Yuhui HUANG[1,*], Dong LIU[1], Weiwei XU[1,2], Wenpeng YU[1]

Key Laboratory of Ministry of Education of Power transmission and power conversion control School of Electrical (1)
Hangzhou Power company (2)

Distribution Networks State Estimation Based on Topological Contraction and Quality Tag

***Abstract**. By processing a set of raw measurement data, it provides a real-time system state solution which is the basis of the advanced applications. Currently, the incomplete and poor quality of measurement data is the key factor to impacting state estimation of distribution network. With the help of decision tree theory, the bad data identification method which combines the historical data and remote data from IDP (Integrated Data Platform) and a data processing method is proposed. Algorithm uses the laws of the distribution network data logic to determine the logic of the distribution network. With the help of the theoretical framework of decision tree we can establish the data quality assessment system and fix those bad data so as to increase the quality of the input data. On the basis of data evaluation and fixing, by collecting grid-based data of distribution network and using topology analysis and contracting technology, we can contract the actual network to the state estimation calculation network which meets the observation demand. Quality labels are used to modify the weight of the least squares state estimation algorithm to improve the accuracy of state estimation. Based on the status of data in city distribution network, a proper measurement data detecting and fixing method as well as topology contracting method is proposed. Two real cases of certain central power supply area in Shanghai have verified the effectiveness and feasibility of the method.*

***Streszczenie.** W artykule zaproponowano metody badania stanu sieci dystrybucyjnej w czasie rzeczywistym. Metodę zilustrowano na przykładzie dwóch obszarów sieci zasilającej w Szanghaju. (**Określanie stanu sieci dystrybucyjnej na podstawie topologicznej kontrakcji i tagów jakości**)*

Keywords: Distribution Network; State Estimation; Topological Contraction; Decision Tree; Quality Tag
Słowa kluczowe: sieć zasilająca, jakość energii.

1 Introduction

State estimation[1](SE) plays an important role in modern power energy management system. By processing a set of raw measurement data, it provides a real-time system state solution which is the basis of the advanced applications for system security monitoring and control. State estimation was first studies in 1970s. Huge effects were devoted to state estimation on the basis of the software and hardware conditions at past decades, and significant advances on transmission networks were made, especially those on state estimation criteria. [2][3].

State estimator has to consider certain issues such as detection and identification of bad data, identification of topology data and system observation. Static state estimator identifies bad data according to the spatial connections between different measurements at constant sampling interval. Dynamic state estimator identifies abnormal events according to the real time measurements of two-dimensional state estimator based on both time and space. However, there are still several issues need to be addressed. Nowadays, the monitoring system in distribution networks has been greatly improved with the development of Distribution Automation (DA). As advanced applications are to be utilized in distribution networks, it is necessary and possible to employ SE in distribution networks.

In the past, due to lack of data acquisition and monitoring facilities, state estimation for distribution networks is transformed into a series of power flow matching problems under certain assumptions. Reference [4]~[6] provides ideal solutions for power flow in distribution networks. With the implementation of power system automation, new challenges emerge. The conception of "multi-source data" was used to demonstrate the current data status in transmission networks and was subsequently employed in the selection of correct data from different data sources according to their quality signs and priority[7][8].

Two types of data are required for distribution network SE: the measurement data and the network data. The measurement data includes remote data and telemetry data. The former is mainly from remote devices of SCADA system for distribution network (DSCADA), the basis for distribution network topology. The later is composed of not only real time measurements from DSCADA, but also historical data from other data acquisition systems and electrical degree data from electricity metering system, which are regarded as redundant data[9][10][11].

Since the data volume is huge, and may contain bad data, it is important that a data pre-process is conducted, including bad data detection and identification, and data repairing[13]. At the same time, innovative technology and theory keep immerging in data processing area, such as expert system, neural network, genetic algorithm, machine learning, et al. which may provide solutions for intelligent algorithms of SE problems. However, due to the large number of network nodes, data after pre-process can hardly meet the demand of SE. Therefore, we need to conduct observability analysis to calculation network[14].

Masurement configuration is uesd to calculate the range of grid state in observability analysis. The quality of the algorithm will greatly effect the operating performance of SE[15].

