WILEY | Hindawi

Research Article

An Extensible Gradient-Based Optimization Method for Parameter Identification in Power Distribution Network

Chuanjun Wang [ID],[1] **Kehao Fei** [ID],[1] **Xinle Xu** [ID],[2] **Haoran Chen,**[3] **Ke Hu** [ID],[4] **Shihe Xu** [ID],[5] **and Jiayang Ma** [ID][1]

[1]*Nanjing Institute of Technology, Nanjing 211167, Jiangsu, China*
[2]*University of California, Davis 1 Shields Ave, Davis, CA 95616, USA*
[3]*School of Information and Communication, National University of Defense Technology, Wuhan 430019, Hubei, China*
[4]*Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
[5]*University of Science and Technology of China, Hefei 230027, Anhui, China*

Correspondence should be addressed to Ke Hu; huke@cqupt.edu.cn

Accurate parameter identification of power distribution network (PDN) has attracted remarkable attention recently. However, power device parameters usually show an instability attributed to both the operating status and manual entry. Therefore, it is urgent to develop reliable algorithms for identifying PDN parameters with both high accuracy and high efficiency. Most of the existing algorithms are gradient-free and based on the heuristic schemes, resulting in an unstable numerical calculation. Herein, based on our previous work about the adaptive gradient-based optimization (AGBO) method, we propose an extensive version, namely, AGBO-Pro model. In this method, both the numerical and categorical features of experimental observations are utilized and incorporated with each via a weighted average. By comparing the proposed method with several heuristic algorithms, it is found that the errors in RMSE, MAE, and MAPE criteria via AGBO-Pro are all about 2 times lower with a much faster and more stable convergence of the loss function. By further taking a linear transformation of the loss function, the AGBO-Pro model achieves a more robust performance with a much lower variance in repeat numerical calculations. This work shows great potential in possible extension of gradient-based optimization methods for parameter identification in PDN.

1. Introduction

Acquiring accurate and reliable device parameters is crucial in the context of power distribution networks (PDNs) due to their multifaceted implications [1]. However, the lack of in situ measurement techniques poses challenges in directly obtaining certain PDN parameters, which are typically assumed to be static in real situations. These parameters include line resistance, line reactance, transformer resistance, transformer reactance, transformer conductance, and transformer electrical susceptance. This limitation often leads to poor estimation in parameter identification for PDN [2]. To address these challenges, numerous approaches have been developed to enhance numerical efficiency and reduce

residuals in parameter estimation. These approaches include supervisory control and data acquisition, power management unit (PMU), and advanced metering infrastructure. They can be classified into various methods, such as the full-scale approach [3], PSOSR [4], normalized Lagrange multiplier (NLM) test [5], finite-time algorithm (FTA) [6], residual method, sensitivity analysis method, Lagrange multiplier method [7], Heffron-Phillips method [8], and specialized Newton–Raphson iteration [9]. Additionally, recent advancements in machine learning and deep learning techniques have led to the proposal of smart methods, including artificial neural network [10], graph convolution network (GCN) [11], support vector machine (SVM) [12], multihead attention network [13], deep reinforcement

learning [14], estimation using synchrophasor data [15], PSCAD simulation [16], multimodal long short-term memory deep learning [17], and edge computing [18]. While these methods show effectiveness with simulation data, they often require specialized measuring devices.

To overcome the challenges posed by the lack of required data and measuring devices, a mathematical approach called the power flow model can be employed. The power flow model establishes relationships between PDN parameters and easily obtainable data, such as active power, reactive power, and voltage. The challenging parameters mentioned earlier can be optimized using algorithms in combination with the static parameters, namely, active power, reactive power, and voltage, on the low-voltage side. By utilizing the power flow model, the voltage values on the high-voltage side can be calculated, and the residuals between the calculated and true values can be used to construct a loss function for optimization methods. The optimization methods for parameter identification can be generally classified into two categories, namely, the gradient-free methods [19–25] and gradient-based methods [26, 27]. First, in gradient-free methods, the heuristic or biomimetic optimization rules are designed such as particle swarm optimization (PSO), genetic algorithm (GA), ant colony algorithm, Aquila optimizer (AO) [28], nuclear reaction optimization (NRO) [29], and Pareto-like sequential sampling heuristic method (PSS) [30]. The performance of these heuristic algorithms is largely dependent on the initialization. On the other hand, the gradient-based method is usually constructed by combining the physical model of PDN and the neural network with the backward propagation of the loss function. Beneficial from the chain rule and automatic derivative of the loss function with respect to the parameters to be optimized, the gradient-based method is a deterministic approach. Therefore, many advanced gradient descent algorithms can be utilized to accelerate the optimization. In addition, researchers pay more attention to data preprocessing methods, including the utilization of clustering algorithms and hypothesis testing [31, 32].

In previous work, we proposed an adaptive gradient-based optimization (AGBO) algorithm for parameter identification in PDN. In AGBO, a physical model based on the power flow calculation is incorporated into the neural network, and the input data are the numerical features obtained from experimental measurement. However, besides the numerical features, the experimental observations of PDN also contain many categorical features such as recording duration and peak-valley electricity. It should be noted that such categorical features are usually hard to be utilized directly in heuristic algorithms, but the neural network-based methods have a great advantage on feature embedding and extraction. Therefore, in this work, based on the analysis of the physical model in PDN, we propose an extensive gradient-based optimization method for parameter identification in PDN, and we call this new model AGBO-Pro. The proposed model can utilize not only the numerical features as before but also the categorical information, which is rarely used in previous works. Based on the abovementioned points, this paper mainly has the following contributions:

(1) An extensible gradient-based optimization method is proposed, which is constructed with customized neural network layer and loss function, and it achieves a higher and more robust performance in parameter identification problems of PDN.

(2) In the physical-informed multihead neural network, we separate the experimental measurements into the numerical features and categorical features. After several manipulations, the categorical features are transformed to a weight distribution and incorporated into the numerical features via a linear transformation. Such a treatment is rarely studied or neglected by other investigations.

(3) The performance in three evaluation functions of this hybrid method during the numerical calculation is much better than that of several individual optimization algorithms.

This paper is organized as follows. Section 2 introduces the identification equations of power flow model in PDN and proposes the AGBO-Pro optimization method. The experimental data and calculation details are given in Section 3. The results and discussions are given in Section 4. Finally, Section 5 gives a brief conclusion.

2. Materials and Methods

2.1. Power Flow Model Calculation. The fundamental principles of PDN analysis can be found in references [20, 31–33]. To streamline the computational process, the assumption of balanced three-phase condition is made as a prerequisite for power flow calculations in this work. The schematic diagram of power flow calculation circuit model is shown in Figure 1.

