energies

MDPI

Article

Combined Cluster Analysis and Global Power Quality Indices for the Qualitative Assessment of the Time-Varying Condition of Power Quality in an Electrical Power Network with Distributed Generation

Michał Jasiński [1],*[ID]**, Tomasz Sikorski** [1][ID]**, Paweł Kostyła** [1][ID]**, Zbigniew Leonowicz** [1][ID] **and Klaudiusz Borkowski** [2]

[1]   Department of Electrical Engineering Fundamentals, Faculty of Electrical Engineering, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland; tomasz.sikorski@pwr.edu.pl (T.S.); pawel.kostyla@pwr.edu.pl (P.K.); zbigniew.leonowicz@pwr.edu.pl (Z.L.)
[2]   KGHM Polska Miedź S.A, 59-301 Lubin, Poland; klaudiusz.borkowski@kghm.com
*   Correspondence: michal.jasinski@pwr.edu.pl; Tel.: +48-713-202-022

check for **updates**

Abstract: This paper presents the idea of a combined analysis of long-term power quality data using cluster analysis (CA) and global power quality indices (GPQIs). The aim of the proposed method is to obtain a solution for the automatic identification and assessment of different power quality condition levels that may be caused by different working conditions of an observed electrical power network (EPN). CA is used for identifying the period when the power quality data represents a different level. GPQIs are proposed to calculate a simplified assessment of the power quality condition of the data collected using CA. Two proposed global power quality indices have been introduced for this purpose, one for 10-min aggregated data and the other for events—the aggregated data index (*ADI*) and the flagged data index (*FDI*), respectively. In order to investigate the advantages and disadvantages of the proposed method, several investigations were performed, using real measurements in an electrical power network with distributed generation (DG) supplying the copper mining industry. The investigations assessed the proposed method, examining whether it could identify the impact of DG and other network working conditions on power quality level conditions. The obtained results indicate that the proposed method is a suitable tool for quick comparison between data collected in the identified clusters. Additionally, the proposed method is implemented for the data collected from many measurement points belonging to the observed area of an EPN in a simultaneous and synchronous way. Thus, the proposed method can also be considered for power quality assessment and is an alternative approach to the classic multiparameter analysis of power quality data addressed to particular measurement points.

1. Introduction

Over the years, global electric energy consumption has increased from 440 Mtoe in 1973 to 1737 Mtoe in 2015 [1]. This has resulted in electricity becoming a specific product that is subject to market regulation in both quantitative and qualitative terms. Quantitative analysis is mainly focused on the balance between energy that is produced, transmitted, stored, consumed or lost. The current issues connected to quantitative aspects of energy consumption are related to demand-side

response, the integration of renewable energy sources, and energy storage systems with electrical power systems [2–5]. The qualitative approach mainly uses power quality analysis. The issues related to power quality (PQ) include definitions of the parameters, and the methods of measurements and assessment, which are already standardized, among others [6–8]. The methods described in the standards of power quality analysis are based on power quality parameters measured during a representative period of time, normally one week, which corresponds to the normal working conditions of the observed network [9]. The parameters which characterize power quality include: frequency variation, voltage variation, voltage fluctuation, voltage asymmetry, and voltage waveform distortion. These parameters are collected during the period of observation, with the aggregation time interval usually equal to 10 min; however, 1 min aggregation intervals are also currently studied [10]. Using these parameters creates a significant number of data to be considered in the analysis. Moreover, PQ data depends on the network conditions, load changes, generation level, or configurations of the network. For this reason, a rational approach is to search for data mining techniques able to extract and classify vectors of the power quality data that represent different features. This would allow the range of qualitative analyses to be extended by correlating the information of the network, environment or market condition.

There are many works dedicated to power quality disturbance extraction, PQ events recognition, and the classification of PQ events and disturbances, which are all directly focused on the measured voltage and current signals. Most of the works propose wavelet transforms, S-transforms, empirical mode decomposition, and other different decomposition techniques, which are supported by artificial neural networks in order to find valuable methods for PQ disturbance extraction and recognition [11–13]. However, a different problem can be formulated when there is a need for identifying and extracting some of the data that represent different features from the long-term aggregated power quality data. This necessitates a comprehensive method for the automatic classification of long-term power quality data into groups that represent similar features. This task is essentially a data mining area of interest. One of the data mining techniques that can meet these requirements is cluster analysis (clustering). The general application of data mining techniques in power systems is presented in [14]. Specific applications of cluster analysis to electrical power networks include:

- Detection and classification of disturbances [15–19];
- Detection and analysis of events [20–25];
- Load characteristics and classification [26–30];
- Power flow problems [31–34];
- Power quality evaluation and monitoring [35–38].

This article extends the cluster analysis (CA) proposed by the authors in [9]. Jasiński et al. [9] present the results of the application of CA in order to achieve a desirable division of the long-term 10-min aggregated power quality data into groups of data representing similar features. The collection of the PQ data comes from four real points of measurement in the supply network of a copper mine. The significant elements of the investigated power network are combined heat and power (CHP) plants with gas-steam turbines working as a local distributed generation (DG), and also a welding machine (WM) as the main time-varying load. Time-varying PQ conditions were intentionally created. The distributed generation was switched on and off for a period of time, and a network reconfiguration was also performed. The results discussed in [9] confirm the possibility of using cluster analysis for the extraction of power quality data into groups related to the different working conditions of an electrical network, including the influence of DG, reconfiguration of the network, working days, and holiday time. In [9], the methodology of application of the cluster analysis, including the preparation of the database structure, was also described. The idea presented in [9] leads to efficient classification of the power quality data, but it does not provide a suitable method for the assessment of collected clusters of the data. Searching for (1) a comprehensive solution that provides automatic classification of the multipoint measurement data, and (2) a method for comparative evaluation of the collected data, remains a desirable aim for wide-area monitoring systems and smart grids. Thus this article

is an extension of the previously obtained classification [9] in order to a achieve quality assessment of obtained clusters using global power quality indices. This leads to an automatic classification of working conditions of an electrical power network (EPN), and the possibility of an easy comparison using global values, that incorporate the impact of different PQ elements.

Both cluster analysis of PQ data and global power quality index (GPQI) application may be found in the literature:

- Sacasqui et al. [39] present an application of grey clustering with entropy weight methodology. The proposed solution was used to calculate a unified quality index of distributed electricity. Their research is based on [40], where a new unified index was proposed, as well as a network model. The model consists of a 138 kV system, wind energy system, hybrid wind-photovoltaic-fuel cell system and the load. The PQ data consist of current total harmonic distortion, voltage total harmonic distortion, sag, frequency deviation, instantaneous flicker level, and power factor. The unified index is calculated for different working conditions using gray CA and entropy weight for the measurement points separately. The research is based on simulations.

- The work of Song et al. [41] concerns the application of cluster analysis combined with a support vector machine for the prediction of PQ indexes. The real measurement data from a 35 kV substation are processed. The database contains selected PQ parameters including frequency deviation, voltage unbalance, and total harmonic distortion (*THD)* in voltage, as well as weather conditions and data on other associated factors. In the described article, CA was used to obtain implicit classifications of indexes. The analysis concerns a single measurement point.

- Florencias-Oliveros et al. [42] present the analysis of recorded signals representing different disturbances. The proposed index realizes a comparison of the variance values, skewness, and kurtosis connected with each cycle, versus the ideal signal. Then, the CA is used to create a classification of the disturbances using proposed PQ index.

The aspect that distinguishes the solution proposed in this paper from the methods described in quoted works is the area-based approach to the PQ assessment, involving all measurement points for the cluster analysis, as well as development of a new synthetic power quality index. Novel aspects of the method proposed in this article include:

- Application of cluster analysis for the data collected from several measurement points distributed in the supply network of a mining industry in order to achieve suitable identification of different working conditions of the observed network. This approach treats the collected data as a common database more representative of the observed area than particular measurement points.

- New synthetic global power quality indices are used for the assessment of groups of PQ data identified by cluster analysis. The proposed definition of the GPQI consists of a set of classical PQ parameters based on a 10-min aggregation interval; however, it is also extended by selected parameters based on a 200-ms aggregation interval. The aim of extending the proposed GPQI definition with parameters related to a 200-ms aggregation interval is to enhance the sensitivity of the obtained global index. This proposed approach is tested by investigating the influence of the factors which comprise the proposed global power quality index on the sensitivity of the assessment.

- The proposed approach of using GPQIs leads to a straightforward comparison of the clusters in terms of a generalized assessment of the power quality conditions, which in turn finally allows a comparative assessment of different working conditions of the investigated network to be performed. The indicated clusters, which represent different working conditions, may be easily compared using a single GPQI for each of the measurement points.

