DOI 10.24425/aee.2024.150891

# Research on modeling method for digitizing distribution substation areas based on big data from smart meters

Hao Bai[1]✉, Yu Fu[2], Weichen Yang[1], Yue li[2], Yongxiang cai[2], Min Xu[1]

[1] *Electric Power Research Institute, China South Power Grid, Ltd.*
*Guangzhou, China*

[2] *Guizhou Power Grid Co. Ltd.*
*Guiyang, China*

*e-mail:* ✉ *2965326985@qq.com*

(Received: 24.11.2023, revised: 23.08.2024)

**Abstract:** With the wide application of smart meters, real-time data collection in power grid operation can be carried out in real time, which provides big data support for the digital modeling of distribution station areas. At the same time, the digital transformation of the distribution area can be promoted through the digital modeling of the distribution area. In this situation, how to use the big data of smart meters to achieve the digital modeling of the distribution area needs to be further studied. Firstly, this paper briefly introduces the connotation of digitalization in the distribution station area. At the same time, aiming at the digitalization of the distribution station area, it analyzes the digitalization of station features, user features, new energy features, network parameters and operation features, and obtains the digitalization model. Finally, on the basis of the above, select an application scenario to introduce the application of digitalization modeling. The research results can provide a reference for the combination of big data application of smart meters and digitalization of the distribution area, and help the digital transformation of the distribution area.

**Key words:** digital model, low voltage distribution network, numerous measurement data, smart meters

## 1. Introduction

With the implementation of goals “carbon neutral” and “carbon peak”, a distribution network characterized by low-carbon, digital and distributed has being developed, and in order to speed up the green transformation of energy, large-scale distributed power is connected to the grid [1, 2]. The previous power grid management cannot cope with the challenges brought by distributed energy access, and it is urgent to explore new management methods and establish a new power system that can meet the new trend of energy development [3, 4]. Therefore, it is necessary to build a perfect new power system with the support of digital means to meet the development of new business forms of energy and power [5]. Smart electricity meters play an important role in building new power systems. Smart electricity meters based on the Internet of Things read the user's electricity quantity in real time from residential household appliances, and the generated data usually shows the large volume, time-varying and fast speed characteristics of big data [6, 7]. Smart meters can read a large number of different formats of electricity data in a short time, and these data can provide powerful data support for the construction of a new digital power system if it is effectively applied [8–10].

In fact, digital technology is not a new proposition in the field of power distribution, but a comprehensive and cross-cutting technology field that continues to develop with the evolution of power distribution technology forms. Its connotation and extension are constantly expanding, and it has new characteristics and demands under the current situation of building a new power system with new energy as the main body [11, 14].Related to the digital research of the power distribution platform area, Reference [15] constructs the evaluation index system of "obtaining power" based on the demand side under the background of digital transformation. Reference [16] with digital-driven management value-added, strives to achieve a win-win situation for customers, power supply enterprises and society. At the same time, based on the data integration of the distribution network automation system, References [17, 18] explore a kind of digital intelligent distribution and transformation platform area suitable for the development of new power systems. At the same time, there are some problems in the current distribution platform area, such as insufficient data collection breadth and depth, a low intelligence degree of low-voltage equipment, and a limited coverage range of equipment communication capacity. Reference [19] puts forward four development strategies for power supply reliability under the background of digital transformation. Reference [20] studies the key technology and application of digitalization in the distribution platform area, and puts forward the panoramic digital monitoring scheme. In contrast, Reference [21] briefly summarizes the digital model of a 0.4 kV grid in asymmetric mode. At the same time, there are few studies on the application of smart electricity meter big data. Among them, Reference [22] puts forward the application idea of big data technology in the quality research of electricity meters. References [23, 24] show that it is possible to detect the abnormal power consumption using the big data method.

To sum up, the relevant literature only considers one aspect of smart meter big data application or distribution area digitalization, and does not quantify the digital characteristics of each index in a unified standard. The analysis of the connotation of distribution area digitalization is not clear, and systematic application research on how to use smart meter big data to realize the digital modeling of the distribution area is lacking. In order to solve the above problems, this paper briefly introduces the digital connotation of the distribution station area and systematically and comprehensively proposes a digital modeling method of the distribution station area based on the big data of smart meters according to the characteristics of the distribution network. The five index calculation methods of platform-type feature digitalization, user feature digitalization, new energy feature digitalization, network parameter digitalization, and operation feature digitalization were refined. Finally, a platform area is selected for modeling application to verify the accuracy of the index calculation.

## 2. Digital connotation of power distribution platform area

With the vigorous implementation of new infrastructure and new energy, elements such as electric vehicle charging stations, distributed energy sources, energy storage facilities, 5G base stations, and data centers in the vicinity of the distribution station area show a rapid growth trend. These new elements need to be incorporated into the distribution station area through power electronic conversion devices, and a large number of such devices have been introduced to significantly increase the degree of power electrification in the distribution station area. Distributed power electronic resources represented by distributed photovoltaic systems, energy storage units, data centers, electric vehicles, frequency conversion air conditioning units, etc., are abundant in the distribution area. The distribution area is gradually transitioning from a single "load" nature to a coexistence nature of "source, network, load, and storage", with some loads exhibiting flexible development. This presents a challenge to the efficient utilization of controllable resources of source network load and storage in the distribution station area. Additionally, the distributed power electronic resources represented by electric vehicles and new energy sources have significant uncertainty characteristics, which will increase the uncertainty and complexity of the operating status of the station area. Therefore, it is necessary to leverage digital technology to open up information on all aspects of the source network, load, and storage, promote the upgrade of key infrastructure, and realize the coordination and mutual benefit of various energy sources such as electricity, heat, cold, and gas. To achieve "comprehensiveness and observability, accuracy, and high controllability", effectively aggregate massive adjustable resources to support real-time dynamic response, optimize the operation scheduling technology of the distribution area, and improve the power and energy balance capability of the distribution area and its flexible operation level.