In Figure 1, P_d, Q_d, and U_d represent the active power, reactive power, and voltage on the high-voltage side of the transformers at bus D, respectively. These three parameters can be obtained directly by real-time measurements. Other parameters, such as transformer electrical R_d, transformer resistance X_d, transformer conductance G_d, transformer electrical susceptance B_d, line resistance R_{cd}, and line reactance X_{cd}, are in general hard to be detected in PDN calculation and satisfy the following equations:

$$P_d = P_{\mathrm{Ld}} + \frac{P_{\mathrm{Ld}}^2 + Q_{\mathrm{Ld}}^2}{U_{\mathrm{Ld}}^2} R_d^T + U_{\mathrm{Ld}}^2 G_d^2, \tag{1}$$

$$Q_d = Q_{\mathrm{Ld}} + \frac{P_{\mathrm{Ld}}^2 + Q_{\mathrm{Ld}}^2}{U_{\mathrm{Ld}}^2} X_d^T + U_{\mathrm{Ld}}^2 B_d^2, \tag{2}$$

$$U_d = \sqrt{\left(U_{\mathrm{Ld}} + \Delta U_d^T\right)^2 + \left(\delta U_d^T\right)^2}, \tag{3}$$

where ΔU_d^T and δU_d^T in equation (3) are the longitudinal and transverse components of the transformer impedance voltage drop at bus D. P_{Ld}, Q_{Ld}, and U_{Ld} represent the active power, reaction power, and voltage on the low-voltage side of the transformers at bus D, respectively. ΔU_d^T and δU_d^T can be obtained by the following equations:
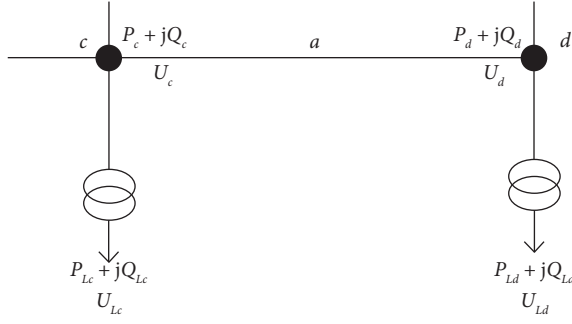
FIGURE 1: Power flow calculation circuit model.

$$\Delta U_d^T = \frac{P_{Ld} R_d^T + Q_{Ld} X_d^T}{U_{Ld}}, \tag{4}$$

$$\delta U_d^T = \frac{P_{Ld} X_d^T - Q_{Ld} R_d^T}{U_{Ld}}. \tag{5}$$

The equation of bus C can be expressed as equations (6)–(8):

$$U_c = \sqrt{\left(U_d + \Delta U_{cd}^T\right)^2 + \left(\delta U_{cd}^T\right)^2}, \tag{6}$$

$$\Delta U_{cd}^T = \frac{P_d R_{cd}^T + Q_d X_{cd}^T}{U_{Ld}}, \tag{7}$$

$$\delta U_{cd}^T = \frac{P_d X_{cd}^T - Q_d R_{cd}^T}{U_d}, \tag{8}$$

where ΔU_{cd}^T and δU_{cd}^T in equation (6) are the longitudinal and transverse components of the transformer impedance voltage drop at bus C.

2.2. Theoretical Framework of AGBO-Pro for Parameter Identification in PDN.

The schematic diagram is shown in Figure 2, where it is combined with gradient-based neural network (NN) and gradient-free optimization method. The inputs of the framework are experimental measurements of PDN, which can be divided into several blocks (or fields) including numerical features and categorical features. Each block can be processed via a customized way and then connected with customized layers. The loss function for gradient-based optimization is also flexible.

2.2.1. Physical-Informed Multihead Neural Network.

The experimental measurements of PDN can be classified into two categories, namely, the continuous (or numerical) features and discrete (or categorical) features such as measurement time, primary time type, and secondary time type. The numerical and categorical features are processed with different treatments by introducing a multihead neural network, which is shown in Figure 3. It is shown that the input layer of NN is separated into two blocks; the first includes numerical features, which are defined as $\mathbf{X}_1 = (P_{Ld,i}, Q_{Ld,i}, U_{Ld,i})$ representing the set of the active

power, reaction power, and voltage on the high-voltage side of the transformers, and the subscript i represents the sample points for $i = (1, 2, \cdots, N)$. The other includes categorical features $\mathbf{X}_2 = (T_i, P_i, S_i \cdots)$, where T_i and PV_i represent the measurement and peak-valley period, respectively. It is noted that the recording duration has 24 categories, and the peak-valley electricity has 3 categories (i.e., high, medium, and low).

After the input layer of NN, the numerical features $\mathbf{X}_1$ are fed into the PDN model, while the categorical features are first encoded (i.e., one-hot encoding) and then imported into an embedding layer to transform the sparse feature matrix into a dense matrix. After the embedding layer, we utilize the max-min normalization to scale the categorical features into $\omega_1 = [\omega_{1,0}, \omega_{1,2}, \cdots, \omega_{1,i}] \in [0, 1]$, which can be seen as a probability distribution. Then, the numerical and categorical features are merged with a linear combination as follows:

$$\mathbf{Y}_{merge} = \boldsymbol{\omega}_1 \cdot \mathbf{X}_1 + \eta, \tag{9}$$

where η represents the noise term, which is subject to the normal distribution, i.e., $\eta \sim N(0, 1)$. With the help of maximum likelihood estimation, the loss function of this work is defined as

$$L = \frac{1}{2N} \sum_{i=1}^{N} \left[y_i - \widehat{y}_i \right]^2 = \frac{1}{2N} \sum_{i=1}^{N} \left[\left(\omega_{1,i} U_{c,i} + \omega_{0,i} \right) - \widehat{U}_{c,i} \right]^2, \tag{10}$$

where y_i and $\widehat{y}_i$ represent the theoretical calculation value of PDN and true value by experimental measurement, respectively. In addition, to avoid gradient vanishing during the backward propagation of NN, a nonlinear transformation, namely, sigmoid activation function rather than linear rectification function (ReLU), is utilized:

$$\sigma(x) = \frac{1}{1 + e^{-x}}. \tag{11}$$

Therefore, the loss function is further modified as

$$L = \frac{1}{2N} \sum_{i=1}^{N} \left[y_i - \widehat{y}_i \right]^2 = \frac{1}{2N} \sum_{i=1}^{N} \left\{ \sigma\left[\left(\omega_{1,i} U_{c,i} + \omega_{0,i} \right) \right] - \widehat{U}_{c,i} \right\}^2, \tag{12}$$

where $\widehat{U}_{c,i}$ represents the true value of voltage on the high-voltage side.

The above loss function is also known as the Euclidean distance measuring the difference between theoretical calculation and experimental observation. In this work, we also utilize a Pearson correlation loss function, which is defined as

$$L_{cos} = \frac{\sum_{i=1}^{N} \left(U_{c,i} - \overline{U}_{c,i} \right) \times \left\{ \sigma\left[\left(\omega_{1,i} U_{c,i} + \omega_{0,i} \right) \right] - \overline{y}_i \right\}}{\sqrt{\sum_{i=1}^{N} \left\{ \sigma\left[\left(\omega_{1,i} U_{c,i} + \omega_{0,i} \right) \right] - \overline{y}_i \right\}^2} \times \sqrt{\sum_{i=1}^{N} \left(U_{c,i} - \overline{U}_{c,i} \right)^2}}. \tag{13}$$

