IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 25 May 2022, accepted 21 June 2022, date of publication 24 June 2022, date of current version 11 July 2022.

*Digital Object Identifier 10.1109/ACCESS.2022.3186092*

RESEARCH ARTICLE

# Microgrid Working Conditions Identification Based on Cluster Analysis–A Case Study From Lambda Microgrid

**MICHAŁ JASIŃSKI[1], (Member, IEEE), LUIGI MARTIRANO[2], (Senior Member, IEEE), ARSALAN NAJAFI[1], (Senior Member, IEEE), OMID HOMAEE[1], ZBIGNIEW LEONOWICZ[1], (Senior Member, IEEE), AND MOSTAFA KERMANI[3], (Member, IEEE)**

[1]Faculty of Electrical Engineering, Wocław University of Science and Technology, 50-370 Wocław, Poland
[2]Department of Astronautical, Electrical and Energy Engineering (DIAEE), Sapienza University of Rome, 00184 Rome, Italy
[3]Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden

Corresponding author: Michał Jasiński (michal.jasinski@pwr.edu.pl)

This work was supported by the National Science Center of Poland under MINIATURA 4—Grant 2020/04/X/ST8/00194.

**ABSTRACT** This article presents the application of cluster analysis (CA) to data proceeding from a testbed microgrid located at Sapienza University of Rome. The microgrid consists of photovoltaic (PV), battery storage system (BESS), emergency generator set, and different types of load with a real-time energy management system based on supervisory control and data acquisition. The investigation is based on the area-related approach - the CA algorithm considers the input database consisting of data from all measurement points simultaneously. Under the investigation, different distance measures (Euclidean, Chebyshev, or Manhattan), as well as an approach to the optimal number of cluster selections. Based on the investigation, the four different clusters that represent working conditions were obtained using methods to define an optimal number of clusters. Cluster 1 represented time with high PV production; cluster 2 represented time with relatively low PV production and when BESS was charged; cluster 3 represents time with relatively high PV production and when BESS was charged; cluster 4 represents time without PV production. Additionally, after the clustering process, a deep analysis was performed in relation to the working condition of the microgrid.

**INDEX TERMS** Microgrid, area-related approach, cluster analysis, different measurement distances, optimal number of clusters.

## I. INTRODUCTION

Recently, the approach to energy supply has tended to decentralized control [1], [2]. Thus, different ways to obtain those solutions are proposed, such as microgrids, power plants, energy communities [3], [4] or smart grids [5].

Microgrids are small-scale power systems consisting of distributed power generation, power storage, and load [6]. Recently many real cases of microgrids are indicated in the literature. The examples of microgrids in Europe are located in:

- UK [7], [8]
- Germany [9], [10],
- Austria [11], [12],
- France [13], [14],
- Poland [15], [16],
- Italy [17], [18].

In this paper, the application of cluster analysis (CA) to real data from a microgrid is presented. Thus, the recent articles in this area are discussed. Article [19] proposes a demand data reduction method based on k-means clustering. In the article, the reduced demand data is considered separately for weekdays and weekends. The proposed data reduction approach is tested on the basis of real data from a microgrid with a 1 hour resolution. Article [20] proposed using k-mean clustering to identify suitable candidates in a local microgrid for demand-side management who have insignificant influence on microgrid peaks. The paper [21] proposes an optimal

The associate editor coordinating the review of this manuscript and approving it for publication was Zhan Bu.

coordination approach for directional overcurrent relays for the microgrid using the k-means algorithm. The final number of clusters is defined in the basement of setting groups number of commercially available relays. The article [22] proposes the method for optimal network planning of hybrid microgrid. In this case, the optimal clustering partition was defined by a combination of k-means cluster analysis with the particle swarm optimization algorithm. The article proposes a scheme for island detection of microgrid using data mining. The method is based on fuzzy c-means clustering [23]. In article [24], a support management system based on stochastic mixed-integer linear programming was proposed. To deal with uncertainty issues, the reduction of scenario was realized. This task was performed using a clustering technique of k-means and a fast backward scenario reduction method. In the paper [25], the optimal location to place the data of the one-year electric vehicle charging station from the real microgrid is investigated. In this case, k-means clustering was applied to divide the load data into representatives. Only those representatives (clusters) were used in simulations. The final number of clusters was selected using the elbow method. As can be observed from the literature review, the application of CA, but the main aim of its application is to reduce the size of the analyzed data [26], as well as to obtain general knowledge with the maintenance of data features. [27].

In this article, the main aim of the CA application is to divide the long-term data from the real microgrid to define its working conditions. The investigated microgrid is the testbed microgrid located at Sapienza University of Rome. The microgrid consists of photovoltaic (PV), battery storage system (BESS), emergency generator set, and different types of load. There are several points observed at the point of energy flow in the microgrid. The investigation is based on four Lambda Microgrid measurement points - point of common coupling (PCC), the connection point of the 12 kWh peak 3-phase PV system, the 6.5 kWh battery energy storage system and the University laboratory as a specific load. Therefore, to connect these measurement points, an area-related approach [28] was proposed for long-term data, 18 months with a resolution of 15 minutes of the data. It is worth noting that the CA algorithm considers the input database that consists of data from all measurement points simultaneously. In this paper, the different aspects of CA are considered such as distance measure selection (Euclidean, Chebyshev, or Manhattan), as well as the optimal number of cluster selection (v-fold validation test). However, the main element is that after CA application, a deep analysis is performed at the point of qualitative assessment of clusters. It's important to notice that each cluster represents different microgrid working conditions defined based on multipoint measurement to assess the system globally (all measurement points) not locally (each point separately).

The article is organized into six sections. Section II presents the methodology part with details about CA, selection of distance measurements, and selection of the optimal number of clusters. Section III present the investigated object – Lambda Microgrid. Section IV presents the results of CA application and analysis in point of microgrid working conditions. Section V contains a discussion of the results. Finally, Section VI concludes the paper

## II. METHODOLOGY

The presented in this section methodology is based on cluster analysis, as a representant of data mining techniques. This technique assures the division of data based on their features. Thus obtained clusters (groups), when obtained from a set of synchronized and related data assures the general knowledge about the object considering all input parameters. In this section, we presented the general aim of clustering and commonly used algorithms. Then we introduced the problem of distance selection, which is a measure of similarity/dissimilarity of data. Finally, we discussed the main problem of clustering – the selection of the final (optimal) number of clusters.