The remaining structure of this paper is as follows: Section 2 reviews the present application of global power quality indices in the electrical power network, and also proposes a new definition of the GPQIs proposed in our assessment of clustered PQ data. Section 3 describes the proposed algorithm

methodology for the comparative assessment of the power quality conditions using a combination of clustering and global power quality indices. The first step of the algorithm is the identification and allocation of the power quality data into groups that represent similar features. This part is based on previous experience with CA application described in [9]. The second step is the assessment of the collected data using the proposed GPQI. The results of the assessment are presented using real multipoint power quality measurements in a medium voltage electrical network supplying the mining industry. Additionally, this section also contains a sensitivity analysis of the proposed GPQI in terms of the selection of the power quality parameters used to construct the GPQI. The presented results are towered to realize one of the article's aims—to highlight the impact of DG on PQ in the industry network. The obtained clusters represent different conditions of PQ indices which are directly associated with impact of the DG. Qualitative assessment of the PQ data collected in the identified clusters using the proposed global power quality indices allows us to confirm several relations between DG impact on PQ condition. Section 4 contains the discussion of the obtained results. Section 5 formulates the conclusions, interpretations in perspective studies, and implications for the future.

2. Global Power Quality Indices

Classical power quality assessment is a multi-criteria analysis approach that is independently applied to particular power quality parameters. The idea of a simplified and generalized assessment of the power quality condition uses a single index, known as a global, unified, total or synthetic index. In this paper, we decided to use global power quality indices (GPQIs) as a unified name. Before new definitions of GPQIs are introduced, it is relevant to have a review of the knowledge concerning the development of GPQIs. Singh et al. [43] present the application of a unified power quality index that uses the matrix method. The index, corresponding to voltage sag severity, was highlighted as a suitable proposition for power quality assessment, and is carried out in a three-stage approach. The first stage requires the preparation of a graphical system model (attribute digraph). The second step is the conversion into an attribute matrix. The next step is the presentation of the matrix as a variable permanent function. Ignatova and Villard [44] define green-yellow-red indicators for all PQ problems. The proposed algorithm obtains the green-yellow-red indicators for both events and disturbances. The index consists of all individual PQ parameters, which are expressed as a percentage in a range from 0% to 100%, where 0% denotes the worst PQ and 100% the optimal PQ. The index may be defined for each single point or for the whole facility. The benefit of the proposed generalization is the possibility to easily understand the interpretation of the PQ condition in the monitoring systems. Nourollah and Moallem [45] present the application of data mining to determine the unified power quality index which corresponds to all power quality parameters, with further classification, normalization, and incorporation. The proposed fast independent component analysis algorithm was proposed to determine the power quality level of each distribution site. The mentioned article proposes two indexes: the Supply-side Power Performance Index, which expresses the impact of six voltage indices; and the Load-side Power Performance Index. The second index corresponds to three current PQ indices. Raptis et al. [46] present artificial neural networks as a sufficient tool to support PQ assessment using an index called Total Power Quality Index. The index is the artificial neural network combination of eight power quality values used as input variables. The presented method uses a multilayer perceptron artificial neural network. Lee et al. [47] propose another power quality index. This index includes the power distortion, which concerns non-linear loads. The indicated aim of the proposed PQI is to support harmonic pollution determination in a distributed power system. The work [47] proposes a new distortion power quality index. The application of this index is a determination of the harmonic pollution ranking for different non-linear loads. It is realized by multiplication of the load composition rate and the load currents' total harmonic distortion. Hanzelka et al. [48] propose the idea of a synthetic PQ index. This index is based on the maximum values of traditional PQ parameters. These parameters are slow voltage change, harmonic content in voltage (represented by total harmonic distortion in voltage, and a particular harmonic from 2nd to 40th), unbalance, and voltage fluctuation

(represented by long-term flicker severity). The proposed assessment provided only satisfactory or unsatisfactory results.

In the present work, two definitions of GPQIs are proposed—one for 10-min aggregated data, and the other for the events. The proposed indices are inspired by the synthetic approach described in [48,49]. Some elements of the GPQI definitions, in terms of the multipoint measurements, were also proposed by the authors in [50]. Typical for the generalization process is that global indices are usually less sensitive due to synthetization. In order to enhance the sensitivity, the global indices proposed in this work are not only based on classical 10-min aggregated power quality parameters, but they are also extended by other parameters like an envelope of voltage changes based on 200-ms values. In order to demonstrate the proposed approach, we also present an analysis of how selected parameters comprising the global index influence its sensitivity.

The first proposed global power quality index is called the aggregated data index (*ADI*), and is expressed in (1).

$$ADI = \sum_{i=1}^{7} k_i \cdot W_i \tag{1}$$

ADI—aggregated data index;

i—number of the factor ranging from 1 to 7;

W_i—the particular power quality factors which create a synthetic aggregated data index;

k_i—the importance rate (weighted factors) of the particular power quality factor constituting the synthetic aggregated data index, range of [0, 1], where $\sum_{i=1}^{7} k_i = 1$.

The *ADI* utilizes five classical 10-min aggregated PQ parameters, including: frequency (*f*), voltage (*U*), short-term flicker severity (P_{st}), asymmetry factor (k_{u2}), total harmonic distortion in voltage (*THDu*), and also two additional parameters which are responsible for the enhancement of the sensitivity of the proposed global index. The first additional parameter is represented by an envelope of voltage deviation obtained by the difference between the maximum and minimum of 200-ms voltage values identified during the 10-min aggregation interval. The second is a maximum of the 200-ms value of the total harmonic distortion in voltage, similarly identified in the 10-min aggregation interval. The mentioned parameters are calculated and refer to standard IEC 61000-4-30 [7]. Three phase values, like *U*, P_{st}, and *THDu* are reduced to one using the mean value of the three phase values. To be more specific, particular factors that create the proposed *ADI* index are based on the differences between the measured 10-min aggregated power quality data and the recommended limits stated in the standards. The differences are expressed as a percentage in relation to the limits. The final values of the factors taken in the *ADI* calculation are the mean values of the time-varying factors during the time period of observation. Additionally, the contribution of the particular power quality factors in global indices can be controlled by the importance factors, which serve as the weight of the contribution of particular parameters. The values of weighting factors are normalized to one. Selection of importance factors makes it possible to check the impact of single parameters as well as groups of parameters. The selection of parameters may be defined by a priori analysis of EPN problems (e.g., harmonics, voltage variations). No a priori statements were conducted in this work, so the weight of all parameters is the same and the priorities of particular parameters were the same. The aim of the introduced weighted factors is to open the possibility to make the analysis more focused on particular PQ parameters and neglect others—in other words, to obtain an analysis that is more sensitive for selected PQ phenomena controlled by weighted factors. For example, to justify adding 200-ms values, analyses with and without them were conducted.

Particular factors which create the global *ADI* index are defined as follows [50]:

$$W_1 = W_f = \frac{mean\left(\left|f_m - f_{nom}\right|\right)}{\Delta f_{limit}} \tag{2}$$

$W_1 = W_f$—factor of frequency change;

f_m—10-min measured value of frequency;

f_{nom}—nominal value of frequency;

$mean(|f_m - f_{nom}|)$—mean of frequency deviations in the observation time period;

Δf_{limit}—limit value of frequency change as a %.

$$W_2 = W_U = \frac{mean(|U_m - U_c|)}{\Delta U_{limit}} \tag{3}$$

$W_2 = W_U$—factor of the voltage level;

U_m—mean of 10-min measured values of voltage from three phases;

U_c—declared voltage;

$mean(|U_m - U_c|)$—mean of voltage deviations in the observation period of time;

ΔU_{limit}—limit value of voltage change in volts.

$$W_3 = W_{Pst} = \frac{mean(Pst_m)}{Pst_{limit}} \tag{4}$$

$W_3 = W_{Pst}$—factor of voltage variation;

Pst_m—mean of 10-min measurement value of the short-term flicker severity index from three phases;

$mean(Pst_m)$—mean of voltage variations in the observation time period;

Pst_{limit}—limit value of short-term flicker severity.

$$W_4 = W_{ku2} = \frac{mean(ku2_m)}{ku2_{limit}} \tag{5}$$

$W_4 = W_{ku2}$—factor of voltage unbalance;

$ku2_m$—10-min measured values of voltage unbalance;

$mean(ku2_m)$—mean value of voltage unbalance in the observation time period;

$ku2_{limit}$—limit level of voltage unbalance.

$$W_5 = W_{THDu} = \frac{mean(THDu_m)}{THDu_{limit}} \tag{6}$$

$W_5 = W_{THDu}$—factor of total harmonic distortion factor of voltage supply;

$THDu_m$—mean of 10-min measurement values of the total harmonic distortion factor of the voltage supply from three phases;

$mean(THDu_m)$—mean value of the total harmonic distortion factor in the observation time period;

$THDu_{limit}$—limit level of the total harmonic distortion factor of the voltage supply.

$$W_6 = W_{Uenv} = \frac{\frac{mean(|U_{max} - U_{min}|)}{U_c}}{2 \times \Delta U_{limit}} \tag{7}$$

$W_6 = W_{Uenv}$—factor of voltage deviation envelope;

U_{max}—mean value of 200-ms voltage maximum values from three phases allocated in 10-min data;

U_{min}—mean value of 200-ms voltage minimum values from three phases allocated in 10-min data;

U_c—declared voltage;

$mean(|U_{max} - U_{min}|)$—mean of voltage envelope width in the observation time period;

ΔU_{limit}—limit level of voltage change.

$$W_7 = W_{THDumax} = \frac{mean(THDu_{max})}{THDu_{limit}} \tag{8}$$

$W_7 = W_{THDu\text{max}}$—factor of the maximum 200 ms value of the total harmonic distortion factor of voltage supply;

$THDu_{\text{max}}$—mean value of 200-ms maximum values of the total harmonic distortion factor of voltage supply from three phases;

$mean(THDu_{\text{max}})$—mean of the total harmonic distortion factor in the observation time period;

$THDu_{\text{limit}}$—limit level of the total harmonic distortion factor of the voltage supply.