In the previous work, we have derived the gradients of the loss function with respect to R_d, X_d, G_d, B_d, R_{cd}, and X_{cd}. According to the PDN model, the loss function can be
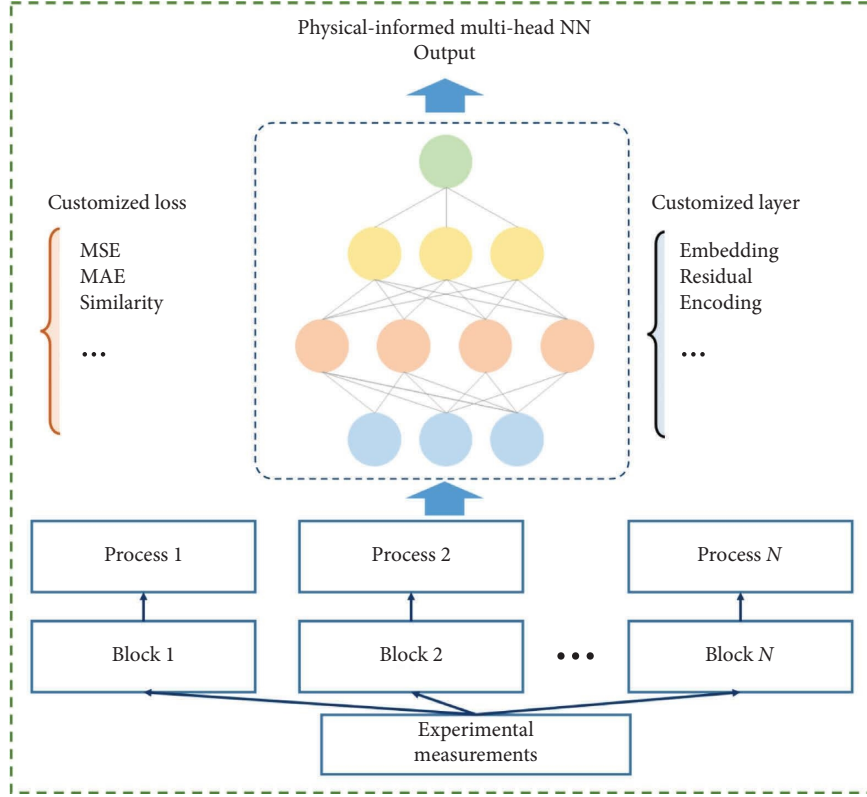
FIGURE 2: Schematic diagram of extensible gradient-based optimization for parameter identification.

calculated with forward calculation; then, the back propagation of the gradient of the loss function can be applied to update the connection weight in NN.

2.2.2. Gradient-Based Optimization Algorithm.

Once we have the above gradients of the loss function with respect to the parameters, then the gradient-based optimization can be implemented. The pseudocode of optimization method of this work is shown in Algorithm 1.

2.3. Evaluation Functions of the Parameter Identification Algorithm.

The underlying three functions are employed to estimate the performance of the proposed algorithm:

(1) Mean absolute error (MAE):

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \widehat{y}_i|. \tag{14}$$

(2) Root mean square error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \widehat{y}_i)^2}. \tag{15}$$

(3) Mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{100\%}{N} \sum_{i=1}^{N} \left| \frac{y_i - \widehat{y}_i}{y_i} \right|, \tag{16}$$

where y_i and $\widehat{y}_i$ represent the ground true value and prediction value, respectively.

3. Dataset and Calculation Details

3.1. Data Collection and Description.

In this work, a dataset including 1499 samples is collected via SCADA [33, 34] for the training of the proposed model. The voltage profiles on the high-voltage (U_a, U_b, and U_c) and low-voltage (u_a, u_b, and u_c) sides are presented in Figures 4 and 5, respectively.

From Figures 4 and 5, it is found that the high-voltage sides in the dataset are similar to the three-phase balance satisfying the equations in Section 2.1. In addition, the active power (P_a, P_b, and P_c) and reactive power (Q_a, Q_b, and Q_c) profiles on the low-voltage side are given in Figures 6 and 7, respectively.

It is found from Figures 6 and 7 that the variations of active power and reaction power show a similar trend, indicating that the data collection is stable enough for parameter identification of PDN.

It is noted that all samples collected have four categorical features, namely, measurement time, date type, primary time type, and secondary time type. The measurement time
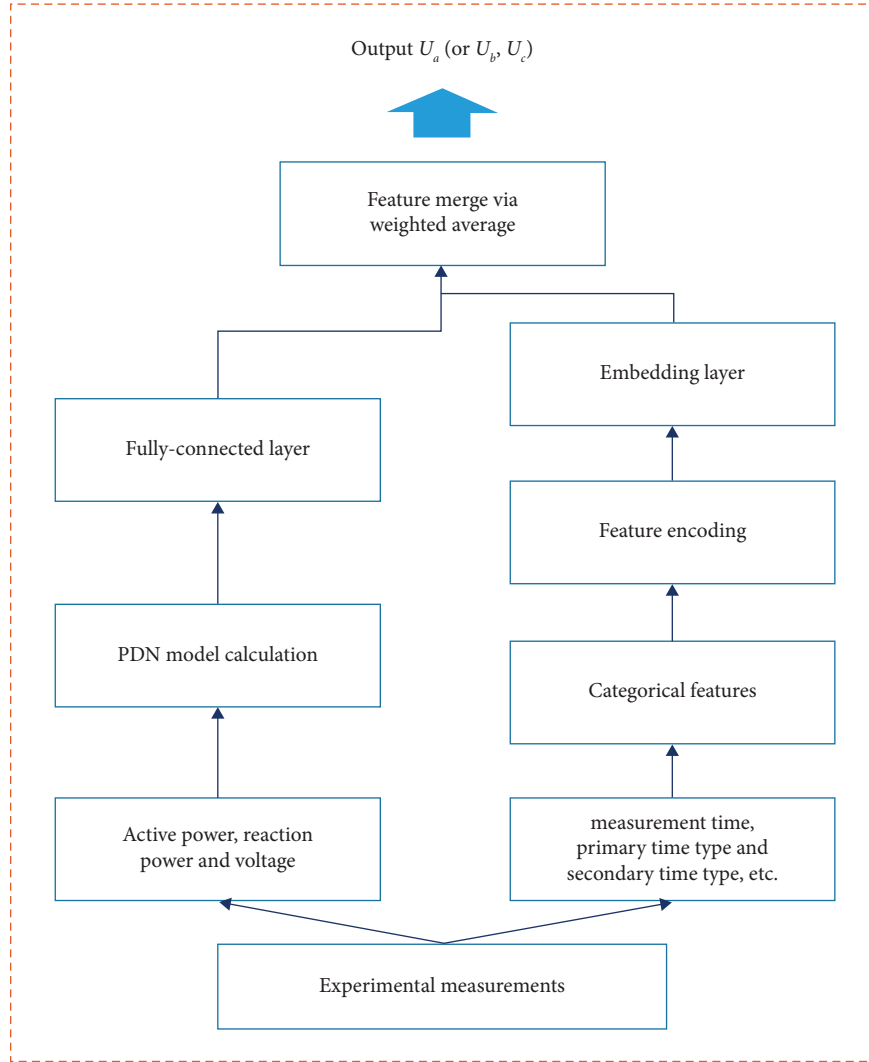
FIGURE 3: Schematic diagram of the proposed physical-informed multihead residual neural network.

represents the time information when the sample is measured, which ranges from 0 to 24 hours. The date type represents whether the measurement time is on a workday and holiday. The primary time type and secondary time type have two different definitions for daytime. The primary time type has three levels: peak mean hours refer to 09:00 to 12:00 and 18:00 to 21:00 daily, plateau mean hours refer to 13:00 to 17:00 and 22:00 to 23:00 daily, and valley represents 00:00 to 08:00. The secondary time type has two levels: peak means 08:00 to 21:00, whereas valley refers to 22:00 to 07:00. The distribution of these four categorical features is shown in Figure 8.

