The New Definitions of Loss Function for the Model-Based Parameter Identification Method in Power Distribution Network

Bin Li, Ke Hu, Jiayang Ma, Shihe Xu, Yuanwei He, Hao Jiao, and Jinning Chen

1 Nanjing Institute of Technology, Nanjing 211167, Jiangsu, China
2 Chongqing University of Posts and Telecommunications, Chongqing 400065, China
3 University of Science and Technology of China, Hefei 230027, Anhui, China
4 Chongqing University, Chongqing 401331, China
5 State Grid Jiangsu Electric Power Co., Ltd., Research Institute, Nanjing 210096, Jiangsu, China

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

Received 24 May 2022; Accepted 9 July 2022; Published 4 August 2022

Academic Editor: Santoshkumar Hampannavar

Copyright © 2022 Bin Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Accurate device parameters play critical roles in calculation and analysis of power distribution network (PDN). However, device parameters are always affected by the operating status and influenced by manual entry; besides, the distribution area of PDN is very wide with many points, which brings more challenges to PDN parameter identification. Most of the proposed algorithms recently assume that the parameters of PDN contribute in a nonlinear probability space and optimize parameters by the power flow model with a loss function. Although these algorithms can achieve satisfying results in PDN analysis, the relationship between the power flow model and loss functions remains unclear. In this paper, the outputs of the power flow model have been analyzed firstly by experimental data, which include the head and end voltages, as well as active and reactive power on the low-voltage side. It is revealed that the loss functions used by current algorithms are not suitable and reasonable for power flow model in PDN calculation, which constitutes one of the main findings of this work. Subsequently, this work proposes four novel loss functions combined with genetic algorithm (GA) and Markov Chain Monte Carlo (MCMC) to identify PDN parameters. Compared with the published algorithms, our experimental results show that the loss function defined in this paper can achieve better and more stable performance with about two times lower in MAE, RMSE, and RMPE evaluation functions to identify PDN parameters.

1. Introduction

Reliable and accurate parameters play important roles in the power distribution network (PDN) for many aspects, such as security analysis, system control, state estimation, line loss calculation, power flow calculation, protection setting, and fault analysis [1]. However, since real-time measuring equipment is always lacking in practice, it is difficult to obtain real-time information directly for PDN in security and stability situation. Some import parameters, such as line resistance, line reactance, transformer resistance, transformer reactance, transformer conductance, and transformer electrical susceptance, are usually assumed to be static in real situation, and it always results in poor calculations for PDN because of the differences of data in actual situation [2]. Some new approaches focusing on calculation efficiency and error reducing have been developed in many fields of PDN, such as supervisory control and data acquisition, power management unit (PMU), power management unit, and advanced metering infrastructure. These methods include 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], and Hefron–Phillips method [8]. Furthermore, some new approaches based on machine learning like particle swarm optimization (PSO) [9], evolutionary strategies [10], estimation using synchro-phasor data [11], PSCAD simulation
needed to be placed in the grid, since these methods require data regarding the voltage, current, active power, and reactive power on both the high-voltage and low-voltage sides.

Identification equations based on power flow model for the equipment parameters in PDN have been established to solve the problems of lacking required data and measuring devices. It builds relationships between the parameters to be identified and the data which can be obtained easily. Generally, active power, reaction power, and voltage on the low-voltage side and high-voltage side in PDN can be obtained easily. Other parameters are lacking in practice, such as transformer electrical, transformer resistance, transformer conductance, transformer electrical susceptance, line resistance, and line reactance. The ranges of these parameters are always optimized by algorithms combined with the static parameters like active power, reaction power, and voltage on the low-voltage side. The calculating values of voltage on the high-voltage can be computed by power flow model with the parameters mentioned before, and the residuals between calculating and true values are always used to build loss function for optimization algorithms. Therefore, the problems of parameters identification can be easily solved by model-free methods, such as the least squares (LS) [14] and Markov Chain Monte Carlo (MCMC) [15]. LS is suitable for convex parameter space and is sensitive to the initial value, and it is difficult to employ LS for obtaining satisfied results in calculating of PDN. MCMC performs better than LS on regular data but it is hard to predict the parameters with outliers or high deviations. Based on this, LS and MCMC are hard to be applied in real situation.