Then, the preparation of the particular factors $W_1 \div W_7$ and the selection of its important rates, the aggregated data index factor expresses the PQ level in a global range. The interpretation of the obtained index values are as natural. A value of "0" represents the ideal PQ; "0–1" represents possible power quality deterioration, but in compliance with the requirements defined in the standards; and finally, a value greater than 1 indicates the permissible parameters level defined in the standard is exceeded.

The second proposed global index relates to events. The classical approach to power quality assessment utilizes a flagging concept, which generally prescribes the extraction of the aggregated values that are affected by events like dips, swells and interruptions. The authors propose to use the information about the number of data which are not considered in classical PQ analysis due to the flagging concept. This is used as the base for a global index called the flagged data index (*FDI*), defined as follows [50]:

$$FDI = \frac{f}{n} \times 100\% \tag{9}$$

FDI—flagged data index;

f—number of 10-min data, which were flagged in the observation time period;

n—number of all 10-min data in the observation time period.

Interpretation of the obtained *FDI* values can be formulated such that "0%" represents the ideal PQ without any event disturbances, and "100%" expresses measurement data where each averaged value is contaminated by voltage events.

The proposed concepts for the generalization of the power quality assessment using GPQIs can be implemented for the fixed time period of observation or for identified periods of time representing different features of the power quality condition of the monitored area of the power system. The identification of such periods can be achieved using cluster analysis.

3. Results of Power Quality Assessment Using Cluster Analysis and Global Power Quality Indices

The idea of combined analysis using CA and GPQIs is presented in Figure 1. In the first step, the clustering is applied to achieve a classification of the power quality data into clusters representing different features. The outcomes of the CA depend on the construction of the PQ database, that is the set of PQ parameters under consideration, as well as the standardization of the formula. The mentioned issues and their impact on the results of the CA were already investigated and presented in [9]. A novelty of this work is the implementation of GPQIs for the group of PQ data identified by CA. We propose using the levels of GPQIs that characterize particular clusters for the comparative analysis.

As was already mentioned, some results of the cluster analysis were described in [9]. However, selected information about the investigated electrical power network is repeated for clarity and to help in understanding the presented application of the global power quality indices. Note that the input PQ data that create the database are the four-week multipoint power quality measurements obtained from a 6 kV power network supplying the mining industry [51]. The points of measurement include a secondary side of 110 kV/6 kV transformers (denoted as "T1", "T2", "T3"), and a 6 kV outcoming feeder supplying a welding machine (denoted as "WM") [9]. Inside the network, distributed generation units are installed (denoted as "DG"), represented by combined heat and power plants (CHP) with gas-steam turbines, denoted as "G1", "G2", and "G3", respectively. The analyzed EPN of the mining industry and placement of the measurement points are presented in Figure 2.

```
                            ┌─────────────────────────┐
                            │       INPUT DATA         │
                            └─────────────────────────┘
                                        │
        ┌───────────────────────────────────────────────────────────────┐
        │                   INITIAL CLUSTERING:                          │
        │    K-MEAN ALGORITHM, CHEBYSHEV DISTANCE, 2 CLUSTERS            │
        └───────────────────────────────────────────────────────────────┘
              │                                           │
    ┌─────────────────────────┐            ┌─────────────────────────┐
    │ CLUSTER 1: NON-FLAGGED   │            │ CLUSTER 2: FLAGGED DATA  │
    │         DATA             │            │                          │
    └─────────────────────────┘            └─────────────────────────┘
```

DATABASE A	DATABASE B	DATABASE C
STANDARDIZATION	STANDARDIZATION	STANDARDIZATION
DATABASE As	DATABASE Bs	DATABASE Cs

```
                            ┌─────────────────────────┐
                            │      PQ DATABASES        │
                            └─────────────────────────┘
                                        │
        ┌───────────────────────────────────────────────────────────────┐
        │                    MAIN CLUSTERING:                            │
        │ K-MEANS ALGORITHM, EUCLIDEAN OR CHEBYSHEV DISTANCE, C-NUMBER    │
        │                        CLUSTERS                                │
        └───────────────────────────────────────────────────────────────┘
                                        │
              ┌─────────────────────────────────────────────┐
              │  CLASSIFICATION REPRESENTED BY C-CLUSTERS     │
              └─────────────────────────────────────────────┘
                                        │
              ┌─────────────────────────────────────────────┐
              │  GLOBAL POWER QUALITY INDICATORS CALCULATION  │
              └─────────────────────────────────────────────┘
                    │                              │
    ┌─────────────────────────┐      ┌─────────────────────────┐
    │ AGGREGATED DATA INDEX    │      │  FLAGGED DATA INDEX       │
    │        (ADI)             │      │        (FDI)              │
    └─────────────────────────┘      └─────────────────────────┘
```
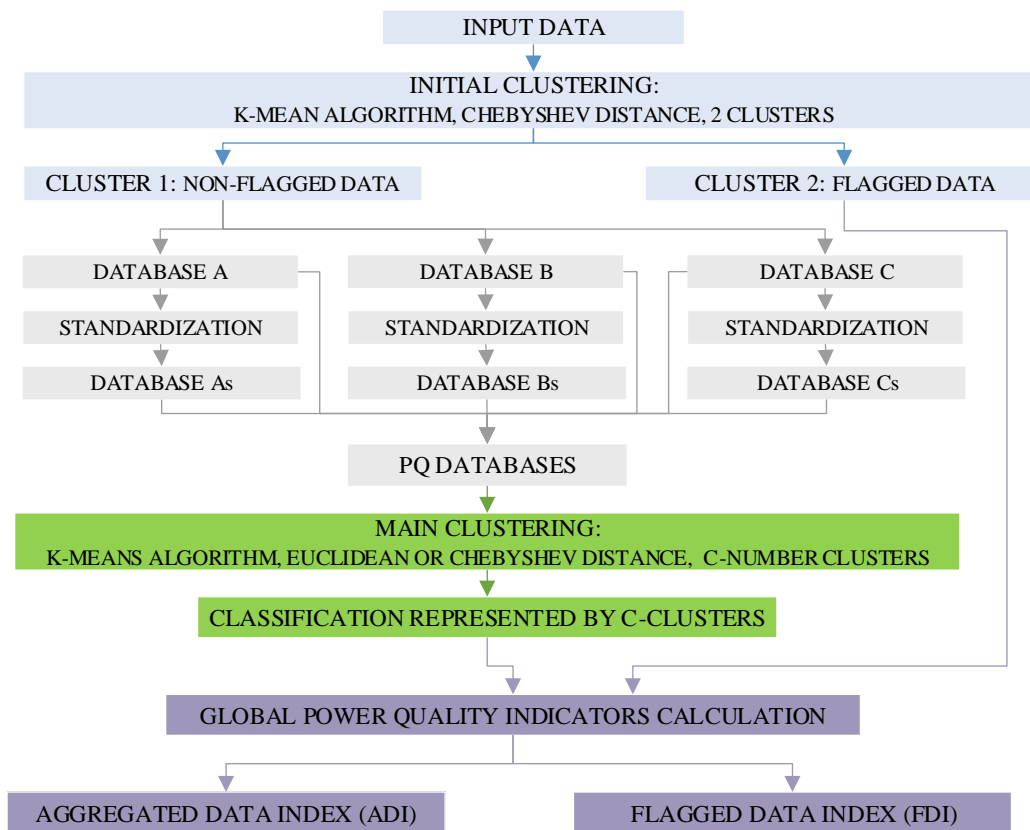
Figure 1. General scheme of the algorithm that performs the cluster analysis and global power quality (PQ) calculation.

