User-transformer connectivity relationship identification based on knowledge-driven approaches

Lai Zhou¹, Fujun Wen¹, Xianfu Yang¹, Yuming Zhong¹
¹School of Intelligent Manufacturing, Guangzhou Panyu Polytechnic, Guangzhou 511483, China

Corresponding author: Fujun Wen (e-mail: 11805914@qq.com).

This work was supported in part by the Science and Technology Project Foundation of Guangzhou Panyu Polytechnic, China, under Grant 2022KJ10

ABSTRACT Accurate user-transformer connectivity relationship (UTCR) plays a key role in fine management of low-voltage distribution network (LVDN) i.e., load expansion, line loss management, and electrical service restoration after outage. Limited data and low discriminability and noise in data increase the difficulty to identify UTCR for the existing data analytics methods. To overcome these hurdles, this paper proposes a novel UTCR algorithm which combining the data preprocessing with multi-dimensional priori knowledge based on voltage characteristics in LVDN. Firstly, the prior knowledge related to UTCR are refined on account of voltage correlation characteristics of users at different locations to provide theoretical foundation. Then, Z-score and principal component analysis are combined to standardize and extract features from the original voltage data to magnify the differences between data and reduce the impact of data noise. Further, on the basis of the prior knowledge of voltage correlation characteristics, a knowledge-driven identification model is proposed to identify users with wrong UTCR and their real UTCR. Finally, the performance of the proposed algorithm is verified on simulated LVNDs. The comparison analysis between the proposed method and other published methods and the impact of the number of principal components on the identification accuracy are also investigated. The results indicate that the proposed method achieves higher recognition accuracy than other published methods with low discriminability and noise in data.

INDEX TERMS User-transformer connectivity relationship identification, low-voltage distribution network, data pre-processing, voltage correlation characteristics, knowledge-driven approaches

I. INTRODUCTION

The massive use of fossil fuels has two drawbacks: resource depletion and climate crisis, which violates the sustainable development goal. To tackle this problem, many countries around the world have taken carbon neutrality into the development plan [1-2]. Under this background, the penetration level of new equipment, i.e., rooftop photovoltaic, electric vehicle, and energy storage in LVDN increases gradually [3-5]. The development of these devices can effectively alleviate the pressure of environmental pollution and energy tension, but it has brought impacts and challenges to the safe operation and power supply quality of LVDN [6-7]. In order to fully dig the potential of distributed energy resources meanwhile operating the grid in an efficient and reliable manner, a high-level operation and maintenance management in LVDN is needed [8]. Of which, accurate low-voltage physical topology connection information is a vital foundation to support the intelligent construction of LVDN [9-10]. Low-voltage topological connections include the connections between distribution transformers, phase sequences, feeders and users. This paper focus on user-transformer connectivity relationship (UTCR) identification, defined as the connection relationship between the terminal user's electricity meter and the distribution transformer of LVDN.

With the development of economy and the continuous advancement of urbanization, the number of users in LVDN increases rapidly, and UTCR changes frequently. However, affected by low efficiency of investigation and untimely update of information, the UTCR information that utilities have usually existed many errors. Accurate UTCR is the key to load expansion, line loss management, electrical service...
restoration after outage, line transformation and other services, and is also the premise to accurately identify the connections among phase sequences, feeders and users [11-12]. The traditional methods relying on manual techniques and signal injection devices are inefficient and need high investment cost, which is hard to afford for grid companies. Besides, it cannot update topology information automatically. Therefore, it is important to study the automatic recognition technology for UTCR.

Nowadays, deploying smart meters in LVDN is development trend [13]. In particular, China has achieved 100% penetration of smart meters in LVDN since 2019. In this context, a large amount of users’ electricity consumption data and network operation data can be obtained. New approaches are intensively investigated to apply the data acquired by smart meters for the planning and operation of distribution systems, i.e., non-technical loss detection, non-intrusive load monitoring, power quality assessment, fault location [14-15]. Similarly, with the advantages of low cost and convenient, new methods using smart meters data have been widely employed for topology connectivity identification in LVDN [16-19]. [16] and [17] focus on the topology and parameter estimation for LVDN with smart meters. [18-19] are methodologies utilizing voltage data and current data from smart meters and transformers to recognize phase connectivity relationship of users. All of the above studies need accurate UTCR as a priori knowledge. By the data types they require, the existing methods for UTCR recognition could be categorized into five sets.

1) Power data: in [11], based on the power data of users and transformers, a quadratic optimization model was established to minimize the network loss fluctuation rate to determine UTCR. In [12], the combination of linear regression and a Dirichlet-Categorical allocation model sampled with Markov chain Monte Carlo was proposed to track and identify low-voltage topology changes, in which the UTCR information was included. In [20], a de-noised differential evolution-based method was proposed for topology identification.

2) Current data: in [21], the high-frequency features from current data were extracted by Discrete Fourier Transform and Inverse Discrete Fourier Transform. Further, an optimization model based on Kirchhoff’s law of current was constructed to get UTCR.

3) Voltage data: in [22], performing Fisher Z transform on Pearson correlation coefficient matrix based on the total harmonic voltage were proposed to determine low-voltage connectivity. In [23], voltage correlation factors between all customers were calculated, and the users with strong correlations were assumed to be on the same transformer. In [24], voltage curve correlation analysis between users and low-voltage buses was employed to verify UTCR.

4) GIS data: in [25], new procedures that exploit the graph theory and data structure properties were presented to detect and correct errors in models of LVDN.

5) Multi-source data: in [26], regression and basic voltage drop relationships based on power data and voltage data were employed to generate secondary connectivity and impedance models. In [27], principal component analysis and independent component analysis were employed to extract features from voltage data. Then, the Pearson correlation analysis between the users’ total current and transfer’s current was used to realize UTCR recognition. In [28], a two-stage approach for UTCR was proposed based on voltage data and power data. At first stage, correlation analysis was employed to ensure transformers with errors in UTCR, then a linear regression formulation was built to correct the errors. In [29], a multiple linear regression model using voltage and power data of customers meters was established to estimate topology, line parameters, and customer and line phasing connections in LVDN.

The methods in the first and second categories using current and power data are suitable for scenarios where consumers’ power consumption characteristics are obvious and there is no electricity theft and unmetered load. However, in practice, it is also common that the obtained power consumption data of consumers is incomplete and cannot fully reflect the power consumption of LVDN, resulting from poor communication quality, human error, unregistered meters, and electricity theft. The methods in the fourth category need GIS data. GIS data for LVDN are not available for many areas such as China. The application scenarios of the methods using GIS data are limited. For the methods based on linear regression in the fifth category, on the one hand, it requires a lot of data, including the voltage, active and reactive power data of all users in LVDN. Nevertheless, it is difficult to provide complete data in many areas, which reduces the effective application scenarios of the methods. On the other hand, there are many parameters involved in the regression model, and parameter thresholds need to be set. How to select appropriate parameter thresholds increases the difficulty of applying the methods.