3.2. Evaluation and Calculation. In this paper, 75% samples (1124) are split randomly as train set to identify PDN's parameters. The best parameters are used to calculate voltage per unit in C bus (denoted as U_{cal}) by the power flow model. After that, the rest of 25% samples (375) are used to evaluate the performance of parameter identification as test set through the three metrics as shown in

equations (14)–(16). Instead of directly calculating these metrics, linear regression should be applied in this paper, and the values of U_c and U_{cal} are regarded as dependent variable and independent variable, respectively. The output values of linear regression are denoted by U_{cal}^*, and the final evaluations of parameter identification are gained between U_c and U_{cal}^*:

$$U_{cal}^* = \mathrm{a}U_{cal} + b, \tag{17}$$

where a and b are denoted as slope and bias of linear regression. In the following discussion, the parameters of linear regression optimized by SMBO methods are signed as RS-LR, TPE-LR, and SA-LR, respectively. The upper bounds and lower bounds of the identified parameters should be determined firstly, and they are listed in Table 1.

To mitigate the impact of randomness associated with AGBO and SMBO-based methods on the results in this study, the dataset was randomly partitioned 25 times to ensure accuracy and stability in the results.

Input: θ_0 initial parameters
   $f(\theta)$ objective function to be optimized
   β_1, β_2 decay rates for moment estimates
   m_0 initialized first-order moment
   v_0 initialized second-order moment
   t time step
   η learning rate
   $l_\infty \leftarrow 2/(1 - \beta_2) - 1$ maximum length of the simple moving average
While θ_t is not converged **do**
   $t \leftarrow t + 1$
   $g_t \leftarrow \nabla_\theta f_t(\theta_{t-1})$ gradient with respect to parameters at time step t
   $m_t \leftarrow \beta_1^t \cdot m_{t-1} + (1 - \beta_1^t) \cdot g_t$ update first-order moment
   $v_t \leftarrow \beta_2^t \cdot v_{t-1} + (1 - \beta_2^t) \cdot g_t^2$ update second-order moment
   $\hat{m}_t \leftarrow m_t/(1 - \beta_1^t)$ biased-corrected first-order moment
   $l_t \leftarrow l_\infty - 2t\beta_2^t/(1 - \beta_2^t)$ the length of the moving average
   If the variance is tractable, i.e., $l_t > 4$ **then**
      $\eta_t \leftarrow \sqrt{(1 - \beta_2^t)/v_t}$ update the adaptive learning rate
      $r_t \leftarrow \sqrt{((l_t - 4)(l_t - 2)l_\infty)/((l_\infty - 4)(l_\infty - 2)l_t)}$ the variance rectification term
      $\theta_t \leftarrow \theta_{t-1} - t\eta_t\hat{m}_t r_t$ update parameters
   Else
      $\theta_t \leftarrow \theta_{t-1} - t\hat{m}_t$
End while
Return θ_t

ALGORITHM 1: Adaptive gradient-based optimization methods.
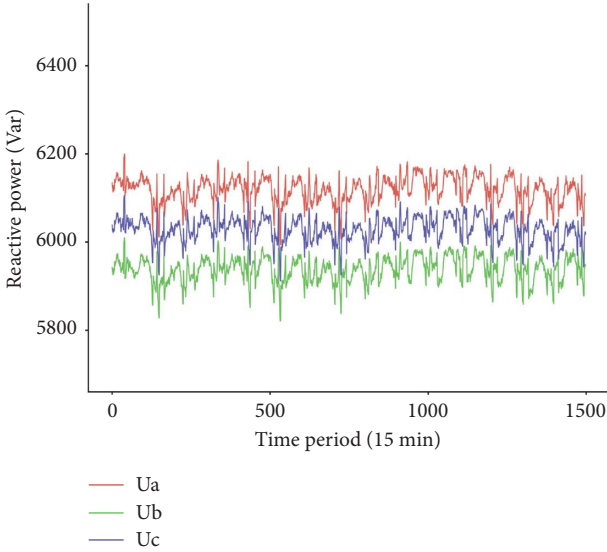
FIGURE 4: U_a, U_b, and U_c on the high-voltage side.
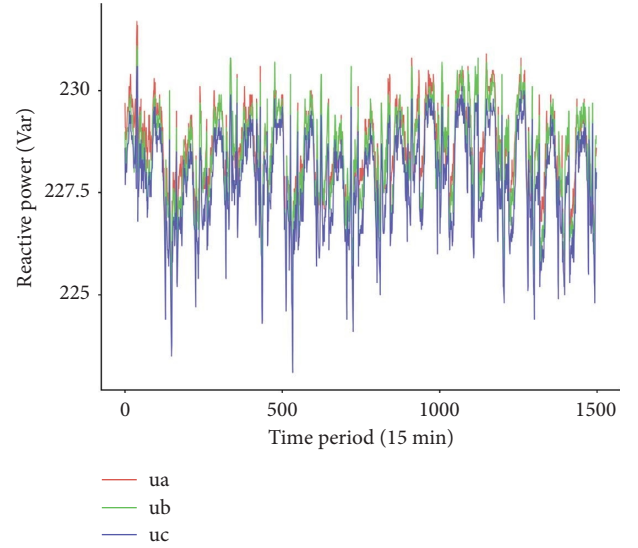
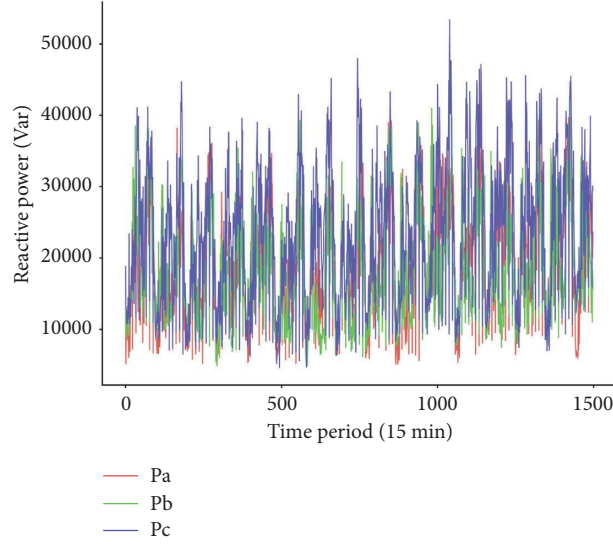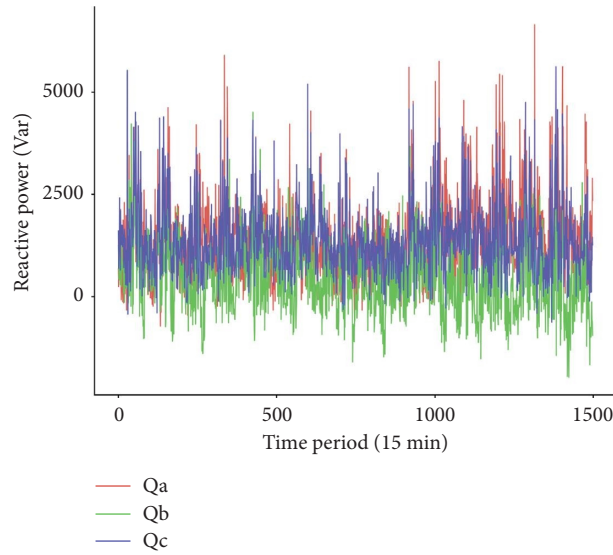FIGURE 5: u_a, u_b, and u_c on the low-voltage side.

