** | **0,58** | **0,37** | **0,31** | **0,37** | **14,52** | **14,75** | **15,20** |
| $k_{u2}$ | Average value | 0,15 | | | 0,15 | | | 0,17 | | | 0,17 | | |
| | Variance | 0,00 | | | 0,00 | | | 0,00 | | | 0,00 | | |
| | Standard deviation | 0,02 | | | 0,02 | | | 0,01 | | | 0,03 | | |
| | **Relative standard deviation** | **0,15** | | | **0,16** | | | **0,08** | | | **0,16** | | |
| f | Average value | 50,00 | | | 50,00 | | | 50,00 | | | 50,01 | | |
| | Variance | 0,00 | | | 0,00 | | | 0,00 | | | 0,00 | | |
| | Standard deviation | 0,02 | | | 0,02 | | | 0,02 | | | 0,02 | | |
| | **Relative standard deviation** | **0,00** | | | **0,00** | | | **0,00** | | | **0,00** | | |

| 10 kV (secondary side of the transformer) | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Aggregated values | I cluster | | | II cluster | | | III cluster | | | IV cluster | | |
| U | Average value | 5,99 | 5,99 | 5,98 | 5,99 | 5,99 | 5,98 | 6,00 | 6,01 | 5,99 | 5,99 | 6,00 | 5,98 |
| | Variance | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| | Standard deviation | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 | 0,03 |
| | **Relative standard deviation** | **0,00** | **0,00** | **0,00** | **0,00** | **0,00** | **0,00** | **0,00** | **0,00** | **0,00** | **0,01** | **0,01** | **0,01** |
| THD | Average value | 1,59 | 1,52 | 1,46 | 1,61 | 1,52 | 1,49 | 1,62 | 1,51 | 1,46 | 1,69 | 1,52 | 1,56 |
| | Variance | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 | 0,02 | 0,03 | 0,03 | 0,03 | 0,02 | 0,02 | 0,03 |
| | Standard deviation | 0,13 | 0,15 | 0,15 | 0,14 | 0,15 | 0,15 | 0,18 | 0,18 | 0,17 | 0,15 | 0,16 | 0,17 |
| | **Relative standard deviation** | **0,08** | **0,10** | **0,10** | **0,09** | **0,10** | **0,10** | **0,11** | **0,12** | **0,12** | **0,09** | **0,10** | **0,11** |
| $P_{lt}$ | Average value | 0,11 | 0,11 | 0,12 | 0,12 | 0,11 | 0,12 | 0,10 | 0,10 | 0,11 | 0,09 | 0,09 | 0,10 |
| | Variance | 0,00 | 0,00 | 0,00 | 0,01 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 | 0,00 |
| | Standard deviation | 0,02 | 0,02 | 0,02 | 0,09 | 0,03 | 0,04 | 0,02 | 0,02 | 0,02 | 0,47 | 0,48 | 0,49 |
| | **Relative standard deviation** | **0,21** | **0,22** | **0,19** | **0,75** | **0,30** | **0,30** | **0,19** | **0,17** | **0,14** | **5,09** | **5,10** | **4,80** |
| $k_{u2}$ | Average value | 0,15 | | | 0,15 | | | 0,17 | | | 0,16 | | |
| | Variance | 0,00 | | | 0,00 | | | 0,00 | | | 0,00 | | |
| | Standard deviation | 0,03 | | | 0,04 | | | 0,03 | | | 0,03 | | |
| | **Relative standard deviation** | **0,24** | | | **0,27** | | | **0,18** | | | **0,22** | | |
| f | Average value | 50,00 | | | 50,00 | | | 50,00 | | | 50,01 | | |
| | Variance | 0,00 | | | 0,00 | | | 0,00 | | | 0,00 | | |
| | Standard deviation | 0,02 | | | 0,02 | | | 0,02 | | | 0,02 | | |
| | **Relative standard deviation** | **0,00** | | | **0,00** | | | **0,00** | | | **0,00** | | |

Legend:
f – frequency variation
$k_{u2}$ – asymmetry
$P_{lt}$ – long-term flicker severity
THD – total harmonic distortion
U – voltage variation