Voltage-based methods in the third category have strong robustness to electricity theft and unmetered load. The voltage correlation characteristics analysis among users or that between users and transformer was employed in the existing voltage-based methods individually. However, the voltage correlation characteristics of users depend on their location. There may be contradictory correlation characteristics between users located in different locations. Hence, the existing voltage-based methods only using one correlation characteristic are not sufficient to accurately identify UTCR.

Besides, the existing voltage-based methods lack data preprocessing, and have less robustness to data discrimination and noise. In practice, the voltage data collected from LVDN with three-phase unbalanced governance tend to be centralized. And the difference between user’s voltage characteristics is small, which affects the accuracy of the algorithm. Moreover, affected by meter measurement errors and communication problems,
the data collected by smart meters often contains some noise. Low discriminability and noise in data affect the performance of the existing voltage-based methods.

In conclusion, despite its importance, how to identify UTCR accurately with voltage data and enhance the robustness of the identification method to data discrimination and noise has not been well investigated. To overcome these hurdles, this paper introduces data preprocessing and multi-dimensional prior knowledge in UTCR algorithm based on voltage characteristics in LVDN. The improvements on existing voltage correlation approaches include:

1) The voltage correlation characteristics of users at different locations are deduced, and the prior knowledge related to UTCR are further refined, which provides a theoretical basis for the recognition algorithm.

2) Z-score and principal component analysis are combined to standardize and extract features from the original voltage data to magnify the differences between data and reduce the impact of data noise.

3) Based on the prior knowledge of voltage correlation characteristics, a knowledge-driven identification model is proposed to identify users with wrong UTCR and their real UTCR.

4) Compared with the existing methods, the proposed algorithm achieves higher UTCR identification accuracy, and has better robustness to data discrimination and noise.

The rest of this paper is organized as follows. Section II describes the problem formulation. Section III deduces the prior knowledge of UTCR based on voltage data. Section IV describes the mathematical model of the proposed UTCR algorithm. The tests and results are illustrated in Section V. At last, Section VI presents the conclusion of the study and the future work.

II. PROBLEM FORMULATION

Distribution system is the final portion of electric power system and feeds power from transmission system to users. It includes medium-voltage distribution network and LVDN. At present, the topology of medium-voltage distribution network is available for grid companies. The information that utilities have about LVDN is limited to UTCR which defined as the connectivity between meters and transformer.

An illustration of a simple distribution network is presented in Fig.1. It can be seen from Fig.1, the meters M1, M2, M3, and M4 are powered by the distribution transformer T1. Hence, these meters are considered to belong to LVDN#1. Similarly, the meters M5, M6, M7, and M8 are powered by the distribution transformer T2, these users are considered to belong to LVDN#2. Smart meters located in consumer side measure consumers’ power consumption, voltage, current, power factor, and other data at regular time intervals.

Data concentrator units (DCU) are installed near the distribution transformer in LVDN to collect smart meter data through wireless and power line communication [12]. Terminal meters installed at the low-voltage busses of distribution transformer measure the total power consumed, voltage, current, power factor of each low-voltage bus. As a result, the power consumption, voltage, current, power factor, and other data of consumer and transformer in LVDN can be available for Grid company by DCU and terminal meters. DCU stores the ID of smart meters needed to be collected. This ID information can be regarded as the UTCR that power grid companies can obtain at present. Ideally, DCU contains only the ID information of all meters powered by the LVDN which it located in. However, in practice, it is quite common that the ID information of smart meters in the DCU is not consistent with the actual UTCR. This may have the following situations:

1) DCU contains the ID information of part of meters powered by the LVDN which it located in;

2) DCU contains not only the ID information of all meters powered by the LVDN which it located in, but also the ID information of meters in other LVDN;

3) DCU contains not only the ID information of part of meters powered by the LVDN which it located in, but also the ID information of meters in other LVDN.

The reasons for the discrepancy between the information in DCU and the actual UTCR include: 1) due to LVDN operation mode adjustment, some users are transferred to other LVDN, but the ID information in the DCU is not updated in time; 2) in the complex wiring area, it is difficult to distinguish UTCR, and the wrong ID information of meters was manually recorded. Accurate UTCR is not only the basis for the refinement of LVDN line loss management and energy saving, but also affects the accurate recognition of the following full topology of LVDN. Hence, it is very important and necessary to investigate how to identify UTCR.
voltage calculation formula. On this basis, the prior knowledge related to UTCR is further analyzed. The details are elaborated as follows.

In practice, the service drop line between feeder and household is short. Hence, the voltage drop between users and feeder line is ignored in the theoretical derivation to better understand the voltage characteristics among users. Moreover, the reverse power flow is not considered in the theoretical derivation. The illustration of a feeder line in LVDN is depicted in Fig.2.

1. Voltage space characteristic analysis

$$U^t_u = U^t_{u-1} - \frac{R^t u_i P^t u_i + X^t u_i Q^t u_i}{U^t_{u-1}}$$

$$= U^t_{u-2} - \frac{R^t u_i P^t u_i + X^t u_i Q^t u_i}{U^t_{u-1}} \quad (1)$$

Where, $U^t u_i$ is the voltage of node $u$ at time $t$, $U^t_{u-1}$ is the voltage of node $u-1$ at time $t$, $R^t u_i$ and $X^t u_i$ are the resistance and reactance of line $u$, respectively, $P^t u_i$ and $Q^t u_i$ are the active power transmitted by line $u$ at time $t$, respectively, including the active and reactive power loss of line $u$. Due to the strong correlation between lines, the voltage drops at adjacent nodes are also considered. Therefore, for nodes near the low-voltage bus, we assume $\Delta U^t u_i \approx 0$.

2. Without consideration of reverse power flow in LVDN, the voltage amplitude of the nodes along the line gradually decreases.

2. Voltage time characteristic analysis

According to (1), the voltage changes of node $u$ at adjacent time $\Delta U^t u_i$ is given by

$$\Delta U^t u_i = U^t_{u-1} - U^t_u = \frac{\sum_{i=1}^{n} \frac{R^t u_i P^t u_i + X^t u_i Q^t u_i}{U^t_{u-1}}}{\sum_{i=1}^{n} \frac{R^t u_i P^t u_i + X^t u_i Q^t u_i}{U^t_{u-1}}}$$

$$u = 1, 2, \ldots, n$$

Where, $U^t u_i$ and $U^t_{u-1}$ are the voltage of node $u$ at time $t$ and $t+1$, respectively, $P^t u_i$ and $Q^t u_i$ are the voltage of low-voltage bus at time $t$ and $t+1$, respectively, $P^t u_i$ and $Q^t u_i$ are the active power transmitted by line $j$ at time $t$, respectively, including the active and reactive power loss of line $j$. According to (3), voltage time characteristics are summarized as follow:

3. $\Delta U^t u_i$ depends on the variation characteristic of the total load and the voltage of low-voltage bus, and a combination of distance of lines between the consumer and the source node i.e., $R^t u_i$, $R^t u_i$, $X^t u_i$, $X^t u_i$, $P^t u_i$, and $Q^t u_i$.