At the same time, the new power system oriented towards the dual carbon goal further integrates the core motivation of achieving the "carbon neutral goal" and the key approach of "digital transformation" based on the key elements of "source, network, load, and storage". Therefore, when considering low carbon emissions, distributed new energy permeability, energy storage, and flexible load proportions under the conditions of new load characteristics, the distribution area focuses on digitalization across five aspects: digital, user characteristics, new energy digitalization, network parameters digitalization, and operational characteristics digitalization. Each of these aspects involves different types of characteristics corresponding to digital indices, as shown in Table 1.

Table 1. Digital indicators corresponding to different types of features

| **Types of characteristics** | **Digital indicators** |
|---|---|
| Platform area features digitization | Power supply area type, transformer characteristics, line type, power supply radius of platform area, grid structure, power supply type and load type |
| Digitalization of the user features | load curve |
| Digitalization of new energy features | Photovoltaic power generation (timing, randomness) |
| Digitalization of network parameters | Line impedance value |
| Run features are digitized | Line load rate, line voltage loss, comprehensive line loss rate, comprehensive voltage qualification rate, harmonic qualification rate, asymmetry factor, voltage fluctuations and rapid voltage changes |

# 3. Digital modeling method of power distribution platform area

## 3.1. Digital model of platform area features

The characteristics of the distribution area will be described from seven aspects: power supply area type, transformer characteristics, line type, power supply radius, grid structure, power supply type and load type.

1. Power supply area characteristics: urban areas and rural areas;
2. Transformer characteristics: model and rated capacity;
3. Line characteristics: cable type (line length accounts for > 90%), overhead type (line length accounts for > 90%), mixed type (line not conforming to the cable type and overhead type) [25];
4. Power supply radius characteristics: short radius (< 250 m), medium radius (250–500 m), medium and long radius (500–800 m), long radius: (> 800 m)[26];
5. Structural characteristics of the network frame: radial network, tree network, hand in hand (chain) network;
6. Characteristics of new power supply: penetration rate of distributed new energy (high, medium and low);
7. Load type: proportion of traditional load types (industrial load, commercial load, agricultural load and residential load, etc.), and proportion of new load types (electric vehicles, energy storage, etc.).

Based on the characteristics of the distribution area, the distribution area is classified. At the same time, the wire type of each branch can be obtained according to the existing topology diagram, and the resistance and reactance of each branch can be derived to realize the digitization of the network structure of the distribution area.

## 3.2. Digital model of user load characteristics

### 3.2.1. Load characteristic index

At present, the general load characteristic description index can be divided into three types: description, comparison, and curve. The curve index can describe the periodic changes in load, including the daily load curve (showing the load change curve by hour throughout the day), monthly load curve (showing the daily maximum load curve), annual load curve (showing the monthly maximum load curve for the year), etc. The peak load duration in the curve class indicator represents the load value at the highest system load during a given period, and the annual continuous load curve indicates the numerical size of the system load and the duration in hours over the year.

The daily load curve describes the time sequence of the load in a day and is suitable for the digitization of the user's load characteristics. However, the monthly load curve and annual load curve can only describe the timing characteristics of the maximum load, and the description of load periodic timing is too rough and not suitable for the digitization of user load characteristics. There is no unified description index for the timing characteristics of the load within a week or a quarter. Additionally, considering the new forms of the new power system, such as distributed photovoltaic permeability, energy storage, and flexible load ratios, it is essential to incorporate these factors into load curve analysis. This ensures that the obtained load curve accurately reflects the actual site situation.

#### 3.2.2. Classification method of daily load curve based on fuzzy clustering algorithm

The sampling frequency of the load curve should be compatible with the feasible operation frequency of the control equipment. For instance, the hourly daily load curve is suitable for the optimal control of the distribution network that can be operated once per hour. Conversely, when operations can occur every 5 minutes, the load curve with a 5-minute interval should be utilized. If only several regulation operations can be carried out daily, it's advisable to use the phased daily load curve, which reflects the general law of intraday load fluctuation.

Different standardized transformation methods are employed when various methods are used to measure the range of data variation. Commonly used methods include:

1. Standard deviation standardization

$$S_j = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_{ij} - \overline{X}_j)^2}, \tag{1}$$

where: $S_j$ represents the standard deviation of the data of the $j$−th variable $n$; $X_{ij}$ represents the observed value of the $i$ $(i = 1, 2, \dots, n)$−th $j$ variables and $\overline{X}_j$ represents the average of the observed value of the $j$ variables.

2. Extreme difference standardization

The extreme difference of the $j$−th variable is $R_j = \max\left(X_{ij}\right) - \min\left(X_{ij}\right)$ $(1 \le i \le n)$, and the $n$-th data of the $j$-th variable is:

$$x_{ij} = \frac{X_{ij} - \overline{X}_j}{R_j}, \quad i = 1, 2, \dots, n, \tag{2}$$

where $X_{ij}$ is the value range normalization of the $n$ data for the $j$−th variable. After the transformation, the mean value of each variable was 0, and the extreme difference was 1.

3. Very poor regularization

$$x_{ij1} = \frac{X_{ij} - (X_{ij})}{R_j}, \quad i = 1, 2, \dots, n, \tag{3}$$

where $x_{ij1}$ is the extreme normalization of the $n$ data for the $j$-th variable. The minimum value of each variable was 0 and the extreme difference was 1.

The normalization or standardization of the load curve varies, and there are different practices without specific literature specifying which method is superior. Typically, the common standard treatment used in power system analysis can be applied, wherein the load data at each moment of the load curve is divided by the peak value of the load over a period of time (day, week, month, year, etc.). This method is feasible and effective.

Digitizing user load characteristics enables the extraction of typical load curves of users, providing crucial support for more accurate and effective mining of users' power load patterns and a more comprehensive analysis of users' power load characteristics. This, in turn, contributes to the design of precise and diversified demand response strategies.

### 3.3. New power supply feature digital model

Distributed new power supply mainly includes photovoltaic, wind energy, biomass, fuel cells, batteries and other types. For biomass power generation, fuel cell power generation, battery output and other distributed new power supplies, output characteristics are not affected by the time cycle.

There is no obvious timing and periodicity that can be based on the actual load demand to develop a reasonable power generation plan and flexible regulation. For the new intermittent distributed power supply such as photovoltaic, its output characteristics are affected by the time period, and have the sequence and randomness. The following takes the characteristics of photovoltaic power generation as an example to establish a digital model 27].

