21th International Conference on Renewable Energies and Power Quality (ICREPQ'23)
Madrid (Spain), 24th to 26th May 2023
Renewable Energy and Power Quality Journal (RE&PQJ)
ISSN 2172-038 X, Volume No.21, July 2023

Phase Identification in Smart Grids – Case Study

Lluc Crespí-Castañer[1], Miquel Roca[1], Josep Lluís Rosselló[1], Lluís Juncosa[2] and Vicente Canals[1]

[1] Department of Industrial Engineering and Construction
University of the Balearic Islands
Carretera de Valldemossa, km 7.5, 07122 Palma (Spain)
Phone/Fax number: +34 971 173 211, e-mail: **lluc.crespi@uib.es**

[2] Vall de Sóller Energia SLU
C/ Sa Mar 146, 07100 Sóller (Spain)
Phone/Fax number: +34 900 373 067, e-mail: oficines@valldesollerenergia.es

Abstract. A widespread problem of the electricity distribution system is determining to which phase a consumer is connected. Generally, utility companies did not record the distribution of the phases over time, which is essential to address the problem of unbalanced voltage distribution networks. In addition, manual phase identification is not feasible due to its high cost. The massive deployment of smart meters allowed periodic readings of energy consumption, voltage, current, etc. Out of this large amount of data, several techniques based on machine learning have emerged, addressing the problem of phase identification automatically while maintaining the existing infrastructure. Hence, phase identification is essential for developing smart-grid solutions. In the present work, we applied an unsupervised machine-learning technique that allows the classification of the time-voltage series recorded by the smart meters of a low-voltage three-phase radial distribution network located in the Balearic Islands (Spain). The results show that there is a correlation between the time series of the feeder voltage and the consumer meters. The proposed method reached a 100% success rate in the case study. In addition, the results obtained open the way to deploy a new grid configuration to minimize the load imbalance.

Key words. Phase identification, Smart grids, Constrained Multi Tree, Voltage correlation

1. Introduction

Nowadays, the challenges of distributed generation and the issue of phase imbalance are driving distribution companies to identify the phases of their meters; however, manual identification is not feasible due to its high cost. Historically, electric utilities have not recorded the phase distribution in the deployment of new supplies, or these records have been deteriorating over time, which causes phase identification to be mandatory.

Phase identification is essential to address the unbalanced load network issue, which carries the following downsides: power losses, life span shortening of the grid assets, and decreasing electric power supply quality [1].

The transition of the energy system towards a more decarbonized and decentralized model involves implementing low-carbon emission technologies that demand an extra load on distribution networks [2]. As a result, operators must manage their power lines optimally and efficiently. To this end, numerous analysis and modeling tools for lines are developing, such as the popular digital twins [3], [4], which require knowledge of the phases of the different meters.

The emerging development of smart grids [5] provides energy companies with a large amount of data. The meters which gather these variables have been mandatory in Spain for ≤15kW supplies since 2018 [6]. Therefore, such monitoring is no longer exclusive to high-voltage grids, as was usual, but now covers all low-voltage grids. Consequently, the emergence of different innovative methodologies [7]–[10], which combines machine-learning techniques and big-data analysis, allows the improvement and optimization of network management. Specifically, these techniques enable the identification of the phases using telemetry data. Their main advantage is that they do not require the deployment of expensive additional hardware systems or knowledge of the network topology. These techniques can be divided into two categories depending on the type of data: consumption-based and voltage-based approaches. Among the different works related to the identification of phases based on the voltage-based approximation, we highlight the following. V. Arya et al. [11] propose a method to identify the phases of smart meters by applying the principle of conservation of electric charge. The obtained equations are simplified and then solved using mathematical optimization. Alternatively, Oliver et al. [12] propose an algorithm called **C**onstrained **M**ulti **T**ree (CMT) that classifies the phases by finding the correlation between the voltage time series of the meters and the feeder. Another approach that is based on the classification of voltage time series, among which we will highlight. W. Wang et al. [13] propose a method based on **A**dvanced **M**easurement **I**nfrastructure (AMI) data, which are centered and normalized. In the second place, the data is size reduced by applying **P**rincipal **C**omponent **A**nalysis (PCA) to extract the most important features of the voltage time series. Finally, the top components are classified using the K-Means algorithm calculating the Euclidean distances between them. Must-link and cannot-link restrictions are applied by the topology grid knowledge. Its main drawback is that the phase identification accuracy of the constrained k-Means decreases as the level of network unbalance does. Although, W. Wang et al. improved this approach on the work [13]. In the first stage, key features of the voltage time series are extracted with a dimensionality reduction method as well. First, the dimension of the raw data is reduced by applying PCA. Then, t-SNE is used to further reduce the dimensionality and extract the nonlinear features. In the second stage, the proposed CHC algorithm groups the data points. This algorithm uses DBSCAN as a density-based clustering algorithm to separate data points in different clusters. Finally, the k-NN (**k**-**N**earest **N**eighbors) algorithm assigns the outliers points that haven not been classified by DBSCAN to an existing cluster. More recently, H. Yu et al. [14] have presented a similar method for phase detection based on the DBSCAN algorithm, reaching 90% accuracy. It should be noted that the results have been obtained from synthetic data obtained from the IEEE European Low Voltage Test Feeder. In contrast to the methods cited above, Blakely et al. [15] use spectral clustering to identify the phases claiming that feeder measurements and topology of the network are not required. This approach is also based on clustering the phases by calculating the correlation between pairs of voltage timeseries. One of its advantages consists of clustering the data within sliding windows of a few days. In this way, voltage profiles containing missing data are simply removed. However, this method requires a lot of measurements. It should be noted that a large part of the phase identification methodologies in distribution networks analyzed have been tested on synthetic data, obtained from simulations, and not on real data from the AMI, which incorporate an additional challenge associated with the treatment of these data (data science).