Further, the prior knowledge related to UTCR are refined on the basis of the voltage space and time characteristics. For nodes near the low-voltage bus, the line distance from it to the low-voltage bus is short. Hence, resistance ($R^t u_i$) and reactance ($X^t u_i$) of each line between these nodes and the low-voltage bus are also small. Therefore, for nodes near the low-voltage bus, we assume $\Delta U^t u_i \approx 0$.

Plugging (4) into (3), we obtain $\Delta U^t u_i \approx 0$. Due to the variations in the voltage of low-voltage bus and the total load of different LVDN, the users near the low-voltage bus have the greatest voltage similarity to the bus which they connect to. For example, the voltage curves of M1 and M5 in Fig.1 are the most similar to that of the low-voltage bus of T1 and T2, respectively.

The Person correlation coefficients of voltage profiles (PCCVP) are introduced to describe the similarity among voltage profiles in this study. The more similar the voltage profiles, the greater the PCCVP. The calculation formula is as follows:

$$\rho_{rs} = \frac{\text{cov}(r,s)}{\sigma_r \sigma_s} = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_r \sigma_s}$$

Where, $\rho_{rs}$ is the Pearson correlation coefficient between voltage series; $\text{cov}(r,s)$ is the covariance of the voltage series of node $r$ and $s$; $\sigma_r$ and $\sigma_s$ are the standard deviations of the voltage series of node $r$ and $s$, respectively; $X$ and $Y$ are the voltage series of node $r$ and $s$; $\mu_X$ and $\mu_Y$ are the mean values of $X$ and $Y$, respectively.

Therefore, for users close to distribution transformer, the PCCVP between them and the low-voltage bus can be
compared to determine their UTCR. The prior knowledge related to UTCR is summarized as follows:

**Prior knowledge 1**: For the users near distribution transformer, the PCCVP value between them and the low-voltage bus to which they are connected is the highest among low-voltage buses

However, for nodes far away from the low-voltage bus, affected by the total load variation and long line distance, $\Delta U$ could be significantly different from $\Delta U'$. The voltage profile similarity between the nodes far away from distribution transformer and the low-voltage bus would be low. It’s uncertain which low-voltage bus has the largest PCCVP value with them. In other words, the transformer connectivity of low-voltage bus which has the largest CCVP value to them may be same as them or not, on a case-by-case basis. Hence, it’s hard to determine the UTCR of users far away from the low-voltage bus by comparing PCCVP between them and low-voltage buses.

According to the voltage time characteristics analyzed in above, the voltage changes of node $u$ at adjacent time depends on the variation characteristic of the total load and the voltage of low-voltage bus, and a combination of distance of lines between the consumer and the source node. Due to the differences in the voltage of low-voltage bus and the total load of different LVDN, the voltage correlation between users will show different characteristics when they connect to different LVDN and phase sequence. Let $\Omega_k$ be the PCCVP between user $k$ and other users, as described as below.

$$\Omega_k = \{\rho_{k1}, \rho_{k2}, \ldots, \rho_{kl}, \ldots, \rho_{kK}\}$$  \hspace{1cm} (6)

Where, $\rho_{kl}$ is the PCCVP between user $k$ and user $l$, $K$ is the number of users.

The following situations exist among users in different LVDN:

1. The PCCVP of users in the same phase is strong, while that of users in different phases is weak. Therefore, the PCCVP sequence between users connected to same LVDN will show significant fluctuation, which means the standard deviation of PCCVP sequence is large. Users in other LVDN have weak voltage correlation with users connected to same LVDN. Therefore, the standard deviation of PCCVP sequence between users in other LVDN and users connected to same LVDN is small.

2. The PCCVP between users connected to same LVDN is strong, while that between users in other LVDN and users connected to same LVDN is weak. Therefore, the mean value of PCCVP sequence between users connected to same LVDN is large, while the mean value of PCCVP sequence between users in other LVDN and users connected to same LVDN is small.

The above two characteristics can be used to determine the UTCR of users far from low-voltage bus, and the prior knowledge related to UTCR is summarized as follows:

**Prior knowledge 2**: The standard deviation and mean of PCCVP sequence between users connected to same LVDN are large, while that between users in other LVDN and users connected to same LVDN is small.

The above prior knowledge is derived from the node voltage formula in the power grid. Since the node voltage formula is freely available, there is no cost to provide this prior knowledge.

**IV. KNOWLEDGE-DRIVEN UTCR IDENTIFICATION ALGORITHM**

Knowledges in this section are the prior knowledge related to UTCR which are deduced in Section III. Knowledge factor $\vartheta$ is defined as empirical rules derived from knowledge, and knowledge factor $\vartheta$ can be expressed as:

$$\vartheta : \{\text{if } M \text{ then } N\}$$  \hspace{1cm} (7)

Formula (7) is an empirical rule of causal logic that if $M$ is true, then $N$ is the conditional event, which is the triggering condition of the knowledge factor. In this paper, conditional event can be set as the voltage correlation between multiple users, the convergence degree of node spatial location distribution or the similarity of high frequency components of load current and other indicators exceeding the threshold. $N$ is the conclusion event, representing the empirical judgment under the truth of the conditional event $M$, such as the judgment of topological relations including upstream and downstream relationship, hierarchical relationship and position relationship of nodes. Knowledge-based UTCR identification algorithm is established in this section based on the prior information in Section III. The input data in the proposed method include the voltage curves of consumer and low-voltage buses in LVDNs to be recognized. The output result is user-transformer connectivity.

The flowchart for the proposed knowledge-driven UTCR identification algorithm is presented in Fig.3. As shown in Fig.3, the proposed method consists of two parts. The first part is the data pre-processing. In this part, Z-score and principal component analysis are combined to standardize and extract features from the original voltage data to magnify the differences between data and reduce the impact of data noise. The second part is the recognition for UTCR, in which **Prior knowledge 1 and 2** are employed to verify the UTCR of users near to low-voltage buses and users far away from low-voltage buses.
Figure 3. Flowchart of the knowledge-driven UTCR identification algorithm

A. DATA STANDARDIZATION AND MAIN FEATURE EXTRACTION

In practice, the voltage data collected from LVDN with three-phase unbalanced governance tend to be centralized. And the difference between user’s voltage characteristics is small, which affects the accuracy of the algorithm. Data standardization process, i.e., Z-score standardization is employed to improve the robustness of the proposed method on data discrimination. Moreover, affected by meter measurement errors and communication problems, the data collected by smart meters often contains some noise. Besides, a long period of data is also required to describe the overall law of voltage for all users in the LVDN. In order to reduce the influence of noise and time complexity of the algorithm, dimensionality reduction technique is used to retain the main characteristics of the voltage data. Of which, the principal component analysis (PCA) algorithm is introduced, since it has better performance in topology recognition compared with other dimensionality reduction technique, i.e., T-SNE.