1. Time sequence

   Since the output power of the photovoltaic system is directly proportional to the solar irradiance, the total output power of the photovoltaic system $P$ is:

   $$P = rNS\eta, \tag{4}$$

   where: $r$ is the radiation intensity of the sun during this period; $N$ is the number of photovoltaic arrays; $S$ is the area of each array; $\eta$ is photoelectric conversion efficiency.

   The output of photovoltaic power generation equipment is affected by local light intensity, and seasonal and weather changes will affect the magnitude of light intensity. Because there is no light at night, photovoltaic output starts as early as 7:00 and ends no later than 19:00 throughout the year. Photovoltaic power generation is directly related to solar radiation. The duration of sunshine is longest in summer, relatively similar in spring and autumn, and shortest in winter. Therefore, power generation duration is longest and highest in summer, while it is shortest and lowest in winter.

2. Random

   Many current studies have shown that the solar irradiance over a certain time period is approximately a continuous *Beta* distribution between [0, 1]:

   $$f\left(\frac{r}{r_{\max}}\right) = \frac{1}{B(\alpha,\beta)}\left(\frac{r}{r_{\max}}\right)^{\alpha-1}\left(1-\frac{r}{r_{\max}}\right)^{\beta-1}, \tag{5}$$

   where: $r_{\max}$ is the maximum radiation intensity of the sun during this period; $\alpha, \beta$ represent the *Beta* shape parameters of the distribution, respectively, and the relevant parameters of the distribution can be calculated from the light intensity $\mu$ and variance $\sigma$ in the *Beta* statistical period:

   $$\alpha = \mu \cdot [\frac{\mu \cdot (1-\mu)}{\sigma^2} - 1], \tag{6}$$

   $$\beta = (1-\mu) \cdot [\frac{\mu \cdot (1-\mu)}{\sigma^2} - 1]. \tag{7}$$

   The *Beta* mathematical expectation of the corresponding distribution is $\frac{\alpha}{(\alpha+\beta)}$, the variance is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$, $B\,(\alpha,\beta)$ is the regularized factor, and the value is:

   $$B(\alpha\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{8}$$

   The probability density function of the output power of the solar cell array can be obtained from the probability density function of the light intensity:

   $$f\,(P) = -\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \left(\frac{P_m}{P_{\max}}\right)^{\alpha-1} \cdot \left(1-\frac{P_m}{P_{\max}}\right)^{\beta-1}, \tag{9}$$

   where $P_{\max} = r_{\max}NS\eta$ is the maximum output power of the square array.

Wind and other new energy power generation exhibit strong randomness, volatility, and intermittency. Under current technical conditions, insufficient power support capacity may result from its volatility, and the tolerance of new energy power generation to extreme weather is relatively fragile. Moreover, due to the low energy density of new energy and the low annual utilization hours of power generation, coupled with the fact that large new energy bases are typically located far from the load center, the relationship between reliable and safe power supply and the economy of the system can be further enhanced through digitalization. By combining the analysis of electricity price trends, the power side can reduce costs and increase efficiency, while the user side can save energy and improve efficiency.

### 3.4. Digital model of network parameters

Many on-site electrical equipment operating outdoors will be affected by the external environment, accelerating the aging of the equipment and resulting in changes in the impedance value of each section of the line (impedance increase). This poses a threat to the operational safety of the power grid. The installation and use of a large number of smart meters provide power companies with a wealth of measured data with high frequency, wider coverage, and longer time scales, enabling the calculation of line impedance in low-voltage power supply networks. Based on the line impedance parameters, line loss analysis, power flow calculation, and electric theft detection can be carried out. Additionally, changes in impedance parameters (impedance increase) can help identify old circuits affected by corrosion and aging, providing a reference for the timely replacement of line equipment with security risks.

The calculation steps are as follows: based on the voltage, active power, and reactive power data collected by the electricity meter, and considering the previous network topology, a mathematical model for estimating line impedance is established according to the line voltage drop equation. Subsequently, the line impedance parameters are estimated in stages.

1. The voltage drop equation of the line is:

$$\dot{U}_1 - \dot{U}_2 = \frac{P_b R + Q_b X}{U_b} + j\frac{P_b X - Q_b R}{U_b}, \tag{10}$$

where: the subscript letter $b \in \{1, 2\}$ in $U_b$, $\{1, 2\}$, 1 and 2 represent the electrical volume at the beginning and end of the line; $P_b$is the active power of the line$b$; $Q_b$is the reactive power of the line $b$; $U_b$ is the voltage of the line $b$; $R$ is the resistance value of the line and $X$ is the reactance value of the line. Considering the characteristic that both $X/R$and the reactive power of the low-voltage power supply line is relatively small, the small voltage angle difference between the voltage is ignored. The voltage drop in the adjacent node circuit is approximately:

$$\dot{U}_1 - \dot{U}_2 \approx \frac{P_b R + Q_b X}{U_b} = I_R R + I_X X, \tag{11}$$

where $I_R$ represents the current value flowing through the resistance and $I_X$ represents the current value flowing through the reactance.

2. Equation of linear regression
   The headend voltage of the circuit is led as a multivariate linear regression equation:

$$\dot{U}_1 - \dot{U}_2 = I_R R + I_X X + \varepsilon, \tag{12}$$

   where $\varepsilon$ represents the error at both ends of Eq. (11). It can be seen from Eq. (12) that the line parameters $X$ and $R$ can be regarded as the multiple linear regression equation regression coefficient. The data of smart meters at different times reflect the state of the low-voltage power supply network at different times. By substituting the sampling values of multiple moments into Eq. (12), the overdetermined set $Z$ for $X$ and the overdetermined set $Y$ for $R$ is formed, respectively. Calculate the least squares estimate of the regression coefficient according to the following equation, (13), that is, the impedance of the line. If there is no distributed generation, then $U_1 > U_2$. The calculated line parameters are correct, otherwise the introduced value is negative.

$$\hat{\beta} = (Z^T Z)^{-1} Z^T Y, \tag{13}$$

   where $\hat{\beta}$ is the regression coefficient.