The proposed method was implemented for the real measurements collected from four measurement points: three transformers T1, T2, T3 which supplied the medium voltage (MV) industrial network and a significant load (i.e., the welding machine—WM). The changes in the power demand of the investigated measurements points T1, T2, T3, and WM during the selected four weeks of observation are presented in Figure 3a. The investigation was aimed to evaluate the influence of the DGs installed inside the observed industrial network, and so Figure 3b presents changes in active power generation of particular DG units denoted as G1, G2, and G3. Generator G1 was permanently switched off during the experiment. G2 and G3 switched off, as can be seen in Figure 3b, due to a planned maintenance break. Additionally, it can be seen that during the experiment, only G2 (connected to the transformer T3 which also supplies the welding machine WM) and G3 (connected to transformer T2) were operating. The power variations of the DG is additional information, representing conditions. The data from the DG do not form the database of measurements taken for the investigation. An analysis of voltage events in the PQ measurements was conducted. Indicated events were voltage dips, rapid voltage changes, swells, and interruptions. Detailed information about the events and number of flagged data is included in Table 1 [52]. In accordance with the flagging concept introduced in the standard [7], the aggregated 10-min data that contained such voltage events were excluded from the power quality analysis. Based on the research presented in [9], it was shown that the best results of the CA with regards to the identification of different PQ conditions caused by the impact of the DGs could be achieved for the PQ databases denoted as C and C_S, where database C is constructed of frequency variation (f), voltage variation (U), short-term flicker severity (P_{st}), asymmetry (k_{u2}), total harmonic distortion in voltage ($THDu$), and active power level (P). Database C_S is the standardized version of database C, obtained by dividing the particular time series by their maximum values to achieve expression of the data in the range 0–1. Thus, for the investigation presented in this paper, database C and its standardized version Cs were taken for consideration.
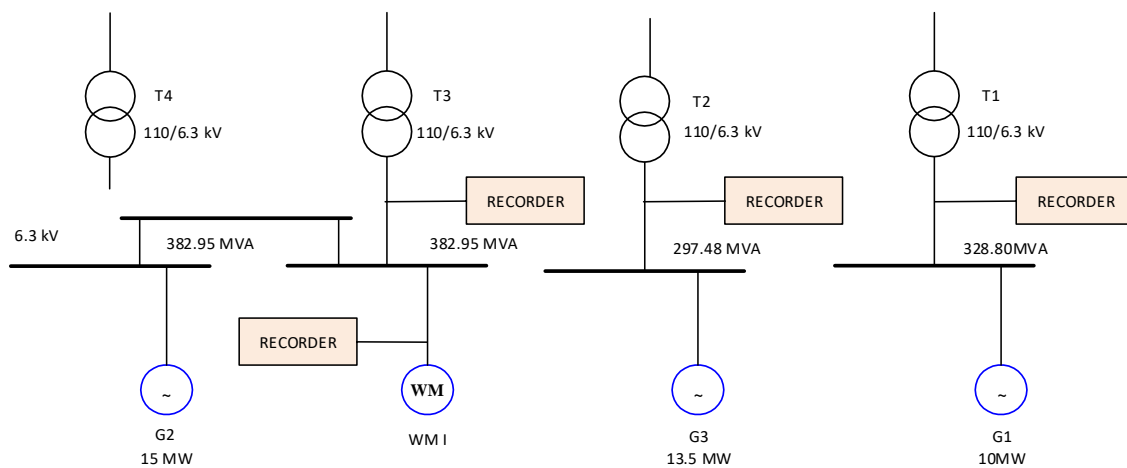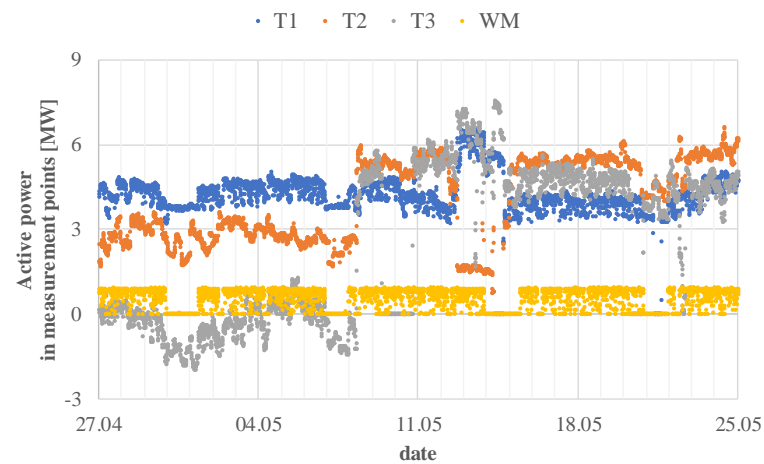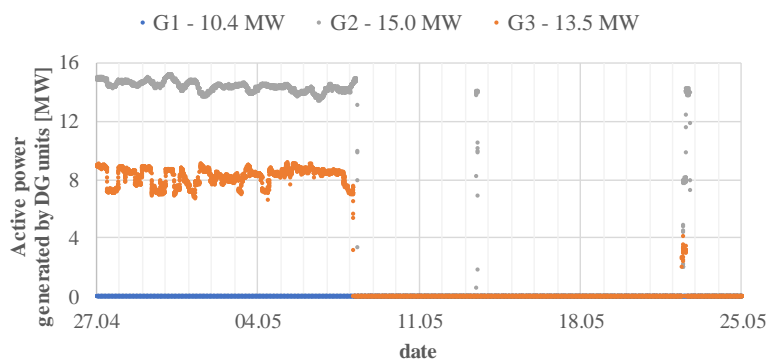
Figure 2. Analyzed mining industry electrical power network (EPN) and placement of distributed generation and PQ recorders [9]. T1—transformer 1; T2—transformer 2; T3—transformer 3; T4—transformer 4; G1—generator 1; G2—generator 2; G3—generator 3; WM—welding machine.

(a)

(b)

Figure 3. The analysis of active power level in electrical power network of mining industry: (**a**) Active power of the observed point of the measurements in the investigated network including the high voltage/medium voltage (HV/MV) transformers T1, T2, T3 and the connection point of the welding machine WM; (**b**) Active power of the distributed generators (DGs) during the investigated period of observation (G2 is connected to transformer T3, G3 is connected to transformer T2).

Table 1. Indication of events and number of flagged data during the measurements.

Point	Type of Event	Flag Start		Flag Stop		Number of 10-min Flagged Data
		Date	Hour	Date	Hour	
T1	long interruption, voltage dip, swell, rapid voltage change	21.05	07:50:00	21.05	15:10:00	45
T2	voltage dip	05.05	05:40:00	05.05	05:40:00	1
	long interruption, short interruption, voltage dip, swell, rapid voltage change	14.05	08:20:00	14.05	16:20:00	49
	voltage dip	05.06	14:00:00	05.06	14:00:00	1
T3	voltage dip	05.05	05:30:00	05.05	05:30:00	1
	rapid voltage change	07.05	22:40:00	07.05	22:40:00	1
	long interruption, voltage dip, swell, rapid voltage change	09.05	09:00:00	10.05	17:20:00	195
	long interruption, voltage dip, transient overvoltage, rapid voltage change	20.05	08:00:00	20.05	19:20:00	69
WM	voltage dip	05.05	05:30:00	05.05	05:30:00	1
	rapid voltage change	07.05	22:40:00	07.05	22:40:00	1
	short interruption, voltage dip, rapid voltage change	20.05	07:40:00	20.05	07:40:00	1

3.1. Cluster Analysis—Identification of the Power Quality Data Representing Different PQ Conditions Due to the Impact of DG

In [9], different results of the clustering were presented using different numbers of clusters (2, 3, 5, 20). It was shown that increasing the number of clusters enabled the identification of data not only related to the impact of the DG (i.e., when the DG was active or switched off), but also for the extraction of data associated with other working conditions (i.e., working day or non-working day, time of the network reconfiguration). This article aims to highlight the influence of distributed generation on power quality in the industry network. Thus, referring to the achievements presented in [9], in this work the scope of the CA was limited to the aim of classifying the data into three groups: cluster 1—DG was active; cluster 2—DG was switched off; cluster 3—other conditions. After the experiences described in [9], we decided to use the K-means algorithm with Euclidean distance.

In order to visualize the association of the obtained clusters with the distributed generation work information, Figure 4, which presents the clustering results, is supported by additional, artificial clusters indicated as cluster −1 and cluster 0, which were created on the basis of external information collected by the control and monitoring systems of particular DGs, as well as the output of the PQ monitoring systems considering the flagged data. Cluster −1 denotes the time series when the DG was active. This approach enables the easy comparison of the CA outcomes with regards to the identification of the working condition of the DGs. As was previously indicated, the databases are comprised only of unflagged data. Cluster 0 concerns flagged data that must be excluded from the main cluster analysis. The main clusters that are the outcomes of the CA analysis are cluster 1, which represents data when the DG was working, and cluster 2, which expresses the time period when the DG was switched off. Comparing the outcomes of the applied clustering with an artificial informative cluster denoted as −1 allows for the conclusion that the applied technique provides an appropriate output for connection of the clusters to different working states of the DG time period. Figure 4 presents the outcomes of the clustering with Euclidean distance when the initial number of clusters is 3. Referring to the information coming from external network dispatcher systems, it was confirmed that the time period indicated by cluster 3 was related to the reconfiguration of the network topology. In this case, increasing the number of clusters ensures the determination of a more sensitive classification of the collected PQ data when a specific working condition of the EPN is indicated. These and other issues concerning the initial number of clusters and the construction of the database were studied in [9]. However, it is important to note that the clustering is the first step in the multipoint long-term measurement analysis, which ensures a classification of the data into groups that are matched with the specific condition of the observed network. It finally leads to the possibility of the qualitative assessment of the data collected into clusters, as well as comparative analysis between the clusters. For this purpose, this paper proposes the use of global power quality indices.
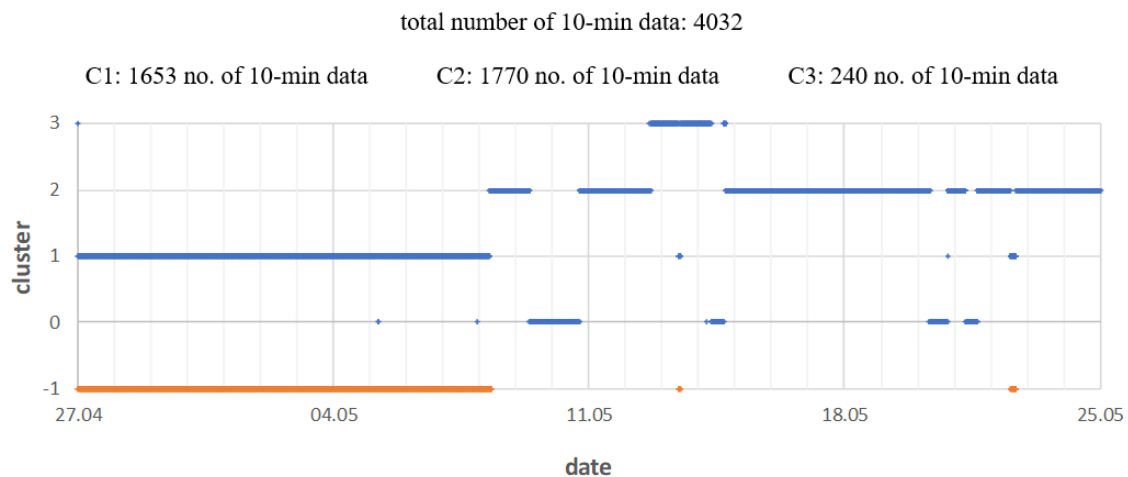
total number of 10-min data: 4032