4. Results and Discussion

Before discussing the results, some hyperparameter settings of each method are described as follows. The prior weight and number of started jobs are set as 1 and 20 for TPE, and the rate of reduction in SA is 0.1 as default value. The learning rate is $5e-4$ in AGBO-based methods. The maximum of iteration step is 1000 for all the methods in this study. The parameter identification results of AGBO and SMBO-based methods with mean square error between U_c and U_{cal} are shown in Table 2.

It can be found in Table 2 that AGBO-Pro has the best performance with significantly low values of MAE, RMSE, and MAPE compared with other metaheuristic algorithms such as AO, NRO, and PSS. AGBO also has better results than SMBO-based methods, but the prediction results do not have remarkable differences since the statistical properties between U_c and U_{cal} are neglected.

FIGURE 6: P_a, P_b, and P_c on the low-voltage side.

FIGURE 7: Q_a, Q_b, and Q_c on the low-voltage side.

Other recent studies also have proposed the prediction results with the same metrics and the same dataset in this paper, such as the methods of MCMC and SMBO combined with clustering and hypothesis testing (denoted as $MCMC_C$ and $SMBO_C$). Li et al. [32] published the best results of MAE values of $MCMC_C$ and $SMBO_C$ being 62.467 ± 0.366 and 61.868 ± 0.322, respectively. In another paper [31], the values of MAE computed by $MCMC_C$ and $SMBO_C$ are 62.136 ± 0.336 and 61.268 ± 0.311, respectively.

Based on the previous study [26], the line transformation should be implemented to U_{cal} before calculating loss function. The parameter identification results with linear transformation are listed in Table 3.

All methods perform better in Table 3 than the results in Table 2, which indicates that the linear transformation between U_c and U_{cal} has an important contribution to identify

PDN's parameters. Moreover, the results between AGBO-Pro and AGBO mean that the supplementary categorical information such as measurement time, date type, primary time type, and secondary time type plays an important role in PDN's parameter identification and the key categorical information can be merged by AGBO-Pro proposed in this work.

Leaning rate, the size of the embedding layer dimension, and the number of hidden layers are three critical hyperparameters of AGBO-Pro; therefore, the PDN's parameter identification performance under different hyperparameters has been investigated in this section. The performances of various learning rates are displayed in Table 4.

It can be found that the learning rate has a remarkable influence on AGBO-Pro; when the learning rate is set to $5e - 3$, the identification performance is optimal, and the
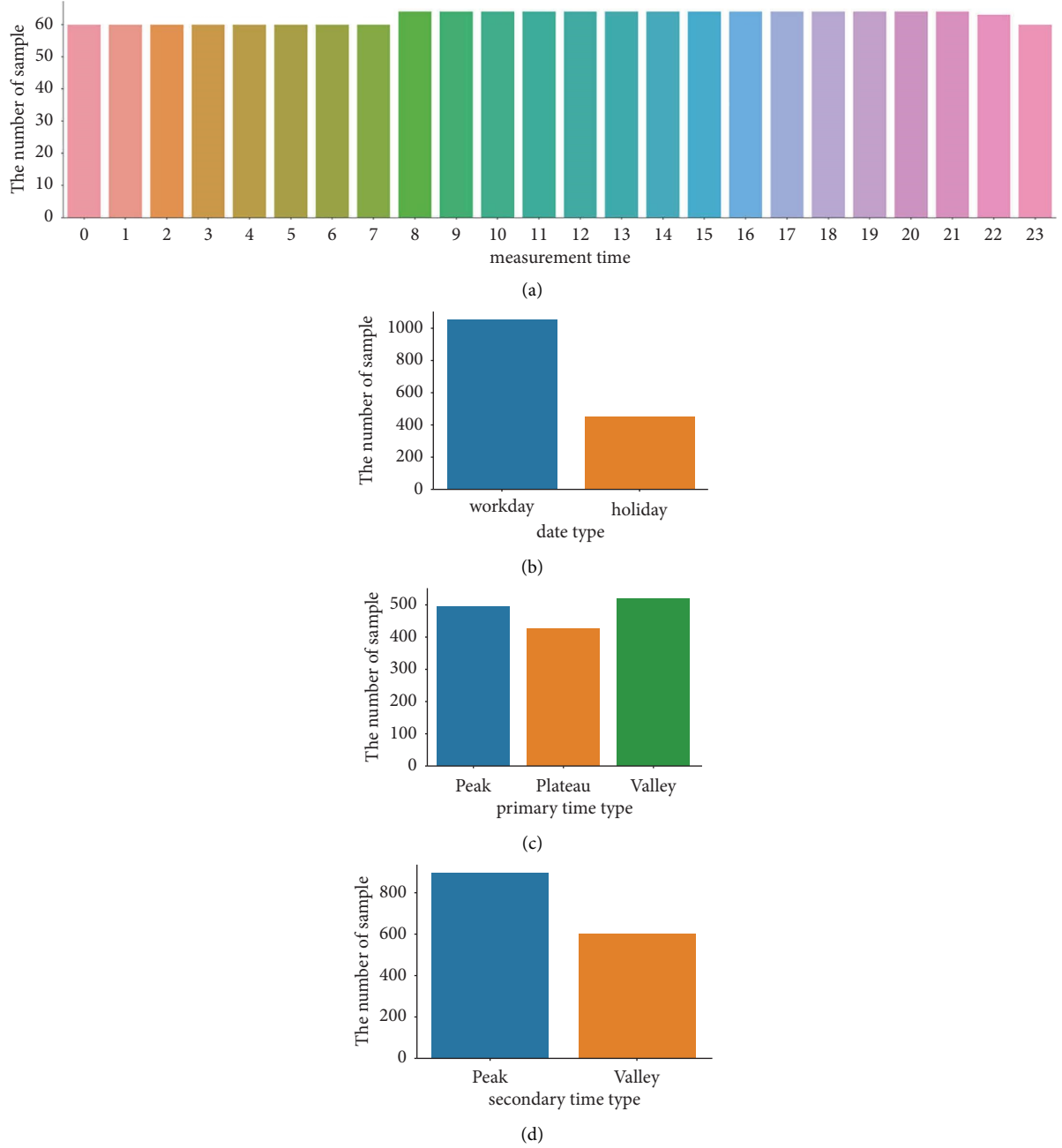
FIGURE 8: The distribution plot of measurement time, date type, primary time type, and secondary time type among samples.

TABLE 1: The upper and lower bounds of the identified parameters.