### 3.5. Run the feature digital model

From the perspective of system security and stability, a line-load rate, line voltage loss, comprehensive line-loss rate, comprehensive voltage-qualification rate, harmonic-qualification rate, asymmetry factor, voltage fluctuations and rapid voltage changes are selected as the indicators to measure the digitalization of the distribution network.

1. Line load rate
   The line load ratio in a system that meets the safety constraints of thermal stability should not exceed 1, which is calculated by:

$$\lambda_{l1} = \frac{I_l}{I_{l,\max}}, \tag{14}$$

   where $I_l$ is the line $l$ current and $I_{l,\max}$ is the upper limit of the line $l$.
2. Line voltage loss
   The line voltage drop is the absolute difference in voltage values between the nodes at both ends of the line, which is a scalar and also one of the important indicators for evaluating the efficiency of power system operation and the stability of the power grid. It can reflect the quality of the power grid, indicate the status of the line, affect the distribution of power load, and guide grid optimization. Therefore, using the line voltage drop as one of the digital models for the operational characteristics of low-voltage distribution substations is essential, and the calculation method is:

$$\lambda_{l2} = |U_i - U_j|, \tag{15}$$

   where $U_i$ and $U_j$ are the voltage values of the node $i$ and $j$, respectively. When the system has a large reactive power transmission, the line voltage loss is relatively large, which can easily induce voltage instability accidents. At the same time, after digitizing the operational characteristics of the distribution platform area, data such as peak and valley differences can be further calculated to promote the digital transformation of the distribution platform area.

3. Comprehensive line loss rate
   The comprehensive line loss rate indicates whether the power grid operation state is economic, and its calculation formula is:
$$X_1 = \frac{P_1 - P_2}{P_1}, \tag{16}$$
   where $P_1$ is the power supply load in the platform area of the low-voltage distribution network and $P_2$ is the total electricity consumption of users in the platform area of the low-voltage distribution network.
4. Qualified rate of comprehensive voltage
   The comprehensive voltage pass rate represents the percentage of voltage reaching the standard during the running time of the station area, and its calculation formula is:
$$K = \left(1 - \frac{t}{T}\right) \times 100\%, \tag{17}$$
   where $t$ is the voltage limit time of the monitoring point and $T$ is the total operating time of the monitoring point.
5. Harmonic pass rate
   There are two numerical measurement methods for evaluating voltage/current distortion, the THD (Total Harmonic Distortion ) factor and the TTHD (True Total Harmonic Distortion) factor, but only the THD factor is considered in the paper. THD is a measure of the harmonic components contained in a current or voltage signal. THD is typically calculated through the following steps:
   (a) Data collection: Initially, voltage or current waveform sampling data needs to be collected, usually by continuously sampling voltage or current over a certain period to obtain waveform data.
   (b) Waveform decomposition: The collected voltage or current waveform is decomposed into fundamental and harmonic components. The fundamental is the basic frequency of the signal, usually 50 Hz or 60 Hz, while harmonics are multiples of the fundamental frequency.
   (c) Calculation of harmonic amplitudes: For each harmonic component, its amplitude (magnitude) is calculated.
   (d) Total harmonic amplitude: The squared amplitudes of each harmonic component are summed, and the square root of the sum is taken to obtain the total harmonic amplitude.
   (e) THD calculation: The total harmonic amplitude is divided by the fundamental amplitude, then multiplied by 100 to obtain the percentage value of THD.
   The calculation formula is as follows:
$$\mathrm{THD}_I\ (\%) = \frac{\sqrt{\sum_{i=2}^{n_{\lim}} I_i}}{I_1} \times 100\%, \tag{18}$$
$$\mathrm{THD}_U\ (\%) = \frac{\sqrt{\sum_{i=2}^{n_{\lim}} U_i}}{U_1} \times 100\%, \tag{19}$$

where: $n_{\mathrm{lim}}$ is the borderline order of the considered harmonics, in most standards $n_{\mathrm{lim}} = 50$; $I_i$ and $U_i$ are the harmonic of current and voltage of the $n_{\mathrm{lim}}$-th order, respectively; $I_1$ and $U_1$ are the fundamental harmonic of current and voltage, respectively.

6. Asymmetry factor

The asymmetry factor can be calculated by computing the difference in magnitude of each phase voltage or current. Assuming the voltages of a three-phase system are *Va*, *Vb*, and *Vc*, the asymmetry factor can be computed as:

$$A = \frac{3\left(\max\{Va, Vb, Vc\} - \min\{Va, Vb, Vc\}\right)}{(Va + Vb + Vc)} \times 100\%, \tag{20}$$

where $\max\{Va, Vb, Vc\}$ and $\min\{Va, Vb, Vc\}$ are the maximum and minimum values among the three-phase voltages.

For further polymerization analysis, it is necessary to calculate the average value of the asymmetry coefficient over a period of time, which is calculated as follows:

$$\overline{A} = \frac{1}{N}\sum_{i=1}^{N} A_i, \tag{21}$$

where $N$ indicates the number of calculated data points and $A_i$ represents the value of the asymmetry coefficient of the $i$-th data point.

7. Voltage fluctuations

Pst (short term flicker severity) and Plt (long term flicker severity) are two indicators used to measure voltage fluctuations, and their calculation formula is as follows:

(a) The formula for calculating Pst (short term flicker severity) is:

$$\mathrm{Pst} = \sqrt{0.0314P_{0.1} + 0.0525P_{1s} + 0.0657P_{3s} + 0.28P_{10s} + 0.08P_{50s}}, \tag{22}$$

where the percentiles $P_{0.1}$, $P_1$, $P_3$, $P_{10}$ and $P_{50}$ are the flicker levels exceeded for 0.1; 1; 3; 10 and 50% of the time during the observation period. The suffix "$s$" in the formula indicates that smoothed values should be used; they are obtained using the following equations:

$$P_{50s} = \left(P_{30} + P_{50} + P_{80}\right)/3, \tag{23}$$

$$P_{10s} = \left(P_6 + P_8 + P_{10} + P_{13} + P_{17}\right)/5, \tag{24}$$

$$P_{3s} = \left(P_{2.2} + P_3 + P_4\right)/3, \tag{25}$$

$$P_{1s} = \left(P_{0.7} + P_1 + P_{1.5}\right)/3. \tag{26}$$

The 0.3 s memory time-constant in the flickermeter ensures that $P_{0.1}$ cannot change abruptly and no smoothing is needed for this percentile.