Figure 4. Results of power quality data clustering using cluster analysis (CA) with K-means and Euclidean distance and three initial clusters. C1—the distributed generation (DG) was working; C2—the DG was switched off; C3—DG was switched off and with a different network topology configuration.

3.2. Qualitative Assessment of the Determined Clusters Based on the Proposed Global Power Quality Indices

As was described in Section 2, the proposed aggregated data index (*ADI*) uses five components based on 10-min aggregated data, and two other components based on 200-ms data. The acceptance levels for the *ADI* components, with regards to aggregated power quality parameters, are presented in Table 2. The values correspond to the demands included in the standard [6].

Table 2. The acceptance level of the components of the *ADI* related to 10-min aggregated power quality parameters in reference to [6].

Parameter	Value
Δf_{limit}	0.5 Hz
ΔU_{limit}	10%
Pst_{limit}	1.2
$ku2_{\text{limit}}$	2%
$THDu_{\text{limit}}$	8%

In the presented results, each importance rate $k_1 \div k_7$ (weighted factors) of the seven parameters comprising the *ADI* were the same and equal to 1/7. This means that the importance of all the parameters was treated equally. The 10-min step *ADI* variation for particular measurement points (T1, T2, T3, WM) in relation to the determined clusters of the PQ data is presented in Figure 5. In order to link the *ADI* variation with the output of the CA analysis, that is time periods which refer to particle clusters, colored backgrounds for particular clusters were inserted in Figure 5.
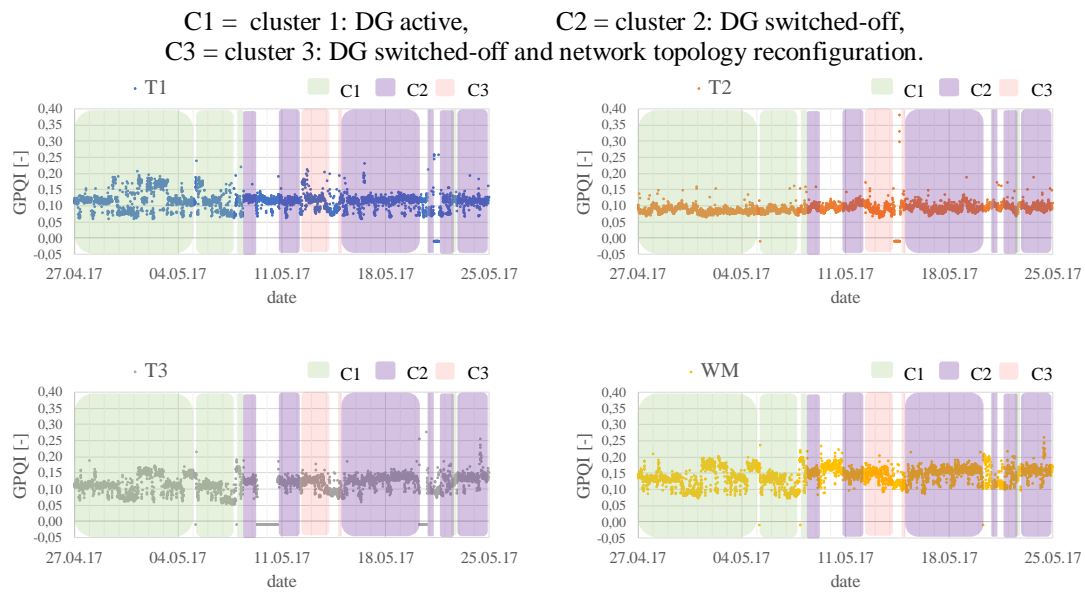
C1 = cluster 1: DG active,   C2 = cluster 2: DG switched-off,
C3 = cluster 3: DG switched-off and network topology reconfiguration.

Figure 5. *ADI* variation for particular measurement points (T1, T2, T3, and WM) with relation to determined clusters of the PQ data.

The lack of a background color means that the data were flagged. It can be noticed that changeability of the *ADI* for different working conditions (represented by clusters) is observable but very faint. Thus, the results of the power quality assessment using the proposed technique combining the CA global power quality indices means that *ADI* can be summarized by statistics of the *ADI* variation for particular clusters and measurement points. The results are collected in Table 3.

Table 3. Results of the assessment of the power quality using the proposed global power quality indices *ADI* and *FDI* for the particular measurement points and with relation to clusters 1–3 when full definition of the *ADI* index is implemented.

Cluster		Cluster 1—DG Working				Cluster 2—DG Switched Off				Cluster 3—DG Switched Off and With a Different Network Topology Configuration			
Measurement Point		T1	T2	T3	WM	T1	T2	T3	WM	T1	T2	T3	WM
ADI	Minimal value	0.119	0.087	0.115	0.130	0.060	0.069	0.060	0.084	0.071	0.062	0.075	0.092
	Mean value	0.121	0.089	0.105	0.137	0.113	0.098	0.124	0.148	0.114	0.091	0.116	0.142
	Maximum value	0.123	0.092	0.115	0.145	0.259	0.188	0.256	0.260	0.212	0.382	0.178	0.203
FDI (%)		0.00				0.15				0.17			

Comparative analysis of the *ADI* levels allows the formulation of the following remarks:

- Transformers T2 and T3, as well as the connection point of the welding machine WM, had the highest level of *ADI* for cluster 2 when the DG was switched off, and the lowest for cluster 1 when the DG was active. Distributed generation units were connected directly to T2 and T3 and the impact of the DGs was identified.
- Transformer T1 had relatively higher *ADI* values for cluster 1 when the DG was active, and the lowest for cluster 2 when the DG was switched off. However, there was active generation directly connected to transformer T1.
- The highest level of *ADI* was recognized in the outcoming feeder that supplies the welding machine which is a significant load with highly time-varying nature.

- Referring to cluster 1 when the DG was active, the *ADI* had the lowest level for T2, then T3, and the highest for T1.
- Referring to cluster 2 when the DG was switched off, the *ADI* had the lowest level for T2, and the highest for T3.
- Cluster 3 represents a short period of time (around 2 days) when all the DGs were switched off and some reconfiguration of the electrical power network connection was made. During the reconfiguration, transformer T1 was more loaded, and transformers T2 and T3 were less loaded. Comparing the *ADI* level during cluster 3, consisting of a period of time when there was a network reconfiguration with cluster 2, when the network was operating in the normal configuration, it can be seen that the values of the *ADI* decreased for T2, T3 and WM, and increased for T1.

To sum up, using the proposed cluster analysis and the proposed global power quality index, the *ADI* can be a suitable tool for the identification and comparative assessment of different conditions of the observed network. We revealed that for the observed transformers T2, T3, and the connection point of welding machine WM, the power quality was better in cluster 1 when the DG was active. The different outcomes of the *ADI* level formulated for transformer T1 could be caused by the fact that there was no DG directly connected to T1. The highest values of *ADI* were identified in the feeder supplying the welding machine.

The next global power quality index proposed in this work is the flagged data index (*FDI*), which is related to the number of aggregated data affected by the events in reference to the periods identified by clusters. Comparative analysis of the *FDI* levels is presented in Table 3. It allows for the formulation of a general remark that the *FDI* level was noticeable for cluster 2 and cluster 3. The high values for cluster 2 and cluster 3 are probably connected with the events caused by changes in the electrical power network topology.

Additionally, correlation calculations between each factor to the *ADI* value were realized for each point separately. Pearson correlation was used and the description of the coefficient was defined as [10]:

- $r_{xy} = 0$—no correlation;
- $0 < |r_{xy}| \leq 0.1$—slight correlation;
- $0.1 < |r_{xy}| \leq 0.4$—poor correlation;
- $0.4 < |r_{xy}| \leq 0.7$—noticeable correlation;
- $0.7 < |r_{xy}| \leq 0.9$—high correlation;
- $0.9 < |r_{xy}|$—strong correlation.