1) Z-score standardization processing

Voltage correlation characteristics are used for UTCR recognition. Hence, in the data re-processing, it is expected to retain the data distribution characteristics in the original data set. In addition, there are differences in voltage fluctuation characteristics between users located in different phases, and the influence of statistical variance needs to be eliminated.

As a feature scaling method, Z-Score standardization transforms the original data into a distribution with a mean value of 0 and a standard deviation of 1. Z-Score standardization does not change the characteristics of data distribution, de-averaging, and standardized variance, which can satisfy the data processing requirements in UTCR recognition.

The smart meters located on low-voltage buses and user side are defined as the metering point. The original voltage data matrix $U$ is constructed based on the voltage data of the low-voltage buses and user of the LVDN to be identified and its neighboring LVDN, $U=[U_{L1}; U_{L2}; \ldots; U_{LC}; U_{C1}; \ldots; U_{Cn}]$, where $U_{L1}$ and $U_{C1}$ respectively represent the voltage matrix of the low-voltage buses and users of the first LVDN, as shown in eq.(8) and eq.(9), $e$ represents the total number of LVDN.

$$U_{L1} = \begin{bmatrix}
    u_{L1A}^1 & u_{L1A}^2 & \cdots & u_{L1A}^T \\
u_{L1B}^1 & u_{L1B}^2 & \cdots & u_{L1B}^T \\
u_{L1C}^1 & u_{L1C}^2 & \cdots & u_{L1C}^T \\
\end{bmatrix}$$

$$U_{C1} = \begin{bmatrix}
    u_{C1}^1 & u_{C1}^2 & \cdots & u_{C1}^T \\
u_{C2}^1 & u_{C2}^2 & \cdots & u_{C2}^T \\
\vdots & \vdots & \ddots & \vdots \\
u_{Cf}^1 & u_{Cf}^2 & \cdots & u_{Cf}^T \\
\end{bmatrix}$$

Where, $u_{L1A}^t$, $u_{L1B}^t$, and $u_{L1C}^t$ represent the voltage values of low-voltage bus of phase A, B and C in the first LVDN at time $t$, respectively; $u_{Cf}^t$ represents the voltage of the $f$-th user at time $T$ in the first LVDN; $f$ is the total user ID number in the DCU of the first LVDN.

The Z-Score standardization calculation formula for voltage data of metering points is shown as below.

$$U_{\Phi}^t = U_{\Phi}^t - \mu(U_{\Phi}^t) \over \sigma(U_{\Phi}^t), u \in \Phi$$

Where, $U_{\Phi}^t$ represents the Z-Score standard value of the $u$-th metering point’s voltage at time $t$, $U_{\Phi}^t$ is a column vector, including the voltage value of all metering points at time $t$, $\mu(U_{\Phi}^t)$ represents the average voltage value of all metering points at time $t$, $\sigma(U_{\Phi}^t)$ represents the voltage standard deviation of all measurement points at time $t$, $\Phi$ is the set of measurement points.

Let $U_{\Phi}^t$ be the data set of measuring points standardized by Z-Score at time $t$. Then, the standardized voltage data set of measuring points $U_{zs}$ can be expressed as below.

$$U_{zs} = [U_{\Phi}^1, U_{\Phi}^2, \ldots, U_{\Phi}^T]$$

The dimensions of the standardized data set $U_{zs}$ are consistent with the dimensions of the original voltage data set $U$.

2) Feature extraction based on PCA

PCA algorithm is an unsupervised dimensionality reduction algorithm based on linear transformation [30]. It
uses orthogonal transformation to transform correlated variables into a group of linearly unrelated variables, so as to obtain the main content that can replace the data and achieve dimensionality reduction and feature extraction by abandoning other minor dimensions.

PCA is widely used to eliminate data redundancy and data noise. The schematic diagram of PCA is shown in Fig. 4. In Fig. 4, \( \mathbf{W} \) is original data matrix of dimension \( m \times l \), \( \mathbf{C} \) is the covariance matrix of \( \mathbf{W} \), \( \mathbf{P} \) is the transformation matrix of dimension \( l \times a \), \( a \) is the number of principal components to be retained, \( \mathbf{Z} \) is the matrix after dimensionality reduction; \( \mathbf{D} \) is the covariance matrix of \( \mathbf{Z} \).

![Schematic diagram of PCA](image)

**FIGURE 4. Schematic diagram of PCA**

Perform PCA dimensionality reduction processing on the voltage standardized data set \( \mathbf{U}_{\text{zt}} \), then obtain a data matrix \( \mathbf{U}_{\text{zm}} \) of dimension \( N \times a \), which retains \( a \)-dimensional main feature, where \( N \) is the total number of measurement points.

**B. UTCR RECOGNITION MODEL BASED ON PRIOR KNOWLEDGE**

There are two problems to be solved in the UTCR identification: 1) which users’ UTCR are error; 2) what the real UTCR of the users with the error are. Prior knowledge 1 and 2 mentioned in Section II contain voltage correlation characteristics of users at different LVDN. Hence, in this section, Prior knowledge 1 and 2 are combined to build the UTCR recognition model to verify the users with the UTCR error and their real UTCR. The details are shown as follows.

Step 1: Calculate the PCCVP between measurement points on the basis of matrix \( \mathbf{U}_{\text{zm}} \), and obtain the PCCVP matrix \( \mathbf{R} \), which can be divided into four block matrices, as shown below.

\[
\mathbf{R} = \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix} \tag{12}
\]

Where, \( R_1 \) is an square matrix with dimension \( N_1 \times N_1 \), which represents the PCCVP between the low-voltage bus of the LVDN to be identified and the low-voltage bus of LVDN adjacent to it, \( N_1 \) represents the total number of low-voltage buses of the LVDN to be identified and the adjacent LVDN; \( R_2 \) is a matrix with dimension \( N_1 \times N_2 \), which represents the PCCVP between the users and the low-voltage buses; \( R_3 \) is a matrix with dimension \( N_2 \times N_1 \), which is the transpose of the matrix \( R_2 \); \( R_4 \) is an square matrix with dimension \( N_2 \times N_2 \), which represents the PCCVP between the users contained in LVDN to be recognized and adjacent LVDN; \( N_2 \) represents the total number of users of the LVDN to be identified and the neighboring LVDN.

Step 2: The column vectors in \( \mathbf{R}_2 \), described as \( \mathbf{R}_2(:,h) \), \( h = 1, 2, \ldots, N_2 \) is the PCCVP value between the user and low-voltage buses of multi-LVDNs. For the user \( h \), the LVDN which the low-voltage bus corresponding to the maximum value in \( \mathbf{R}_2(:,h) \) connect to is used as its initial UTCR.

Step 3: Compare the existing UTCR stored in DCU of LVDNs with the initial UTCR obtained in step 2. For the \( g \)-th LVDN, \( g = 1, 2, \ldots, b \) is the number of LVDN, if the users in it with inconsistent results in the comparison, these users are treated as suspected meter and form suspected user set \( \xi_g \). After this, we obtain a total of \( b \) suspected user sets.