Parameter name	Abbreviation	Upper bound	Lower bound	Unit
Line resistance	R_{cd}	0.5	0.005	Ω/km
Line reactance	X_{cd}	0.5	0.005	Ω/km
Transformer resistance	X_d	20	5	Ω/km
Transformer reactance	R_d	10	0.8	Ω/km
Transformer conductance	G_d	$8e-6$	$4e-6$	S
Transformer electrical susceptance	B_d	$8e-5$	$2e-5$	S
Slope of linear regression	a	-5	5	—
Bias of linear regression	b	500	1000	—

TABLE 2: The results of parameter identification with the loss function of mean square error.

Method	MAE	RMSE	MAPE
AGBO-Pro[*]	**25.415 ± 0.855**	**32.511 ± 1.050**	**0.415 ± 0.014**
AGBO	64.100 ± 0.474	65.791 ± 0.478	1.047 ± 0.008
RS	65.570 ± 0.746	67.148 ± 0.699	1.071 ± 0.012
TPE	64.689 ± 0.648	66.355 ± 0.633	1.057 ± 0.011
SA	65.894 ± 0.983	67.439 ± 0.913	1.077 ± 0.016
AO	64.5001 ± 0.7016	66.1795 ± 0.6854	1.0539 ± 0.0115
NRO	64.5174 ± 0.6998	66.1954 ± 0.6837	1.054 ± 0.0114
PSS	64.6944 ± 0.6984	66.3599 ± 0.6847	1.0571 ± 0.0114

[*]AGBO-Pro uses mean square loss to ensure fairness in comparison. The bold values indicate that the AGBO-Pro method gains the lowest values in all three evaluation functions, viz. MAE, RMSE and MAPE, indicting its best performance.

TABLE 3: The results of parameter identification with linear transformation.

Method	MAE	RMSE	MAPE
AGBO-Pro-LR[*]	**5.131 ± 0.093**	**6.514 ± 0.152**	**0.084 ± 0.002**
AGBO-LR	5.247 ± 0.079	6.593 ± 0.111	0.086 ± 0.001
RS-LR	6.447 ± 0.801	8.054 ± 0.958	0.105 ± 0.013
TPE-LR	6.078 ± 0.753	7.589 ± 0.830	0.099 ± 0.012
SA-LR	6.970 ± 1.111	8.682 ± 1.318	0.114 ± 0.018

[*]The results of AGBO-Pro with Pearson correlation coefficient loss. After the linear transformation labeled as "AGBO-Pro-LR," the optimization method proposed in this work still has the best performance.

TABLE 4: The performance of AGBO-Pro under different learning rates.

Learning rate	MAE	RMSE	MAPE
$5e - 2$	5.850 ± 0.192	7.365 ± 0.266	0.0956 ± 0.00315
$1e - 2$	5.486 ± 0.242	6.916 ± 0.266	0.0897 ± 0.0028
$5e - 3$	**5.131 ± 0.093**	**6.514 ± 0.152**	**0.0839 ± 0.0015**
$1e - 3$	5.486 ± 0.172	6.916 ± 0.242	0.0897 ± 0.0028
$1e - 4$	8.231 ± 0.327	11.121 ± 0.493	0.135 ± 0.0053
$1e - 5$	8.545 ± 0.057	11.454 ± 0.326	0.140 ± 0.0009

The bold values indicate that with the learning rate $5e - 3$, the model has the lowest values in three evaluation functions.

TABLE 5: The performance of AGBO-Pro under different sizes of the embedding layer dimension.

Embedding dimension	MAE	RMSE	MAPE
Measurement time: 32 Date type: 16 Primary time type: 16 Secondary time type: 16	5.370 ± 0.201	6.751 ± 0.279	0.0878 ± 0.003
Measurement time: **64** Date type: **32** Primary time type: **32** Secondary time type: **32**	**5.131 ± 0.093**	**6.514 ± 0.152**	**0.0839 ± 0.001**
Measurement time: 128 Date type: 64 Primary time type: 64 Secondary time type: 64	5.224 ± 0.120	6.596 ± 0.209	0.0854 ± 0.002

value of learning rate between $1e - 2$ and $1e - 3$ is suggested in this paper. The size of embedding layer dimension is investigated subsequently under the optimal value of learning rate and listed in Table 5.

Since the dimension of categorical features is small, the embedding dimension of the neural network is less than 128 in this work. According to the results in Table 5, the change of the embedding dimension has only a minor impact on the identification performance, and the optimal size of the embedding dimension is chosen as 64, 32, 32, and 32 for the four categorical features, respectively. AGBO-Pro include the hidden layer to leaning the information of categorical features after embedding, and the influence of the number of the hidden layers are shown in Table 6.

Having more hidden layers in the network implies a larger number of parameters, slower computation speed, and a higher risk of overfitting. Combining the results from Table 6, it can be found that a single hidden layer achieves

TABLE 6: The performance of AGBO-Pro under different number of the hidden layers.

Embedding dimension	MAE	RMSE	MAPE
1 hidden layer	**5.131 ± 0.093**	**6.514 ± 0.152**	**0.084 ± 0.002**
2 hidden layers	5.574 ± 0.285	7.048 ± 0.344	0.091 ± 0.005
3 hidden layers	5.545 ± 0.279	7.052 ± 0.397	0.091 ± 0.005

The bold values indicate that with one hidden layer, the model gains the best performance.
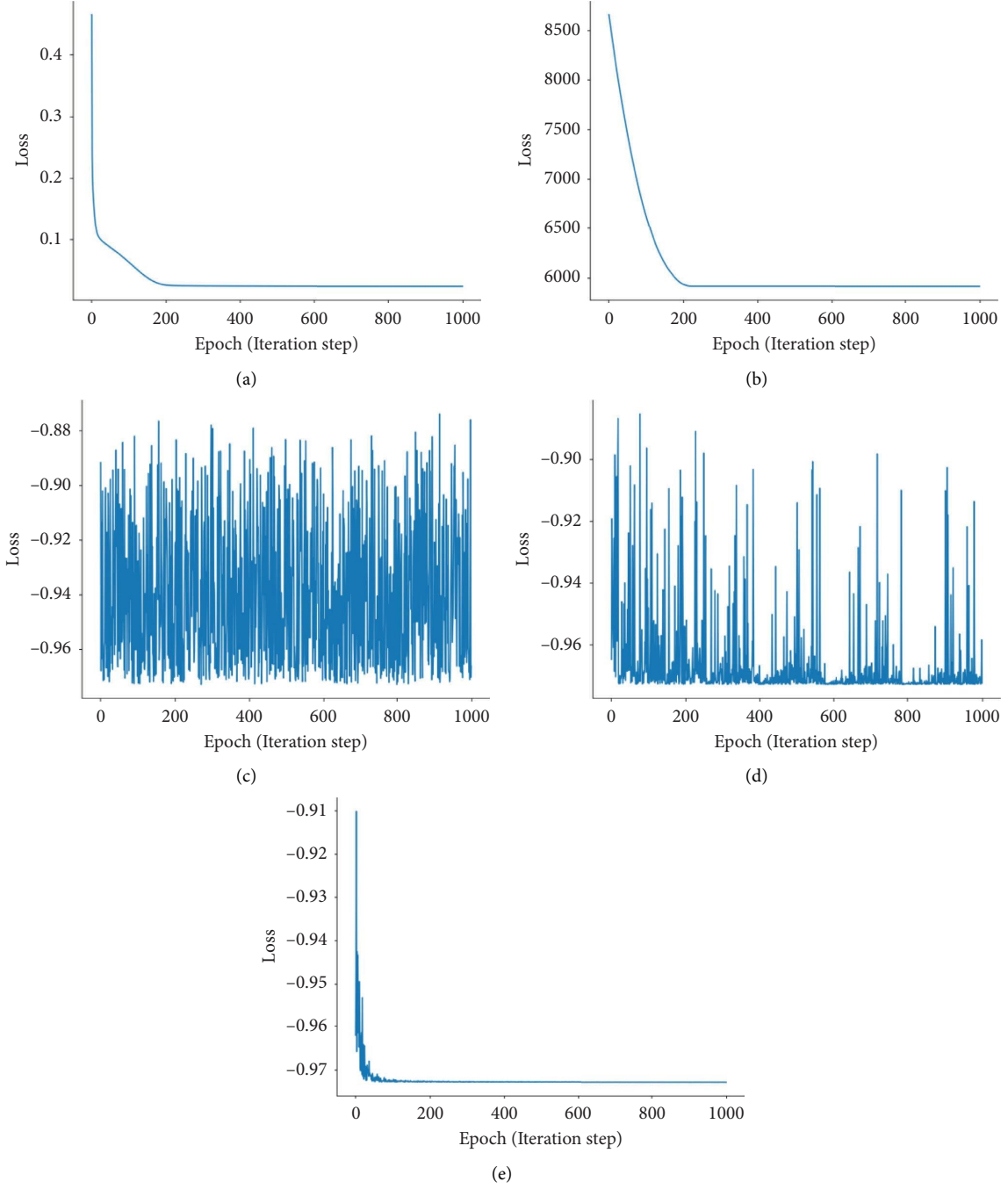