### A. INTRODUCTION TO CLUSTER ANALYSIS

Cluster analysis (CA) is a representative technique of data mining [29], [30]. In general, CA aims to assure the division of the data in the point of their features [31], [32]. Cluster analysis can be performed using both a hierarchical and nonhierarchical approach [33], [34]. The hierarchical approach is represented as n classes of m observations. The non-hierarchical method concerns the assignment of all observations to the earlier assumed number of groups (clusters). In the case of non-hierarchical clustering is not assured the tree as a classification result as in the case of the hierarchical clustering [35], [36]. This approach results in the division into groups of the database to maximize/minimize the selected evaluation criteria [37]. In this article, the application of the non-hierarchical approach is discussed. The most used in the literature nonhierarchical methods are based on e.g. [38]–[40]:

- K-mean algorithm,
- K-median algorithm,
- Fuzzy C-means Method
- Expectation maximization (EM) algorithm.

In this paper, the authors suggest using the nonhierarchical with the k-mean algorithm for different distance measures. Generally, the K-mean algorithm's main aim is to find the extremum value of the specific objective function. The mentioned objective is defined as [41]:

$$ObjecFun\,(B, MC) = \sum\nolimits_{\mathrm{a}=1}^{x} \sum\nolimits_{\mathrm{b}=1}^{y} e_{ab} * \text{distance} \quad (1)$$

where:

- ObejcFun – objective function,
- B – matrix of the object belonging to a cluster,
- MC – matrix in which a row vector represents the centroids of clusters,
- a = 1, 2, 3, . . . , x – number of objects,
- b = 1, 2, 3 . . . , y – number of classes (clusters),

- eab – element indicating the fact of assignment of a-th object to the b-th class (cluster),
- distance – a measure of distance.

### B. DIFFERENT DISTANCE MEASURES

As it was indicated in equation 1 for k-mean clustering different measures of distance can be applied. Under this investigation, three different measures were investigated Euclidean (equation 2), Manhattan (equation 3), and Chebyshev (equation 4).

$$EuclideanD\,(a,b) = \sqrt{\sum_i (a_i - b_i)^2}, \tag{2}$$

$$ManhattanD\,(a,b) = \sum_i |a_i - b_i| \tag{3}$$

$$CzebyshevD\,(a,\mathrm{b}) = Maksimum\,|a_i - b_i| \tag{4}$$

where:

- EuclideanD – Euclidean distance,
- ManhattanD – Manhattan distance,
- CzebyshevD – Chebyshev distance,
- a = 1, 2, 3, . . . , x – number of objects,
- b = 1, 2, 3 . . . , y – number of classes (clusters),
- ai – vector of observations, that belong to cluster a,
- bi – vector of observations, that belong to cluster b.

Note that the Euclidean distances (and the squares of the Euclidean distances) are calculated from the raw data rather than from standardized data. This method has some advantages (e.g., the distance between any two objects is not affected by adding new objects to the analysis, which may be outliers). However, distances are greatly affected by unit differences between the dimensions from which distances are calculated. For example, if one dimension represents a length measured in centimeters when converted to millimeters (by multiplying the corresponding values by 10), we will most often get distinctly different Euclidean distances and Euclidean distance squares (calculated for multiple dimensions). This can result in completely different cluster analysis results. In the general case, it is good to use standardization so that we have data on comparable scale. Manhattan distance is simply the sum of differences measured along dimensions. In most cases, this distance measure produces similar results as the ordinary Euclidean distance. However, note that with this measure the effect of individual large differences (outlier cases) is suppressed (because they are not squared). The Chebyshev distance measure is appropriate in those cases where we want to define two objects as ‘‘other’’, then when they differ in one arbitrary dimension [42].

### C. OPTIMAL NUMBER OF CLUSTER DEFINITIONS

The noticeable and indicated in literature disadvantage of the nonhierarchical approach is a-prior defining the final number of clusters. The literature indicated different methods to define optimal number of clusters e.g.:

- k-fold cross-validation test [42]
- gap statistic [43],
- an entropy-based initialization method [44],
- u-control chart [45].

In the following article, the k-fold cross-validation test was chosen. This type of cross-validation is useful for a situation where there is no knowledge about the test sample. The user defines the ‘v’ value for v-fold cross-validation. Normally, v is equal to a value of 1 to 10. The value v refers to the number of random subsamples that are used for the learning part of the data. After that, the tree (with specified size) is computed v times. After each of the iteration steps, it leaves out one of the subsamples from the computations.

The next step is based on using those subsamples as the test sample for cross-validation. The cross-validation cost is computed for each of the v test samples. After that, this cost is averaged to give the v-fold estimate of the cross-validation costs [42]. Finally, by analyzing cross-validation cost, the optimal number of clusters is defined.

## III. LAMBDA MICROGRID

The Lambda Microgrid operates at the Department of Astronautical, Electrical and Energy Engineering Sapienza University of Rome in Italy. The main components of that microgrid are:

- 12 kW three-phase PV system,
- 5.2 kW emergency generator set,
- 6.5 kWh BESS,
- local load including laboratory loads.

In summary about the Lambda MG operation as shown in Fig. 1, the Com’X 510 (energy server) and Acti Smartlink collect data from smart meters and PowertTags. Moreover, PLC is used for the proper as well as faster operation of Lambda MG. Finally, all data are transferred into the data center.

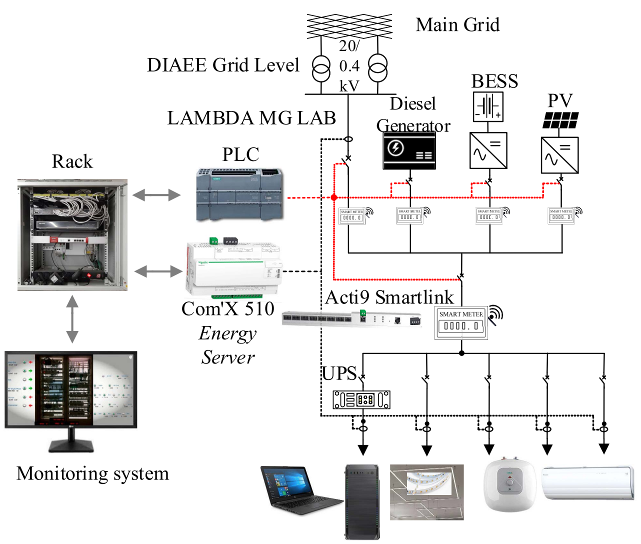



**FIGURE 1.** Load/generation changeability in observed time for different measurement points.

During normal operation, local sources work synchronously with the power system. PV management is based on maximum power point tracking control in grid mode.