With the rapid development of computer science and artificial intelligence technology, some new methods have been published in the field of PDN parameter identification. The method based on global optimization [16], including random search approach (RS) [17], tree-structured Parzen estimator approach (TPE) [18], and simulated annealing (SA) [19], have been published recently [20] and shown better performance of parameter identification than LS and MCMC. Genetic algorithms (GA) combined with transient measurements have been proposed to identify electrical parameters in a power network [21]. Besides, some methods based on machine learning, such as support vector regression [22], ensemble Kalman filtering [23], and interior point method, are combined with discrete particle swarm optimal [24]. Sun et al. [25] take advantage of the time invariant characteristics of PDN line parameters in a short time and proposed a method of convolutional neural network (CNN) for regression calculations for line parameter identification. These methods have proven to be effective for simulation data; however, only part of the parameters can be obtained, and the precise values of voltage and voltage phase angle are necessary.

The definition of loss function is critical for optimization algorithms since it determines the optimization objectives in PDN calculation. Nevertheless, most of the published algorithms define loss function by the residuals [20], the integral of the absolute error [21], or weighted least absolute value [24] between calculating and true values of a parameter directly (e.g., voltage on the high-voltage side). This approach of definition is straightforward, but it may neglect some statistical properties of parameters. If there exist the statistical properties, the calculating value of the parameters can be further transformed to reduce values of the loss function. In this paper, the value of voltage on the high-voltage side is regarded as the output from power flow model. Firstly, we investigate the statistical property between calculating and real values of this parameter, and an approximate linear relationship has been found by experimental data. Subsequently, a linear transformation is used after obtaining the calculating values of the parameters, and it is shown that this linear transformation can significantly reduce the residual between the calculating and real values of voltage on the high-voltage side. Finally, four linear statistical metrics, viz. Person’s correlation coefficient, Spearman’s correlation coefficient [26], Akaike information criterion [27], and Bayesian information criterion [28], have been used to build new loss functions for MCMC and GA to identify PDN’s parameters. Based on the abovementioned points, this paper mainly has the following contributions:

1. The statistical properties between the calculating and true values of PDN’s parameters are investigated, which are rarely studied and always neglected by other researchers.
2. The linear statistical property is found in our experiment, and this property can improve parameter identification in PDN’s calculation.
3. Four new loss functions are proposed by linear statistical metrics and combined with GA as well as MCMC to identify PDN’s parameters.

The main work of this paper is displayed in Figure 1, and the structure of the paper is organized as follows: Section 2 introduces the identification equations of power flow model in PDN and proposes the new definition of loss functions. The experimental data and calculation details are given in Section 3. The results and relevant discussions can be found in Section 4. Finally, Section 5 gives a brief conclusion.

2. Methodology

2.1. Power Flow Model Calculation. The basic theory of analysis in PDN can be found in [15, 20]. To simplify the computation, the three phases are assumed to be balanced as the premise for calculating the power flow in this paper. The schematic diagram of power flow calculation circuit model is shown in Figure 2.

In Figure 2, $P_d$, $Q_d$, and $U_d$ represent the active power, reaction power, and voltage on the high-voltage side of the transformers at bus $D$, respectively. And these 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 hard to be detected in PDN calculation and satisfy the following equations:

$$
P_d = P_{Ld} + \frac{P_{Ld}^2 + Q_{Ld}^2}{U_{Ld}^2} U_{Ld}^2 G_{cd}^2,
$$

$$
Q_d = Q_{Ld} + \frac{P_{Ld}^2 + Q_{Ld}^2}{U_{Ld}^2} U_{Ld}^2 B_{cd}^2,
$$

$$
U_d = \sqrt{(U_{Ld} + \Delta U_d^T)^2 + (\delta U_d^T)^2},
$$

where $\Delta U_d^T$ and $U_d^T$ in (3) are the longitudinal and transverse components of the transformer impedance voltage drop at bus $D$ in $V$ and can be expressed by (4) and (5), respectively:

$$
\Delta U_d^T = \frac{P_{Ld} R_{cd}^T + Q_{Ld} X_{cd}^T}{U_{Ld}},
$$

$$
\delta U_d^T = \frac{P_{Ld} X_{cd}^T + Q_{Ld} R_{cd}^T}{U_{Ld}}.
$$

The equation of bus $C$ can be expressed as (6)–(8), whereas the final equation are calculated by (9):

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

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

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

$$
f_c = \sqrt{(U_d + \Delta U_{cd}^T)^2 + (\delta U_{cd}^T)^2} - \sqrt{(U_{Le} + \Delta U_{Le}^T)^2 + (\delta U_{Le}^T)^2}.
$$

The parameters in the line and transformer can be calculated based on (1)–(9) with the measured power and voltage data.