The correlation between factors and *ADI* are presented in Table 4. The generally noticeable correlations in each measurement point were indicated for P_{st} (W_3), *THDu* (W_5), U_{env} (W_6), and THD_{Umax} (W_7).

Table 4. Results of correlation analysis between each factor and the global index.

Measurement Point	W_1	W_2	W_3	W_4	W_5	W_6	W_7
T1	slight	slight	high	slight	high	noticeable	high
T2	slight	poor	noticeable	poor	noticeable	noticeable	high
T3	poor	slight	noticeable	poor	high	high	high
WM	poor	poor	noticeable	poor	high	noticeable	high

3.3. Influence of the Factors Comprising the Proposed Global Power Quality Indices on the Sensitivity of the Assessment

The construction of a global power quality index, *ADI*, understood as a weighted sum of component factors related to power quality parameters, inclines us to discuss the impact of individual factors on the assessment results. It is possible to select weighting coefficients in a way that favors the selected

parameters in the assessment and moves the center of gravity of the global assessment in the direction of the favorite parameters. The opposite direction is to enhance the sensitivity of the assessment by including additional parameters in the definition of global indices. This work proposes the construction of a global index using five basic 10-min parameters of power quality (frequency, root mean square (RMS) voltage, asymmetry, voltage fluctuations, and total harmonic distortion in voltage) and to extend the definition with two other parameters which are close to 200-ms values (i.e., the envelope of voltage changes and the maximum value of the total harmonic distortion in voltage identified during 10-min aggregation intervals). The aim of extending the *ADI* definition with parameters related to 200-ms intervals is to enhance the sensitivity of the obtained global index. In order to investigate the impact of the proposed 200-ms parameters on the sensitivity of the assessment, a differential approach is proposed. The *ADI* values for particular clusters and points of measurements when the full definition is involved are presented in Table 3. The results represent a scenario where all seven factors with the same weighting factors equal to $k_1 \div k_7 = 1/7$ were applied in the *ADI* calculation. Application of the full definition of *ADI* allowed us to conclude that for the observed transformers T2, T3, and the connection point of the welding machine WM, the power quality was better in cluster 1 when the DG was active. The obtained *ADI* values were generally smaller in cluster 1 than in cluster 2, and the differences of *ADI* between clusters 1 and 2 were consistently mostly negative. In order to perform a differential comparison between the *ADI* obtained using the full definition and the *ADI* based on a reduced definition, new values of the *ADI* were calculated where the parameters related to 200-ms values were neglected (i.e., when weighting factors were equal to $k_1 \div k_5 = 1/5$, $k_6 = 0$ and $k_7 = 0$, respectively). The obtained values of the *ADI* calculated without the 200-ms parameters are presented in Table 5.

Table 5. Results of the power quality assessment using the proposed global power quality index *ADI* for the selected measurement points, with relation to the revealed clusters when 200-ms parameters are neglected in the *ADI* definition.

Cluster		Cluster 1—DG Working				Cluster 2—DG Switched Off				Cluster 3—DG Switched Off and with a Different Network Topology Configuration			
Measurement point		T1	T2	T3	WM	T1	T2	T3	WM	T1	T2	T3	WM
ADI	Minimal value	0.106	0.095	0.099	0.124	0.052	0.064	0.056	0.090	0.071	0.062	0.075	0.092
	Mean value	0.110	0.095	0.097	0.133	0.099	0.099	0.111	0.144	0.114	0.091	0.116	0.142
	Maximum value	0.114	0.096	0.104	0.142	0.221	0.193	0.219	0.231	0.212	0.382	0.178	0.203

Instead of calculating the direct differences between the *ADI* values obtained for both scenarios (which actually differ very slightly), we propose a comparison between interpretations of the results. In other words, the sensitivity analysis was redirected to formulate the question of whether neglecting the 200-ms parameters in the *ADI* definition would change the interpretation of the assessment. Changes in the interpretation of the results can be identified if the signs of the difference of the *ADI*s applied for full and reduced definitions are different. For example, we found that the *ADI* obtained using the full definition decreased when the DG is active (C1—cluster 1) and increased when the DG was switched off (C2—cluster 2). The difference of the *ADI*s between C1 and C2 was negative because the values of the *ADI* in C2 predominated. If a reduction of the *ADI* parameters has an influence on the sign of the differences between the clusters, it means that the interpretation is not coherent and is dependent on the *ADI* construction. Table 6 contains information about the assessment results between the *ADI* with 200-ms factors ($k_1 \div k_7 = 1/7$) and without 200-ms factors ($k_1 \div k_5 = 1/5$, $k_6 = 0$, and $k_7 = 0$). Additionally, an interpretative logical comparative assessment index is introduced in the table. A value equal to 1 means that the assessment and interpretation of the results are the same for the full and reduced definitions of the *ADI*. A value equal to −1 means that the interpretations using full and reduced definitions of the *ADI* are not coherent.

Table 6. Assessment of the influence of removing two parameters related to 200-ms intervals in the *ADI* definition on the power quality assessment. The interpretation differences were highlighted.

ADI	Delta of Clusters	Min	T1 Mean	Max	Min	T2 Mean	Max	Min	T3 Mean	Max	Min	WM Mean	Max
with 200 ms	Δ C1−C2	0.059	0.008	−0.136	0.017	−0.009	−0.097	0.055	−0.019	−0.141	0.046	−0.010	−0.114
	Δ C1−C3	0.048	0.007	−0.089	0.025	−0.001	−0.290	0.040	−0.012	−0.063	0.038	−0.004	−0.058
	Δ C2−C3	−0.012	−0.001	0.047	0.007	0.007	−0.193	−0.014	0.008	0.078	−0.008	0.006	0.056
without 200 ms	Δ C1−C2	0.053	0.011	−0.107	0.030	−0.004	−0.097	0.043	−0.014	−0.115	0.034	−0.011	−0.089
	Δ C1−C3	0.034	−0.004	−0.097	0.033	0.005	−0.286	0.024	−0.020	−0.075	0.032	−0.009	−0.061
	Δ C2−C3	−0.019	−0.015	0.010	0.002	0.009	−0.188	−0.019	−0.006	0.041	−0.002	0.002	0.028
Logical comparative assessment		1	1	1	1	1	1	1	1	1	1	1	1
no change in assessment: 1		1	−1	1	1	−1	1	1	1	1	1	1	1
change in assessment: −1		1	1	1	1	1	1	1	−1	1	1	1	1

Based on the analysis presented in Table 6, it can be generally concluded that among the 36 assessments of the clusters, 3 differ in terms of the interpretation after a reduction of the *ADI* definition. In other words, a reduction of the *ADI* components introduced an 8% difference in the assessment. Alternatively, this means that including the parameters associated with the 200-ms values in the *ADI* definition enhances the sensitivity of the assessment.

To be more precise, the comparison of the differences of the *ADI* values constructed on seven and five parameters addressed to particular clusters were seen to deliver additional observations. For clusters 1 and 2, it can be concluded that interpretation results based on the *ADI*s were not sensitive to a reduction of *ADI* components, and the interpretation results were the same. This is due to the substantial differences between the power quality condition in clusters 1 and 2, which are reflected in the *ADI* values. However, when comparing clusters containing similar data, the reduction of the *ADI* components may cause differences in the assessment. An example of this can be seen with the data associated with transformer T3 in cluster 2 (DG switched off) and cluster 3 (DG switched off and with network reconfiguration). In this case, there was a significant impact of the DG; the power quality conditions were similar, and the reduction of *ADI* components brought differences in the interpretation in Table 6. This is denoted by the logical value −1. Another example can be seen in the differences between cluster 1 and cluster 3 in the case of transformer T2. The network configuration result was more loaded in transformer T1 and less loaded in transformer T2. In terms of transformer T2, it was generally a similar condition as for the impact of DG when the reduction of the load demand was also achieved. In this case, the power quality condition was similar for both clusters 1 and 3, and the reduction of the *ADI* components introduced uncertainty to the assessment.

It can be concluded that the reduction of the parameters comprising the synthetic *ADI* index influences the sensitivity of the assessment. In the case of the presented investigation, this inherent relation was more significant when the differences between the power conditions in the compared clusters were insignificant.

4. Discussion

This work presents the possibility of connecting CA and GPQIs. As indicated by the authors in a previous work [9] PQ measurements are an appropriate input to cluster analysis. Note that the aim of CA is to divide data based on its features. The proposed method was implemented for the real measurements collected from four measurement points in an industry network: three transformers T1, T2, T3 which supplied the MV industrial network, and a significant load (a welding machine, WM). The investigation aimed to evaluate the influence of the DGs installed inside the observed industrial network. However the power variations of the DGs are additional information, representing conditions. The data from the DGs do not create the database of measurements taken for the investigations. Naturally the same classification can be obtained using time identification representing different conditions of the DGs, but the point of the method is to obtain automatic classification of the PQ data based on its features, and then to find the reasons explaining the automatic classification. The presented approach has a crucial meaning when the number of monitored points is increased.