This work addresses the problem of phase identification in a real environment through non-intrusive techniques such as those based on Machine Learning as a basic element for the determination of phase imbalance in distribution networks, in areas with a high seasonality in electrical demand, as it happens in the touristic areas of the Spanish coast. Specifically, we have developed an algorithm based on the CMT methodology, written in Python, to process the data files obtained from a low-voltage distribution grid by the AMI of the company Vall de Sóller Energía S.L.U. The present work has been structured in study is structured in 5 sections. This first section serves as an introduction to the problem of phase identification using Machine Learning techniques. The second section introduces the low voltage power distribution sectors to be studied in this work. The third section presents the methodology used to carry out this study. In the fourth section the results obtained are presented and discussed. Finally, the conclusions of this work are presented in the last section.

2. Case study

The phase identification and load imbalance of a 3-phase low-voltage radial distribution grid is addressed in this work. The low-voltage distribution grid analyzed is in Sóller, the Balearic Islands (Spain), as shown in Fig. 1. Sóller is a beautiful town on the Northwest Coast of the island of Majorca. It is located about 3 kilometers from its port, Puerto de Sóller. The town stands on a fertile valley surrounded by mountains. Sóller divides its valley between the town of Fornalutx and the village of Biniaraix, whose population amounts to about 14,000 people.

Fig. 1. The distribution grid areas studied locations, in Sóller the Balearic Islands (Spain).

Sóller is an eminently tourist municipality, visited by 407,234 tourists throughout 2017. In turn, in this period 66.7% of tourists stayed in hotels, 15.6% in rental accommodation (tourist houses) and 17.7% in private residences. To this great flux of floating population must be added the problem of seasonality. Given that during the low season months (November, December, January, and February) almost all tourist establishments remain closed or with very low occupancy rates, concentrating the activity in the high season. This has a great impact on the electricity demand of the low voltage distribution networks and their imbalances.

The local utility company, Vall de Soller Energia S.L.U, has provided the telemetry data and has made several field measurements of the phase connectivity to compare the results of the applied methodology. It has provided voltage and consumption data from two distribution areas named Binidorm and Dalt des Traves. The first area consists of one feeder, ten 3-phase meters, and 37 single-phase meters. The second consists of one feeder, 33 3-phase meters, and 14 single-phase meters. The recording period of the voltage data is 01/07/2022 to 07/07/2022 for the first area and 02-06-2022 to 07-06-2022 for the second area, with a 15 min period. After filtering the data, we obtained 501 and 415 voltage measurements per phase and meter with a 17 min sample period for Binidorm and Dalt des Traves, respectively. Regarding consumption data, local utility company has provided the hourly consumption of each meter for the year 2022.