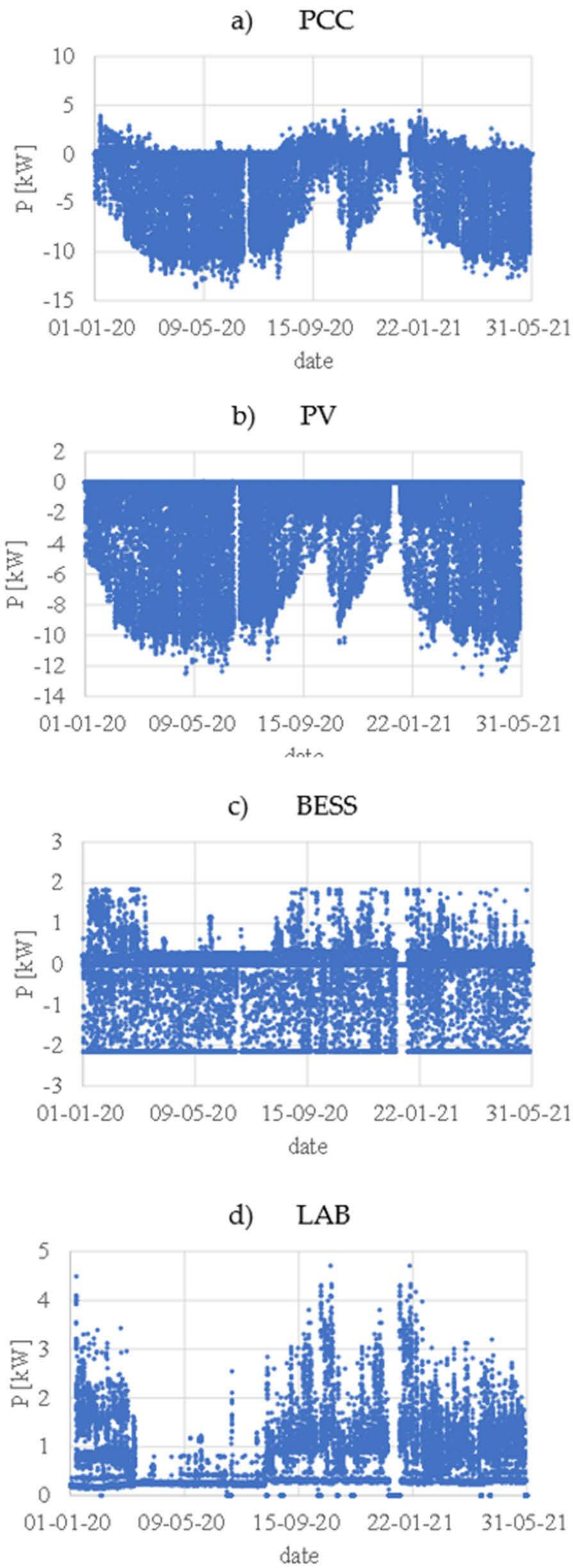



**FIGURE 2.** Load/generation changeability in observed time for different measurement points.

BESS operations are mounted to avoid power flux from the microgrid to the external system. The BESS consists of one battery with an inverter inside, Varta's model. The setups for BESS are the depth of discharge in this battery set to 90%, 2.2 kW of possible charging, and 1.8 kW of discharging capacity. For emergency issues such as external system failure or blackout, the requirements of the IEEE 1547 standard are applied [46] - PV and BESS are automatically disconnected. Then the emergency generators are started to ensure the off-grid microgrid mode. Then after 5 minutes, the PV and BESS are connected, respectively. During the operation of the microgrid, different monitoring systems are used [47]. The source for data in this paper was Schneider Electric iEM3150 meters, which are installed in different points of Lambda microgrid. The location of the meters is presented in Fig. 1. The data collection was realized within 18 months from 01.01.2020 to 31.05.2021. The simplified scheme of Lambda microgrid with the location of smart meters.

## IV. RESULTS

### A. CHARACTERIZATION OF THE OBJECT IN THE LONG-TERM APPROACH

Based on the long term measurements (18 months with 15 minutes aggregation) for multipoint measurement, the following changeability diagrams of active power level are performed (Fig. 2.):

- PCC of Lambda microgrid (a),
- PV system (b),
- BESS (c),
- Laboratory Load - LAB (d).

As it can be observed in Fig. 2, for such a long period of time different working conditions are indicated based on the changeability of all parameters. To define each point in measurement point in the mean of active power level in the long term approach the statistical analysis based on normality charts was performed (Fig. 3). The results are presented in Fig. 3. As it can be observed for each of the points the specific similar groups can be indicated. Thus authors decided to apply cluster analysis to assure the correct division of them into groups (clusters). The division will be realized in area related approach – together for all measurement points.

### B. COMPARISON IN POINT OF DIFFERENT DISTANCE DEFINITION

The first step of the investigation was to define the optimal number of cluster for the investigated multipoint dataset. Thus the v-fold cross-validation was performed for k-mean algorithm with different measurement distances: Euclidean, Manhattan, Chebyshev. The following assumption based on literature suggestions were performed:

- minimal number of clusters equal to 2 and maximal number of clusters equal to 10 [48];
- minimal percentage decrease 10% [49];
- the number of subsets: 10 [50].

The results for the v-fold test were presented in Fig. 4 for Euclidean (a), Manhattan (b), and Chebyshev (c). The v-fold cross-validation for all investigated distances indicated that the optimal number of clusters is equal to 4.