2.2. The Optimization Algorithm for PDN. There are three critical difficulties pointed out by [22] in parameter identification task of PDN:

(1) The parameter space of PDN may be nonconvex.

(2) The calculation efficiency is low since there are too many parameters needed to be identified owing to the large scale of lines and nodes.
(3) The numerical levels of parameters are exponentially different, which can affect the identification results due to the measurement errors.

These three critical points can lead some commonly used algorithms to suffer in parameter identification of PDN, such as least squares (LS). Therefore, [15] proposed Markov Chain Monte Carlo (MCMC) to solve these problems, and the methods based on sequential model-based global optimization (SMBO) [20] have also been proposed and shown their excellent performance to overcome these problems. In this study, MCMC is considered as the optimization algorithm for PDN. Besides, genetic algorithm (GA) is also introduced in this paper as a classical intelligent optimization algorithm.

2.2.1. Markov Chain Monte Carlo (MCMC). MCMC has been firstly introduced into PDN calculation by [15], where the detailed theory of MCMC can also be found, and the main processes of MCMC in parameter identification are given as follows:

(1) Determining the initial distribution of the parameters, and the distribution can be set as a normal distribution or a continuous distribution since the selection of the initial values does not affect the smooth convergence.

(2) Random samples generated through Monte Carlo sampling.

(3) Obtaining the value of loss function (fitness function) by power flow model.

(4) Obtaining the probability of parameters and calculating Markov chain.

(5) Optimal approximate solution of parameters.

(6) Updating parameter distribution by means of Bayesian posterior and updating the joint distribution of parameters.

(7) Repeating steps (2)∼(6) until the maximum value of iterations is reached or the convergence conditions are met.

2.2.2. Genetic Algorithm (GA). Stochastic optimization techniques such as GA are gaining increasing popularity in various fields. GA is especially beneficial to the complex search space with many local optimal solutions compared with conventional methods to find the global optimal solution. The main processes of GA can be listed as follows:

(1) Coding of the parameters needed to be identified and randomly generating plenty of chromosomes with different initial codes.

(2) Evaluating each chromosome by the loss function.

(3) Selecting a number of chromosomes randomly by the probability associated values which are calculated by loss function in step (2). These chromosomes are randomly paired and create 2 new chromosomes by randomly assigning to each of the genes of one of the parents. This step is also called reproduction of GA.

(4) Randomly selecting some chromosomes and flips their codes; this step is also called mutation of GA.

Repeat steps (2)∼(4) until a criterion is reached.

2.3. The New Definitions of Loss Function

2.3.1. The Statistical Properties in PDN Parameters. Instead of identifying parameters to PDN by optimization algorithms, the static parameters of the standard 10 kV feeder network are used to calculate the standard voltage value per unit for the voltage data, which is signed as \(U_{std}\). The difference between \(U_{std}\) and \(U_c\) can be seen as a baseline result. The static parameters of the standard 10 kV network are shown in Table 1.

With the help of the parameters in Table 1, the calculated values of and \(U_c\) and \(U_{std}\) are displayed in Figure 3.

The values of \(U_c\) and \(U_{std}\) have a large discrepancy in Figure 3, and we can also find that \(U_c\) and \(U_{std}\) have a similar tendency with the series of time period. The scatter plots of \(U_c\) and \(U_{std}\) are displayed in Figure 4.

The distributions of \(U_c\) and \(U_{std}\) are obviously linear. If the values of \(U_{std}\) are regarded as dependent variable, and \(U_c\) is seen as dependent variable, the linear regression can be applied to these two variables, and the statistical results are summarized as follows.