(b) The formula for calculating Plt (long term flicker severity) is:

$$\mathrm{Plt} = \left(\frac{1}{T}\int_0^T P_{\mathrm{inst}}^3 \,\mathrm{d}t\right)^{\frac{1}{3}}. \tag{27}$$

These indicators can be used to assess the severity of voltage fluctuations to determine the voltage quality of a power system.

8. Rapid voltage changes
   The rapid change of voltage can be described by the amplitude of the voltage change, the voltage derivative during the change and the number of changes that occurred in a given period. The corresponding calculation formulas for different indicators are as follows:
   (a) Amplitude of the voltage change: describes how much the voltage changes per unit time.

$$\Delta V = |V_{\max} - V_{\min}|, \tag{28}$$

   where $V_{\max}$ and $V_{\min}$ represent the maximum and minimum voltage values, respectively, during the observation period.
   (b) Voltage derivative during the change: describes the magnitude of the voltage change rate.

$$\frac{\mathrm{d}V}{\mathrm{d}t} = \frac{\Delta V}{\Delta t}, \tag{29}$$

   where $\Delta t$ represents the time the change has taken.
   (c) Number of changes that occurred in a given period: describes how often the voltage changes over a given period of time.

$$N_c = \frac{N}{t}, \tag{30}$$

   where $N$ represents the total number of voltage changes that occur during the observation period and $t$ represents the length of the observation period.

## 4. Digital modeling applications

Take a new energy distribution network as an example, the digital modeling method is verified.

1. Platform area features digitization
   Taking a distribution network without new energy as an example, the digital results of its platform area features digitization are shown in Table 2.

Table 2. Digital results of platform features

| **Power supply area** | **Transformer** | **Circuit** | **Power supply radius** | **Grid structure** | **New energy characteristics** | **Load type** |
|---|---|---|---|---|---|---|
| rural area | 100 kVA | Overhead type | Medium and long radius | radial | Low permeability | Traditional load |

2. Digitalization of user load characteristics
   Taking four typical loads in a new energy distribution network as an example, the daily load curve in spring is drawn. In this case, the daily load curve is divided into five sections by the fuzzy clustering method, and the time cut-off points are marked in Fig. 1.

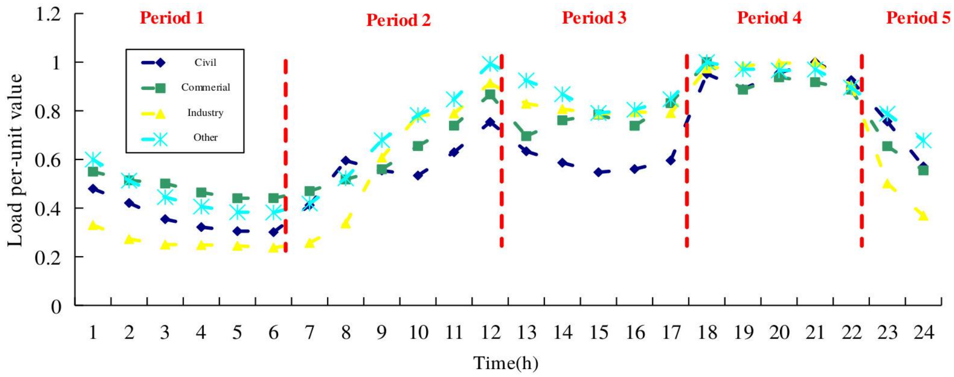


Fig. 1. The segmented results of daily load curves of 4 types of typical loads in spring

After fuzzy clustering of the daily load curve of multiple nodes of the distribution network, the unified daily load curve segmentation mode is determined as 1–6, 7–12, 13–17, 18–22, and 23–24 points. The equivalent load of each time period was obtained using four value modes: cluster center, mean, median, and maximum, and extended at the equivalent load level in each time period to obtain the equivalent load curve as shown in Fig. 2.

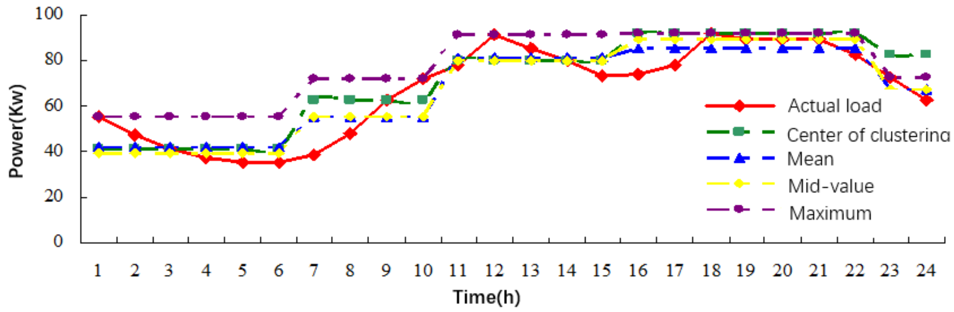


Fig. 2. Equivalent load curve and actual load curve

Using actual load data as the object to be identified, fuzzy pattern recognition is performed based on proximity calculation to determine the value pattern of equivalent loads. Proximity is a measure of the degree of closeness between two fuzzy sets, with higher proximity indicating closer proximity to the actual data. The load equivalent pattern and the results of fuzzy recognition proximity calculation are shown in Table 3.

As seen in Table 3, among the four modes, the representative load curve generated as the representative load value is the closest to the actual load curve, and the fitting effect is better than the other three modes, indicating that the median load in the time period can be used as the equivalent load for this period. Therefore, the 24-point daily load curve is transformed into a 5-point daily load curve with the time period as the horizontal coordinate, which can reflect the general pattern of intraday load fluctuation.