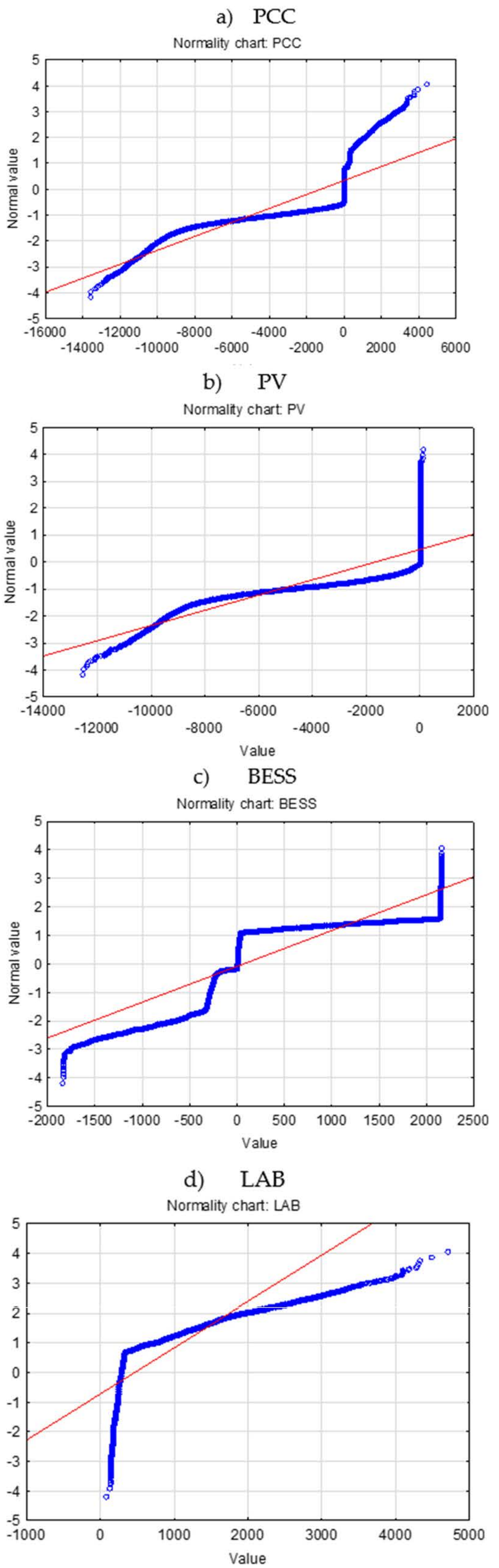



**FIGURE 3.** Normality charts for the monitored measurement points- a) PCC, b) PV, c) BESS, d) LAB.

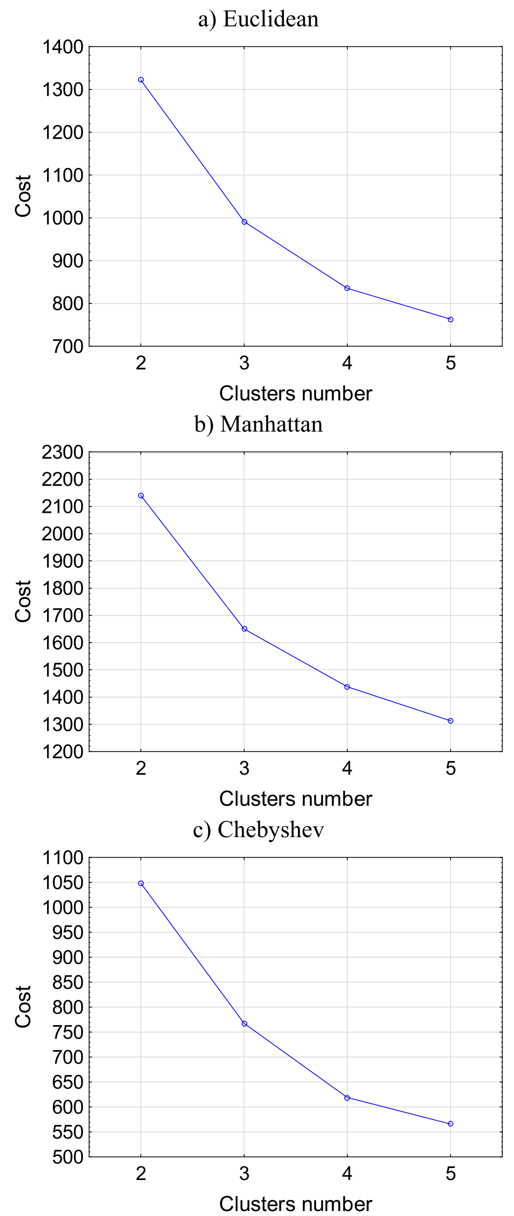



**FIGURE 4.** Defining the optimal number of clusters based on a v-fold cross-validation test for different distance measurements.

However, to assure the correctness of indicated division in the next step was a validation of data assignment to the same representative cluster based on different distances was performed. Table 1 presents the mean values of data that are assignment to clusters based on different distance measurements. Based on the results indicated in Table 1 it can be assumed that for all distances the obtained classification is similar. The differences between number of single data assigned to the same clusters does not exceed 5%. Additionally the mean vale of active power level in clusters do not exceed 100 W for all measurement points.

**TABLE 1.** Comparison of load cluster values for each measurement point for different distance measurements in case of the optimal number of clusters equal to 4.

| Cluster | PCC | PV | BESS | LAB | Number of single data |
|---|---|---|---|---|---|
| *Euclidean distance* | | | | | |
| 1 | -9039 | -8701 | 56 | 479 | 4313 |
| 2 | -4926 | -5809 | 723 | 554 | 3646 |
| 3 | -1160 | -2686 | 1044 | 665 | 5048 |
| 4 | 128 | -133 | -149 | 421 | 33552 |
| *Manhattan distance* | | | | | |
| 1 | -8943 | -8632 | 63 | 485 | 4511 |
| 2 | -4786 | -5710 | 760 | 548 | 3505 |
| 3 | -1202 | -2728 | 1077 | 632 | 4730 |
| 4 | 127 | -145 | -144 | 428 | 33813 |
| *Chebyshev distance* | | | | | |
| 1 | -8911 | -8588 | 64 | 473 | 4590 |
| 2 | -4599 | -5583 | 781 | 568 | 3692 |
| 3 | -917 | -2450 | 954 | 750 | 5383 |
| 4 | 131 | -107 | -156 | 403 | 32894 |

**TABLE 2.** Comparison of load cluster values for each measurement point for different distance measurements in case of the optimal number of clusters equal to 4.

| Cluster | PCC | PV | BESS | LAB | Number of single data |
|---|---|---|---|---|---|
| Final number of clusters equal to 2 | | | | | |
| 1 | -7241 | -7467 | 370 | 516 | 7768 |
| 2 | -58 | -483 | 7 | 452 | 38791 |
| Final number of clusters equal to 3 | | | | | |
| 1 | -3192,79 | -4489 | 990 | 593 | 5019 |
| 2 | -8537,08 | -8344 | 116 | 494 | 5326 |
| 3 | 82,57 | -270 | -67 | 440 | 36214 |
| Final number of clusters equal to 4 | | | | | |
| 1 | -9039 | -8701 | 56 | 479 | 4313 |
| 2 | -4926 | -5809 | 723 | 554 | 3646 |
| 3 | -1160 | -2686 | 1044 | 665 | 5048 |
| 4 | 128 | -133 | -149 | 421 | 33552 |