3. Methodology

The phase identification issue may approach as a machine-learning classification problem. The voltages of every meter vary in function of the impedance of the line that links them. Moreover, the variation among the voltages of the three phases increases on the feeder and throughout the lines owing to the load imbalance. Thereby, the voltages of the meters closer and connected to the same phase tend to be more correlated than the voltages of meters connected to different phases, given a time window. In this way, we implemented a methodology based on the CMT algorithm to classify the phases of the meters by the correlation between voltages. The CMT algorithm is based on graph theory, specifically on Prim's algorithm. The main advantages of this method are that the classification is carried out considering the radial structure of the network, it is designed for the features of European LV distribution networks, it requires relatively small time-windows, and the performance reaches almost 100%. Furthermore, this method reports better performance than the K-Means algorithm [12].

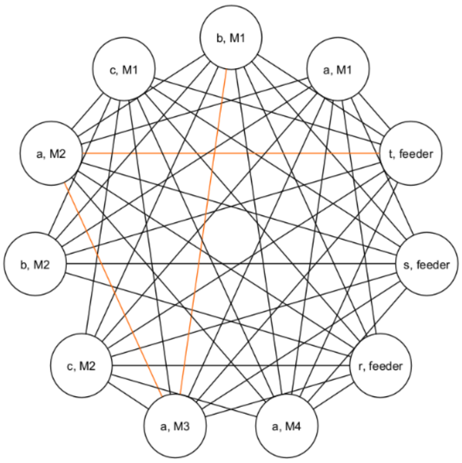

Fig. 2. Initial graph

Initially, a graph connecting all the nodes, except restricted ones, is created. The nodes are the phases of the meters. Meanwhile, the weights of the edges are a function of the correlation between its nodes. Specifically, this method computes the Pearson correlation between the voltage time series of the meters and the feeder. The nature of the problem requires that some nodes cannot be bound, such as the phases (nodes) of the same meter. The Fig. 2 shows an example graph of a distribution grid made up of one feeder, two 3-phase meters (M1 and M2) and two single phase meters (M3 and M4). As it can be seen, all the phases are interlinked except those belonging to the same meter. Once the graph is set up, the CMT algorithm finds three minimum spanning trees corresponding to the three phases of the network. In the Fig. 2 the tree corresponding to the phase *t* of the feeder is orange colored. Unlike Olivier et al. [12], we first classified the phases or nodes of 3-phase meters and then the others, obtaining better results.

Due to the technical limitations of the communication system used by the meters, based on the PLC (Power-Line Carrier) protocol, there are delays in reading measurements and data loss. Because of these limitations, the data have been treated in the following way. On one hand, the voltage measurements are interpolated to fill in the gaps caused by reading errors. On the other hand, the data is resampled with the same timestamp. Thus, for each voltage measurement of the feeder, there is a measurement of every meter at the same timestamp.

A detailed description of the methodology developed in this work to identify the phases is presented in the flowchart of Fig. 3.