The primary goal is to decide whether the system can be observed. If not, there are two general treatment abroad and domestic:

1) Perform SE only to the observable area.

2) By determining the system observable island and adding least additional measurement, we can make those unobservable parts observable.

A three-phase SE algorithm suitable for distribution network is proposed in paper [16], simple analysis is also made in this paper. The measurement of current can be used as an estimator so as to extends the range of observable islands[17]. In considering the distribution network topology, using the system observability as the objective function, use intelligent algorithms to optimize the measurement point distribution[18]. In [19], an algorithm is proposed to find observable island based on the measurement data of the calculating grid. Despite the rapid development of the monitoring ability, the proportion of the distribution network or power supply islands which can strictly meet the demand of observability is still very low, thus making the SE of distribution network not practically used.

This paper proposed a new methodology for distribution network SE based on the current situation. In order to achieve the object of a practical solution, the work considered practical engineering problems. And the decision tree theory was adopted to classify the data. By transforming the actual network to observable network using topology contraction, the distribution SE can meet the demand in practical use.

2 Distribution networks data process based on quality tag

Energy acquisition system provides the database for a set of advanced functions for monitoring and control in distribution networks. The integrity, accuracy and sample frequency of the data has an impact on the efficiency of the advanced functions.

The current distribution network is also faced with the status of multiple-source data. Strategies have to be considered to process the raw data, in order to guarantee the quality of input data for advanced applications, and the efficiency of the monitoring and management of distribution networks.

2.1 Quality tag

Thanks to the wide application of data acquisition systems in distribution networks, it is possible to collect real time data in various ways. This offers a number of advantages, but still suffers from drawbacks such as the difference of accuracy of different systems, which may influence the advanced application. In this way, it is essential to judge the reliability of the data acquisition system and select high quality data for subsequent applications. A quality tag is attached to each data source, representing the current data quality.

The quality tag, denoted by $Q(i)$, is ranging from 0 to 1. Bigger $Q(i)$ is, more reliable the referring data is. $Q_U(i)$, $Q_I(i)$ and $Q_P(i)$ represents the quality of node voltage, current and power respectively.

The initiate value of $Q(i)$ is 1. $Q(i)$ is obtained from the following equation:

(1) $$Q(i) = 1-(1-flag)*W$$

Flag is referring to each of the 6 rules in the next clause. If the rule is satisfied, then flag=1; else, flag=0; W is the corresponding weight.

2.2 Bad data detection basis

For bad data detection in radial distribution networks, the following rules are used:

- ✓ Rule 1 : the voltage should be in the standard range
- ✓ Rule 2: the voltage gradually reduced along the feeder lines[11]
- ✓ Rule 3: the KCL rule[11]:

(2) $$I_m > \sum_{n \in M} I_n$$

where M represents the downstream neighbor node set of m.

- ✓ Rule 4: based on the particular research object of this paper, the degrees of ammeters and the active power satisfy the following equation:

(3) $$PM_i = [(P_{i-1}+P_i)/2 \pm \xi]/(3.6\times10^6)$$

- ✓ Rule 5: the classic measurement sudden-change detection:

(4) $$\frac{|z_{i-1}-z_i|}{z_i} < \zeta$$

- ✓ Rule 6: verification with history data:

(5) $$|z_{di}-z_{da}| < \omega$$

z_{da} is the average value of the measurements of the latest 7 days at the same sampling interval.

The quality tags derived from the bad data detection module are used as input for data repairing module.

2.3 Decision tree algorithm in bad data detection

The most influential algorithm in decision tree learning is ID3, proposed by Quinlan in 1986[10]. The basic principle of ID3 is to recursively divide the domain which the input variables cross, in order to categorize the variables. Probability function is adopted to represent the uncertainty; entropy is to represent the impurity; information gain is to represent the amount of information which is provided when event A occurs.

Set S as a group of training sample which consists of two types of data A and data B, the ratio of which are represented by p_i^+ and p_i^-. Then the entropy E will be calculated as follows:

(6) $$E(S_i) = -p_i^+ \log_2 p_i^+ - p_i^- \log_2 p_i^-$$

Set A as an attribute of the data, V is a value set of A. The decrease of entropy is represented via information gain G according to the categorization of value A:

(7) $$G(S,A) = E(S) - \sum_{v \in V(A)} \frac{S_v}{S} E(S_v)$$

The optimal decision is made at the maximum of G. The same procedure is then repeated until each terminal node in the decision tree is completely generated.

