Distribution Network State Estimation Based on Historical Data Model

Xiaoping Yang¹, Min Wei²

¹Institute of Water Resource and Hydro-electric Engineering, Xi’an University of Technology, Xi’an 710048, Shaanxi, China
²Corresponding author’s e-mail: 2423256013@qq.com

Abstract. Distribution network state estimation provides an important support for the analysis of distribution network operation status. Considering the problems of fewer measurement devices and lower redundancy in distribution network, which make it difficult to ensure the observability of the system and the uncertainty of distribution network load, this paper proposes a distribution network state estimation method based on historical data modeling. This method is to analyze and model the historical data in distribution network, and obtain a set of high precision state variables as pseudo-measurements. Expanding measurement information of distribution network. The weighted least squares state estimation and differential evolution intelligent algorithm are used to estimate the distribution network state. The simulation results show that the proposed method can effectively improve the estimation accuracy compared with the traditional state estimation method.

1. Introduction
As the core part of distribution management system, distribution network state estimation provides basic data support for the analysis of distribution network operation. However, at present, the distribution network measurement device configuration is less, measurement redundancy is low, it is difficult to ensure the observability of the system. Therefore, the estimation accuracy cannot meet the requirements. In order to obtain the observability of the whole network without enough real-time measurement devices, the research in this field at home and abroad mainly focuses on how to optimize the measurement configuration or add different types of new pseudo-measurements. Reference [1] discusses how to select some key measurements, which can not only ensure the observability of distribution network, but also reduce equipment input. In reference [2], it is pointed out that the multi-sampling period measurement data of three measurement systems, RTU, PMU and AMI, coexist for medium and long term in distribution network. Based on this situation, a hybrid distribution network state estimation algorithm based on three kinds of mixed measurement data is proposed.

The actual distribution system is not a complete measurement system. The real-time measurement type is mainly composed of a few node voltage, branch power, branch current and some node injection power measurements. The state information of other non-measurement devices is composed of pseudo-measurements. Literature [3] establishes a three-phase distribution network model with multi-type distributed power grid connection, and adds AC extended pseudo-measurements to the distributed power access nodes without real-time measurement devices to simulate real-time measurements. Document [4] analyses the uncertainty of traditional synchronous measurement and the correlation between pseudo-
measurement, and proves its advantages in the application and accuracy of distribution network state estimation.

Distribution network generates a large amount of historical data during its operation, which are often used as data samples in load forecasting models [5-6]. In this paper, a multivariate regression mathematical model of distribution network operation status data at historical time is established. A set of more accurate state information is obtained by solving the model, which is input into the state estimator as pseudo-measurement, thus expanding the measurement information of distribution network. The traditional weighted least squares state estimation and differential evolution algorithm are used to estimate the distribution network state.

2. Analysis of the observability of distribution network based on historical data modeling

2.1. Multivariate Regression Model of Historical Data

In this paper, the state information data of historical time is taken as the independent variable of the model, then the multivariate regression model of historical data can be expressed as follows:

$$s_i(x) = b_0 + b_1x_{i1} + b_2x_{i2} + \cdots + b_kx_{ik} + \epsilon_i$$ (1)

Among them, $b_0$, $b_1$, $b_2$, ..., $b_k$ are constant, it is a parameter of the model. $\epsilon_i$ mutual independence, and $\epsilon_i \sim N(0, \sigma^2)$. The model parameters are solved by least square method. Its essence is to find the smallest sum of squares of errors and find the best function matching of historical data. After calculating the parameters of the model, the historical data model is substituted, and a set of state information obtained as pseudo-measurements is more accurate and comprehensive than traditional pseudo-measurements.