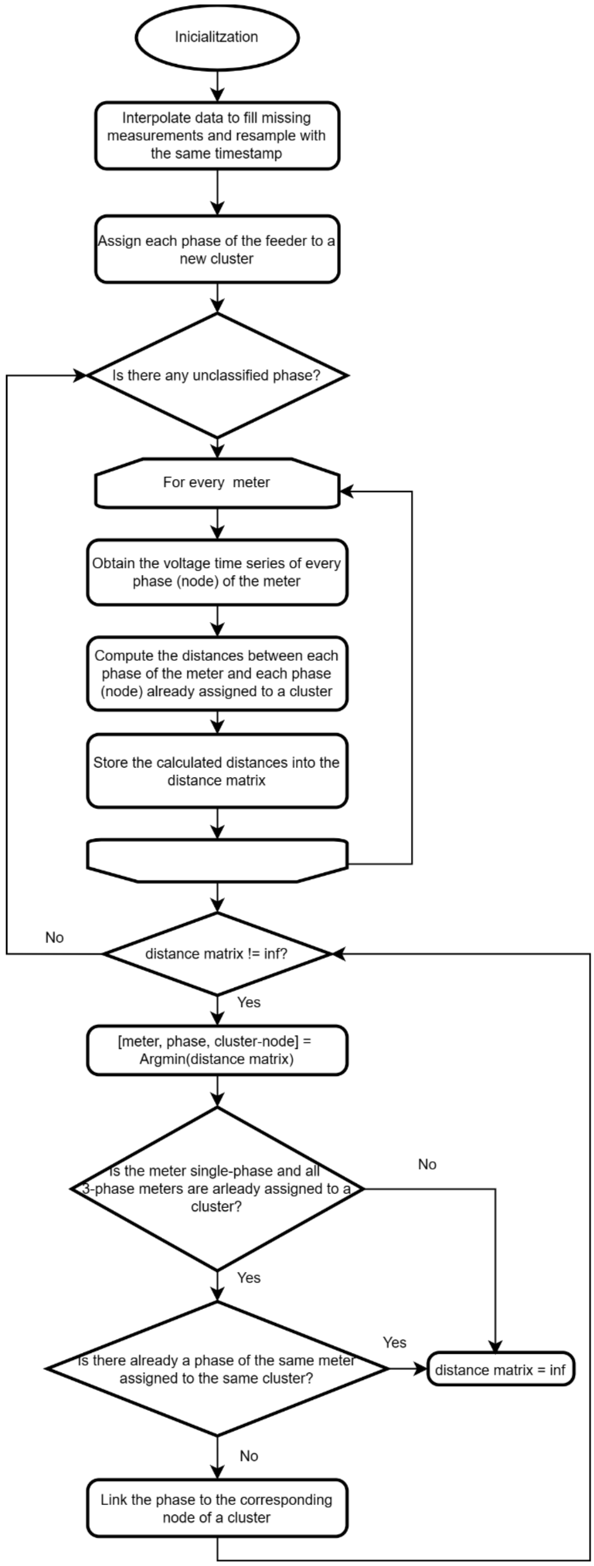

Fig. 3. Methodology flow chart

4. Results

The classification results obtained by applying the methodology described above were compared to a manual identification to evaluate the algorithm's real performance. For this purpose, Vall de Soller Energia SLU, the local utility, carried out a visual inspection to identify the phases of 12 smart meters in the Binidorm area and 11 meters in the Dalt des Traves area. The results of the CMT algorithm on these datasets were found to be consistent with the manual identification in both areas. Therefore, these results support the performance indices claimed by the authors of the proposed method. The manual and automatic phase identification of each meter are shown in Tables I and II.

Since the proposed method is based on the calculation of correlations between voltage measurements, the number of measurements and the period with which they are taken influences the clustering result. In this way, the longer the length of the time series, the more variability between the voltages of each phase can be detected, thus improving the classification. However, voltage-based methods, specifically the one proposed, do not require a great number of samples compared to consumption-based methods in general.

Table I. - Results of Binidrom

Meter identifier	Manual identification	CMT algorithm
34648582	2	2
34648440	0	0
34648577	0	0
34648272	1	1
34648438	1	1
37417618	0	0
34648274	2	2
34701714	1	1
34648252	1	1
34952257	2	2
34648439	1	1
34648444	0	0

Table II. - Results of Dalt des Través

Meter identifier	Manual identification	CMT algorithm
34701915	0	0
50116464	1	1
50114758	1	1
50116483	2	2
36240068	0, 2, 1	0, 2, 1
50116726	0	0
36240105	0, 2, 1	0, 2, 1

This fact is important due to utilities often switch the phases of the meters when connecting new supplies or for maintenance reasons. In this context, the phase identification results of the evaluated meters of tables I and II remained constant with only two days of samples.

In addition, the phase identification of all the meters and the consumption data enables the analysis of the network load unbalances. To do this, the hourly consumption of every meter has been aggregated per phase according to the results of the proposed method. In the case of Dalt des Través, the detected phase imbalance is negligible. While in the case of the Binidorm area, as shown in Fig. 4(a), the shapes of the distributions do not overlap, which implies that the demand for the different phases has been quite different among them in the period analyzed. In this way, a Box-Cox transformation was applied to the data obtaining the mean and deviation per phase. The average demand of phase L2 becomes 6kW higher than the one of phase L3 and the range between ±3σ is above 40kW. Moreover, the imbalance level is higher than 10% for the phases L1 and L2, between a range of 10 – 17kW, which is an elevated demand that could cause damage to the grid. The power demand of each lateral, shown in Fig. 5, can also be analyzed, since it is known to which lateral is connected each meter. The level of imbalance in the first and third laterals is worrying, while the second lateral is well-balanced. The demand for phase L2 of the third lateral is significantly greater than the demand for phases L1 and L3, which could result in extra losses on the power lines. Thanks to phase identification, we discovered that the number of single-phase meters connected to phases L1 and L3 was significantly higher than those connected to phase L2. Additionally, the demand for phase L3 of the first lateral is very low most of the time, which is due to the lack of single-phase meters connected to that phase.