The input database consists of many different parameters, leading to a multielement assessment. Thus, in this work we proposed the use of global indices to simplify the process. The proposed indices consist of power quality parameters that represent frequency, voltage, flicker asymmetry factor, and harmonics in voltage. To classical 10-min aggregated data, we proposed adding the extremum values of voltage and harmonics. Thus, we conducted an analysis of the impact of extending the global indices to such values. Results indicated that our synthetic *ADI* index influenced the sensitivity of the assessment. In the case of the present investigation, this inherent relation was more significant when the differences between the power conditions in the compared clusters were insignificant. The composition of *ADI* index is based on classical 10-min PQ parameters as well as 200-ms parameters. Weighting factors were implemented for particular parameters. In order to reveal the influence of the DGs, all weighting factors were set to one in order to obtain maximum sensitivity of the analysis on every PQ

parameter collected in *ADI*. However, the weighted factors make it possible to focus the analysis more on particular PQ parameters and neglect others (i.e., to obtain an analysis more sensitive for selected PQ phenomena controlled by using different values of the weighting factor).

The proposed combination of CA and GPQIs was indicated as a suitable tool for the identification and comparative assessment of different conditions of the observed mining industry network. Among other things it was revealed that for the observed transformers T2, T3, and the connection point of welding machine WM, the power quality was better in cluster 1 when the DG was active. The different outcomes of the *ADI* level for transformer T1 could be caused by the fact that there was no DG directly connected to T1. The highest values of *ADI* were identified in the feeder supplying the welding machine, which is a high variable load. It can be concluded that obtained method is also technically reasonable.

We also proposed the flagged data index (*FDI*), which is related to the number of aggregated data affected by the events. It was used to compare clusters. Results concerning the use of the proposed global power quality index dedicated to voltage events (*FDI*) showed that the *FDI* was higher in cluster 2 than in cluster 1, which can be attributed to the fact that in the period of time when DG was active (cluster C1) there was relatively fewer detected voltage events than in the period when the DG was switched off (cluster 2). The sense of the *FDI* is general. Detailed analysis of particular voltage events requires separate investigations.

5. Conclusions

This article presents a method of analyzing long-term PQ data using a combined technique based on cluster analysis and newly proposed global power quality indices. The presented investigations were based on multipoint synchronized real measurements performed in a medium voltage electrical power network with distributed generation supplying the mining industry. Time-varying PQ conditions were intentionally created during the experiment when the distributed generation was switched on and off for some period of time, with a network reconfiguration also being performed.

The cluster analysis is the first step of the proposed method and is used for identification of the PQ data which represent different conditions. It was shown that cluster analysis with K-means and Euclidean distance successfully allowed for the identification of portions of PQ data related to the impact of distributed generation (switched on and switched off) and changes to the network configuration. Basic investigations of the application of cluster analysis in an electrical power network were presented by the authors in a previous work [9]. The extension of the mentioned work and the novelty involved in the proposed method lies in extending the cluster analysis by assessing the obtained portions of PQ data using global power quality indices. In order to achieve the goal, newly proposed global power quality indices were provided, including the aggregated data index and flagged data index. The proposed aggregated data index has a synthetic formula and is based on five classical 10-min aggregated power quality parameters and two parameters that demonstrate 200-ms values, including the envelope of voltage changes and the maximum of total harmonic distortion in the voltage. In this work, the proposed global indices were used for comparative assessment of identified clusters, which in turn demonstrated different states of the network condition: active distributed generation, switched off generation, and network reconfiguration when the generation was switched off. It was shown that the use of the proposed global power quality indices resulted in the comparative analysis between particular clusters being successfully performed.

Additionally, a sensitivity analysis of the synthetic aggregation data index was also proposed. It can be concluded that a reduction of the parameters comprising the synthetic global power quality index may influence the results of the assessment. In the case of the presented investigation, this inherent relation was more significant when the differences between power conditions in the compared clusters were insignificant.

The presented approach can be treated as an effective tool (not only related to power quality) for the assessment of long-term multipoint measurements. The advantages of the proposed method are the

automatic classification of the data into clusters and the assessment of the condition of the identified group of data in a parametric global sense, which makes the comparative assessment easier and more intuitive. The proposed technique has the potential for further implementation in the analysis and optimization of energy processes, and also in the development of sustainable energy systems.

Nomenclature:

ADI	aggregated data index
C	database for non-standardized data
Cs	database for standardized data
C	number of classes or clusters
DG	distributed generation
GPQI	global power quality index
k_i	importance rate (weighted factors) of a particular power quality factor constituting the synthetic aggregated data index, range of [0, 1]
k_{u2}	asymmetry
P	active power
P_{lt}	long-term flicker severity
P_{st}	short-term flicker severity
PQ	power quality
THD	total harmonic distortion
U	voltage variation
W_i	particular power quality factors comprising the synthetic aggregated data index
WM	welding machine

References

1. International Energy Agency. *Key World Energy Statistics*; International Energy Agency: Paris, France, 2017.
2. Nieto, A.; Vita, V.; Ekonomou, L.; Mastorakis, N.E. Economic Analysis of Energy Storage System Integration with a Grid Connected Intermittent Power Plant, for Power Quality Purposes. *WSEAS Trans. Power Syst.* **2016**, *11*, 65–71.
3. Nieto, A.; Vita, V.; Maris, T.I. Power Quality Improvement in Power Grids with the Integration of Energy Storage Systems. *Int. J. Eng. Res. Technol.* **2016**, *5*, 438–443.
4. Ceaki, O.; Seritan, G.; Vatu, R.; Mancasi, M. Analysis of power quality improvement in smart grids. In Proceedings of the 2017 10th International Symposium on Advanced Topics in Electrical Engineering (ATEE), Bucharest, Romania, 23–25 March 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 797–801.
5. Seritan, G.; Tristiu, I.; Ceaki, O.; Boboc, T. Power quality assessment for microgrid scenarios. In Proceedings of the 2016 International Conference and Exposition on Electrical and Power Engineering (EPE), Iasi, Romania, 20–22 October 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 723–727.
6. *EN 50160: Voltage Characteristics of Electricity Supplied by Public Distribution Network*; British Standards: London, UK, 2015.
7. *IEC 61000 4-30 Electromagnetic Compatibility (EMC)–Part 4-30: Testing and Measurement Techniques–Power Quality Measurement Methods*; International Electrotechnical Commission: Geneva, Switzerland, 2015.
8. *IEEE P1159: Monitoring and Definition of Electric Power Quality*; IEEE: Piscataway, NJ, USA, 2009.

9.   Jasiński, M.; Sikorski, T.; Borkowski, K. Clustering as a tool to support the assessment of power quality in electrical power networks with distributed generation in the mining industry. *Electr. Power Syst. Res.* **2019**, *166*, 52–60. [CrossRef]

10. Jasiński, M.; Sikorski, T.; Kostyła, P.; Kaczorowska, D.; Leonowicz, Z.; Rezmer, J.; Szymańda, J.; Janik, P.; Bejmert, D.; Rybiański, M.; et al. Influence of Measurement Aggregation Algorithms on Power Quality Assessment and Correlation Analysis in Electrical Power Network with PV Power Plant. *Energies* **2019**, *12*, 3547. [CrossRef]

11. Borges, F.A.S.; Fernandes, R.A.S.; Silva, I.N.; Silva, C.B.S. Feature Extraction and Power Quality Disturbances Classification Using Smart Meters Signals. *IEEE Trans. Ind. Inform.* **2015**, *12*, 824–833. [CrossRef]

12. Erişti, H.; Demir, Y. Automatic classification of power quality events and disturbances using wavelet transform and support vector machines. *IET Gener. Transm. Distrib.* **2012**, *6*, 968–976. [CrossRef]

13. Jamali, S.; Farsa, A.R.; Ghaffarzadeh, N. Identification of optimal features for fast and accurate classification of power quality disturbances. *Meas. J. Int. Meas. Confed.* **2018**, *116*, 565–574. [CrossRef]

14. CIGRE. *Broshure 292: Data Mining Techniques and Applications in the Power Transmission Field*; CIGRE: Paris, France, 2006.

15. Xi, Y.; Li, Z.; Zeng, X.; Tang, X.; Liu, Q.; Xiao, H. Detection of power quality disturbances using an adaptive process noise covariance Kalman filter. *Digit. Signal Process. A Rev. J.* **2018**, *76*, 34–49. [CrossRef]

16. Kanirajan, P.; Suresh Kumar, V. Power quality disturbance detection and classification using wavelet and RBFNN. *Appl. Soft Comput. J.* **2015**, *35*, 470–481. [CrossRef]

17. Mahela, O.P.; Shaik, A.G. Power quality recognition in distribution system with solar energy penetration using S-transform and Fuzzy C-means clustering. *Renew. Energy* **2017**, *106*, 37–51. [CrossRef]

18. Gu, T.; Kadurek, P.; Cobben, J.F.G.; Endhoven, A.W. Power quality data evaluation in distribution networks based on data mining techniques. In Proceedings of the 2013 12th International Conference on Environment and Electrical Engineering, Wroclaw, Poland, 5–8 May 2013; IEEE: Piscataway, NJ, USA, 2013; pp. 58–63.