2.3.2. The Definitions of Loss Function Based on Linear Property. According to the results in Table 2, there is a significant linear relationship between \(U_c\) and \(U_{std}\), and it verifies the correctness of the model for power flow calculation in Section 2.1. The reason why the values of \(U_c\) and \(U_{std}\) have a large discrepancy might be the assumption of three-phase balance, and the error of real-time measurement can also affect the calculation of \(U_c\). Therefore, directly evaluating the performance of optimization algorithm and defining loss function by residuals-based metrics are unsuitable and unreasonable. In this section, four metrics based on linear regression are proposed as loss functions to optimize algorithms, which are listed as follows:

\[
R_1 = \frac{n \sum x_i y_i - \left( \sum x_i \right) \left( \sum y_i \right)}{\sqrt{n \sum x_i^2 - \left( \sum x_i \right)^2} \sqrt{n \sum y_i^2 - \left( \sum y_i \right)^2}}
\]

(10)

\[
R_2 = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}
\]

(11)

\[
AIC = 2k - 2ln(L),
\]

(12)

\[
BIC = kln(n) - 2ln(L).
\]

(13)

\(R_1\) and \(R_2\) are Person’s correlation coefficient and Spearman’s correlation coefficient, where \(n\) denotes the number of samples. The correlation coefficient ranges from −1 to 1. An
the regression slope: a value of +1 implies that all data points lie on a line for which $y$ increases as $x$ increases and vice versa for −1. AIC indicates Akaike information criterion, $In$ and $L$ are natural logarithm and likelihood function, and $k$ is the number of variables. BIC is denoted as Bayesian information criterion, which considers the influence of sample number.

In this study, the values of voltage calculated by MCMC and GA are denoted as $U_{cal}$, which can be seen as independent variable like $U_{std}$. Instead of regarding the residuals between $U_c$ and $U_{cal}$ as loss function directly, in this study, the new loss functions are based on the degree of linear property. Therefore, $R_1$, $R_2$, $AIC$, and $BIC$ are used to build loss functions for MCMC and GA. It should be noted that the larger absolute values of $R_1$ and $R_2$ mean the linear relationship are closer, and the smaller values of AIC and BIC indicate the better fitted result of linear regression.

3. Experiment

3.1. The Descriptions of Experimental Data. In this paper, 1499 samples in the raw dataset were collected by SCADA with the sampling period of 15 minutes. The three-phase first section voltages on the high-voltage side (denoted as $U_{a}$, $U_{b}$, and $U_{c}$) are shown in Figure 5, and the corresponding low-voltage sides of them (denoted as $u_{a}$, $u_{b}$ and $u_{c}$) are displayed in Figure 6.

It can be found in Figures 5 and 6 that the high-voltage sides in the dataset are closed to the three-phase balance, and this dataset satisfied the requirements of the equations in Section 2.1. In addition, the active power (denoted as $P_{a}$, $P_{b}$, and $P_{c}$) and reactive power (denoted as $Q_{a}$, $Q_{b}$, and $Q_{c}$) of three-phase on low-voltage side are shown in Figures 7 and 8, respectively.

The trend of changes of active power and reactive power on low-voltage side is consistent in Figures 7 and 8, and it indicates that the samples in this dataset are stable and can be used to perform parameter identification.

3.2. Evaluation and Calculation. In this paper, three metrics, mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE), are used to evaluate the performance of parameter identification and listed as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|,$$  
(14)

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2},$$  
(15)

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right|.$$  
(16)

The true values and calculating values of voltage per unit in C bus (denoted as $U_c$ and $U_{cal}$) are used to calculate the metrics in (14)–(16). Instead of directly calculating these metrics, the linear regression should be applied in this paper, and the values of $U_c$ and $U_{cal}$ are regarded as dependent

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Abbreviation</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line resistance</td>
<td>$R_{cd}$</td>
<td>0.1263</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Line reactance</td>
<td>$X_{cd}$</td>
<td>0.1665</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Transformer resistance</td>
<td>$R_d$</td>
<td>2.825</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Transformer reactance</td>
<td>$X_d$</td>
<td>10.000</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Transformer conductance</td>
<td>$G_d$</td>
<td>5.7e-5</td>
<td>S</td>
</tr>
<tr>
<td>Transformer electrical</td>
<td>$B_d$</td>
<td>3.2e-4</td>
<td>S</td>
</tr>
</tbody>
</table>

Table 2: The statistical results of linear regression.