Taking the categorization of bad data into consideration; it should be based on the six rules previously mentioned. The six corresponding attributes are: per-unit rule, potential drop, KCL, power check, break-variable detection and historical data check.

Ideally, if the data satisfy all the six rules, it's definitely of high quality, vice versa. However, the existence of bad data makes the situation complicated. Since some of the rules are based on comparison of different data, the bad data will have a negative impact on the result. To some extent, the sample data reflects the relationship between data attribute and data quality within certain distribution network.

Table 1 Subset of sample data for Decision Tree

No	QT	rule1	rule2	rule3	rule4	rule5	rule6
1	0.2	N	N	U	N	Y	Y
2	0.2	N	N	U	U	N	Y
3	0.6	Y	N	U	U	Y	N
4	0.2	U	U	N	U	N	N
5	1	Y	N	U	U	Y	Y
6	1	Y	Y	U	U	Y	Y
7	0.2	U	U	N	N	Y	Y
8	0.6	Y	Y	U	N	Y	Y
9	1	U	U	Y	Y	Y	Y
10	1	U	U	Y	Y	N	N
11	0.2	N	N	U	Y	Y	N
12	0.6	U	U	Y	Y	Y	N
13	0.6	N	N	U	U	Y	Y
14	0.2	N	N	U	N	N	N

*Y=satisfied; N=not satisfied; U=unknown

ID3 generates minimum decision tree according to the attributes and classification of the actual sample data. This decision tree accurately categorizes the acquainted mass data with its shortest path within distribution network. Table 1 lists a subset of sample data for decision tree.

The decision tree in Figure 1 shows the categories in Table 1. This tree is able to make accurate data categorization.

```
attribute#1 = 1:
|   attribute#2 = 1: 1 (9.0)
|   attribute#2 = 0: 0 (36.0/5.0)
attribute#1 = 0:
|   attribute#2 = 1:
|   |   attribute#5 = 0: 0 (14.0)
```

```
|   |   attribute#5 = 1:
|   |   |   attribute#3 = 1: 1 (8.0/3.0)
|   |   |   attribute#3 = 0: 0 (4.0/1.0)
|   attribute#2 = 0:
|   |   attribute#5 = 1: 1 (31.0/4.0)
|   |   attribute#5 = 0:
|   |   |   attribute#3 = 1:
|   |   |   |   attribute#6 = 0: 1 (6.0)
|   |   |   |   attribute#6 = 1:
|   |   |   |   |   attribute#4 = 1: 1 (3.0)
|   |   |   |   |   attribute#4 = 0: 0 (3.0/1.0)
|   |   |   attribute#3 = 0:
|   |   |   |   attribute#6 = 1: 0 (3.0/1.0)
|   attribute#2 = 1: 1 (9.0/1.3)
    |   |   |   |   attribute#6 = 0: 1 (7.0/3.0)
```

Fig 1 Decision tree grow from the sample data

In the decision tree, each internal node is associated to one of the attributes (e.g. the per-unit, potential drop etc). Every possible value of this attribute is mapped to a tree branch. A leaf represents a class (e.g. low quality, medium quality). Data of unknown type is categorized according to the traverse of the tree: test the attribute value of each data and take the appropriate branch. The category of the object is determined by repeating the procedures till a leaf node is reached, where the categorization of the objected is finished .

2.4 Data repairing based on quality tag

Set the lower bound of quality tag as χ_{bad} . If the measured quality tag is lower than χ_{bad} , the data is suspicious, and the data repairing process continues. Set the upper bound of quality tag as χ_{good} . If the measured quality tag is higher than χ_{good} , the data is reliable, and the data repairing process breaks.