19. Gao, Q.; Pan, F.; Yuan, F.; Pan, J.; Zhang, J.; Zhang, Y. The Classification of Multiple Power Quality Disturbances Based on Dynamic Event Tree and Support Vector Machine. In *International Conference on Advanced Machine Learning Technologies and Applications*; Springer: Cham, Switzerland, 2020; pp. 321–329.

20. Bagheri, A.; Bollen, M.H.J.; Gu, I.Y.H. Improved characterization of multi-stage voltage dips based on the space phasor model. *Electr. Power Syst. Res.* **2018**, *154*, 319–328. [CrossRef]

21. Ahire, P.G.; Patil, P.D.; Patil, R.A.; Golande, A.L. An Approach to Clustering of Power Quality Events Data. *Proc. Int. J. Eng. Innov. Technol.* **2012**, *2*, 164–168.

22. Mittal, D.; Mahela, O.P.; Jain, R. Classification of Power Quality Disturbances in Electric Power System: A Review. *Iosr-Jeee* **2012**, *3*, 6–14. [CrossRef]

23. Shaik, A.G.; Mahela, O.P. Power quality assessment and event detection in hybrid power system. *Electr. Power Syst. Res.* **2018**, *161*, 26–44. [CrossRef]

24. Meier, R.; McCamish, B.; Cotilla-Sanchez, E.; Landford, J.; Bass, R.B.; Chiu, D. Event Detection Using Correlation within Arrays of Streaming PMU Data. In Proceedings of the 2018 IEEE Power & Energy Society General Meeting (PESGM), Portland, OR, USA, 5–10 August 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.

25. Guder, M.; Salor, O.; Çadirci, N.; Ozkan, B.; Altintas, A. Data Mining Framework for Power Quality Event Characterization of Iron and Steel Plants. *Proc. Ind. Appl. Soc. Annu. Meet* **2015**, *51*, 3521–3531. [CrossRef]

26. Dent, I.; Aickelin, U.; Rodden, T. The Application of a Data Mining Framework to Energy Usage Profiling in Domestic Residences Using UK Data. *SSRN Electron. J.* **2011**, 1–10. [CrossRef]

27. Asheibi, A.; Stirling, D.; Robinson, D. Identification of Load Power Quality Characteristics using Data Mining. In Proceedings of the 2006 Canadian Conference on Electrical and Computer Engineering, Ottawa, ON, Canada, 7–10 May 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 157–162.

28. Zhou, K.; Yang, S.; Shen, C. A review of electric load classification in smart grid environment. *Renew. Sustain. Energy Rev.* **2013**, *24*, 103–110. [CrossRef]

29. Alfares, H.K.; Nazeeruddin, M. Electric load forecasting: Literature survey and classification of methods. *Int. J. Syst. Sci.* **2002**, *33*, 23–34. [CrossRef]

30. Kotriwala, A.M.; Hernandez-Leal, P.; Kaisers, M. Load Classification and Forecasting for Temporary Power Installations. In Proceedings of the 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Sarajevo, Bosnia-Herzegovina, 21–25 October 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.

31. Azizi, E.; Ghaemi, S.; Mohammadi-Ivatloo, B.; Piran, M.J. Application of comparative strainer clustering as a novel method of high volume of data clustering to optimal power flow problem. *Int. J. Electr. Power Energy Syst.* **2019**, *113*, 362–371. [CrossRef]

32. Stankovic, S.; Soder, L. Optimal Power Flow Based on Genetic Algorithms and Clustering Techniques. In Proceedings of the 2018 Power Systems Computation Conference (PSCC), Dublin, Ireland, 11–15 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–7.

33. Huynh, V.; Ngo, V.; Le, D.; Nguyen, N. Probabilistic Power Flow Methodology for Large-Scale Power Systems Incorporating Renewable Energy Sources. *Energies* **2018**, *11*, 2624. [CrossRef]

34. Galvani, S.; Choogan, M. Data clustering based probabilistic optimal power flow in power systems. *IET Gener. Transm. Distrib.* **2019**, *13*, 181–188. [CrossRef]

35. Florencias-Oliveros, O.; Aguera-Perez, A.; Gonzalez-De-La-Rosa, J.J.; Palomares-Salas, J.C.; Sierra-Fernandez, J.M.; Montero, A.J. Cluster analysis for Power Quality monitoring. In Proceedings of the 2017 11th IEEE International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), Cadiz, Spain, 4–6 April 2017.

36. Lin, S.; Xie, C.; Tang, B.; Liu, R.; Pan, A.; Grid, S.; Municipal, S.; Power, E. The Data Mining Application in the Power Quality Monitoring Data Analysis. In Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications (ICIEA), Hefei, China, 5–7 June 2016; pp. 338–342.

37. Seera, M.; Lim, C.P.; Loo, C.K.; Singh, H. A modified fuzzy min-max neural network for data clustering and its application to power quality monitoring. *Appl. Soft Comput. J.* **2015**, *28*, 19–29. [CrossRef]

38. Seera, M.; Lim, C.P.; Loo, C.K.; Singh, H. Power Quality Analysis Using a Hybrid Model of the Fuzzy Min–Max Neural Network and Clustering Tree. *IEEE Trans. Neural Networks Learn. Syst.* **2016**, *27*, 2760–2767. [CrossRef] [PubMed]

39. Sacasqui, M.; Luyo, J.; Delgado, A. A Unified Index for Power Quality Assessment in Distributed Generation Systems Using Grey Clustering and Entropy Weight. In Proceedings of the 2018 IEEE ANDESCON, Santiago de Cali, Colombia, 22–24 August 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–4.

40. Elbasuony, G.S.; Abdel Aleem, S.H.E.; Ibrahim, A.M.; Sharaf, A.M. A unified index for power quality evaluation in distributed generation systems. *Energy* **2018**, *149*, 607–622. [CrossRef]

41. Song, J.; Xie, Z.; Zhou, J.; Yang, X.; Pan, A. Power quality indexes prediction based on cluster analysis and support vector machine. *CIRED-Open Access Proc. J.* **2017**, *2017*, 814–817. [CrossRef]

42. Florencias-Oliveros, O.; Agüera-Pérez, A.; González-de-la-Rosa, J.J.; Palomares-Salas, J.-C.; Sierra-Fernández, J.-M. A novel instrument for power quality monitoring based in higher-order statistics: A dynamic triggering index for the smart grid. *Renew. Energy Power Qual. J.* **2017**, *1*, 43–48. [CrossRef]

43. Singh, L.P.; Deswal, S.S.; Bhatia, R.; Jain, D.K. A New Approach for Analysing Voltage Sag Severity Based on Power Quality Indices. *Int. J. Emerg. Technol.* **2004**, *4*, 1–5.

44. Ignatova, V.; Villard, D.; Hypolite, J.M. Simple indicators for an effective Power Quality monitoring and analysis. In Proceedings of the 2015 IEEE 15th International Conference on Environment and Electrical Engineering, EEEIC 2015-Conference Proceedings, Rome, Italy, 10–13 June 2015; pp. 1104–1108.

45. Nourollah, S.; Moallem, M. A Data Mining Method for Obtaining Global Power Quality Index. In Proceedings of the 2011 2nd International Conference on Electric Power and Energy Conversion Systems (EPECS), Sharjah, United Arab Emirates, 15–17 November 2011; pp. 1–7.

46. Raptis, T.E.; Vokas, G.A.; Langouranis, P.A.; Kaminaris, S.D. Total Power Quality Index for Electrical Networks Using Neural Networks. *Energy Procedia* **2015**, *74*, 1499–1507. [CrossRef]

47. Lee, S.; Park, J.; Member, S.; Venayagamoorthy, G.K. New Power Quality Index in a Distribution Power System by Using RMP Model. *IEEE Trans. Ind. Appl.* **2010**, *46*, 1204–1211.

48. Hanzelka, Z.; Firlit, A.; Błajszczak, A. Syntetyczne miary jakości napięcia. *Autom. Elektr. Zakłócenia* **2012**, *3*, 62–70.

49. Caramia, P.; Carpinelli, G.; Verde, P. *Power Quality Indices in Liberalized Markets.pdf*; John Wiley & Sons: Hoboken, NJ, USA, 2009; ISBN 9780470033951.

50. Jasiński, M.; Sikorski, T.; Kostyla, P.; Borkowski, K. Global power quality indices for assessment of multipoint power quality measurements. In Proceedings of the Electronics, Computers and Artificial Intelligence, Iasi, Romania, 28–30 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6.

51. Jasiński, M.; Borkowski, K.; Sikorski, T.; Kostyla, P. Cluster Analysis for Long-Term Power Quality Data in Mining Electrical Power Network. In Proceedings of the 2018 Progress in Applied Electrical Engineering (PAEE), Koscielisko, Poland, 18–22 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–5.
52. Jasiński, M.; Sikorski, T.; Borkowski, K. Application of cluster analysis to identification flagged power quality measurements in area-related approach. [Zastosowanie eksploracji danych do identyfikacji oznaczonych wyników pomiaru jakosci energii elektrycznej w ujeciu obszarowym]. *Prz. Elektrotechniczny* **2020**, *3*, 9–12.