<table>
<thead>
<tr>
<th>Slope</th>
<th>Bias</th>
<th>$F$-statistic</th>
<th>$P$ value</th>
<th>$R^2$</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01</td>
<td>0.52</td>
<td>2.27e4</td>
<td>0.00</td>
<td>0.969</td>
<td>1.023e4</td>
<td>1.024e4</td>
</tr>
</tbody>
</table>

$^1$Person’s correlation coefficient.

$\hat{y}$ absolute value of 1 implies that a linear equation describes the relationship between $x$ and $y$ perfectly, with all data points lying on a line. The correlation sign is determined by

Figure 3: The values of $U_c$ and $U_{std}$.

Figure 4: The scatter plot of $U_c$ and $U_{std}$.
variable and independent variable, respectively. The output values of linear regression are signed as $U^*_\text{cal}$, and the final evaluations of parameter identification are gained between $U_c$ and $U^*_\text{cal}$ as

$$U^*_\text{cal} = aU^*_\text{cal} + b,$$

(17)

where $a$ and $b$ are denoted as slope and bias of linear regression, and they can also be optimized like other parameters by MCMC and GA. In the following discussion, the parameters of linear regression optimized by MCMC and GA are signed as MCMC-LR and GA-LR, respectively. The upper bounds and lower bounds of MCMC and GA should be determined before parameter identification, and they are listed in Table 3.

To avoid the impact of randomness on GA and MCMC, each method is repeated 25 times to guarantee the correctness and stability of results.

4. Result and Discussion

MCMC and GA with residual-based loss function are applied for parameter identification firstly, the residual-based loss function refers to [22]

$$\sum_{i=1}^{n} (U_{ci} - U^*_{\text{cal},i})^2,$$

(18)

where $U_{ci}$ and $U^*_{\text{cal},i}$ are the true value and calculating value of voltage for the $i$ sample. The comparisons between baseline algorithm and the algorithms proposed in this work on three evaluation functions are listed in Table 4.

The results in Table 4 show that the optimization algorithm can significantly improve the performances of parameter identification of PDN, and the three metrics’ values have a larger decrease than baseline parameters. The
comparisons between $U_c$ and $U_{cal}$ by MCMC and GA algorithms are exhibited in Figures 9 and 10, respectively.

The results in Figures 9 and 10 show that $U_c$ and $U_{cal}$ have a relatively close linear relationship, and the linear regression results are obtained in Table 5 with the values of $U_c$ and $U_{cal}$. Compared with the baseline results gained by the static parameter in Section 2.3, the results in Table 5 are worse since the optimal objective is decreasing residuals directly.

To evaluate the performances by the method in Section 3.2, linear regression will be applied for $U_c$ and $U_{cal}$ firstly and the results can be found in Table 6 by MCMC and GA with linear regression of $U_c$ and $U_{cal}$.

The results in Table 6 indicate the parameter identification performance of static parameter is better than MCMC and GA from the linear regression perspective, and these results show that the loss function based on linear regression is more suitable and reasonable for PDN calculation. The parameter identification results based on the loss functions in Section 2.3 are listed in Table 7.

There are two main points that can be found in Table 7; the first point is that the parameter identification performances based on the proposed loss functions are all better than the residual-based ones, which show the effectiveness and rationality of the loss function’s definition in this paper. Taking the results of MCMC-R1 as an example, the line plot and scatter plot between $U_c$ and $U_{cal}$ are shown in Figure 11. The line profiles of $U_c$ and $U_{cal}$ are almost overlapping, and their degree of linear relationship has a large improvement. The second point is that the performances of $R_1$, $R_2$, AIC, and BIC do not have significant differences, and this result reflects that the metrics selection of linear regression is not top priority; what matters is that the loss function needs to be defined by these metrics.