Table 3. Load equivalent mode and its fuzzy recognition closeness

| **Pattern and object** | **Time period 1** | **Time period 2** | **Time period 3** | **Time period 4** | **Time period 5** | **Closeness to actual load** | |
|---|---|---|---|---|---|---|---|
| time quantum | 1:00~ 6:00 | 7:00~ 10:00 | 11:00~ 15:00 | 16:00~ 22:00 | 23:00~ 24:00 | Hamming close to degree | Riemann proximity |
| Cluster center | 41.1 | 62.7 | 80.1 | 92.3 | 82.6 | 91.4% | 89.2% |
| average value | 41.9 | 55.4 | 81.6 | 85.1 | 67.7 | 92.6% | 90.4% |
| mid-value | 39.3 | 55.4 | 80.1 | 89.1 | 67.6 | 92.9% | 90.8% |
| crest value | 55.3 | 72.2 | 91.6 | 92.3 | 72.7 | 88.3% | 86.2% |

3. Digitalization of the new power supply features

   The daily output timing and seasonal characteristic curve of photovoltaic power generation equipment are shown in Fig. 3. Because there is no light at night, photovoltaic output starts at 7:00 at the earliest and ends at 19:00 at the latest.

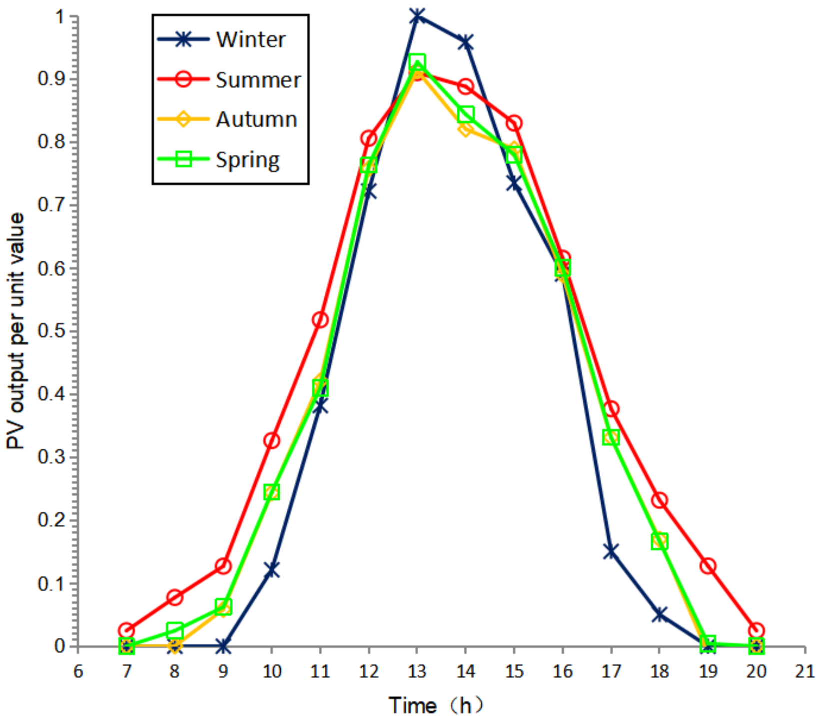


Fig. 3. Seasonal characteristic curve of sunrise force time series of photovoltaic power generation equipment

Photovoltaic power generation is directly related to the amount of solar radiation. Summer has the longest sunshine duration, while spring and autumn have similar sunshine durations. Winter has the shortest sunshine duration. Therefore, in the figure, summer power generation time is the longest with the highest power generation, while winter power generation time is the shortest with the least power generation. Another factor influencing photovoltaic power generation is the temperature of photovoltaic panels. There exists a negative temperature coefficient relationship between the power generating capacity and the temperature of photovoltaic panels. For every 10 degrees Celsius increase in temperature, the power generation capacity of photovoltaic panels decreases by 1%. In summer, the temperature of photovoltaic panels in the area may reach 50–60 degrees Celsius at noon, which is higher than in winter. Due to the higher photovoltaic panel temperature, the midday photovoltaic power generation in winter may be higher than that in summer.
The daily light intensity characteristic curve, which can be divided into three conditions of rainy, cloudy, and sunny days according to meteorological conditions, is shown in Fig. 4.

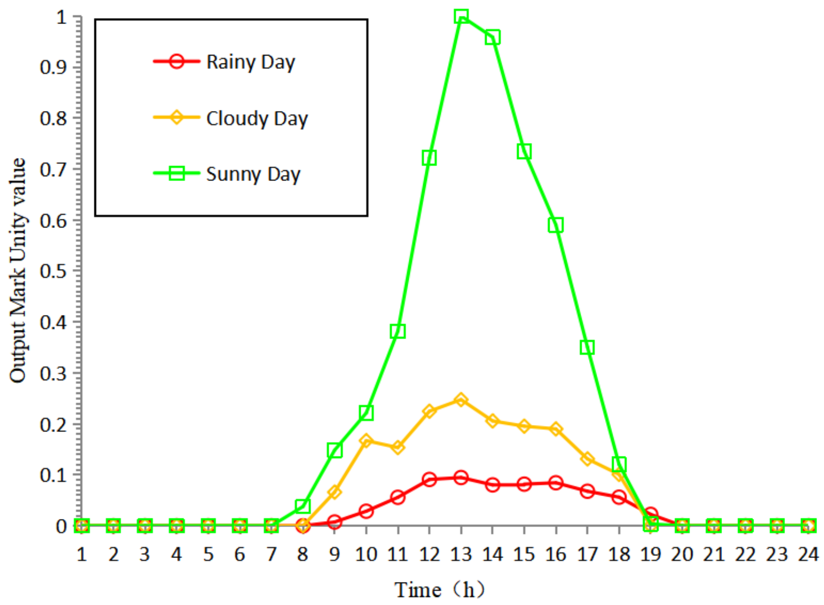


Fig. 4. Meteorological characteristic curve of sunrise force time series of photovoltaic power generation equipment

It can be seen from the figure that the output of photovoltaic power generation equipment is maximum on sunny days, while the output is very small or even non-existent during most of the time on rainy or cloudy days.

4. Digitizing of the network parameters
   The results of the network parameters from the topology are shown in Table 4.
5. Digitizing of operation features
   The results of the digital operational characteristics of the station area obtained through simulation are presented in Table 5.