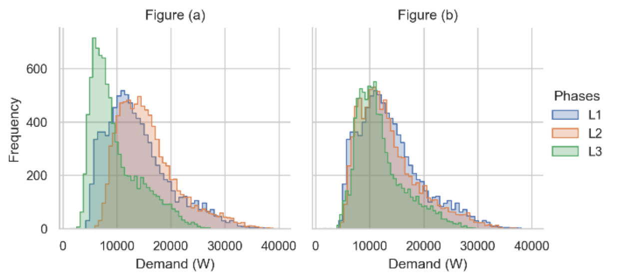

Fig. 4. Aggregated demand per phase of the Binidorm area.

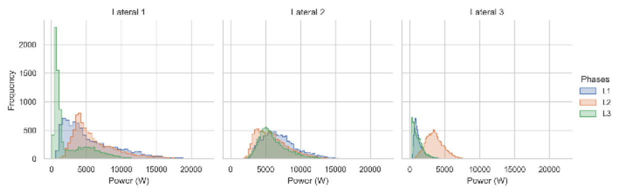

Fig. 5. Aggregated demand per phase and lateral of the Binidorm area.

Consequently, based on the results of the identification of phases and analysis of customer consumption, it is possible to address new load distribution configurations to minimize phase imbalance. Based on the proposed phase identification methodology, we have developed an automated methodology for the optimal redistribution of single-phase consumers in a low voltage grid. The developed methodology makes use of a genetic algorithm [16] specially developed to minimize the imbalances between bases.

To do this, a supervised classification algorithm has been implemented for 3 categories, based on the K-NN algorithm, to identify: triphasic supplies, single-phase supplies with a seasonal profile and single-phase supplies with a uniform annual profile.

Once the supplies have been classified, a genetic algorithm provided with a cost function, which minimizes imbalances between phases, to redistributing the single-phase supplies in an optimal way, since the three-phase supplies are connected to each of the three phases. Specifically, the genetic algorithm firstly redistributes the single-phase supplies that do not present a seasonal demand, to then redistribute the seasonal supplies among the three phases. As a result, six single-phase meters from phase L2 in the Binidorm area have been reconnected to phase L3, reducing the load imbalance, as shown in Fig. 4(b). As can be seen in the previous figure, the algorithm has managed to significantly reduce the imbalance of the phases due to the large number of seasonal supplies, typical of coastal tourist areas in Spain.

5. Conclusions

This work has presented a methodology based on the CMT algorithm, an unsupervised machine learning technique, for phase identification in two (real) low-voltage distribution networks and has analyzed the supply imbalances of one of them. The proposed method classifies the phases of the supplies based on the temporal correlation of the voltages of the smart meters. The results show a successful performance of the algorithm in the two distribution zones studied, since its identification coincides with that physically carried out by the operators of the local public service company. Reaching 100% success in identifying the phases of the analyzed areas, reinforcing the validity of the proposed methodology, which in turn coincided with the conclusions of Olivier et al. [12]. In addition, the analysis of the consumption of each smart-meter together with the identification of phases makes it possible to detect load imbalance problems in the low-voltage network. Since phase identification reveals physical information about the network structure, it is possible to detect unbalanced load problems on the feeder and other nodes in the low-voltage network. Since phase identification reveals physical information about the structure of the network, this has allowed us to identify phases mislabeled by the local utility, as well as to identify unbalanced load problems on the feeder and other nodes in the low voltage network. Finally, based on the phase identification methodology presented in this paper, a methodology based on genetic algorithms has been implemented for the redistribution of single-phase supplies in networks that present a highly seasonal demand, as occurs in tourist areas.

Acknowledgement

This work has been funded by the Spanish Ministry of Science and Innovation (MICINN) and the European Regional Development Funds (ERDF) under Grants PID2020-120075RB-I00, PDC2021-121847-I00 and CIN/AEI/10.13039/501100011033/ by the "FEDER una manera de hacer Europa"; in part by the "European Union NextGenerationEU/PRTR"; and partly by the "Programa SOIB Investigación e Innovación" for the period 2022-2025 through the employment contract, with reference RI-01.33/22-2, related to the job offer with reference "042022003852"; all this within the framework of the Recovery and Resilience Mechanism aimed at executing investment projects of "*Nuevas políticas públicas para un mercado de trabajo dinámico, resiliente e inclusivo*" included in the Recovery, Transformation and Resilience Plan financed by the European Union (Next Generation EU), with the participation of the Spanish Ministry of European Funds, University and Culture, through the General Directorate of University Policy and Research.