<table>
<thead>
<tr>
<th>Parameter name</th>
<th>Abbreviation</th>
<th>Upper bound</th>
<th>Lower bound</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line resistance</td>
<td>$R_{cd}$</td>
<td>0.5</td>
<td>0.005</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Line reactance</td>
<td>$X_{cd}$</td>
<td>0.5</td>
<td>0.005</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Transformer resistance</td>
<td>$X_d$</td>
<td>20</td>
<td>5</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Transformer reactance</td>
<td>$R_d$</td>
<td>10</td>
<td>0.8</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Transformer conductance</td>
<td>$G_d$</td>
<td>$8e-6$</td>
<td>$4e-6$</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Transformer electrical susceptance</td>
<td>$B_d$</td>
<td>$8e-5$</td>
<td>$2e-5$</td>
<td>$\Omega$/km</td>
</tr>
<tr>
<td>Slope of linear regression</td>
<td>$a$</td>
<td>-5</td>
<td>5</td>
<td>*</td>
</tr>
<tr>
<td>Bias of linear regression</td>
<td>$b$</td>
<td>500</td>
<td>1000</td>
<td>*</td>
</tr>
</tbody>
</table>

65.6009 ± 0.3562 indicates that the mean of MAE is 65.6009 with the standard deviation of 0.3562.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>90.1969</td>
<td>90.4946</td>
<td>1.4740</td>
</tr>
<tr>
<td>MCMC</td>
<td>65.6009 ± 0.3562</td>
<td>67.1831 ± 0.3202</td>
<td>1.0719 ± 0.0058</td>
</tr>
<tr>
<td>GA</td>
<td>63.8382 ± 0.3479</td>
<td>65.4615 ± 0.3198</td>
<td>1.0431 ± 0.0057</td>
</tr>
</tbody>
</table>

Figure 9: The line plot (a) and scatter plot (b) of $U_c$ and $U_{cal}$ by MCMC.
GA-LR and MCMC-LR contain the parameters of linear regression and optimize them with other PDN’s parameters simultaneously. However, their performances are worse than other methods shown in Table 7. The main reason is that PDN’s parameters have impact on linear regression in optimization process; therefore, applying linear regression after PDN’s calculation is a better choice in practice. The bar charts with error bars of each method are shown in Figure 12. The upper and lower bounds of error bar are the mean values of metrics (such as MAE), add or subtract standard variance obtained by 25 repeated experimental times, and it can be found that the methods proposed in this paper are robust with lower values of standard variance even if MCMC and GA present random-like properties.

Table 5: The statistical results of linear regression for MCMC and GA.

<table>
<thead>
<tr>
<th>Method</th>
<th>Slope</th>
<th>Bias</th>
<th>F-statistic</th>
<th>P value</th>
<th>R</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCMC</td>
<td>0.866</td>
<td>877.21</td>
<td>5191</td>
<td>0.00</td>
<td>0.881</td>
<td>1.216e4</td>
<td>1.217e4</td>
</tr>
<tr>
<td>GA</td>
<td>0.869</td>
<td>856.21</td>
<td>5303</td>
<td>0.00</td>
<td>0.883</td>
<td>1.214e4</td>
<td>1.215e4</td>
</tr>
</tbody>
</table>

Table 6: The results of parameter identification.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>5.8043</td>
<td>7.3323</td>
<td>0.0949</td>
</tr>
<tr>
<td>MCMC</td>
<td>11.3743 ± 0.0733</td>
<td>13.9676 ± 0.0886</td>
<td>0.1859 ± 0.0012</td>
</tr>
<tr>
<td>GA</td>
<td>11.2988 ± 0.0956</td>
<td>13.8778 ± 0.1146</td>
<td>0.1847 ± 0.0016</td>
</tr>
</tbody>
</table>

Table 7: The results of parameter identification by MCMC and GA based on the proposed loss function.