2.2. Model Verification

2.2.1. Resolvable Coefficient. The resolvable coefficient $r^2$ is a non-negative coefficient. The greater the value, the better the goodness of fit between the regression line of historical data samples and the selected historical data samples, which indicates that the historical time operation data have stronger ability to explain and reflect the operation status of distribution network. The resolvable coefficient can be expressed as:

$$r^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$ (2)

In the formula, $RSS$ represents the sum of regression squares; $TSS$ represents the sum of total squares; $ESS$ represents the sum of residual squares;

2.2.2. Significance Test. It is necessary to test the significance of the model to determine whether the linear correlation between historical data and modeling results is significant on the whole. Construction statistics:

$$F_a = \frac{(RSS / k) / [ESS / (n - k - 1)]}{1 - \frac{ESS}{TSS}}$$ (3)

Sample data are substituted into formula (8) and the calculated value is compared with F value table (test F value of 95% confidence level). If $F_a \geq F_{table}$, there is a significant linear relationship between the operation status data of historical time and the operation status of distribution network[7].

2.2.3. IEEE33 Node System History Data Model. In order to ensure the voltage quality, the electric power department should keep abreast of the voltage levels in all parts of the system at all times. However, only a few important nodes in the actual distribution network have installed node voltage measurement devices, while the nodes without measurement devices can only rely on off-line power flow calculation to obtain their operation status information, and the data obtained can not meet the requirements of real-time and accuracy. Therefore, this paper chooses the node voltage amplitude at the historical moment of distribution network as an independent variable to establish a multivariate regression model. The mathematical model of historical data needs to take into account the influence of random fluctuation load in actual distribution network. So the data sample source selected in this paper
is to divide a day into 24 periods to select Load data, and calculate the load flow of each period to get the data samples of historical state information of distribution network at different times. The sampling interval of load data is one hour. Selecting the load data of seven days in a week, the power flow calculation values of 7 x 24 groups at different times are obtained. Then, the Gauss random distribution error is added to the results of power flow calculation. The result is that the information of distribution network operation status at historical time is taken into account when the load changes. For the IEEE33 bus system shown in Figure 1, the node voltage model at historical time is established. The historical time model of node i is:

\[ U_i = \alpha_{i,0} + \alpha_{i,1}U_{i,1} + \alpha_{i,2}U_{i,2} + \cdots + \alpha_{i,k}U_{i,k} \]  

(4)

By substituting the data samples into the historical data model mentioned above, as shown in Figure 2, the comparison between the historical data modeling value and the true value of power flow shows that although there are some errors in the modeling results, the overall trend is consistent with the true value of power flow of the system, which shows that the historical data model established in this paper can reflect the operation status of the system to a certain extent.

![Figure 1. IEEE33 Node System](image1)

![Figure 2. Comparison of Truth Value and Modeling Value](image2)

Figure 1. IEEE33 Node System

Figure 2. Comparison of Truth Value and Modeling Value

The resolvable coefficients of the model \( r^2 = 0.92 \) can be obtained by calculating formula (2). Therefore, the node voltage amplitude at the historical time of the system has a strong ability to explain the operation state of the system. The saliency test value of the model \( F_\alpha = 9.15 \) can be obtained by calculating formula (3). Table lookup is available: \( F_\alpha \geq F_{table} \). It is considered that there is a significant linear correlation between independent variables and dependent variables, At that time, there was no significant linear correlation. Sample data are substituted into formula (3), and the calculated value is compared with the \( F \) value table (test \( F \) value of 95% confidence level). When \( F_\alpha \geq F_{table} \), there is a significant linear relationship between the operation state data of historical time and the operation state of distribution network; otherwise, there is no significant linear correlation.

3. Distribution Network State Estimation Model and Solution

3.1. Distribution Network State Estimation Model

The weighted least squares estimation method aims at minimizing the sum of squares of the residual of the estimated value[8]. Its objective function can be expressed as:

\[ \min J(x) = \left[ z - h(x) \right]^T W \left[ z - h(x) \right] = \sum_{i=1}^{n} W_i \left[ z_i - h_i(x) \right]^2 \]  

(5)

