A Case Study on a Hierarchical Clustering Application in a Virtual Power Plant: Detection of Specific Working Conditions from Power Quality Data

Michał Jasiński [1,*], Tomasz Sikorski [1], Dominika Kaczorowska [1,*], Jacek Rezmer [1], Vishnu Suresh [1], Zbigniew Leonowicz [1], Paweł Kostyła [1], Jarosław Szymańda [1], Przemysław Janik [2], Jacek Bieńkowski [2] and Przemysław Prus [2]

[1] Faculty of Electrical Engineering, Wroclaw University of Science and Technology, 50-370 Wroclaw, Poland; tomasz.sikorski@pwr.edu.pl (T.S.); jacek.rezmer@pwr.edu.pl (J.R.); vishnu.suresh@pwr.edu.pl (V.S.); zbigniew.leonowicz@pwr.edu.pl (Z.L.); pawel.kostyla@pwr.edu.pl (P.K.); jaroslaw.szymanda@pwr.edu.pl (J.S.)

[2] TAURON Ekoenergia Ltd., 58-500 Jelenia Góra, Poland; przemyslaw.janik@tauron-ekoenergia.pl (P.J.); jacek.bienkowski@tauron-ekoenergia.pl (J.B.); przemyslaw.prus@tauron-ekoenergia.pl (P.P.)

[*] Correspondence: michal.jasinski@pwr.edu.pl (M.J.); dominika.kaczorowska@pwr.edu.pl (D.K.); Tel.: +48-713-202-022 (M.J.); +48-713-202-901 (D.K.)

Abstract: The integration of virtual power plants (VPP) has become more popular. Thus, research on VPP for different issues is highly desirable. This article addresses power quality issues. The presented investigation is based on multipoint, synchronic measurements obtained from five points that are related to the VPP. This article provides a proposition and discussion of using one global index in place of the classical power quality (PQ) parameters. Furthermore, in the article, one new global power quality index was proposed. Then the PQ measurements, as well as global indexes, were used to prepare input databases for cluster analysis. The mentioned cluster analysis aimed to detect the short-term working conditions of VPP that were specific from the point of view of power quality. To realize this the hierarchical clustering using the Ward algorithm was realized. The article also presents the application of the cubic clustering criterion to support cluster analysis. Then the assessment of the obtained condition was realized using the global index to assure the general information of the cause of its occurrence. Furthermore, the article noticed that the application of the global index, assured reduction of database size to around 74%, without losing the features of the data.

Keywords: virtual power plant; power quality; data mining; clustering; distributed energy resources; energy storage systems; short term conditions

1. Introduction

The concept of renewable energy sources (RES) and energy storage systems (ESS) integration into virtual power plants (VPP) includes different areas. This investigation concerns power quality (PQ) and data mining (DM) issues in VPP.

The article [1] is concerned with standard IEC 61850. It proposed an extension to this standard as a step to enhance the interaction between the controller of RESs and VPPs. As one of the elements, the power quality recorders' issues were included. The demands for them in point of IEC 61850 are discussed. Finally, the proposed methodology was verified in the virtual power plant that consists of HPP, PV, and wind power plant as well as storage systems. The indicated VPP operates on a medium voltage (MV) level. In Pudjianto et al. [2], the virtual power plant is treated as an instrument to enable a cost efficient integration of RES with the present power systems. The article includes the performance analysis of a VPP system from the point of view of different indicators such as energy efficiency, power quality, and security. The analyzed case consists of fuel cells, a

wind microturbine PV system, a fly wheel, a combined heat and power plant, and a storage system. Caldon et al. [3] propose the applicable framework for harmonizing operations of different VPP units. The authors indicated that decisions and profits although complying with power quality requirements and real network constraints must be a significant part of VPP operation strategies. Zhang et al. [4] considered the coordinated operation of VPPs. The bi objective dispatch model was applied for the performance optimization of multi energy VPP in terms of both economic and power quality issues. The article concerned both simulations based on a 118-node IEEE test system and a real case from Hongfeng Eco-town in China. Gong et al. [5] proposed the optimization issues concerning VPP management strategies. The customers' satisfaction, system stability, and PQ were used together with the economic objective to formulate a multi-objective optimization problem. The fuzzy multiple objective optimizations were applied to solve optimization problems. The proposed approach was verified in a test system. Beguin et al. [6] presented simulations for an islanded grid model of the virtual power plant. The VPP integrates a 0.2 GW wind power plant, a 0.1 GW PV power plant, and a 0.25 GW pumped storage system. The outcomes of the research were control strategies of the storage plant. The control strategies were investigated to highlight VPP on PQ.

Another element of the investigation is using data mining (DM) techniques to obtain knowledge from dataset. In Ref. [7], one of the DM techniques available is cluster analysis (CA). The current research connecting CA and VPP are as follows. Luo et al. [8] proposed demand response schemes based on data mining for electricity trading between VPPs and their participants. Participants' bid offers were used to categorize them using ordering points to identify the clustering structure (OPTICS) algorithm. Yi et al. [9] presented multi time scale scheduling for VPP. The scheduling contains both a day ahead bidding and real-time operations. The proposed models were based on the deferrable loads' aggregation connected with the k-mean cluster analysis. The proposed strategy enables efficient management of massive deferrable loads. Then it results in a reduction of the energy management complexity, and increases the general economics. This approach seems promising for s efficient scheduling of VPPs. Kong et al. [10] presented the robust stochastic optimal dispatching approach to solve scheduling issues. K-medoids cluster analysis was used to define typical scenarios of different units, that were integrated into VPP. Ai et al. [11] proposed a VPP load curve cluster analysis approach. It uses the principal-component dimension-reduction analysis, aggregation level clustering, and fc-means clustering. The principal component analysis is applied to determine different loads, that are aggregated into a VPP. Then hierarchical aggregation and fc-means cluster analysis was applied to divide all load output curves participating in the aggregation process. The last step of the investigation was the analysis of clustering results to establish an evaluation system. The article [12] proposes the methodology concerning distributed energy sources management. It concerns resource scheduling, aggregation, as well as remuneration. To realize the aggregation process, the k-means algorithm was proposed. The calculations of the clusters optimum number were used to determine the best number of demand response programs to be implemented by the VPP units. The research was based on 20,000 consumers and 500 DG units. Then, in article [13], the proposed methodology was extended to investigate other 2592 operation scenarios.