<table>
<thead>
<tr>
<th>Method</th>
<th>MAE</th>
<th>RMSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCMC</td>
<td>11.3743 ± 0.0733</td>
<td>13.9676 ± 0.0886</td>
<td>0.1859 ± 0.0012</td>
</tr>
<tr>
<td>MCMC-LR</td>
<td>33.1434 ± 27.6397</td>
<td>34.9127 ± 27.0669</td>
<td>0.5417 ± 0.4518</td>
</tr>
<tr>
<td>MCMC-R1</td>
<td>5.4171 ± 0.0056</td>
<td>6.8198 ± 0.0061</td>
<td>0.0885 ± 0.0001</td>
</tr>
<tr>
<td>MCMC-R2</td>
<td>5.4254 ± 0.0106</td>
<td>6.8254 ± 0.0090</td>
<td>0.0887 ± 0.0002</td>
</tr>
<tr>
<td>MCMC-AIC</td>
<td>5.4174 ± 0.0053</td>
<td>6.8184 ± 0.0037</td>
<td>0.0885 ± 0.0001</td>
</tr>
<tr>
<td>MCMC-BIC</td>
<td>5.4196 ± 0.0079</td>
<td>6.8214 ± 0.0078</td>
<td>0.0886 ± 0.0001</td>
</tr>
<tr>
<td>MCMC-R1</td>
<td>5.4171 ± 0.0056</td>
<td>6.8198 ± 0.0061</td>
<td>0.0885 ± 0.0001</td>
</tr>
<tr>
<td>GA</td>
<td>11.2988 ± 0.0956</td>
<td>13.8778 ± 0.1146</td>
<td>0.1847 ± 0.0016</td>
</tr>
<tr>
<td>GA-LR</td>
<td>7.1978 ± 0.7169</td>
<td>8.9569 ± 0.8580</td>
<td>0.1177 ± 0.0117</td>
</tr>
<tr>
<td>GA-R1</td>
<td>5.4135 ± 0.0031</td>
<td>6.8144 ± 0.0018</td>
<td>0.0885 ± 0.0001</td>
</tr>
<tr>
<td>GA-R2</td>
<td>5.4237 ± 0.0063</td>
<td>6.8206 ± 0.0054</td>
<td>0.0886 ± 0.0001</td>
</tr>
<tr>
<td>GA-AIC</td>
<td>5.4125 ± 0.0029</td>
<td>6.8137 ± 0.0024</td>
<td>0.0885 ± 0.0000</td>
</tr>
<tr>
<td>GA-BIC</td>
<td>5.4129 ± 0.0034</td>
<td>6.8137 ± 0.0027</td>
<td>0.0885 ± 0.0001</td>
</tr>
<tr>
<td>GA-R1</td>
<td>5.4135 ± 0.0031</td>
<td>6.8144 ± 0.0018</td>
<td>0.0885 ± 0.0001</td>
</tr>
</tbody>
</table>

Figure 10: The line plot (a) and scatter plot (b) of $U_c$ and $U_{cal}$ by GA.
Figure 11: The line plot (a) and scatter plot (b) of $U_c$ and $U_{cal}$ by MCMC-R1.

Figure 12: Continued.
The results of MCMC-LR are neglected in Figure 12 since their values are too large to exhibit appropriately in one figure with others.

5. Conclusion

Parameter identification plays a key role in PDN calculation and analysis; therefore, some methods have been proposed to improve the accuracy of parameters in PDN, such as LS, MCMC, and sequential model-based global optimization. However, these proposed methods pay more attention to the residuals between the true values and calculating values of voltage. In this paper, the linear relationship between these two values has been revealed firstly, which indicates that calculating metrics based on residual is not suitable in PDN, and the four new loss functions based on linear regression have been proposed to combine with optimization algorithms, which are Pearson’s correlation coefficient, Spearman’s correlation coefficient, AIC, and BIC, respectively. The averaged values in MAE and RMSE evaluation functions in parameter identification by MCMC and GA with the residual-based loss are about 65, which is much lower than 90 by traditional algorithms, and the value in MAPE evaluation function is about 1.05, which is also much lower than 1.47 by baseline algorithms. The results show that the accuracy and stability of parameter identification have a significant improvement when the loss function proposed in this paper is combined with GA and MCMC, and these loss functions can also be used in other optimization algorithms to identify the PDN’s parameters. In addition, it should be noted that the heuristic algorithms utilized in this work strongly depended on the initialization, which may lead to the local optimization. And the optimization speed for parameter identification should also be improved in the future works.

Data Availability

The data used to support the findings of this study are currently under embargo while the research findings are commercialized. The data used to support the findings of the study can be obtained from the corresponding author, 6/12 months after publication of this article, upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Nanjing Institute of Technology Scientific Research Start-up Fund for High-level Introduced Talents, Grant no. YKJ202046, and the State Grid Jiangsu Electric Power Co., Ltd., Grant no. J2020097.

References

International Transactions on Electrical Energy Systems