In the formula, \( W \) denotes the weight coefficient matrix of the measurement. Generally, it is the inverse matrix \( R^{-1} \) of the measurement error variance matrix. The weight of the measurement determines the impact of the measurement on the estimation results.
3.2. Differential Evolution Algorithms
Differential evolution algorithm is relatively simple in structure and easy to operate, but its performance is very superior. It has good adaptability and optimization ability for distribution network state estimation problem [8]-[9]. The process of distribution network state estimation using differential evolution algorithm is as follows: Step 1: The power flow calculation of distribution network is used as the real-time measurement value of state estimation; Step 2: Initialize control parameters. Step 3: The initial population is randomly generated and the fitness of each individual is calculated. Step 4: Judging the fitness of each individual in the initial population, i.e. whether the sum of squares of the estimated residuals is the smallest, terminating the evolution if the end condition is satisfied. Step 5: Variation operation, mutation operation of the first individual in the G Generation state variables to obtain new variation vectors. Step 6: Crossover operation, which generates a new generation of intermediate population by crossover operation of parent individual \(x_i(g)\) and mutation vector \(v_i(g+1)\).Step 7: Selection operation, choosing a group of state variables with the smallest sum of squares of estimated residuals to form a new generation of population. Step 8: Evolutionary algebra \(g = g + 1\), go back to step 4.

4. Examples to Verify
Real-time measurement data and pseudo-measurement data constitute the data base of distribution network state estimation. Because the measurement information in the actual system will produce errors in its transmission process, in the simulation, for the nodes with measurement devices, the simulation value of the actual measurement is generated by the power flow calculation results plus the Gauss random distribution error with mean value of 0 and standard deviation of delta, and the weight value of the state estimation is \(1/\delta^2\). The standard deviation of voltage amplitude and branch power is 0.01 and 0.02 respectively. For the nodes without measurement devices, the pseudo-measurements with lower accuracy need to set smaller weights to reduce their impact on the estimation results. The standard deviation is generally 0.4. The historical data modeled by the historical data model in this paper, as a pseudo-measurement with high accuracy, is set to the reciprocal of the error variance of the node voltage modeled value at the historical time above. The measurement configuration of the system is shown in Figure 1, where the voltage amplitude measurement is labeled with triangle, the real line represents the line with branch power measurement, and the dotted line represents the line without branch power measurement.

The following three indicators are used to evaluate the accuracy of state estimation:

- Maximum absolute estimation error: 
  \[ E_m = \max_{i=1}^n |x_i - x_{true}| \]  
  (6)

- Average absolute estimation error: 
  \[ E_a = \frac{1}{n} \sum_{i=1}^n |x_i - x_{true}| \]  
  (7)

- Error deviation rate: 
  \[ E = \frac{1}{n} \sum_{i=1}^n |x_i - x_{true}| / \sum_{i=1}^n x_{true} \times 100\% \]  
  (8)

\(x_i\) represents the state estimation value and \(x_{true}\) represents the true value of power flow.

4.1. Impact of Historical Data Model on Distribution Network State Estimation
In this section, the weighted least squares method is used to validate the data provided by the historical data model above. Figures 3 and 4 show the estimation results of voltage amplitude and phase angle under traditional pseudo-measurements and historical data model pseudo-measurements of IEEE33 bus system. The comparison of the three estimation error indices in the two cases is shown in Table 1.
Figure 3. Voltage amplitude estimation result   Figure 4. Voltage phase angle estimation result

Table 1. Error comparison of two methods

| Method            | $E_m$ | $E_r$ | $E_U$ | $E_\theta$ |
|-------------------|-------|-------|-------|------------|
|                  | $U_{pu}$ | $\theta/\degree$ | $U_{pu}$ | $\theta/\degree$ |
| Traditional       | 0.021  | 0.387 | 0.017 | 0.163       |
| Method            |       |       |       |             |
| Article method    | 0.013  | 0.111 | 0.009 | 0.055       |

From Figure 3, Figure 4 and Table 1, we can see that the estimated values of voltage amplitude and phase angle are more close to the real operation state of the system, and the error is obviously reduced. It is shown that adding historical data model values as pseudo-measurements to the distribution network state estimation can effectively improve the system observability and thus improve the estimation accuracy.