Under this investigation, a VPP that operates at both low voltage (LV) and medium voltage (MV) distribution networks in Poland was selected. The investigated part of the VPP consists of a 1250 kW hydropower plant (HPP) and associated 500 kWh battery energy storage system (BESS) as well as low voltage loads. The investigation is concerned with power quality issues in the selected part of the VPP. The PQ measurements were performed synchronically in five measurement points of the VPP area. The measurement points include the HPP, the BESS, the associated MV line, and two LV loads. The measurements were conducted for 182 days: from 1 May 2020 to 28 October 2020. Therefore, they are 26 weeks long from the point of view of classical PQ assessment [14].

The single parameter analysis of each measurement point for such a period of time would be very time-consuming. Thus, the concept of the global index was introduced. Such an approach, in the literature, is known under different names e.g., global power quality index [15,16]; unified power quality index [17,18]; total power quality index [19,20]; or synthetic power quality index [21,22].

This article applied the ADI index proposed in [16], and the newly proposed power quality pollution index (PQPI). Then both the classic PQ parameters and the global indexes were used to define datasets for cluster analysis. Global indexes were applied to reduce the size of the input database, without losing features of the PQ data. Then for those databases, hierarchical clustering was performed. Cluster analysis aimed to detect short term specific working conditions of the VPP from the point of power quality. As a tool to realize this, the cubic clustering criterion was applied for results assessment of hierarchical cluster analysis with the Ward algorithm for indicated databases. Finally, the application of PQPI was presented to highlight the difference from the point of PQ level for clusters.

The contributions of this research are as follows:

- The source of the data was multipoint, synchronic, and long-term power quality measurements, that were obtained from a real VPP.
- The global index approach for PQ issues was discussed, and a new index is proposed.
- The proposed input databases to cluster analysis are concerned with raw PQ data and global indexes. Global indexes were proposed to reduce the size of the input databases but the reduction has been realized while maintaining existing features of the PQ data.
- The cubic clustering criterium for hierarchical cluster analysis results was used to detect short-term working conditions of VPP, that were specific in point of power quality.
- The global index was used for comparative assessment between clusters.

To realize those contributions, the article is organized into five sections. In Section 2 the virtual power plant description, global index proposition, clustering methods, and input databases are presented. Section 3 presents the results of specific working conditions detection in view of power quality using hierarchical clustering. Section 4 presents a discussion of the results. Section 5 concludes the article.

2. Methodology and Research Object Description

This section is based on four elements. The first element is a description of the real VPP, that operates in Poland. This VPP became a source for long term, synchronic power quality measurements. Then, the global index approach was discussed. Next the long-term measurements and the global index were combined to obtain different datasets. These datasets consisted of classic power quality parameters and global power quality indexes. Then those datasets were used as an input for hierarchical clustering. The assessment of cluster assignment for those measurement data was realized using cubic clustering criterion (CCC). The CCC were applied to select the adequate number of clusters, that will indicate short-term specific working conditions from a power quality point of view. To summarise the proposed approach, the simplified methodology scheme was presented in Figure 1.
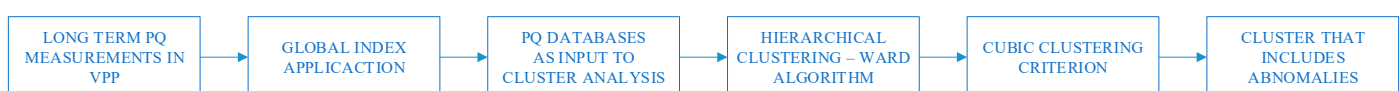

Figure 1. Simplified methodology scheme.

2.1. Virtual Power Plant That Operates in Poland as a Source of Power Quality Measurements

This article deals with a real VPP that operates in Poland, in a region called Lower Silesia. The virtual power plant consists of a fragment of the distribution network on both medium voltage (MV) and low voltage (LV) [23]. The two substations 110/20 kV are used

as points of connection to the 110 kV polish grid. However, under this investigation one MV network was selected. The network fed from the 110/20 kV station is an overhead cable network. The selected network has earth fault current compensation [24]. The main distributed energy resources that are integrated into the virtual power plant are a 1.25 MW HPP and a 0.5 MW battery ESS. Both are connected to a medium voltage level.

The scheme of the investigated fragment of the VPP is presented in Figure 1. The analyzed fragment consists of a 20 kV distribution network with a 1.25 MW hydropower plant (HPP_MV) and an 0.5 MW battery ESS (ESS_MV). Those energy sources are connected with the HV/MV substation by a 20 kV line (Line_MV). Additionally, representatives of low voltage loads are indicated: LOAD1_LV and LOAD2_LV. LOAD1_LV is connected with the indicated medium voltage line (LINE_MV). LOAD2_LV is connected with the node of the HPP_MV and ESS_MV. This fragment of the VPP is monitored by power quality recorders. Power quality recorders are indicated as "R". The location of these recorders is also included in Figure 2. The HPP_MV and ESS_MV are connected to one node and their PQ recorders use the same voltage transformer. Thus, in this research, they are treated as one point for the PQ level (HPP and ESS_MV) and another for the active power level (HPP_MV and ESS_MV).
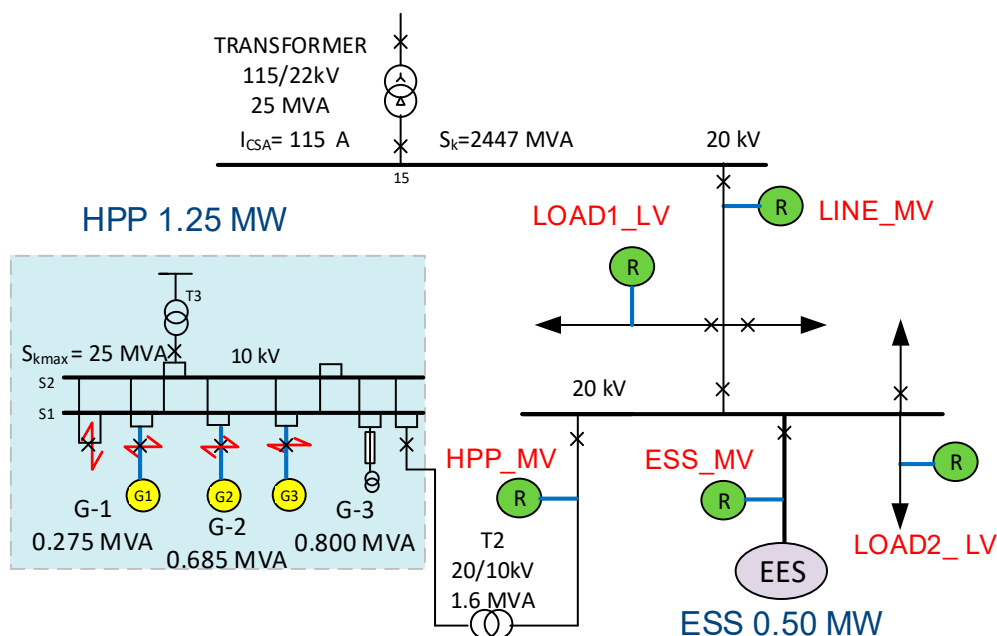