FIGURE 9: The convergence plot of AGBO-Pro (a), AGBO (b), RS (c), TPE (d), and SA (e).
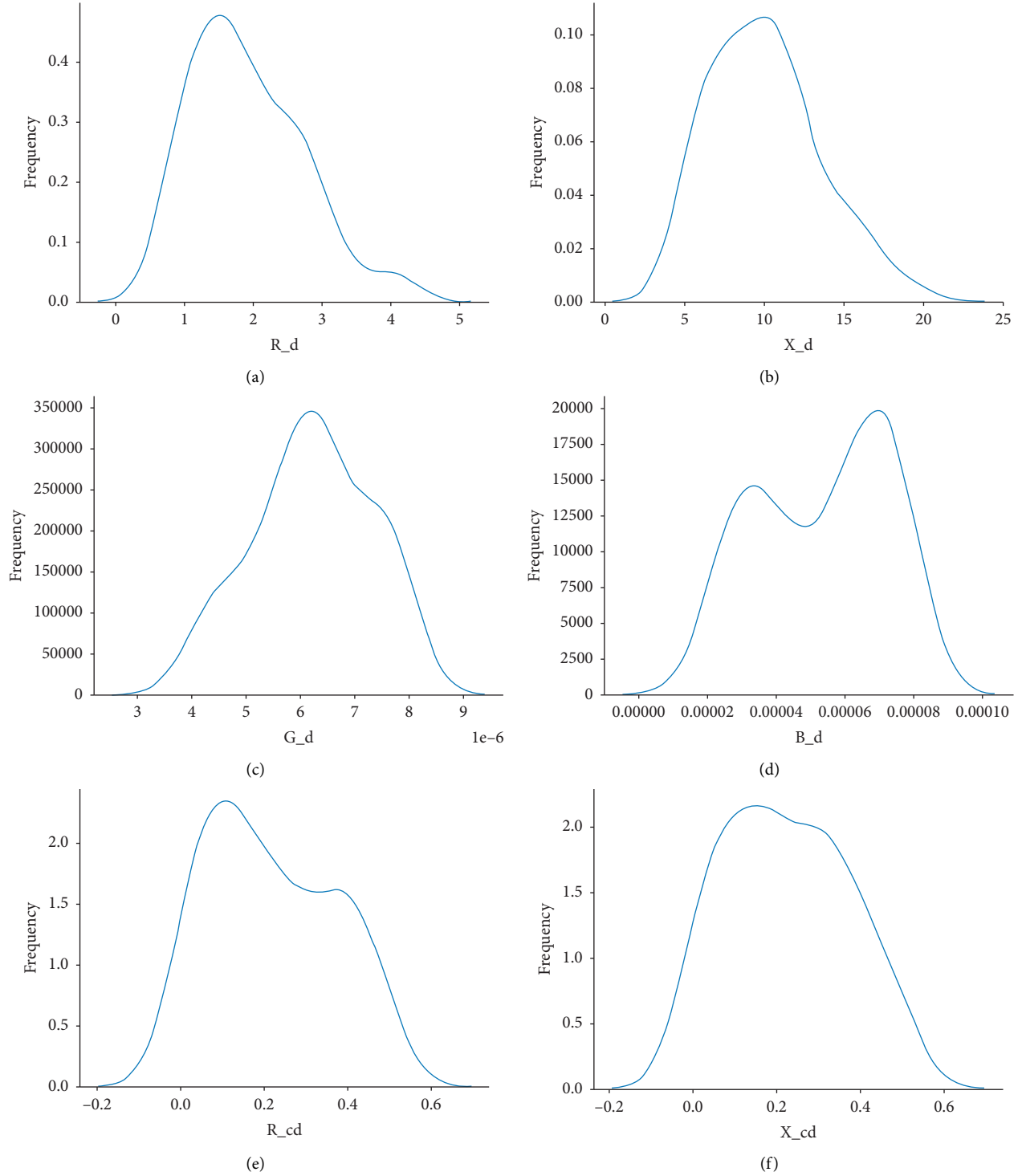
FIGURE 10: The distribution plot of R_d (a), X_d (b), G_d (c), B_d (d), R_{cd} (e), and X_{cd} (f).

better identification performance. The convergence plots of AGBO and SMBO-based methods are displayed in Figure 9.

The AGBO-based methods converge after 200 iterations; compared with the SMBO-based methods, the convergence plots of AGBO-based methods are much smoother and stable, since the searching direction for parameter update is deterministic to the gradient-based optimization method,

such as AGBO and AGBO-Pro. After 25 repeated splitting datasets, the distribution plots of the identified PDN's parameters from AGBO-Pro-LR are shown in Figure 10. It can be found that all the identified parameters are roughly distributed within a relatively fixed range, providing a data foundation for the subsequent parameter analysis in future research.

5. Conclusion

In this work, we propose an extensible gradient-based optimization method for parameter identification in PDN calculation and analysis. A physical-informed multihead neural network is adopted to treat the numerical features and categorical features separately. The two kinds of features are merged via a weighted average. After several forward-backward calculations, the similarity loss function with respect to the six parameters to be identified achieves a fast convergence.

We compare the proposed method (namely, AGBO-Pro model) with the original AGBO model and several heuristic algorithms such as RS, TPE, SA, AO, NRO, and PSS. The numerical calculations show that the errors by AGBO-Pro are the lowest in all three evaluation functions, i.e., MAE, RMSE, and MAPE, with a faster and more stable convergence of the loss function. By further taking a linear transformation of the loss function, the method of this work has a lower variance in 25 repeat experiments, showing a much more robust performance in parameter identification.