The larger the voltage amplitude of the user is, the closer it is to the low-voltage bus. On this basis, a location index \( \zeta \) is developed to determine the users close to distribution transformer.

Step 4: For each LVDN, perform the following steps.

1. Average users’ voltage value during measurement period by (13)

\[
U_{\text{ave}}^u = \frac{1}{T} \sum_{t=1}^{T} U_{a}^u \tag{13}
\]

Where, \( U_{\text{ave}}^u \) is the average voltage value of consumer \( u \) in the measurement period, \( T \) is the number of intervals in the measurement period.

2. Sort the users by average voltage value obtained in above from the highest to the lowest. The sorting result reflects the sorting of users by electrical distance between users and the low-voltage buses from nearest to farthest.

3. Extract top \( \tau \) users from the sorted result in step 2-1 to form set \( \eta_g \) as the set of consumers near the low-voltage buses in the \( g \)-th LVDN. After this, we obtain a total of \( b \) consumer sets near the low-voltage buses.

5. For each user in set \( \eta_g \), perform the following steps.

6. Divide the users in each LVDN into a set of suspected users and a set of non-suspected users. For example, the set of suspected users and the set of non-suspected users in the \( g \)-th LVDN are represented by \( \xi_{1g} \) and \( \lambda_{1g} \), respectively. For each user in set \( \xi_{1g} \), extract the PCCVP value between it and non-suspected users in the LVDNs from the matrix \( \mathbf{R}_2 \). Each user has a total of \( b \) voltage correlation coefficient series. The \( g \)-th voltage correlation coefficient
A. DATA DESCRIPTION FOR CASE STUDY

In order to verify the effectiveness of the proposed method, two LVDNs model based on real LVNDs in Guangdong are established in MATLAB to simulate the adjacent LVDN scenario. The connectivity of two adjacent LVDN on 10kV line is shown in Fig.5, and the network topology of two LVDN are shown in Fig.6 and Fig.7.

The LVDN1 has 9 low-voltage feeders and serves 170 consumers including 150 single-phase consumers and 20 three-phase consumers. The LVDN2 has 9 low-voltage feeders and serves 315 consumers including 274 single-phase consumers and 41 three-phase consumers. The smart meter in three-phase consumers can record the power consumption, voltage, current of each phase. Hence, a three-phase consumer can be treated as three single-phase consumers. In other words, there are 210 single-phase consumers in the LVDN1 and 397 single-phase consumers in the LVDN2. In two LVDNs, BLV-150 ×4 overhead wire is used in the feeders, BLV-50×2 overhead wire is used in branch lines, and BLV-16×2 overhead wire is used in the service drop line between feeder and household.

V. CASE STUDY

The proposed method was modelled in MATLAB R2019a. Simulations for case study were run on 11th Gen Intel(R) Core (TM) i5-1135G7 @ 2.40GHz with 16.0 GB memory. The case study includes five parts. At first, the data used in case study are described. Then, the identification procedure is given to show how the proposed method identifies the consumer phase in detail. Further, the performance of the proposed method is evaluated. After that, the comparisons between the proposed method and other published methods are carried out. Finally, the influence of the number of principal components defined in Section III.A on the identification accuracy is investigated.
B. IDENTIFICATION PROCEDURE

The 1-day voltage measurement data with 96 measurements of users and low-voltage buses of two LVDNs are taken from the database in this case. The user number in LVDN1 and LVDN2 starts with G and H, respectively. The UTCR of LVDN1 is set to be recognized. Assuming that there are 10 users in the adjacent LVDN2 mixed into the LVDN1’s DCU file, the user names are H3, H5, H10, H12, H20, H32, H55, H56, H70 and H120 connect to LVDN2. The UTCR of two LVDNs is obtained from Step 1 and Step 2 in Section III.B.

In this initial UTCR, 210 users of LVDN1 except H3, H5, H10, H12, H20, H32, H55, H56, H70 and H120 connect to LVDN1, and 387 users of LVDN2 connect to LVDN2. The PCCVP values between the above 10 users and the 6 low-voltage buses in the two LVDNs are given in Tab.1. Where, B1~B3 are low-voltage buses of LVDN1 with phase A, B and C, respectively, B4~B6 are low-voltage buses of LVDN2 with phase A, B and C, respectively. And the value marked in red is the maximum CCVP between user and low-voltage buses.
It can be seen from Tab.1 that the low-voltage buses with the largest PCCVP values of user H3, H5, H10, H12, H20, H32, H55, H56, H70, H120 all belong to LVND2. However, the ID information of these 10 users is in LVDN1. In other words, for these 10 users, the existing UTCR stored in DCU of LVDNs and the initial UTCR obtained by Step 2 in Section III.B are inconsistent, so these 10 users are included in the suspected user set $\hat{C}_1$. Further, perform Step 4 in Section III.B, two consumer sets near the low-voltage buses are obtained, as shown in Tab.2.

It can be seen from Tab.2 that the above 10 suspected users are not in consumer sets near the low-voltage buses. According to Step 5 in Section III.B, the updated suspected user set $\hat{C}_1$ is still equal to the set of suspected users set $\hat{C}_1$. Then, perform Step 6 and Step 7 in Section III.B for these 10 suspected users and the mean value $E_{lg}$ and standard deviation series $F_{lg}$ are obtained, as shown in Tab.3.

It can be seen from Tab.3 that the mean and standard deviation of PCCVP values between the 10 suspected user and non-suspected users in the LVDN2 are higher than those in LVND1. Therefore, these 10 suspected users are
confirmed as users with wrong UTCR, and their real UTCR is connected to LVDN2. The recognition result is consistent with the real situation, and the recognition accuracy is 100%, which fully verifies the effectiveness and accuracy of the proposed method.

### Table 2 USER SETS OF TWO LVDNS NEAR THE LOW-VOLTAGE BUSES

| LVND1 | Users          | LVND2 | Users          |
|-------|----------------|-------|----------------|
| G1, G2, G199, G205, G202, G196, G154, G16, G160, G157, G163, G17, G46, G45, G52, G61, G53, G62, G69, G67, G68, G66, G4, G184, G5, G203, G69, G70, G187, G6, G197, G200, G206, G71, G8, G7, G195, G104, G114, G115, G116, G117, G198 |
|       | H346, H349, H343, H352, H355, H358, H39, H40, H37, H344, H350, H353, H347, H318, H36, H51, H141, H139, H52, H49, H5, H317, H48, H55, H50, H54, H47, H56, H119, H11, H2, H116, H11 | H346, H158, H155, H156, H157, H345, H212, H213, H351, H348, H32, H349, H352, H355, H358, H39, H40, H37, H344, H350, H353, H347, H318, H36, H51, H141, H139, H52, H49, H5, H317, H48, H55, H50, H54, H47, H56, H119, H11, H2, H116, H11 |