Figure 2. The virtual power plant with the power quality (PQ) recorders placements. Where: LINE_MV: medium voltage line that connects the virtual power plant (VPP) to 110/20 substation; HPP_MW: 1.25 MW hydropower plant; ESS_MW: 0.5 MW electric energy storage; LOAD1_LV: low voltage load related with LINE_MV; LOAD2_LV: low voltage load related with the hydropower plant (HPP) and ESS substation.

2.2. Global Power Quality Index

Recent research on power quality considers different areas. One of these areas is a simplification of the assessment using global values. Thus this article is concerned with the application of the global index-aggregated data index (ADI) [16], as well as the proposition of a new power quality pollution index (PQPI). ADI consists of seven 10-min PQ parameters–frequency (f), voltage (U), an envelope of voltage, short term flicker severity (Pst), unbalance (ku2), total harmonic distortion in voltage (THDu), and maximum harmonic distortion.

However, it was decided to exclude frequency as a customization step of the index to VPP issues as proposed in [16]. Thus, the ADI index corresponds to:

- voltage indicator,
- an envelope of voltage deviation obtained by the difference between the maximum and minimum of 200-ms U values identified during the 10-min aggregation interval,
- short term flicker severity indicator,
- asymmetry indicator,
- total harmonic distortion in voltage indicator,
- a maximum of the 200-ms value of total harmonic distortion of voltage indicator, identified in the 10-min aggregation interval [16].

The mentioned indicators are in response to a 10-min aggregation interval proposed by standard IEC 61000-4-30 [25]. They use the mean value from three phase values to calculate one as a representative. Those factors of ADI are based on the differences between the measured 10-min data and the recommended limit as a division. The limits may be taken differently and based on the object. For VPP that operates in the distribution network standard EN 50,160 [26] was selected. The applied limits based on EN 50,160 [26] are:

- voltage—10% of the declared voltage,
- short term flicker severity—1.0,
- unbalance—2%,
- total harmonic distortion in voltage—8%.

ADI index responds to the voltage level as well as an envelope of voltage. The same situation is with total harmonic distortion and maximum total harmonic distortion. So in view of data features, they are similar. Thus, in this article, the power quality pollution index (PQPI) is representative of the ADI factor but with the reduction of voltage and harmonic distortion indicators. This reduction would retain the data features of those parameters using the envelope of voltage and maximal THDu. Thus, PQPI includes following indicators:

- voltage distortion that responds to the envelope of voltage,
- unbalance distortion that responds to the asymmetry of voltage,
- flicker distortion that responds to short term flicker severity,
- harmonic distortion that responds to maximal total harmonic distortion in voltage.

2.3. Input Databases Description

During the investigation, three different databases were analyzed. The data for each parameter or indicator were used in a 10-min aggregation interval. Generally, the applied variables represented classical PQ parameters, global indexes as well as the active power level (P). The indicated databases are presented in Table 1.

Table 1. Description of input databases.

Database	Parameters	Number of Variables for Each Measurement Point That Describe Every 10-Min Data
Database A	PQ parameters data + P: consisting of classical PQ parameters and active power levels	20
Database B	ADI indicators + P: consists of ADI components and active power level	7
Database C	PQPI indicators + P: consists of PQPI components and active power level	5

Database A (PQ parameters + P) consists of classic power quality parameters and active power level:

- 3 values of U,
- 3 values of 200-ms minimum values of U,
- 3 values of 200-ms maximum values of U,
- 3 values of Pst,
- 1 value of ku2,
- 3 values of THDu,
- 3 values of 200-ms maximum values of THDu,
- 1 value of active power level.

The database B (ADI + P) consists of ADI components and active power

- 1 value that represents U,
- 1 value that represents 200-ms minimum and maximum values of U,
- 1 value that represents Pst,
- 1 value that represents ku2,
- 1 value that represents to THDu
- 1 value that represents 200-ms maximum values of THDu,
- 1 value that active power level.

The third database (PQPI + P) consists of PQPI and active power:

- 1 value that represents 200-ms minimum and maximum values-an envelope of U,
- 1 value that represents Pst,
- 1 value that represents ku2,
- 1 value that represents 200-ms maximum values of THDu,
- 1 value that active power level.

To summarize the database construction a simplified scheme is presented in Figure 3.
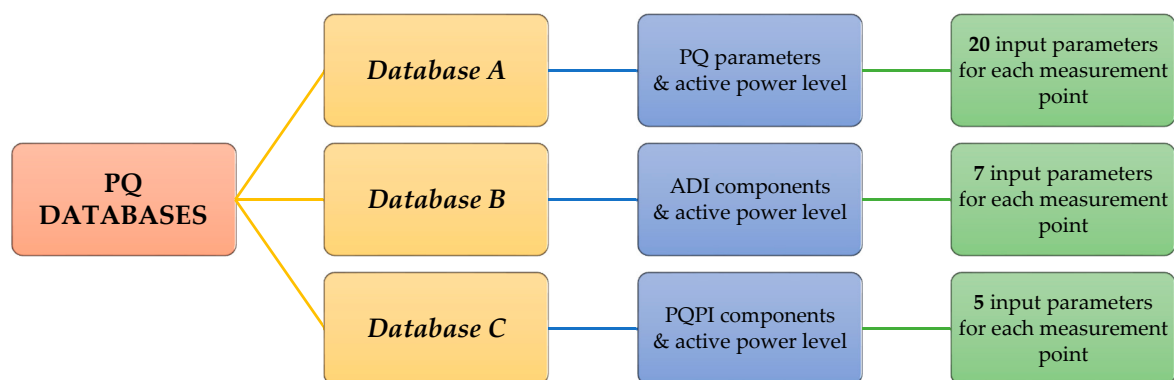

Figure 3. The comparison between investigated databases for hierarchical clustering.