4.2. State Estimation of Differential Evolution Algorithms Based on Historical Data Model

With the increasing complexity of distribution network structure, unbalanced three-phase lines and loads, and large number of distributed power supply access, the traditional weighted least squares state estimation algorithm is more and more difficult to meet the accuracy requirements of distribution network state estimation. Intelligent algorithm has obvious advantages in dealing with this complex optimization and estimation problem. With the addition of historical data modeling values as pseudo-measurements to satisfy the observability of the system, the differential evolution algorithm is used to estimate the distribution network state, and compared with the weighted least squares algorithm. Table 2 shows the comparison of error indices of voltage amplitude and phase angle estimation between the two algorithms under pseudo-measurements based on historical data modeling. The estimation results of the two algorithms are compared as shown in Figure 5. And Figure 6.

Table 2 Error comparison of two algorithms

| Method                     | $E_m$ | $E_r$ | $E_U$ | $E_\theta$ |
|---------------------------|-------|-------|-------|------------|
| Weighted Least Squares    |       |       |       |             |
| Differential Evolution    | 0.013 | 0.111 | 0.009 | 0.055       |
| Algorithms                |       |       |       |             |
| Article method            | 0.008 | 0.038 | 0.005 | 0.017       |
Figure 5. Voltage amplitude estimation result
Figure 6. Voltage phase angle estimation result

From the comparison results of Figure 5, Figure 6 and Table 2, it can be seen that the state estimation error of the differential evolution algorithm is smaller and closer to the real state of the system.

5. Conclusion

Because of the shortage of measuring devices in distribution network, it is difficult to ensure the observability of the system. In this paper, a distribution network state estimation method based on historical data modeling is adopted. Using the modeled value of node voltage at historical time as pseudo-measurement, the node information without measuring device can be supplemented. On the one hand, the measurement redundancy is improved. On the other hand, the pseudo-measurements with higher accuracy are provided for the distribution network state estimation. Thus, the estimation accuracy can be improved. With the addition of historical data modeling values as pseudo-measurements, the weighted least squares algorithm and differential evolution algorithm are used to estimate the state of distribution network respectively. Compared with the two algorithms, the estimation accuracy of differential evolution algorithm is better.

References
[1] Gedandan, Zou Yuzhi. Observability Analysis of Distribution Network Based on Key Measurement Identification [J]. Electrical Technology, 2018 (Phase 5).
[2] Wang Shaofang, Liu Guangyi, Huang Renle, Qin Shuai. Active Distribution Network State Estimation Method in Multi-sampling Period Mixed Measurement Environment [J]. Power System Automation, 2016, (19).
[3] Wei Zhinong, Chen Sheng, Sun Guoqiang, Wang Dan, Sun Yonghui. Distributed three-phase state estimation of active distribution network with multi-type distributed generators [J]. Power system automation, 2015, (9).
[4] Paolo Attilio Pegoraro;Andrea AngioniMarco Pau;Antonello Monti.Bayesian Approach for Distribution System State Estimation With Non-Gaussian Uncertainty Models[J].IEEE Transactions on Instrumentation and Measurement. 2017: No.66.
[5] Zhang Bingyu. Short-term power load forecasting based on data mining technology [J]. Big Power Data, 2017, (10).
[6] Cai Guowei, Du Yi, Li Chunshan, Gu Xiaoguang, Li You. Medium and long term daily load curve based on support vector machine [J]. Power grid technology, 2006, (23).
[7] Li Qiumin. Mathematical model of probability statistical halo [M]. Beijing: Science Press, 2014.8.
[8] Wu Zaijun, Xu Junjun, Yu Xinghuo, Dou Xiaobo, Gu Wei. Review of Active Distribution Network State Estimation Technology [J]. Power System Automation, 2017 (Phase 13).
[9] Xu Wankai. Optimal Power Flow Computing Based on Differential Evolution Algorithms [D]. North China Electric Power University, 2016.