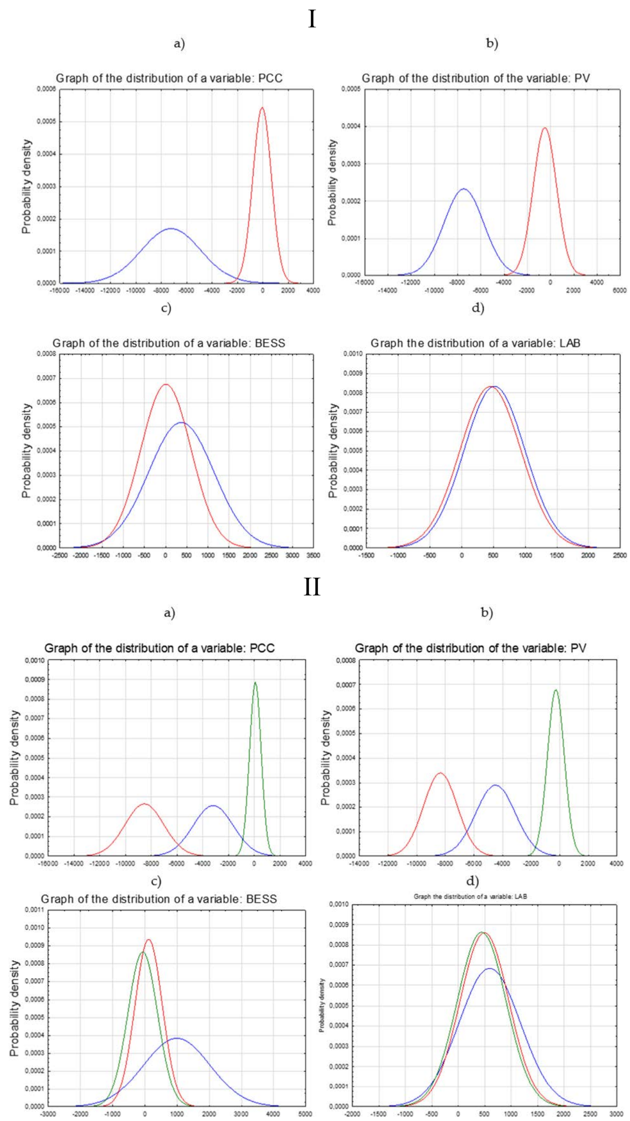



**FIGURE 5.** Distribution graphs for clustering results for final number of cluster equal to I) 2 clusters; II) 3 clusters; measurement points where: a) PCC, b) PV, c) BESS, d) LAB.

### *C. CASE STUDY FOR EUCLIDEAN DISTANCE - DEEP STATISTICAL ANALYSIS*

As a representative distance for CA, the Euclidean distance was selected due to its features as a representative selection of distance measurement value [51]. Thus, the deep analysis for the final number of the cluster from 2 to optimal based on v-fold validation values was performed in this subsection. As indicated in the previous section (section IV.E) the optimal number of clusters is equal to 4, thus the qualitative assessment is performed for the number of clusters equal to 2, 3, and 4. The results in point-of-mean values are presented in Table 2. However, using only mean values, the analysis of the results can be misleading. Thus, to justify that results are correct, the distribution graphs were prepared and presented in Fig. 5 and Fig. 6.

Based on both mean values and distribution graphs prepared for each cluster for clustering with the different final number of clusters (from 2 to 4) using Euclidean measurement of distance, the following observation is indicated:

- CA with the final number of clusters equal to 2: cluster 1 represents time with high PV production; cluster 2 represents time without PV production.
- CA with the final number of clusters equal to 3: cluster 1 represents time with PV production (but not high) and BESS is charged; cluster 2 represents time with high PV production; cluster 3 represents time without PV production.
- CA with the final number equal to 4: cluster 1 represents time with high PV production; cluster 2 represents time with relatively low PV production and BESS is charged; cluster 3 represents time with relatively high PV production and BESS is charged; cluster 4 represents time without PV production.

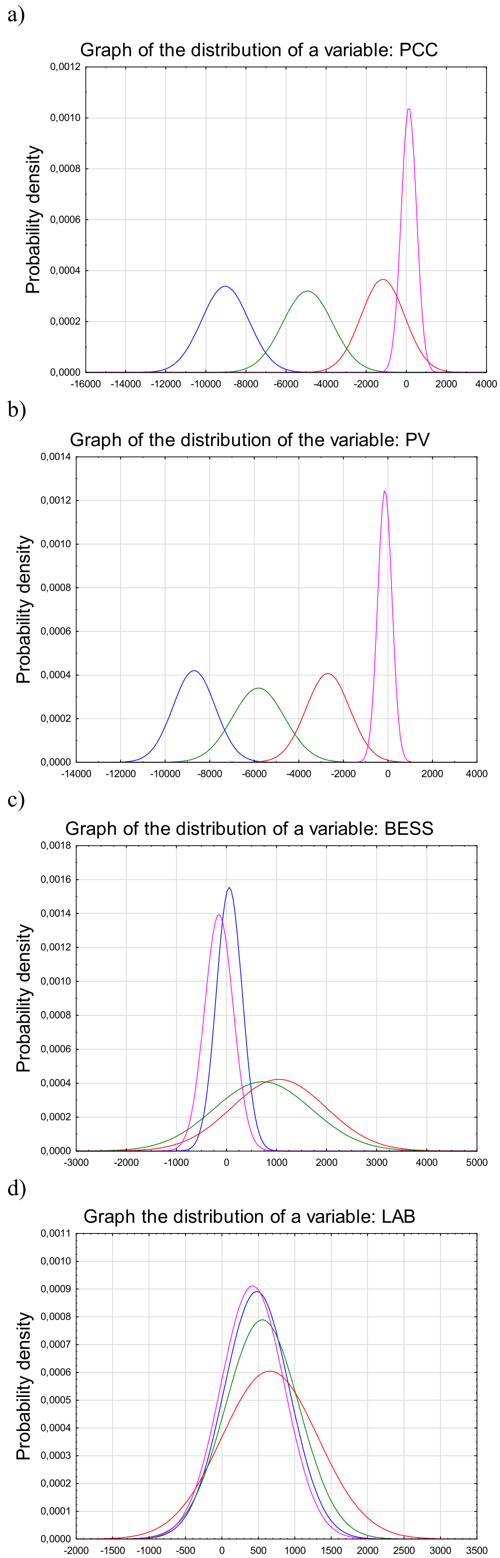



**FIGURE 6. Distribution graphs for clustering results for final number of cluster equal to 4 cluster for all measurement points where: a) PCC, b) PV, c) BESS, d) LAB.**

## V. DISCUSSION

Firstly, it is worth noting that the investigation performed was based on a real testbed microgrid located at Sapienza University of Rome. The microgrid consists of PV, BESS, and different types of load. During the observed time, the microgrid worked only in grid mode, so the emergency generator system had no impact on the results.

Four measurement devices were used as the data source. All the measurement devices were settled in the local microgrid, thus under this investigation the area-related approach was used. The area-relation approach means that we can analyze the object (microgrid) based on the simultaneous analysis of all objects that have an impact on each other. This mutual impact is connected with the fact that all these units are decentralized managed by the microgrid.

In this investigation, the three different measurement distances were investigated – Euclidean, Chebyshev, and Manhattan. In II.B, all of them were compared and highlights about the selection of them in point of data features were indicated. Using different measurement distances for the clustering process, there is also a possibility to verify the character of data. If the clustering results for Euclidean distance are noticeable different from Chebyshev and Manhattan, it means that some specific outliners in any of the measurement points are noticeable. In the case of the investigated object for 18 months period because results for all distances are similar it can be stated that outliners are not indicated in those data.