2.4. Hierarchical Clustering

Clustering is one of the data mining (DM) techniques, that aims to divide data into groups that represent similar features [27]. Clustering may be realized in two approaches: a hierarchical or a nonhierarchical [28].

The nonhierarchical algorithms aim toward assigning all observations to the earlier known number of clusters [29]. The most commonly used methods are e.g., K-mean, K-median, or expectation maximization [30].

The hierarchical methods constitute x classes of y observations. Hierarchical methods are also realized by two approaches: agglomerative or divisive. In this research, the agglomerative approach was selected. Generally, the agglomerative approach represents a set of observations, when each piece of data is treated as a separate cluster during the first step. Then, the data are connected into new clusters until one single cluster is established.

This single cluster contains all the data [31]. The agglomerative methods to obtain clusters are single linkage, complete linkage, average linkage, weighted pair-group average linkage, unweighted pair-group centroid linkage, and the Ward method [31,32].

In this article, hierarchical clustering was selected. It determines if the connection is better realized by a single data point or by a group of similar data (achieved in the previous agglomeration) to get a final classification. In this article hierarchical clustering is realized using the Ward algorithm. Ward algorithm cluster analysis is carried out to connect data concentrated in an average value until the data has a similar value (range). The hierarchical CA algorithm, that uses the Ward method of minimal variance, is based on six steps [31,33].

- step 1: Initiate an agglomeration clustering -> divide into x clusters from x data -> calculate the distance between each pair of clusters -> create symmetrical Dis matrix, that consists of distances.
- step 2: find the one pair of clusters that has the smallest squares sum of the distances between adequate object and the related cluster center of the object.
- step 3: create the new cluster, that connects those indicated clusters.
- step 4: update the matrix Dis with the distance between a new cluster and other clusters.
- step 5: check if the number of clusters is equal to 1? YES-go to step 6; NO–back to step 2.
- step 6: final classification when all data are connected to one cluster.

In the Ward method, the indicated "finding pair of clusters which have the smallest sum of squares distance between the object and the cluster center to which this object belongs" is obtained using Equation (1) [31,33].

$$\mathrm{Dis}_{ik} = \frac{n_i + n_k}{n_i + n_j + n_k} * \mathrm{dis}_{ik} + \frac{n_j + n_k}{n_i + n_j + n_k} * \mathrm{dis}_{jk} + \frac{-n_k}{n_i + n_j + n_k} * \mathrm{dis}_{ij}, \tag{1}$$

where: [31,33]

Dis_{pr}: the distance of a new cluster to cluster of number "k",

k: the proceed numbers of cluster from "i" to "j",

dis_{ik}: the distance of a primary cluster "i" from cluster "k",

dis_{jk}: the distance of a primary cluster "j" from cluster "k",

dis_{ij}: the common distance of primary clusters "i" and "j",

n: number of a single object inside each object.

The big problem in cluster analysis, irrespective of the method, is to determine the final number of clusters [34]. The solution in literature for the Ward method is the cubic clustering criterion (CCC). This criterion is obtained by comparing an observed coefficient of determination (R^2) to the approximate expected R^2. It is realized using an approximate variance-stabilizing transformation [35]. The positive values of the cubic clustering criterion inform us that the obtained coefficient of determination is greater than would be expected if sampled from a uniform distribution and, therefore, indicate the possible presence of clusters. The features of the cubic clustering criterion are [35]:

- Extremum of CCC for cluster number greater than two or three indicate good clustering.
- CCC can have several local extremums if the data have a hierarchical structure.
- If CCC values are negative for at least 2 clusters, the distribution is probably unimodal or long tailed.
- Very negative values of the CCC (e.g., −30), could be caused by the outliers.

The last feature was a contribution to apply CCC criteria to detect short-term working conditions of VPP that are specific (outliners) in view of PQ.

3. Hierarchical Clustering of Power Quality Measurement Obtained from the Virtual Power Plant

In this section, the comparison between different databases was performed on the basis of hierarchical clustering with the Ward algorithm. The assessment of cluster assignment

was realized using the cubic clustering criterion. Then for the selected number of the cluster and the PQ, comparison between clusters was performed using PQPI.

3.1. Comparison between Databases Using Cubic Clustering Criterion

The PQ measurement time was from 1 May 2020 to 29 October 2020. The analyzed number of PQ points were five: Line_MV, HPP_MV, ESS_MV, Load1_LV, Load2_LV. However, due to the fact that the HPP_MV and ESS_MV PQ records were obtained from the same voltage transformer they were treated as the same point (HPP and ESS_MV). The only difference was the active power level. So, for the analysis of HPP and ESS_MV, there is one more active power parameter than the number of measurement points. For the observed time period, 26 weeks, there is an analysis of 26,208 single 10-min data points. However, the coverage of data point is equal to 97.7% due to measurement device problems. Thus 25,069 10-min data points were accessible [24]. However, as a preprocessing of PQ data, the voltage events connecting data were also excluded as suggested in [36]. Thus, the final number of 10-min data points was 24,612.

Then, for such defined measurement dataset, the input databases for hierarchical clustering were prepared. The database was a matrix that has 24,612 rows (10-min aggregated data) and a different number of columns. The number of columns was connected with the measurement points and the number of variables for each database (check Figure 2). The size of each database set is as follows:

- Database A: matrix 24,612 × 81 consisting of 1,993,572 single cells,
- Database B: matrix 24,612 × 29 consisting of 713,748 single cells,
- Database C: matrix 24,612 × 21 consisting of 516,852 single cells.