In addition, the variations in hyperparameters of optimization method such as the number of hidden layers and embedding layers, learning rate, and weight decay are also systematically investigated. It is found that the method proposed in this work achieves more stable and robust performance to identify PDN parameters. This work shows an effective exploration in incorporating the numerical and categorical features of experimental measurement into gradient-based optimization method.

Data Availability

Data are available from the corresponding authors upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] J. Li, H. Ma, and H. Dong, "Present development condition and trends of energy storage Technology in the integration of distributed renewable energy," *Transactions of China Electrotechnical Society*, vol. 31, pp. 1–10, 2016.

[2] P. M. Anderson, *Analysis of Faulted Power Systems*, Wiley-IEEE Press, Hoboken, NJ, USA, 1995.

[3] 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, 2019.

[4] Z. Wang, C. Hu, W. Wang, X. Kong, and W. Zhang, "A prognostics-based spare part ordering and system replacement policy for a deteriorating system subjected to a random lead time," *International Journal of Production Research*, vol. 53, no. 15, pp. 4511–4527, 2014.

[5] Y. Lin and A. Abur, "Strategic use of synchronized phasor measurements to improve network parameter error detection," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5281–5290, 2018.

[6] A. Arancibia, C. A. Soriano–Rangel, F. Mancilla–David, R. Ortega, and K. Strunz, "Finite–time identification of the Thévenin equivalent parameters in power grids," *International Journal of Electrical Power & Energy Systems*, vol. 116, Article ID 105534, 2020.

[7] B. V. Krishna and N. P. Srinivasu, "A direct approach for distribution system load flow solutions," *International Journal of Engineering and Advanced Technology*, vol. 8, p. 6s3, 2019.

[8] B. Zaker, G. B. Gharehpetian, and M. Karrari, "Small signal equivalent model of synchronous generator-based grid-connected microgrid using improved Heffron-Phillips model," *International Journal of Electrical Power & Energy Systems*, vol. 108, pp. 263–270, 2019.

[9] V. L. Srinivas and J. Wu, "Topology and parameter identification of distribution network using smart meter and μPMU measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–14, 2022.

[10] L. He, S. Zhao, X. Wang, and C. Lee, "Artificial neural network-based parameter identification method for wireless power transfer systems," *Electronics*, vol. 11, no. 9, p. 1415, 2022.

[11] Z. Wang, M. Xia, M. Lu, L. Pan, and J. Liu, "Parameter identification in power transmission systems based on graph convolution network," *IEEE Transactions on Power Delivery*, vol. 37, no. 4, pp. 3155–3163, 2022.

[12] C. Wang, Z. Wang, J. Wang, and D. Zhao, "SVM-based parameter identification for composite ZIP and electronic load modeling," *IEEE Transactions on Power Systems*, vol. 34, no. 1, pp. 182–193, 2019.

[13] Z. Wang, L. Weng, M. Lu, J. Liu, and L. Pan, "Application of multiattention mechanism in power system branch parameter identification," *Complexity*, vol. 2021, Article ID 1834428, 12 pages, 2021.

[14] Z. Yan and Y. Xu, "Data-driven load frequency control for stochastic power systems: a deep reinforcement learning method with continuous action search," *IEEE Transactions on Power Systems*, vol. 34, no. 2, pp. 1653–1656, 2019.

[15] J. Song, E. Dall'Anese, A. Simonetto, and H. Zhu, "Dynamic distribution state estimation using synchrophasor data," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 821–831, 2020.

[16] Z. Ma, S. Zhang, and J. Li, "A single-phase grounding fault section location method for power distribution network based on parameter identification," *Guangdong Electric Power*, vol. 32, pp. 65–71, 2019.

[17] M. Cui, M. Khodayar, C. Chen, X. Wang, Y. Zhang, and M. E. Khodayar, "Deep learning-based time-varying parameter identification for system-wide load modeling," *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6102–6114, 2019.

[18] J. Xiao, Y. Situ, W. Yuan, X. Wang, and Y. Z. Jiang, "Parameter identification method based on mixed-integer quadratic programming and edge computing in power Internet of Things," *Mathematical Problems in Engineering*, vol. 2020, Article ID 4053825, 10 pages, 2020.

[19] K. Yuan and Z. Wei, "Power line parameter identification based on multi-innovation least square algorithm," *Electric Power Engineering Technology*, vol. 39, pp. 55–160, 2020.

[20] H. Wang, H. Jiao, J. Chen, and W. Liu, "Parameter identification for a power distribution network based on MCMC algorithm," *IEEE Access*, vol. 9, pp. 104154–104161, 2021.

[21] W. Dong, P. Zanchetta, and D. Thomas, "Identification of electrical parameters in a power network using genetic algorithms and transient measurements," *COMPEL: The International Journal for Computation & Mathematics in Electrical & Electronic Engineering*, vol. 29, no. 1, pp. 235–249, 2010.

[22] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," *Lecture Notes in Computer Science*, vol. 6683, pp. 507–523, 2011.

[23] B. James and B. Yoshua, "Random search for hyper-parameter optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281–305, 2012.

[24] B. James, B. Rémi, B. Yoshua, and K. Balázs, "Algorithms for hyper-parameter optimization," *Advances in Neural Information Processing Systems*, pp. 2546–2554, 2011.

[25] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, no. 4598, pp. 671–680, 1983.

[26] B. Li, Y. Jiang, K. Hu, and X. Zhou, "Adaptive gradient-based optimization method for parameter identification in power distribution network," *International Transactions on Electrical Energy Systems*, vol. 2022, Article ID 9300522, 14 pages, 2022.

[27] J. Sun, M. Xia, and Q. Chen, "A classification identification method based on phasor measurement for distribution line parameter identification under insufficient measurements conditions," *IEEE Access*, vol. 7, pp. 158732–158743, 2019.

[28] L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. Al-qaness, and A. H. Gandomi, "Aquila Optimizer: a novel meta-heuristic optimization algorithm," *Computers & Industrial Engineering*, vol. 157, Article ID 107250, 2021.

[29] Z. Wei, C. Huang, X. Wang, T. Han, and Y. Li, "Nuclear reaction optimization: a novel and powerful physics-based algorithm for global optimization," *IEEE Access*, vol. 7, pp. 66084–66109, 2019.

[30] M. Shaqfa and K. Beyer, "Pareto-like sequential sampling heuristic for global optimisation," *Soft Computing*, vol. 25, no. 14, pp. 9077–9096, 2021.

[31] C. Wang, W. Xiang, K. Hu, and B. Li, "A dynamic data prepossessing method for parameter identification in power distribution network," *IET Communications*, vol. 17, no. 8, pp. 1012–1019, 2023.

[32] B. Li, H. Chen, and K. Hu, "A data preprocessing based on cluster and testing of parameter identification method in power distribution network," *Energies*, vol. 15, no. 21, p. 8007, 2022.

[33] B. Li, J. Ma, K. Hu et al., "A method for parameter identification of distribution network equipment based on sequential model-based optimization," *International Transactions on Electrical Energy Systems*, vol. 2022, Article ID 9880284, 12 pages, 2022.

[34] B. Li, K. Hu, J. Ma et al., "The new definitions of loss function for the model-based parameter identification method in power distribution network," *International Transactions on Electrical Energy Systems*, vol. 2022, Article ID 4197043, 11 pages, 2022.