The next element of discussion is based on a comparison of working conditions that are obtained during the clustering process. The most general division (for 2 clusters) assures division in point of PV generation. When the final number of clusters increased into 3 also the operation of BESS was significant to division. The additional increase in the final number to cluster to 4 indicated that both PV and BESS had an impact to indicate additional cluster.

## VI. CONCLUSION

In this article, the case study of cluster analysis for real microgrid data was performed. The results were based on long-term data (18 months), so the different working conditions occurred. The CA performed was realized simultaneously for all points, thus in fact the investigation was toward the whole microgrid not only a separated part of it. The results of the clustering performed are representants of the Lambda Microgrid working conditions. Generally the working conditions are related to PV generation and operation of BESS. By analyzing each cluster separately, it is possible to analyze the general scenarios of working of the microgrid. Such scenarios can be used to define the microgrid operation rather than analyzing the entire dataset with maintaining the data features.

## REFERENCES

[1] Y. Hennane, A. Berdai, J.-P. Martin, S. Pierfederici, and F. Meibody-Tabar, "New decentralized control of mesh AC microgrids: Study, stability, and robustness analysis," *Sustainability*, vol. 13, no. 4, p. 2243, Feb. 2021, doi: 10.3390/su13042243.

[2] J. Tavoosi, A. Mohammadzadeh, B. Pahlevanzadeh, M. B. Kasmani, S. S. Band, R. Safdar, and A. H. Mosavi, "A machine learning approach for active/reactive power control of grid-connected doubly-fed induction generators," *Ain Shams Eng. J.*, vol. 13, no. 2, Mar. 2022, Art. no. 101564, doi: 10.1016/j.asej.2021.08.007.

[3] S. Zwickl-Bernhard and H. Auer, "Open-source modeling of a low-carbon urban neighborhood with high shares of local renewable generation," *Appl. Energy*, vol. 282, Jan. 2021, Art. no. 116166, doi: 10.1016/j.apenergy.2020.116166.

[4] Z. Liu, A. Mohammadzadeh, H. Turabieh, M. Mafarja, S. S. Band, and A. Mosavi, "A new online learned interval type-3 fuzzy control system for solar energy management systems," *IEEE Access*, vol. 9, pp. 10498–10508, 2021, doi: 10.1109/ACCESS.2021.3049301.

[5] J. Cao, Z. Bu, Y. Y. Wang, H. Yang, J. C. Jiang, and H. J. Li, "Detecting prosumer-community groups in smart grids from the multiagent perspective," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 8, pp. 1652–1664, Aug. 2019, doi: 10.1109/TSMC.2019.2899366.

[6] S. Chandak and P. K. Rout, "The implementation framework of a microgrid: A review," *Int. J. Energy Res.*, vol. 45, no. 3, pp. 3523–3547, Mar. 2021, doi: 10.1002/er.6064.

[7] L. Mariam, M. Basu, and M. F. Conlon, "Microgrid: Architecture, policy and future trends," *Renew. Sustain. Energy Rev.*, vol. 64, pp. 477–489, Oct. 2016, doi: 10.1016/j.rser.2016.06.037.

[8] T. T. Teo, T. Logenthiran, W. L. Woo, K. Abidi, T. John, N. S. Wade, D. M. Greenwood, C. Patsios, and P. C. Taylor, "Optimization of fuzzy energy-management system for grid-connected microgrid using NSGA-II," *IEEE Trans. Cybern.*, vol. 51, no. 11, pp. 5375–5386, Nov. 2021, doi: 10.1109/TCYB.2020.3031109.

[9] M. C. Pulcherio, A. A. Renjit, M. S. Illindala, A. S. Khalsa, J. H. Eto, D. A. Klapp, and R. H. Lasseter, "Evaluation of control methods to prevent collapse of a mixed-source microgrid," *IEEE Trans. Ind. Appl.*, vol. 52, no. 6, pp. 4566–4576, Nov. 2016, doi: 10.1109/TIA.2016.2599139.

[10] N. Beg, H. Biechl, and A. Rosin, "Stability issues with inverter loads and their control in low inertia islanded microgrids," in *Proc. 2nd Global Power, Energy Commun. Conf. (GPECOM)*, Oct. 2020, pp. 196–201, doi: 10.1109/GPECOM49333.2020.9247913.

[11] M. Mansoor, M. Stadler, H. Auer, and M. Zellinger, "Advanced optimal planning for microgrid technologies including hydrogen and mobility at a real microgrid testbed," *Int. J. Hydrogen Energy*, vol. 46, no. 37, pp. 19285–19302, May 2021, doi: 10.1016/j.ijhydene.2021.03.110.

[12] M. Mansoor, M. Stadler, M. Zellinger, K. Lichtenegger, H. Auer, and A. Cosic, "Optimal planning of thermal energy systems in a microgrid with seasonal storage and piecewise affine cost functions," *Energy*, vol. 215, Jan. 2021, Art. no. 119095, doi: 10.1016/j.energy.2020.119095.

[13] V. H. Nguyen, Q. T. Tran, H. Buttin, and M. Guemri, "Implementation of a coordinated voltage control algorithm for a microgrid via SCADA-as-a-service approach," *Electr. Eng.*, vol. 104, no. 2, pp. 389–399, Apr. 2022, doi: 10.1007/s00202-021-01247-z.

[14] S. J. Ben Christopher and M. Carolin Mabel, "A bio-inspired approach for probabilistic energy management of micro-grid incorporating uncertainty in statistical cost estimation," *Energy*, vol. 203, Jul. 2020, Art. no. 117810, doi: 10.1016/j.energy.2020.117810.

[15] V. Suresh, P. Janik, J. M. Guerrero, Z. Leonowicz, and T. Sikorski, "Microgrid energy management system with embedded deep learning forecaster and combined optimizer," *IEEE Access*, vol. 8, pp. 202225–202239, 2020, doi: 10.1109/ACCESS.2020.3036131.

[16] V. Suresh and D. Kaczorowska, "Energy savings approach to optimal location of EV charging stations in microgrids," *Renew. Energy Power Qual. J.*, vol. 19, pp. 33–38, Sep. 2021, doi: 10.24084/repqj19.207.

[17] M. Kermani, B. Adelmanesh, E. Shirdare, C. A. Sima, D. L. Carnì, and L. Martirano, "Intelligent energy management based on SCADA system in a real microgrid for smart building applications," *Renew. Energy*, vol. 171, pp. 1115–1127, Jun. 2021, doi: 10.1016/j.renene.2021.03.008.

[18] E. Bompard, S. Bruno, A. Cordoba-Pacheco, C. Diaz-Londono, G. Giannoccaro, M. La Scala, A. Mazza, and E. Pons, "Connecting in real-time power system labs: An Italian test-case," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur. (EEEIC/ICPS Europe)*, Jun. 2020, pp. 1–6, doi: 10.1109/EEEIC/ICPSEurope49358.2020.9160505.