The first aim was the verification of using global indexes (ADI and PQPI) in place of classic PQ parameters. These global indexes aimed to minimize the size of the database and retrain the features of data in point of power quality.

The investigated datasets concern long term measurements. During this measurement time, the specific working conditions (like high/low harmonic content, high low voltage level or asymmetry), are indicated. Such states are mainly, short. Thus, it may be treated as an anomaly in view of the general assessment. The selection of such anomalies (specific short-term working conditions) by analyzing every 10-min data of each PQ parameter separately may be time-consuming. Thus, in this article the application of the cubic clustering criterion (CCC) is proposed.

The CCC was used on the databases (A, B, C) for hierarchical clustering with the Ward algorithm. Under the investigation, the minimum number of clusters was equal to two and maximum was equal to ten. Ten, as the maximum value, was selected on the basis of justification presented in e.g., [37] or [38]. The results were presented in Table 2.

Table 2. The results of the cubic clustering criterion for different databases.

Number of Clusters	Cubic Clustering Criterion		
	Database A (PQ Parameters + P)	Database B (ADI Indicators + P)	Database C (PQPI Indicators + P)
2	−52.14	−41.34	−80.81
3	−81.65	−66.43	−84.94
4	−84.51	−71.51	−86.20
5	−78.08	−69.19	−77.45
6	−67.57	−64.51	−59.78
7	−67.47	−50.28	−40.24
8	−58.11	−35.86	−30.65
9	−57.84	−25.89	−19.91
10	−52.43	−15.95	−8.67

The results in point of CCC values were different for each database for different clusters, but for each database the extremal value was indicated for a final number of

clusters equal to 4. It means that this division assured clusters that represent data which are very different from one another. Thus, four clusters were selected as the most appropriate to detect the specific short-term working conditions for the investigated measurement. Furthermore, as for all databases that indicated extremum for CCC the least numerous input database was selected to further investigate (database C), which uses PQPI indicators and an active power level.

3.2. Results of Hierarchical Clustering

As it was indicated in the previous subsection the optimal selection in view of the size of the input database is those that consist of PQPI indicators and active power level (database C). Thus, in this section, that database was used for hierarchical clustering with the Ward algorithm.

The main result of the hierarchical clustering is a dendrogram. The dendrogram for the selected database for 26 measurement weeks is presented in Figure 4. On the vertical line, the connection distance between clusters is presented. On a horizontal line, each of 24,612 single 10-min data points is indicated.
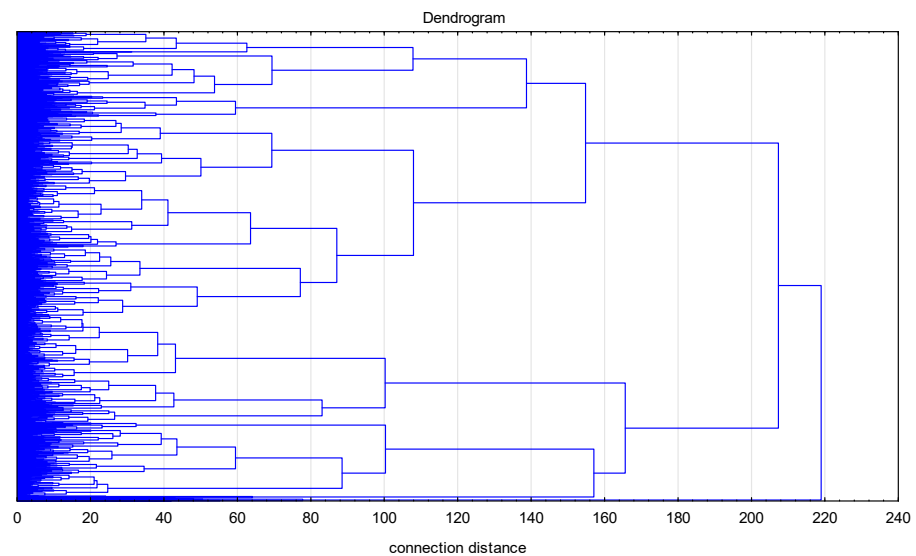

Figure 4. Dendrogram of the hierarchical clustering.

As it was indicated in the previous subsection, division into four clusters has an extremum value for cubic clustering criteria. Based on the dendrogram, it was indicated that for four clusters one is with a small number of 10-min data. The numbers for each cluster are presented and the cluster assignment in the time-domain is presented in Figure 5. It can be observed that cluster 4 is a short-term condition that is represented by only 165 10-min data points. Those 165 10-min data points represent around 0.7% of the measurement time.
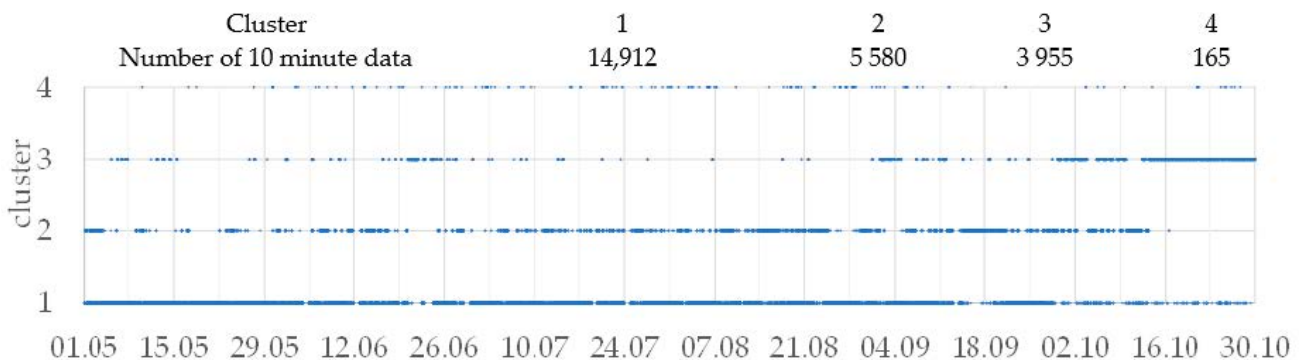

Figure 5. Clustering results in the point of cluster size and time domain.