The repairing equations are as follows:

If data K is of node s is to be repaired, assuming it to be voltage, then:

(8) $$U_k^{re} = U_k^{mea} \times Qtag + \frac{U_j^{mea} + \max_{n\in k} U_n^{mea}}{2} \times (1 - Qtag)$$

Where j is the parent node of k, J is the child node set of j, K is the child node (downstream node) set of k. And U_k^{re} is the modified value of U_k^{mea} .Quality tag of repaired U_k^{re} :

(9) $$Qtag(U_k^{re}) = Qtag^2(U_k^{mea}) + [1 - Qtag(U_k^{mea})] * \min\{\frac{Qtag(U_j^{mea}) + Qtag(\max_{n\in k} U_n^{mea})}{2}\}$$

Assuming data K to be current, then:

(10) $$I_k^{re} = I_k^{mea} \times Qtag + \frac{[(I_j^{mea} - \sum_{m\in J, m\neq k} I_m^{mea}) + \sum_{n\in K} I_n^{mea}]}{2} \times (1 - Qtag)$$

Quality tag of repaired U_k^{re} :

(11) $$Qtag(I_k^{re}) = Qtag^2(I_k^{mea}) + [1 - Qtag I_k^{mea})] * \min\{\frac{Qtag(I_j^{mea}) + Qtag(\max_{n\in k} I_n^{mea})}{2}\}$$

Where j is the parent node [11]of k, J is the child node set of j, K is the child node set of k. And I_k^{re} is the modified value of I_k^{mea} .

3 Observability analysis of the distribution networks based on topological contraction techniques

Existing Observability Analysis algorithms can be divided into two types--numerical algorithm and topological algorithm. The latter is popular among power system researchers with its fast calculation speed and the capability to avoid the influence of rounding error, which may be observed in numerical algorithm.

This paper is written on the basis of distribution network analysis combining topological analysis and monitoring point. It proposed the sufficient constraint conditions for Observability Analysis of the distribution network based on CIM model, apply Topological Contraction Algorithm to the topological analysis, and model the actual distribution network to calculating network which will effectively meet the needs of observability.

3.1 SE observability analysis for distribution networks

An n-node power system can be demonstrated by an n-vertex connected graph G_n . Assume that in each power measurement active and reactive power appear in pairs, then there exists a measurement graph G_m , each line in G_m matches with a line in G_n whose active power is measured.

Search in G_m for spanning tree. If the measurement does not appear in the spanning tree more than once and system reference phase and voltage exist, the sufficient and necessary condition for the observability of the system is: rank of the biggest spanning tree is n , i.e. the number of the spanning tree vertexes is n .

Through the topological analysis based on CIM model, each power island forms an associated electrically connected Topological Island (referred to as topological island). Topological Island is an effective object of state estimation independent on whether the equipments outside the topological island are actually in service.

Assuming all the equipments connected with a certain connectivity node CN_i are measured in power, then this connectivity node can be regarded as an observable node. If a specific topological island connectivity nodes set S is recorded, and all of the connectivity nodes in S can be derived observable according to the sufficient and necessary conditions of observability, this topological island must be measurable.

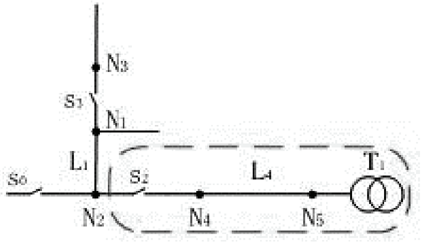
According to Kirchhoff's laws, necessary and sufficient condition for connectivity node a_i to be observable is that $N_i - 1$ equipments connected by terminals out of N_i are observable.

3.2 Networks contraction algorithm based on CIM

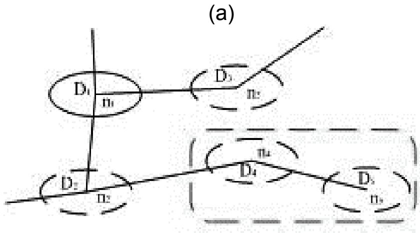
Common measurement in Distribution Network includes: root node voltage, root node injection power, load power, branch power, branch current amplitude, node voltage, etc. Measurements collected on outlet breakers and load transformers are of high integrality and accuracy. Distribution switch acquisition devices are of low popularity. Even on test lines measurement devices are usually installed on inlet and outlet switches, and switches on load branch are usually in loss of measurements.