### Table 3 THE \( E_{\text{ig}} \) AND \( F_{\text{ig}} \) OF PCCVP VALUES BETWEEN THE SUSPECTED USER AND NON-SUSPECTED USERS IN THE TWO LVND

| User | \( E_{\text{ig}} \) LVND 1 | \( E_{\text{ig}} \) LVND 2 | \( F_{\text{ig}} \) LVND 1 | \( F_{\text{ig}} \) LVND 2 |
|------|-----------------|-----------------|-----------------|-----------------|
| H3   | 0.067           | 0.224           | -0.156          | 0.071           | 0.245           | -0.174          |
| H5   | 0.091           | 0.299           | -0.207          | 0.071           | 0.264           | -0.192          |
| H10  | 0.113           | 0.328           | -0.215          | 0.069           | 0.216           | -0.148          |
| H12  | 0.103           | 0.334           | -0.231          | 0.056           | 0.217           | -0.161          |
| H20  | 0.133           | 0.258           | -0.125          | 0.073           | 0.245           | -0.172          |
| H32  | 0.092           | 0.202           | -0.110          | 0.067           | 0.223           | -0.156          |
| H55  | 0.094           | 0.236           | -0.142          | 0.061           | 0.217           | -0.156          |
| H56  | 0.098           | 0.238           | -0.141          | 0.064           | 0.216           | -0.152          |
| H70  | 0.084           | 0.268           | -0.184          | 0.049           | 0.188           | -0.139          |
| H12  | 0.087           | 0.280           | -0.193          | 0.052           | 0.184           | -0.131          |

### C. PERFORMANCE EVALUATION OF THE PROPOSED METHOD

In this section, the performance of the proposed method under the conditions of different UTCR error rate, data measurement error rate, three-phase imbalance level and data length are investigated. In detail, by increasing the number of users in LVDN2 into the DCU reading file of LVDN1, the data scenario with increasing UTCR error rate is constructed. UTCR error rate \( \varepsilon \) is defined as the ratio of the number of users not belonging to LVDN1 to the total number of LVDN1 users, as below.

\[
\varepsilon = \frac{N_{\text{false}}}{N_{\text{LVDN1}}}
\]

Where, \( N_{\text{false}} \) is the number of users not belonging to LVDN1, \( N_{\text{LVDN1}} \) is the total number of LVDN1 users.

To construct the data scenario with an increasing data measurement error rate \( \eta \), every user’s measurement has been added noise by introducing a Gaussian error whose mean value is 0 and standard deviation is one third of the measurement error rate \( \eta \). \( \varepsilon \) is set as the average value of three-phase voltage unbalance in the measurement period to reflect the three-phase unbalance level of LVND, as described below.

\[
\varepsilon = \left[ \sum_{t=1}^{T} \left( \frac{U_{\text{max}}^D(t) - U_{\text{min}}^D(t)}{U_{\text{max}}^D(t) \times 100%} \right) \right] / T
\]

Where, \( U_{\text{max}}^D(t) \) is the maximum voltage in low-voltage buses at time \( t \), \( U_{\text{min}}^D(t) \) is the minimum voltage in low-voltage buses at time \( t \), \( T \) is the number of intervals in a measurement period.

Firstly, a comprehensive calculation is executed by gradually increasing the value of \( \varepsilon \) and \( \eta \) with fixed data length (15 days with 1440 measurements). Under each data scenario, UTCR identification is executed multiple times to obtain average accuracy. The results are shown as below.
different from the LVDN them are connected to, on a case-by-case basis.

Then, a comprehensive calculation is executed by gradually increasing the value of $\varepsilon$ and $\Lambda$ with fixed data length (15 days with 1440 measurements) and fixed measurement error rate of 0.4%. Under each data scenario, UTCR identification is executed multiple times to obtain average accuracy. The results are shown as below.

As shown in Fig.9, with fixed $\Lambda$ value, the identification accuracy increases gradually as $\Lambda$ increases. The reason is that the proposed method depends on the PCCVP value among users and that between consumers and low-voltage buses of LVDN. As three-phase imbalance level ($\Lambda$) increases, the voltage discrimination of users connected to different phases increases so that the identification accuracy of the proposed method is gradually improved. In particular, in the scenario where the $\varepsilon$ is 0.5, the recognition accuracy rate is increased by 8% when the $\Lambda$ is 0.24% compared with when the $\Lambda$ is 0.12%. Therefore, in order to alleviate the influence of measurement error, the data in the period with large three-phase imbalance level can be selected to carry out UTCR identification.

Further, the influence of data length is discussed. A comprehensive calculation is executed by gradually increasing the value of $\eta$ and length of data with fixed UTCR error rate of 0.3. Under each data scenario, UTCR is executed multiple times to obtain average accuracy. The results are shown in Fig.10.

It can be seen from Fig.10 that when there is no data measurement error ($\eta=0$), the recognition accuracy rate for each data length is 100%. When there is a data measurement error ($\eta=0.4\%$~2%), the recognition accuracy rate increases as the data length increases, and the growth rate gradually slows down. This is because the difference in the PCCVP value between users increases with the increase of the data length, thereby improving the recognition accuracy. In particular, when $\eta=0.4\%$, the recognition accuracy in the data scenario of 6-8 days can be increased to 100%. Therefore, in order to alleviate the influence of the measurement error, the data with a longer time length can be selected for UTCR recognition.

**FIGURE 9. UTCR identification accuracy under different $\varepsilon$ and $\Lambda$**

As shown in Fig.9, with fixed $\Lambda$ value, the identification accuracy increases gradually as $\Lambda$ increases. The reason is that the proposed method depends on the PCCVP value among users and that between consumers and low-voltage buses of LVDN. As three-phase imbalance level ($\Lambda$) increases, the voltage discrimination of users connected to different phases increases so that the identification accuracy of the proposed method is gradually improved. In particular, in the scenario where the $\varepsilon$ is 0.5, the recognition accuracy rate is increased by 8% when the $\Lambda$ is 0.24% compared with when the $\Lambda$ is 0.12%. Therefore, in order to alleviate the influence of measurement error, the data in the period with large three-phase imbalance level can be selected to carry out UTCR identification.

Further, the influence of data length is discussed. A comprehensive calculation is executed by gradually increasing the value of $\eta$ and length of data with fixed UTCR error rate of 0.3. Under each data scenario, UTCR is executed multiple times to obtain average accuracy. The results are shown in Fig.10.

It can be seen from Fig.10 that when there is no data measurement error ($\eta=0$), the recognition accuracy rate for each data length is 100%. When there is a data measurement error ($\eta=0.4\%$~2%), the recognition accuracy rate increases as the data length increases, and the growth rate gradually slows down. This is because the difference in the PCCVP value between users increases with the increase of the data length, thereby improving the recognition accuracy. In particular, when $\eta=0.4\%$, the recognition accuracy in the data scenario of 6-8 days can be increased to 100%. Therefore, in order to alleviate the influence of the measurement error, the data with a longer time length can be selected for UTCR recognition.