3.3. Qualitative Assessment of Hierarchical Clustering Results Using the Global Index

In this subsection, the qualitative assessment for indicated clusters is performed. The assessment is realized using PQPI. Thus, the comparison for voltage indicator (a), unbalance indicator (b), flicker indicator (c), and harmonic distortion indicator (d) was presented in Figure 6. The assessment goes towards obtaining knowledge about what PQ parameters and for which measurement point there was a reason for indicating this short-term working condition of VPP. Based on this comparison it may be concluded that cluster 4 represents:

- a problem with voltage indicator in all measurement points. However, a relatively higher value of the indicator is observed for LINE_MV and HPP and ESS_MV;
- a problem unbalance indicator for HPP and ESS_MV, LOAD1_LV, LOAD2_LV;
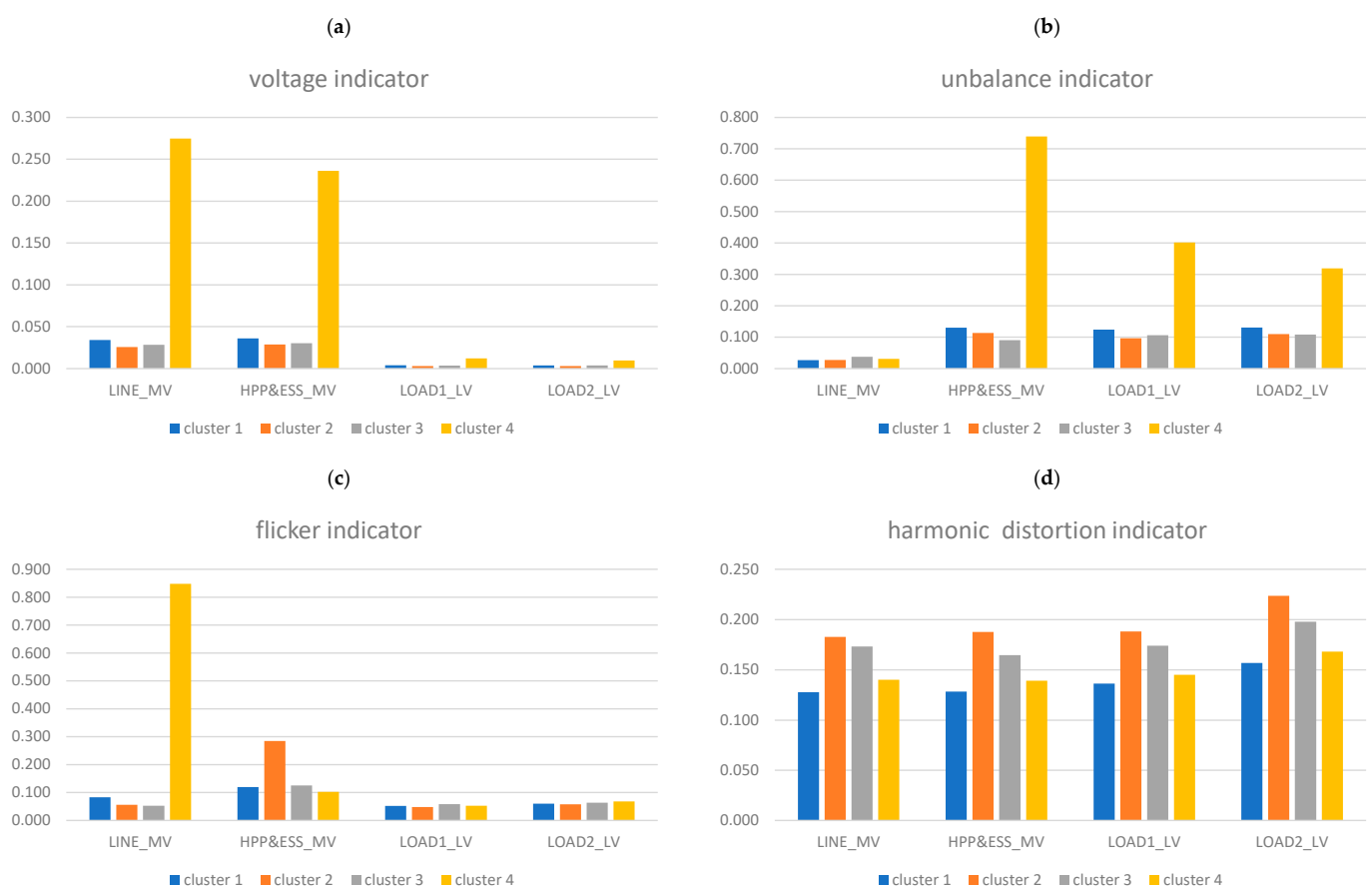- a problem flicker indicator for LINE_MV.

Figure 6. The qualitative comparison between clusters for power quality pollution index (PQPI) indicators: (**a**) voltage indicator; (**b**) unbalance indicator; (**c**) flicker indicator; (**d**) harmonic distortion indicator.

4. Discussion

This article is concerned with a virtual power plant that operates in Poland. The presented research is based on synchronic measurements from five PQ recorders. The PQ measurements were performed at both medium and low voltage levels. The PQ data consists of 182 days (26 weeks) (from 1 May 2020 to 28 October 2020). Thus, these data represent long term data during which different working conditions may occur. This dataset, in work [24], was used to compare the working conditions that were defined a-priory. The states were connected with HPP and ESS working condition, and level of active power. Thus in this work, the conditions were obtained without any predefinition but based on data features.

Those long-term measurements were used as the input to define different PQ databases. The proposed databases were based on classical PQ parameters, PQ global indices as well as the level of active power. Database A represents classical PQ parameters and active power. It concerns 20 variables for each measurement point. Database B uses the ADI components separately and active power level. It concerns seven variables for each measurement point. Database C consists of PQPI and active power level. It concerns five variables for each measurement point. It is important to notice that both global indexes (ADI and PQPI) aim to reduce the amount of data that are analyzed. However, this minimization should not cause one to lose the data features.