Table 4. Results of digitization of network parameters

| By-pass | First end node | Endpoint node | Resistance (Ω) | Reactance (Ω) | Distance (m) |
|---|---|---|---|---|---|
| 1 | 1 | 17 | 0.0059 | 0.002242 | 59 |
| 2 | 17 | 2 | 0.0004 | 0.000152 | 4 |
| 3 | 17 | 18 | 0.0221 | 0.008398 | 221 |
| 4 | 18 | 3 | 0.0012 | 0.000456 | 12 |
| 5 | 18 | 19 | 0.0026 | 0.000988 | 26 |
| 6 | 19 | 4 | 0.0003 | 0.000114 | 3 |
| 7 | 19 | 5 | 0.006 | 0.00228 | 60 |
| 8 | 1 | 20 | 0.006 | 0.00228 | 60 |
| 9 | 20 | 6 | 0.0003 | 0.000114 | 3 |
| 10 | 20 | 21 | 0.006 | 0.00228 | 60 |
| 11 | 21 | 7 | 0.0137 | 0.005206 | 137 |
| 12 | 21 | 8 | 0.0056 | 0.002128 | 56 |
| 13 | 21 | 22 | 0.005 | 0.0019 | 50 |
| 14 | 22 | 9 | 0.001 | 0.00038 | 10 |
| 15 | 22 | 23 | 0.0058 | 0.002204 | 58 |
| 16 | 23 | 10 | 0.0021 | 0.000798 | 21 |
| 17 | 23 | 11 | 0.0024 | 0.000912 | 24 |
| 18 | 23 | 24 | 0.0033 | 0.001254 | 33 |
| 19 | 24 | 12 | 0.0001 | 0.000038 | 1 |
| 20 | 24 | 13 | 0.0033 | 0.001254 | 33 |
| 21 | 24 | 25 | 0.0095 | 0.00361 | 95 |
| 22 | 25 | 14 | 0.0026 | 0.000988 | 26 |
| 23 | 25 | 26 | 0.0074 | 0.002812 | 74 |
| 24 | 26 | 15 | 0.004 | 0.00152 | 40 |
| 25 | 26 | 16 | 0.0088 | 0.003344 | 88 |

Table 5. Digitized results of platform operation characteristics

| Load factor | Voltage loss | Comprehensive line loss rate | Voltage pass rate | Harmonic pass rate | Asymmetry factor | Voltage fluctua-tions | Rapid voltage changes |
|---|---|---|---|---|---|---|---|
| 85.6% | 1.032% | 0.4% | 99.89% | 99.95% | 0.23% | Pass | Pass |

## 5. Conclusions

In this paper, the digital connotation and five aspects of digital modeling of the distribution station area are analyzed. A calculation example is selected to verify the modeling method, and the following conclusions are drawn:

1. This paper proposes a digital modeling method for the distribution station area. By processing massive and complex household smart meter data, it innovatively sorts out five characteristic indicators that can systematically and comprehensively express the characteristics of the distribution station area;
2. In view of the current situation where there is no unified quantification standard for the digital characteristics of each index, the proposed five index calculation methods are refined. These methods include platform-type feature digitalization, user feature digitalization, new energy feature digitalization, network parameter digitalization, and operation feature digitalization. The accuracy of index calculation is verified with examples;
3. Based on a large volume of electrical data generated by smart meters in various formats, and considering the limitations of existing technologies, it is challenging to provide complete and accurate distribution network data and operational data. Thus, it is necessary to generate typical load curves to support the digitalization of distribution station areas. Therefore, this paper considers the typical output curves of the source load at different time scales to provide data support for the load classification of the station area;
4. Based on the analysis of big data from smart meters, the digital model of network parameters is extended. A line configuration calculation method is proposed that relies solely on the measurement data from users' smart meters to backward calculate the grid structure and parameters of the station area. This method provides support for power flow calculations in areas where network configuration is difficult to obtain or where account data is missing;
5. Based on the digital characteristics of the substation, it is possible to conduct an initial analysis of the resource endowment and output characteristics of each substation. Comparing various characteristic indicators can help identify safety hazards within the system, facilitating substation operators in optimizing the scheduling of the substation and effectively controlling power quality issues. This further supports the digital transformation of distribution substations.

**Acknowledgements**

This work was financially supported by the Science and Technology Support Program of Guizhou Province ([2022] General 012) and China Southern Power Grid Co., Ltd. (GZKJXM20220043).

**References**

[1] Caixia W., Zhiyong S., Zhifeng L., Qinmiao L., Bowen H., Bibin Huang., Liping J., *Key technologies and prospects of demand-side resource utilization in power system based on new energy*, Automation of Electric power Systems, vol. 45, no. 16, pp. 37–48 (2021), DOI: 10.7500/AEPS20210323004.

[2] Xiandong T., Jun L., Zhicheng X., Li Y., Guoqiang J., Baoguo S., *Electricity supply and demand situation in the "14th Five-Year Plan" under the "Double carbon" goal*, China Electric Power (in Chinese), vol. 54, no. 5, pp. 1–6 (2021), DOI: 10.11930/j.issn.1004-9649.202103047.

[3] Jun L., Wanxing S., Riliang L., Peng W., Guangxian L., Jiaying L., Zhiming Z., *Design and application of Distribution Internet of Things*, High Voltage Technology (in Chinese), vol. 45, no. 6, pp. 1681−1688 (2019), DOI: 10.13336/j.1003-6520.hve.20190604001.

[4] Kai Z., Haowen Key S., *Technologies and practices of 10kV distribution iot smart station*, Electrical Technology, vol. 21, no. 7, pp. 80–84 (2020), DOI: 10.3969/j.issn.1673-3800.2020.07.016.

[5] Yongjun Z., Yingqi Y., Liwu L., Kangye Z., Qinhao L., Siliang L., Huangqing H., *Technology Prospect of new low-voltage Distribution system driven by dual-carbon target*, Automation of Electric Power Systems, vol. 46, no. 22, pp. 1–12 (2022), DOI: 10.7500/AEPS20210922006.