References

[1] J. Zhu, M. Y. Chow, and F. Zhang, "Phase balancing using mixed-integer programming," *IEEE Transactions on Power Systems*, vol. 13, no. 4, pp. 1487–1492, 1998, doi: 10.1109/59.736295.

[2] European Commission, "A Roadmap for moving to a competitive low carbon economy in 2050," Mar. 2011. Accessed: Jan. 10, 2023. [Online]. Available: https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0112:FIN:en:PDF

[3] M. Atalay and P. Angin, "A Digital Twins Approach to Smart Grid Security Testing and Standardization," in *2020 IEEE International Workshop on Metrology for Industry 4.0 & IoT*, Jun. 2020, pp. 435–440. doi: 10.1109/MetroInd4.0IoT48571.2020.9138264.

[4] A. Francisco, N. Mohammadi, and J. E. Taylor, "Smart City Digital Twin–Enabled Energy Management: Toward Real-Time Urban Building Energy Benchmarking," *Journal of Management in Engineering*, vol. 36, no. 2, Mar. 2020, doi: 10.1061/(ASCE)ME.1943-5479.0000741.

[5] M. L. Tuballa and M. L. Abundo, "A review of the development of Smart Grid technologies," *Renewable and Sustainable Energy Reviews*, vol. 59, pp. 710–725, Jun. 2016, doi: 10.1016/j.rser.2016.01.011.

[6] E. y T. Ministerio de Industria, "Orden IET/290/2012, de 16 de febrero, por la que se modifica la Orden ITC/3860/2007, de 28 de diciembre, por la que se revisan las tarifas eléctricas a partir del 1 de enero de 2008 en lo relativo al plan de sustitución de contadores.," Madrid, Feb. 2012. Accessed: Jan. 13, 2023. [Online]. Available: https://www.boe.es/diario_boe/txt.php?id=BOE-A-2012-2538

[7] T. A. Short, "Advanced Metering for Phase Identification, Transformer Identification, and Secondary Modeling," *IEEE Trans Smart Grid*, vol. 4, no. 2, pp. 651–658, Jun. 2013, doi: 10.1109/TSG.2012.2219081.

[8] H. Padullaparti, S. Veda, J. Wang, M. Symko-Davies, and T. Bialek, "Phase Identification in Real Distribution Networks with High PV Penetration Using Advanced Metering Infrastructure Data," in *2022 IEEE Power & Energy Society General Meeting (PESGM)*, Jul. 2022, pp. 01–05. doi: 10.1109/PESGM48719.2022.9916986.

[9] K. Montano-Martinez *et al.*, "Detailed Primary and Secondary Distribution System Model Enhancement Using AMI Data," *IEEE Open Access Journal of Power and Energy*, vol. 9, pp. 2–15, 2022, doi: 10.1109/OAJPE.2021.3125900.

[10] Z. S. Hosseini, A. Khodaei, and A. Paaso, "Machine Learning-Enabled Distribution Network Phase Identification," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 842–850, Mar. 2021, doi: 10.1109/TPWRS.2020.3011133.

[11] V. Arya *et al.*, "Phase identification in smart grids," in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Oct. 2011, pp. 25–30. doi: 10.1109/SmartGridComm.2011.6102329.

[12] F. Olivier, A. Sutera, P. Geurts, R. Fonteneau, and D. Ernst, "Phase Identification of Smart Meters by Clustering Voltage Measurements," in *2018 Power Systems Computation Conference (PSCC)*, Jun. 2018, pp. 1–8. doi: 10.23919/PSCC.2018.8442853.

[13] W. Wang, N. Yu, and Z. Lu, "Advanced Metering Infrastructure Data Driven Phase Identification in Smart Grid."

[14] Y. Zhou *et al.*, "Distribution Network Electrical Topology Identification Based on Edge Computing and Improved KNN," in *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering*, Nov. 2020, pp. 806–813. doi: 10.1145/3443467.3443858.

[15] L. Blakely, M. J. Reno, and W. C. Feng, "Spectral Clustering for Customer Phase Identification Using AMI Voltage Timeseries," in *2019 IEEE Power and Energy Conference at Illinois, PECI 2019*, Apr. 2019. doi: 10.1109/PECI.2019.8698780.

[16] L.-M. Ionescu, N. Bizon, A.-G. Mazare, and N. Belu, "Reducing the Cost of Electricity by Optimizing Real-Time Consumer Planning Using a New Genetic Algorithm-Based Strategy," *Mathematics*, vol. 8, no. 7, p. 1144, Jul. 2020, doi: 10.3390/math8071144.