The research aimed to detect short-term working conditions in view of power quality. This task was realized using hierarchical clustering with the Ward algorithm. For those three databases, the clustering was realized using the cubic clustering criterion to realize the qualitative assessment of the clustering process. As it is known from the literature if CCC has a negative value it may mean that clustered data has anomalies. For all databases, the CCC was negative for final clusters equal to the range of 2 to 10. Additionally, the extremum for each database was the same and equal to four clusters. Thus, for further investigation, the database with the smallest number of variables (database C) was selected. Furthermore, the final number of clusters was selected as four.

Then, for the indicated circumstances, the comparison between clusters for each measurement point was performed. The PQPI indicators such as voltage indicator, unbalance indicator, flicker indicator and harmonic distortion indicator was used. Based on a comparison of the above mentioned parameters it was concluded that this short term condition (anomaly) was connected with problems of voltage, unbalance, and flicker. The harmonic distortion does not have a significant impact on this condition. It is very important to notice that this short-term condition was not connected with the voltage event. All data that contained voltage events were excluded during the preprocessing of measurements.

The appliances of PQPI index and hierarchical clustering indicated the short-term condition and measurement points, for which it occurs. Thus, the general information about the outcome in PQ problems was also indicated. However, using all this information, it is impossible to define the reason for this condition. Thus, the proposed solution seems essential as a first step of the investigation to define the time period of specific working condition occurrence.

Furthermore, to generalize the discussion of the results, it is important to notice that:

- The investigation was realized in real VPP that operates in Poland, but it also may be applied to other VPP, only if long term power quality data would be available.
- The investigation was based on four measurement points but it may be conducted also for other numbers of points. The minimum number is one, and maximum is limited to the computer computing capabilities.
- In the investigation the extremal value for CCC was for four clusters. However, if the other measurement data would be applied to this methodology, another number would be obtained. However, the most crucial aspect is to select the division when CCC has an extremum. So, the results should be treated in view of proposition of the methodology as well as investigation of the real case study.
- The proposed global index was directed only to selected voltage issues (voltage level, flicker, unbalance and harmonic distortion) and active power level. However, there is a possibility to add other parameters to the global index to make the division more sensitive to other phenomes like current parameters or reactive power.

5. Conclusions

The article proposes the application of clustering to long term power quality data obtained from a virtual power plant. The synchronic, multipoint measurements were used as common input to prepare different databases. The databases were based on both classic PQ parameters and global indexes as well as active power level. The selected PQ global

indexes (ADI and PQPI) enabled reduction of the size of the input databases and retrain the features of the PQ data.

The application of the global indexes for clustering the input dataset reduced the size by around 74%. The results for 26 weeks clustering in view of cubic clustering criterion had an extremum for the same number of clusters and indicated specific short-term working conditions of the virtual power plant in view of PQ.

Additionally, using the proposed global index PQPI helped decide which measurement points and which group of parameters had caused the specific working condition. However, it is not possible to define the reason for such a situation. Thus, the single parameter for a single measurement point assessment is still needed even though the time period of this short condition is strictly defined. So, the application of global index and hierarchical clustering may be treated as a first step for deeper analysis.

Author Contributions: Conceptualization, M.J. and T.S.; methodology, M.J. and T.S.; software, M.J. and T.S.; validation, D.K., J.R. and V.S.; formal analysis, M.J., T.S., D.K., J.R.; investigation, M.J. and T.S.; resources, P.K., P.J., J.B. and P.P.; data curation, V.S. and J.S.; writing—original draft preparation M.J. and T.S.; writing—review and editing, D.K. and J.R.; visualization, D.K. and J.R.; supervision, T.S., J.R., Z.L.; project administration, T.S. and P.J.; funding acquisition, T.S. and P.J. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are not publicly available due to policy of associate company.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Etherden, N.; Vyatkin, V.; Bollen, M.H.J.J. Virtual Power Plant for Grid Services Using IEC 61850. *IEEE Trans. Ind. Inform.* **2016**, *12*, 437–447. [CrossRef]
2. Pudjianto, D.; Ramsay, C.; Strbac, G. Microgrids and virtual power plants: Concepts to support the integration of distributed energy resources. *Proc. Inst. Mech. Eng. Part A J. Power Energy* **2008**, *222*, 731–741. [CrossRef]
3. Caldon, R.; Patria, A.; Turri, R. Optimal Control of a Distribution System with a Virtual Power Plant. In Proceedings of the Bulk Power System Dynamics and Control-VI, Cortina d'Ampezzo, Italy, 22–27 August 2004; pp. 278–284.
4. Zhang, J.; Xu, Z.; Xu, W.; Zhu, F.; Lyu, X.; Fu, M. Bi-Objective Dispatch of Multi-Energy Virtual Power Plant: Deep-Learning-Based Prediction and Particle Swarm Optimization. *Appl. Sci.* **2019**, *9*, 292. [CrossRef]
5. Gong, J.; Xie, D.; Jiang, C.; Zhang, Y. Multiple Objective Compromised Method for Power Management in Virtual Power Plants. *Energies* **2011**, *4*, 700–716. [CrossRef]
6. Beguin, A.; Nicolet, C.; Kawkabani, B.; Avellan, F. Virtual power plant with pumped storage power plant for renewable energy integration. In Proceedings of the 2014 International Conference on Electrical Machines (ICEM), Berlin, Germany, 2–5 September 2014; pp. 1736–1742. [CrossRef]
7. CIGRE. *Broshure 292: Data Mining Techniques and Applications in the Power Transmission Field*; CIGRE: Paris, France, 2006.
8. Luo, Z.; Hong, S.; Ding, Y. A data mining-driven incentive-based demand response scheme for a virtual power plant. *Appl. Energy* **2019**, *239*, 549–559. [CrossRef]
9. Yi, Z.; Xu, Y.; Gu, W.; Wu, W. A Multi-Time-Scale Economic Scheduling Strategy for Virtual Power Plant Based on Deferrable Loads Aggregation and Disaggregation. *IEEE Trans. Sustain. Energy* **2020**, *11*, 1332–1346. [CrossRef]
10. Kong, X.; Xiao, J.; Liu, D.; Wu, J.; Wang, C.; Shen, Y. Robust stochastic optimal dispatching method of multi-energy virtual power plant considering multiple uncertainties. *Appl. Energy* **2020**, *279*, 115707. [CrossRef]
11. Ai, X.; Yang, Z.; Hu, H.; Wang, Z.; Peng, D.; Zhao, Z. A load curve clustering method based on improved k-means algorithm for virtual power plant and its application. *Dianli Jianshe/Electr. Power Constr.* **2020**, *41*, 28–36. [CrossRef]
12. Silva, C.; Faria, P.; Vale, Z. Multi-Period Observation Clustering for Tariff Definition in a Weekly Basis Remuneration of Demand Response. *Energies* **2019**, *12*, 1248. [CrossRef]