[6] Lingchao G., Liming Y., Zhiwei Y., Fei Z., *Research on Hybrid Index Method of Double-LayerB+Tree for Power Big Data Considering Knowledge Graph*, Journal of Physics: Conference Series, vol. 1771, no. 1, 012004 (2021), DOI: 10.1088/1742-6596/1771/1/012004.

[7] Xiaohua L., Juncheng G., Shaolin W., Hairong Z., *Smart grid abnormal power consumption detection framework*, Automation Technology and Application, vol. 39, no. 12, pp. 124−128 (2019), DOI: 10.3969/j.issn.1003-7241.2020.12.028.

[8] Xiaodong X., Qianyun L., Tao L., Yu Z., Qiyu W., *Station area recognition Based on smart meter data and fuzzy C-means algorithm*, Journal of Nanjing Institute of Technology (Natural Science Edition), vol. 18, no. 4, pp. 1–7 (2019), DOI: 10.13960/j.issn.1672-2558.2020.04.001.

[9] Yi S., Yehua M., Zekun L., Xudong Z., Fei L., *Comprehensive clustering method of user load characteristics and adjustable potential for power big data*, Proceedings of the CSEE (in Chinese), vol. 41, no. 18, pp. 6259–6271 (2021), DOI: 10.13334/j.0258-8013.pcsee.201928.

[10] Zhiyong Y., Zekun X., Li Y., Quan X., Yuehuan L., Peiqiang L., Xi H., *Review of smart grid big data research*, Guangdong Electric Power, vol. 34, no. 1, pp. 1–12 (2021), DOI: 10.3969/j.issn.1007-290X.2021.001.001.

[11] Chen S., Yunni C., Mengshuo J., Ying C., Shaowei Huang., *Concept, characteristics and application prospect of digital twin of power system*, Proceedings of the CSEE (in Chinese), vol. 42, no. 2, pp. 487–498 (2022), DOI: 10.13334/j.0258-8013.pcsee.211594.

[12] Tianjiao P., Sheng C., Qi Z., Xinying W., Dongxia Z., *Framework design and application prospect of digital twin system of energy Internet*, Proceedings of the CSEE, vol. 41, no. 6, pp. 2012−2028 (2021), DOI: 10.13334/j.0258-8013.pcsee.201757.

[13] Chengshan W., Bo D., Hao Y., Jianzhong W., Jinpeng Y., Peng L., *Digital twin technology and application of smart city integrated energy system*, Proceedings of the CSEE, vol. 41, no. 5, pp. 1597–1607 (2021), DOI: 10.13334/j.0258-8013.pcsee.201804.

[14] Zhelin L., Chengshan W., Peng L., Hao Y., Li Y., *Estimation and application of distribution network state based on multi-source measurement data fusion*, Proceedings of the CSEE (in Chinese), vol. 41, no. 8, pp. 2605–2615 (2021), DOI: 10.13334/j.0258-8013.pcsee.201416.

[15] Dan L., Geng Y., Guohong Z., Guangzhi D., Rui M., *Construction of evaluation index system of "access to electricity" based on demand side in the context of digital transformation*, Power supply, vol. 40, no. 2, pp. 60–67 (2023), DOI: 10.19421/j.cnki.1006-6357.2023.02.009.

[16] Liyun C., Hanqi Y., *Digital transformation and upgrading management of distribution network operation and maintenance for power grid enterprises*, Enterprise Management, iss. S1, pp. 158–159 (2022).

[17] Qilu S., *Application research of digital Intelligent operation and Maintenance Technology of distribution network*, Internet of Things Technology, vol. 11, no. 11, pp. 93–95 (2021), DOI: 10.16667/j.issn.2095-1302.2021.11.027.

[18] Lei Q., Gang A., Jie Y., Jianping H., Wei D., Jie T., Bing L., *Digital intelligent power distribution area construction research*, Rural Electrification, vol. 3, pp. 27–31 (2023), DOI: 10.13882/j.cnki.ncdqh.2023.03.005.

[19] Rui L., Mingyang L., Yanzhang G., *Development strategy of power supply reliability of China Southern Power Grid under Digital Transformation*, Power Supply and Electricity, vol. 38, no. 3, pp. 38–44 (2021), DOI: 10.19421/j.cnki.1006-6357.2021.03.006.

[20] Hongquan L., Hao D., Cong W., Helong S., Yanchao C., Aichun G., Dingli L., *Research and Application of Digital Technology Scheme in Distribution Station Area*, Electrotechnical Engineering, vol. 22, pp. 143–146 (2021), DOI: 10.19768/j.cnki.dgjs.2021.22.048.

[21] Vasiliev N.V., Kartashev D.A., Yu Krishtopa N., *Analysis of operating modes of three-phase distribution networks using their digital models of 0.4kV*, IOP Conference Series: Earth and Environmental Science, vol. 979, no. 1, 012110 (2022), DOI: 10.1088/1755-1315/979/1/012110.

[22] Guang C., Shaobing C., Zekun L., Xiaoyi C., *Research on big data analysis technology for smart meter quality in modern supply chain*, Automation and Instrumentation, vol. 5, no. 16, pp. 180–183+188 (2021), DOI: 10.14016/j.cnki.1001-9227.2021.05.180.

[23] Huaying L., *Big data analysis and abnormal detection of smart electricity meters*, Master Thesis, Changchun University of Technology (2020).

[24] Lei P., Yan Y., Hao L., Longquan F., *Abnormal power consumption detection based on smart meter big data*, Computing Technology and Automation, vol. 39, no. 2, pp. 177–183 (2020), DOI: 10.16339/j.cnki.jsjsyzdh.202002035.

[25] Limin J., Guixiong H., *GB/T 31367-2015, Guidelines for Energy Efficiency Evaluation of medium and low voltage distribution Network*, China Electricity Council (2015).

[26] Jian D., *Research on low-voltage distribution network planning in Chongshanchang Village, Anji County*, Master Thesis, North China Electric Power University (Beijing) (2016).

[27] Andruszkiewicz J., Lorenc J., Weychan A., *Distributed generation as efficient measure to improve power generation adequacy*, Archives of Electrical Engineering, vol. 68, no. 2, pp. 373–385 (2019), DOI: 10.24425/aee.2019.128274.