**FIGURE 10. UTCR identification accuracy under different data length and $\eta$**

**D. COMPARISON ANALYSIS WITH OTHER PUBLISHED METHODS**

In this section, multiple data scenarios are constructed by gradually increasing the value of $\varepsilon$ and $\eta$ with fixed data length of one day to compare the performance of different methods. At present, there are two methods to identify UTCR based on AMI measurement data. One is to compare voltage curve correlation between users [23], and the other is to compare voltage curve correlation between users and low-voltage buses [24]. The comparative analysis of the following four identification methods is executed.

- **Method 1**: comparing voltage curve correlation between users [23];
- **Method 2**: comparing voltage curve correlation between users and low-voltage buses [24];
- **Method 3**: removing Z-score standardization processing and PCA feature extraction from the proposed method;
- **Method 4**: the proposed method.

The UTCR recognition accuracy rates of the four methods under different $\varepsilon$ and $\eta$ are shown in below.
As illustrated in Fig. 11, several findings are:

1) Comparing Method 1 and Method 4, when the data measurement error $\eta=0$, the recognition accuracy of the two methods is both 100%. But when there is a measurement error i.e., $\eta=0.8\%$, $\eta=1.6\%$, the recognition accuracy of Method 4 is significantly higher than that of Method 1 in every scenario. This indicates that it is not enough to accurately identify UTCR by comparing the voltage curve correlation alone when there is measurement error in data. On the basis of Method 1, Method 4 taking the correlation between users as supplementary verification has better performance.

2) Comparing Method 2 and Method 4, when the data measurement error $\eta=0$ and $\epsilon$ is less than 0.2, the recognition accuracy of the two methods is both 100%. However, when $\epsilon$ is greater than 0.2, the recognition accuracy of Method 2 decreased obviously with the increase of $\eta$, while that of Method 4 remained at 100%. The reason is that Method 2 identify UTCR by comparing the voltage curve correlation between users. When there are many users in LVDN that do not belong to it, it is easy for Method 2 to identify these users wrongly. The more users who do not belong to LVDN in the initial meter reading file of LVDN, the higher the error recognition rate of Method 2. When there is data measurement error in LVDN, there is no linear relationship between recognition accuracy and UTCR error rate in Method 2. This is because the superposition of measurement errors changes the PCCVP value between different voltage curves. Specifically, the similarity of voltage curves between users located in the same LVDN is reduced. In this case, the results of Method 2 are stochastic. Except for the scenario where $\eta=0.8\%$ and $\epsilon=0.3$, the accuracy of Method 4 is higher than that of Method 2 in every scenario. This demonstrates that the recognition accuracy of Method 4 is more stable and superior.

3) Comparing Method 3 and Method 4, when the data measurement error $\eta=0$, the recognition accuracy of the two methods is both 100%. But when there are measurement errors i.e., $\eta=0.8\%$, $\eta=1.6\%$, the recognition accuracy of Method 4 is higher than that of Method 3 in all scenarios. This fully illustrate that the Z-score and PCA dimensionality reduction links enhance the robustness of the recognition method against data measurement errors.

Further, to show the role of data processing in the proposed method clearly, the comparation analysis about the identification results with and without data standardization and with different dimensionality reduction techniques are carried out. Define the proposed method without data standardization as Method 4_1, the proposed method with T-SNE dimensionality reduction as Method 4_2, the proposed method without PCA dimensionality reduction as Method 4_3. Multiple data scenarios are constructed by gradually increasing the value of data measurement error rate $\eta$ with UTCR error rate of 0.3 to compare the identification results, as shown in Fig. 12.

As shown in Fig.12, the recognition accuracy of Method 4 with Z-score standardization and PCA dimensionality reduction is higher than that of other 3 Methods in all scenarios. Comparing Method 4_1 and Method 4_3 with Method 4, it is clearly that Z-score standardization and PCA dimensionality reduction are beneficial to improve the robustness of the recognition method against data measurement errors, respectively. And PCA dimensionality reduction perform better than T-SNE dimensionality reduction method comparing Method 4_2 and Method 4.

### E. Sensitivity Analysis for the Threshold Coefficient

As shown in Fig. 11, the recognition performance of 4 methods in different scenarios is illustrated. The recognition accuracy of Method 4 is significantly higher than that of Method 1 in every scenario. This indicates that it is not enough to accurately identify UTCR by comparing the voltage curve correlation alone when there is measurement error in data. On the basis of Method 1, Method 4 taking the correlation between users as supplementary verification has better performance.

As shown in Fig. 12, the recognition accuracy of Method 4 with Z-score standardization and PCA dimensionality reduction is higher than that of other 3 Methods in all scenarios. Comparing Method 4_1 and Method 4_3 with Method 4, it is clearly that Z-score standardization and PCA dimensionality reduction are beneficial to improve the robustness of the recognition method against data measurement errors, respectively. And PCA dimensionality reduction perform better than T-SNE dimensionality reduction method comparing Method 4_2 and Method 4.
In this section, a comprehensive calculation is executed by gradually increasing the value of $\eta$ and PCA main feature numbers $a$ with fixed UTCR error rate of 0.3 and data length of one day. The results are shown as below.

![Fig. 13: Recognition accuracy under different $\eta$ and PCA main feature values](image)

**FIGURE 13.** Recognition accuracy under different $\eta$ and PCA main feature values

It can be seen from Fig.13 that the number of PCA main features have different effects on the recognition accuracy with different $\eta$. When $\eta=0$, the increase in the number of PCA main features can increase the recognition accuracy to 100%. When $0<\eta\leq0.4\%$, the recognition accuracy first increases and then decreases with the increase of the number of PCA main features. When $\eta\geq0.8\%$, the recognition accuracy first increases with the increase of the number of PCA main features and then tends to be flat. This demonstrates that the number of PCA main features for optimal recognition accuracy is affected by the measurement error rate, and the higher the measurement error rate, the greater the number of PCA main features is needed. The above results verify the effectiveness of the PCA dimensionality reduction method on retaining a few main features to replace high-dimensional data.

### VI. CONCLUSION

To identify UTCR intelligently and strengthen the algorithm's robustness to data discrimination and noise, in this paper, statistical data processing methods and the prior knowledge of voltage correlation characteristics in LVDN are combined to develop a knowledge-driven UTCR identification method. The performance of the proposed method is evaluated under various conditions and compared with other published methods. Further, the influence of PCA main feature numbers is investigated. From the study, the conclusions are elaborated as follows.

1) The proposed method can effectively distinguish the users with wrong UTCR and identify their correct UTCR. The data processing process with Z-score and PCA feature extraction can enhance the robustness of the proposed method to the data measurement error.

2) The recognition accuracy of the proposed method decreases as UTCR error rate and measurement error rate increase. However, it can be improved by selecting data with high three-phase voltage imbalance level or long length.

3) In the scenario where there is measurement error in data, the proposed method outperforms the method in reference [24] that only compares the voltage curve correlation between users and low-voltage buses and the method in reference [23] that only compares the voltage curve correlation between users.

4) The number of PCA main features to achieve the best recognition accuracy is affected by the measurement error rate, and the higher the measurement error rate, the greater the number of PCA main features is needed.