13. Faria, P.; Spínola, J.; Vale, Z. Distributed Energy Resources Scheduling and Aggregation in the Context of Demand Response Programs. *Energies* **2018**, *11*, 1987. [CrossRef]

14. Klajn, A.; Bątkiewicz-Pantua, M. *Application Note–Standard EN 50 160: Voltage Characteristics of Electricity Supplied by Public Electricity Networks*; European Copper Institute: Brussels, Belgium, 2017; pp. 1–42.

15. 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, UAE, 15–17 November 2011; pp. 1–7.

16. Jasinski, M.; Sikorski, T.; Kostyla, P.; Borkowski, K. Global power quality indices for assessment of multipoint Power quality measurements. In Proceedings of the 2018 10th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), Iasi, Romania, 28–30 June 2018; pp. 1–6.

17. Serpak, M. A unified index and system indicator for global power quality assessment. *Sci. Int.* **2016**, *28*, 1131–1136.

18. Lee, B.; Sohn, D.; Kim, K.M. Development of Power Quality Index Using Ideal Analytic Hierarchy Process. In *Information Science and Applications (ICISA) 2016*; Springer: Singapore, 2016; pp. 783–793.

19. 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]

20. Langouranis, P.A.; Kaminaris, S.D.; Vokas, G.A.; Raptis, T.E.; Ioannidis, G.C.; General, A. Fuzzy Total Power Quality Index for Electric Networks. In Proceedings of the MedPower 2014, Athens, Greece, 2–5 November 2014; pp. 1–6.

21. De Capua, C.; De Falco, S.; Liccardo, A.; Romeo, E. Imporvement of New Synthetic Power Quality Indexes: An Original Approach to Their Validation. In Proceedings of the 2005 IEEE Instrumentationand Measurement Technology Conference Proceedings, Ottawa, ON, Canada, 16–19 May 2005; Volume 2, pp. 819–822. [CrossRef]

22. Ge, B.; Pan, T.; Li, Z. Synthetic assessment of power quality using relative entropy theory. *J. Comput. Inf. Syst.* **2015**, *11*, 1323–1331. [CrossRef]

23. Sikorski, T.; Jasiński, M.; Ropuszyńska-Surma, E.; Węglarz, M.; Kaczorowska, D.; Kostyla, P.; Leonowicz, Z.; Lis, R.; Rezmer, J.; Rojewski, W.; et al. A Case Study on Distributed Energy Resources and Energy-Storage Systems in a Virtual Power Plant Concept: Technical Aspects. *Energies* **2020**, *13*, 3086. [CrossRef]

24. Jasiński, M.; Sikorski, T.; Kaczorowska, D.; Rezmer, J.; Suresh, V.; Leonowicz, Z.; Kostyla, P.; Szymańda, J.; Janik, P. A Case Study on Power Quality in a Virtual Power Plant: Long Term Assessment and Global Index Application. *Energies* **2020**, *13*, 6578. [CrossRef]

25. IEC 61000 4-30. *Electromagnetic Compatibility (EMC)–Part 4-30: Testing and Measurement Techniques–Power Quality Measurement Methods*; International Electrotechnical Commission: London, UK, 2003.

26. *EN 50160: Voltage Characteristics of Electricity Supplied by Public Distribution Network*; British Standards: London, UK, 2010; Available online: https://orgalim.eu/position-papers/en-50160-voltage-characteristics-electricity-supplied-public-distribution-system (accessed on 8 February 2021).

27. Vehkalahti, K.; Everitt, B.S. *Multivariate Analysis for the Behavioral Sciences*, 2nd ed.; CRC Press: Boca Raton, FL, USA, 2019; ISBN 9781351202275.

28. Roiger, R.J. *Data Mining*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2017; ISBN 9781315382586.

29. Wierzchoń, S.; Kłopotek, M. *Modern Algorithms of Cluster Analysis*; Studies in Big Data; Springer International Publishing: Cham, Switzerland, 2018; Volume 34, ISBN 978-3-319-69307-1.

30. 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]

31. Wierzchoń, S.; Kłopotek, M. *Algorithms of Cluster Analysis*; Institute of Computer Science Polish Academy of Sciences: Warsaw, Poland, 2015; Volume 3, ISBN 9789638759627.

32. Sneath, P.H.; Sokal, R.R. *Numerical Texonomy*; Freeman: San Franciso, CA, USA, 1973; ISBN 9780716706977.

33. Statsoft Polska StatSoft Electronic Statistic Textbook. Available online: https://www.statsoft.pl/textbook/stathome.html (accessed on 15 December 2020).

34. Chowdhury, K.; Chaudhuri, D.; Pal, A.K. An entropy-based initialization method of K-means clustering on the optimal number of clusters. *Neural Comput. Appl.* **2020**. [CrossRef]

35. Sarle, W. Cubic clustering criteria. In *SAS Technical Report A-108*; SAS Institute Inc.: Cary, NC, USA, 1983; p. 51.

36. Jasiński, M.; Sikorski, T.; Leonowicz, Z.; Borkowski, K.; Jasińska, E. The Application of Hierarchical Clustering to Power Quality Measurements in an Electrical Power Network with Distributed Generation. *Energies* **2020**, *13*, 2407. [CrossRef]

37. Claeys, R.; Azaioud, H.; Cleenwerck, R.; Knockaert, J.; Desmet, J. A Novel Feature Set for Low-Voltage Consumers, Based on the Temporal Dependence of Consumption and Peak Demands. *Energies* **2020**, *14*, 139. [CrossRef]

38. Kang, J.; Lee, J.-H. Electricity Customer Clustering Following Experts' Principle for Demand Response Applications. *Energies* **2015**, *8*, 12242–12265. [CrossRef]