[19] K. Fahy, M. Stadler, Z. K. Pecenak, and J. Kleissl, "Input data reduction for microgrid sizing and energy cost modeling: Representative days and demand charges," *J. Renew. Sustain. Energy*, vol. 11, no. 6, Nov. 2019, Art. no. 065301, doi: 10.1063/1.5121319.

[20] G. H. Philipo, Y. A. Chande Jande, and T. Kivevele, "Clustering and fuzzy logic-based demand-side management for solar microgrid operation: Case study of ngurudoto microgrid, Arusha, Tanzania," *Adv. Fuzzy Syst.*, vol. 2021, pp. 1–13, Feb. 2021, doi: 10.1155/2021/6614129.

[21] S. D. Saldarriaga-Zuluaga, J. M. López-Lezama, and N. Muñoz-Galeano, "Optimal coordination of over-current relays in microgrids using principal component analysis and $K$-means," *Appl. Sci.*, vol. 11, no. 17, p. 7963, Aug. 2021, doi: 10.3390/app11177963.

[22] T. Wang and X. Yang, "Optimal network planning of AC/DC hybrid microgrid based on clustering and multi-agent reinforcement learning," *J. Renew. Sustain. Energy*, vol. 13, no. 2, Mar. 2021, Art. no. 025501, doi: 10.1063/5.0034816.

[23] A. Damanjani, M. Hosseini Abardeh, A. Azarfar, and M. Hojjat, "A novel scheme for island detection in microgrids based on fuzzy c-means clustering technique," *Int. Rev. Appl. Sci. Eng.*, vol. 12, no. 2, pp. 157–165, May 2021, doi: 10.1556/1848.2021.00215.

[24] I. L. R. Gomes, R. Melicio, and V. M. F. Mendes, "A novel microgrid support management system based on stochastic mixed-integer linear programming," *Energy*, vol. 223, May 2021, Art. no. 120030, doi: 10.1016/j.energy.2021.120030.

[25] V. Suresh, N. Bazmohammadi, P. Janik, J. M. Guerrero, D. Kaczorowska, J. Rezmer, M. Jasinski, and Z. Leonowicz, "Optimal location of an electrical vehicle charging station in a local microgrid using an embedded hybrid optimizer," *Int. J. Electr. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 106979, doi: 10.1016/j.ijepes.2021.106979.

[26] D. J. Vergados, I. Mamounakis, P. Makris, and E. Varvarigos, "Prosumer clustering into virtual microgrids for cost reduction in renewable energy trading markets," *Sustain. Energy, Grids Netw.*, vol. 7, pp. 90–103, Sep. 2016, doi: 10.1016/j.segan.2016.06.002.

[27] M. Jasiński, T. Sikorski, D. Kaczorowska, J. Rezmer, V. Suresh, Z. Leonowicz, P. Kostyła, J. Szymańda, P. Janik, J. Bieńkowski, and P. Prus, "A case study on a hierarchical clustering application in a virtual power plant: Detection of specific working conditions from power quality data," *Energies*, vol. 14, no. 4, p. 907, Feb. 2021, doi: 10.3390/en14040907.

[28] M. Jasiński, "Combined correlation and cluster analysis for long-term power quality data from virtual power plant," *Electronics*, vol. 10, no. 6, p. 641, Mar. 2021, doi: 10.3390/electronics10060641.

[29] R. J. Roiger, *Data Mining*. Boca Raton, FL, USA: CRC Press, 2017.

[30] I.-C. Chang, T.-K. Yu, Y.-J. Chang, and T.-Y. Yu, "Applying text mining, clustering analysis, and latent Dirichlet allocation techniques for topic classification of environmental education journals," *Sustainability*, vol. 13, no. 19, p. 10856, Sep. 2021, doi: 10.3390/su131910856.

[31] K. Vehkalahti and B. S. Everitt, *Multivariate Analysis for the Behavioral Sciences*. 2nd ed. Boca Raton, FL, USA: CRC Press, 2019.

[32] P. Tavallali, P. Tavallali, and M. Singhal, "$K$-means tree: An optimal clustering tree for unsupervised learning," *J. Supercomput.*, vol. 77, no. 5, pp. 5239–5266, May 2021, doi: 10.1007/s11227-020-03436-2.

[33] X. Ai, Z. Yang, H. Hu, Z. Wang, D. Peng, and Z. Zhao, "A load curve clustering method based on improved $k$-means algorithm for virtual power plant and its application," *Dianli Jianshe/Electr. Power Constr.*, vol. 41, no. 5, pp. 28–36, 2020, doi: 10.12204/j.issn.1000-7229.2020.05.004.

[34] Z. Bu, H.-J. Li, C. Zhang, J. Cao, A. Li, and Y. Shi, "Graph $K$-means based on leader identification, dynamic game, and opinion dynamics," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 7, pp. 1348–1361, Jul. 2020, doi: 10.1109/TKDE.2019.2903712.

[35] S. Payne, E. Fuller, G. Spirou, and C.-Q. Zhang, "Automatic quasi-clique merger algorithm—A hierarchical clustering based on subgraph-density," *Phys. A, Stat. Mech. Appl.*, vol. 585, Jan. 2022, Art. no. 126442, doi: 10.1016/j.physa.2021.126442.

[36] C. Wu, H. Li, and J. Ren, "Research on hierarchical clustering method based on partially-ordered Hasse graph," *Future Gener. Comput. Syst.*, vol. 125, pp. 785–791, Dec. 2021, doi: 10.1016/j.future.2021.07.025.

[37] P. Filzmoser, K. Hron, and M. Templ, *Applied Compositional Data Analysis*. Cham, Switzerland: Springer, 2018.

[38] M. Jasiński, T. Sikorski, and K. Borkowski, "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.*, vol. 166, pp. 52–60, Jan. 2019, doi: 10.1016/j.epsr.2018.09.020.

[39] S. Wierzchoń and M. Kłopotek, *Modern Algorithms of Cluster Analysis*, vol. 34. Cham, Switzerland: Springer, 2018.

[40] M. Jadidbonab, B. Mohammadi-Ivatloo, M. Marzband, and P. Siano, "Short-term self-scheduling of virtual energy hub plant within thermal energy market," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3124–3136, Apr. 2021, doi: 10.1109/TIE.2020.2978707.

[41] S. Wierzchoń and M. Kłopotek, *Algorithms of Cluster Analysis*, vol. 3. Warsaw, Poland: Institute of Computer Science Polish Academy of Sciences, 2015.

[42] S. Polska. (2016). *StatSoft Electronic Statistic Textbook*. [Online]. Available: http.:/www.statsoft.pl/textbook/stathome.html

[43] J. Yang, J.-Y. Lee, M. Choi, and Y. Joo, "A new approach to determine the optimal number of clusters based on the gap statistic," in *Proc. Int. Conf. Mach. Learn. Netw.*, 2020, pp. 227–239.