The probability to reverse power flow on LVDN will be increased as the penetration of renewable micro-generation such as photovoltaic increases. Thus, how to recognize UTCR with the reverse power flow will be investigated in the future.

### REFERENCES

[1] X. Chen and B. Lin, “Towards carbon neutrality by implementing carbon emissions trading scheme: Policy evaluation in China,” *Energy Policy*, vol. 157, p. 112510, Oct. 2021.

[2] X. Wu, Z. Tian, and J. Guo, “A review of the theoretical research and practical progress of carbon neutrality,” *Sustainable Operations and Computers*, vol. 3, p. 54-66, 2022.

[3] L. Zhou, Y. Zhang, X. Lin, et al, “Optimal Sizing of PV and BESS for a Smart Household Considering Different Price Mechanisms,” IEEE Access, vol. 6, pp. 41050-41059, 2018.

[4] L. Dong, C. Wang, M. Li, and et al, “User Decision-based Analysis of Urban Electric Vehicle Loads,” *CSEE Journal of Power and Energy Systems*, vol. 7, no. 1, pp. 190-200, Jan. 2021.

[5] C. Byers and A. Botterud, “Additional Capacity Value From Synergy of Variable Renewable Energy and Energy Storage,” *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 1106-1109, April 2020.

[6] F. M. Camilo, R. Castro, M. E. Almeida, and et al, “Probabilistic load elasticity analysis in low voltage distribution networks with high penetration of photovoltaic micro generation,” *International Journal of Electrical Power & Energy Systems*, vol. 113, pp. 782–791, Dec. 2019.

[7] Y. H, M. Wang, Y. Jia, and et al. “Low-voltage ride-through control for photovoltaic generation in the low-voltage distribution network.”, *IET Renewable Power Generation*, vol.14, pp.2727-2737, 2020.

[8] B. Liu, K. Meng, Z. Y. Dong, and et al, “Unbalance Mitigation via Phase-Switching Device and Static Var Compensator in Low-Voltage Distribution Network,” *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4856-4869, Nov. 2020.

[9] S. J. Pappu, N. Bhatt, R. Pasumarthy and et al, “Identifying Topology of Low Voltage Distribution Networks Based on Smart Meter Data,” *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5113-5122, Sept. 2018.

[10] M. Lave, M. J. Reno and J. Peppanen, “Distribution System Parameter and Topology Estimation Applied to Resolve Low-Voltage Circuits
on Three Real Distribution Feeders,” *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1585-1592, July 2019.

[11] P. Li, Y. Yan, M. Xu, and et al, “The Research of Users-transformer Relationship Verification Method Based on Data-driven,” *2019 IEEE Innovative Smart Grid Technologies - Asia (ISGT Asia)*, May 2019.

[12] M. Lisowski, R. Masnicki and J. Mindykowki, “PLC-Enabled Low Voltage Distribution Network Topology Monitoring,” *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6436-6448, Nov. 2019.

[13] P. V. Aubel and E. Poll, “Smart metering in the Netherlands: What, how, and why,” *International Journal of Electrical Power & Energy Systems*, vol. 109, pp. 719–725, Jul. 2019.

[14] J. B. Leite and J. R. S. Mantovani, “Detecting and Locating Non-Technical Losses in Modern Distribution Networks,” *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1023-1032, March 2018.

[15] Y. Wang, Q. Chen, T. Hong and C. Kang, “Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges,” *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125-3148, May 2019.

[16] K. Soumalas, G. Messinis and N. Hatziargyriou, “A data driven approach to distribution network topology identification,” *2017 IEEE Manchester PowerTech*, pp. 1-6, 2017.

[17] M. Lave, M. J. Reno and J. Peppanen, “Distribution System Parameter and Topology Estimation Applied to Resolve Low-Voltage Circuits on Three Real Distribution Feeders,” *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1585-1592, July 2019.

[18] L. Zhou, Y. Zhang, S. Liu, and et al, “Consumer phase identification in low-voltage distribution network considering vacant users,” *International Journal of Electrical Power & Energy Systems*, vol. 121, pp. 106079, October 2020.

[19] L. Zhou, Qin. Li, Y. Zhang, and et al, “Consumer phase identification under incomplete data condition with dimensional calibration,” *International Journal of Electrical Power & Energy Systems*, vol. 129, pp. 106851, July 2020.

[20] Y. Hu, Q. Cao, L. Wang, and et al, “Low Voltage Transformer Topology Identification Method Based on De-noised Differential Evolution,” *2020 10th International Conference on Power and Energy Systems (ICPES)*, pp. 356-360, 2020.

[21] W. Hu, Y. Liu, Q. Guo, and et al, “A Data-driven Method of Users-transformer Relationship Identification in the Secondary Power Distribution System,” *2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 585-590, 2020.

[22] J. D. Watson, J. Welch, and N. R. Watson, “Use of smart meter data to determine distribution system topology,” *The Journal of Engineering*, vol. 2016, no. 5, pp. 94-101, May 2016.

[23] W. Luan, J. Peng, M. Maras, and et al, “Smart Meter Data Analytics for Distribution Network Connectivity Verification,” *IEEE Transactions on Smart Grid*, vol. 6, no. 4, pp. 1964-1971, July 2015.

[24] Y. Xiao, Y. Zhao, Z. Tu, and et al, “Topology checking method for low voltage distribution network based on improved Pearson correlation coefficient,” *Power System Protection and Control*, vol. 47, pp. 37-43, 2019.

[25] A. Guzman, A. Arguello, J. Quiros-Tortos, and et al, “Processing and Correction of Secondary System Models in Geographic Information Systems,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3482–3491, Jun. 2019.

[26] T. A. Short, “Advanced Metering for Phase Identification, Transformer Identification, and Secondary Modeling,” *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 651–658, Jun. 2013.

[27] W. Hu, Y. Liu, Q. Guo, and et al, “Detection of users-transformer relationship in the secondary power distribution system with smart meter data,” *2020 15th IEEE Conference on Industrial Electronics and Applications (ICCEA)*, Nov. 2020.

[28] L. Blakely and M. J. Reno, “Identification and Correction of Errors in Pairing AMI Meters and Transformers,” *2021 IEEE Power and Energy Conference at Illinois (PECI)*, Apr. 2021.

[29] V. C. Cunha, W. Freitas, F. C. L. Trindade, and et al, “Automated Determination of Topology and Line Parameters in Low Voltage Systems Using Smart Meters Measurements,” *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5028–5038, Nov. 2020.

[30] A. Rehman, A. Khan, M. A. Ali, and et al, “Performance Analysis of PCA, Sparse PCA, Kernel PCA and Incremental PCA Algorithms for Heart Failure Prediction,” *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*, pp. 1-5, 2020.

[31] D. R. R. Penido, L. R. de Araujo, S. Carneiro, and et al, “Three-Phase Power Flow Based on Four-Conductor Current Injection Method for Unbalanced Distribution Networks,” *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 494–503, May 2008.