[44] K. Chowdhury, D. Chaudhuri, and A. K. Pal, "An entropy-based initialization method of $K$-means clustering on the optimal number of clusters," *Neural Comput. Appl.*, vol. 33, pp. 6965–6982, Nov. 2020, doi: 10.1007/s00521-020-05471-9.

[45] J. Silva, O. B. P. Lezama, N. Varela, J. G. Guiliany, E. S. Sanabria, M. S. Otero, and V. Á. Rojas, "U-control chart based differential evolution clustering for determining the number of cluster in $k$-means," in *Proc. Int. Conf. Green, Pervasive, Cloud Comput.*, 2019, pp. 31–41.

[46] *Standard for Interconnection and Interoperability of Distributed Energy Resources With Associated Electric Power Systems Interfaces*, IEEE Standard 1547-2018 (Revision of IEEE Standard 1547-2003), IEEE Standard Association, 2018.

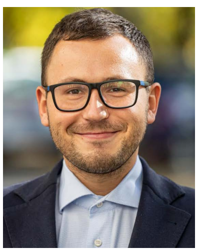
[47] L. Martirano, M. Jasinski, A. Najafi, V. Cocira, and Z. Leonowicz, "Integration of supervision and monitoring systems of microgrids—A case study from lambda microgrid for correlation analysis," in *Proc. IEEE Int. Conf. Environ. Electr. Eng. IEEE Ind. Commercial Power Syst. Eur. (EEEIC/ICPS Europe)*, Sep. 2021, pp. 1–4, doi: 10.1109/EEEIC/ICPSEurope51590.2021.9584689.

[48] R. Claeys, H. Azaioud, R. Cleenwerck, J. Knockaert, and J. Desmet, "A novel feature set for low-voltage consumers, based on the temporal dependence of consumption and peak demands," *Energies*, vol. 14, no. 1, p. 139, Dec. 2020, doi: 10.3390/en14010139.

[49] J. Lei, "Cross-validation with confidence," *J. Amer. Stat. Assoc.*, vol. 115, no. 532, pp. 1978–1997, Oct. 2020, doi: 10.1080/01621459.2019.1672556.

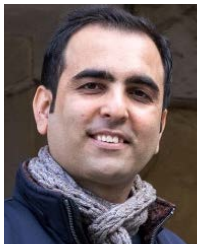
[50] B. G. Marcot and A. M. Hanea, "What is an optimal value of $k$ in $k$-fold cross-validation in discrete Bayesian network analysis?" *Comput. Statist.*, vol. 36, no. 3, pp. 2009–2031, Sep. 2021, doi: 10.1007/s00180-020-00999-9.

[51] S. Kapil and M. Chawla, "Performance evaluation of $K$-means clustering algorithm with various distance metrics," in *Proc. IEEE 1st Int. Conf. Power Electron., Intell. Control Energy Syst. (ICPEICES)*, Jul. 2016, pp. 1–4, doi: 10.1109/ICPEICES.2016.7853264.



**MICHAŁ JASIŃSKI** (Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the Wrocław University of Science and Technology, in 2016 and 2019, respectively. Since 2018, he has been with the Electrical Engineering Faculty, Wrocław University of Science and Technology, where he is currently an Assistant Professor. He is an author and coauthor of over 100 scientific publications. His research interests include using big data in power systems especially in point of power quality, as well as optimization in multicarrier energy systems. Currently, he is a Guest Editor for the special issues of *Energies*, *Electronics*, *Sustainability*, and *Frontiers in Energy Research*.

**LUIGI MARTIRANO** (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the Sapienza University of Rome, Rome, Italy, in 1998 and 2003, respectively. In 2000, he joined the Department of Electrical Engineering, Sapienza University of Rome, where he is currently an Associate Professor of electrical power systems. He is currently the Coordinator of the Ph.D. course in "engineering and applied science for energy and industry" and the Secretary of the National Research Group in Power Systems GUSEE. He has authored more than 180 papers in international journals and conference proceedings. He is the Founder of a startup, operating in the fields of building automation and energy monitoring systems. His research interests include safety and reliability in power systems, MV/LV power networks, building automation, domotics, lighting, smart buildings, and microgrids.



**ARSALAN NAJAFI** (Senior Member, IEEE) received the B.S. degree in electrical engineering from the University of Kurdistan Hewlêr, Sanandaj, Iran, in 2009, and the M.S. and Ph.D. degrees in electrical engineering from the University of Birjand, Birjand, Iran, in 2011 and 2016, respectively. He is supported by the Polish National Agency for Academic Exchange (NAWA) under the Ulam Program Grant. He is currently an Assistant Professor with the Wrocław University of Science and Technology, Wrocław, Poland. His research interests include operation and planning of multi-energy systems, stochastic programming, electricity market, as well as optimization theory and its application in power systems.

**OMID HOMAEE** received the B.Sc. degree in electrical engineering from the University of Birjand, Birjand, Iran, in 2011, and the M.Sc. and Ph.D. degrees in electrical engineering from the Iran University of Science and Technology, Tehran, Iran, in 2013 and 2020, respectively. He is currently a Researcher with the Wocław University of Science and Technology, Wocław, Poland, supported by the Polish National Agency for Academic Exchange (NAWA) under the Ulam Program Grant. His current research interests include smart grids and electromagnetic transient analysis.

**ZBIGNIEW LEONOWICZ** (Senior Member, IEEE) received the M.S. and Ph.D. degrees in electrical engineering from the Wocław University of Science and Technology, in 1997 and 2001, respectively, and the Habilitation degree from the Bialystok University of Technology, in 2012. Since 1997, he has been with the Electrical Engineering Faculty, Wocław University of Technology. He also received the two titles of Full Professor from the President of Poland and the President of the Czech Republic, in 2019. Since 2019, he has been a Professor with the Department of Electrical Engineering, where he is currently the Head of the Chair of Electrical Engineering Fundamentals.

**MOSTAFA KERMANI** (Member, IEEE) received the M.S. degree in electrical engineering from the University of Birjand, Birjand, Iran, in 2013, and the Ph.D. degree in electrical engineering from the Sapienza University of Rome, Rome, Italy, in 2019. He is currently a Postdoctoral Researcher with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg, Sweden. He has authored several papers in international journals and conferences. His research interests include power systems operation and planning, energy storages systems, energy optimization, electrical and energy systems for port, and intelligent energy management systems. He is a member of the Scientific Committee in EEEIC conference. He is also a Reviewer of IEEE TRANSACTIONS ON INDUSTRY APPLICATIONS, IEEE OPEN JOURNAL OF INDUSTRY APPLICATIONS, and Elsevier and MDPI journals.