According to the sufficient and necessary conditions of observability, even in test lines in Shanghai where automation is of a relatively high level, it is still difficult to implement SE in the whole network. To remove the lines, the switches without measurements acquisition as well as numerous branch data, and to maintain the measurement node with higher acquisition accuracy among the giant database is key to the bad data detection and the practicality of SE in distribution network.

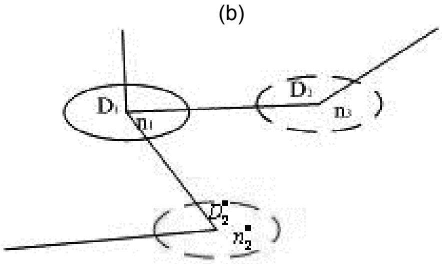
Observability algorithm based on Topo-Contraction is proposed combined with the setting principles of measurement node and unobservable network algorithm at home and abroad. The kernel is to combine the unobservable points/area as equivalent node, and contract the distribution network into a calculation network formed by only observable nodes and equivalent nodes.



(a)



(b)



(c)

Fig 2 Topology Contraction Process

In Fig (a), nodes N_1, N_2, N_3 are observable. Load switch S_2 and line L_4 are in lack of measurement data. Transformer T_1 is available in active and reactive power measurement but lack of measurement of voltage. Obviously, node N_4, N_5 are unobservable, whose branch lines are unacceptable to SE.

If each equipment is regarded as a set to be estimated, then Fig a can be transformed into Fig (b) which shows the calculation model. Set D_4, D_5 are unobservable due to lack of measurement.

Combine the unobservable area D_4, D_5 with the observable area D_2 . Regardless of the line loss, the combined equivalent area D_2^* is still observable.

After the equivalence, the system is contracted from 5 nodes to 3 equivalent nodes. The previous measurements associated with D_4, D_5 are considered redundant.

This Topo-Contraction Algorithm process the electrical topology by combining the unobservable area with the observable area which can provide inputs before the detection and the SE program, thus the convergence of the SE program is ensured by reducing the number of observable nodes. Fig 3 is a simple flow chart of Topo-Contraction.

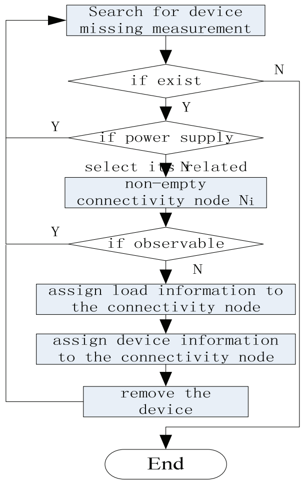


Fig 3 Topo Contraction Process

3.3 Observability analysis based on contracted network

In topologically contracted distribution network (referred to as contracted network), each equivalent node is considered as an observable node. For a specific time t , given a set of measurements Z_t , the whole network status x_t , observation equation is as follow:

(12) $$Z_t = h(x_t) + v_t$$

where $h(x_t)$ is a non-linear network function, v_t is measurement random error which meets the normal distribution $E(v_t \cdot v_t^T) = R$.

Linear iterative equation of state variable correction obtained by weighted least square estimation is:

(13) $$H^T R^{-1} H \Delta x = H^T R^{-1} [Z_t - h(x_t)]$$

where H is an $m \times m$ order Jacobian matrix; coefficient matrix $G = H^T R^{-1} H$ is a n order square matrix; Δx is state variable correction; m is measuring points number; n is network nodes number.

Topological contraction process is essentially a Ward equivalent transformation. Through topological contraction, the whole network is divided in to internal system, external system and boundary system, among which the internal system is to be studied.

Divide the node injection power, node voltage and Node admittance matrix into blocks according to the internal, external and boundary system. The relationship between node voltage and node injection power is:

(14) $$\begin{bmatrix} \dot{V}_E & 0 & 0 \\ 0 & \dot{V}_B & 0 \\ 0 & 0 & \dot{V}_i \end{bmatrix} \begin{bmatrix} Y_{EE} & Y_{EB} & 0 \\ Y_{BE} & Y_{BB} & Y_{Bi} \\ 0 & Y_{iB} & Y_{ii} \end{bmatrix} \begin{bmatrix} \dot{V}_E \\ \dot{V}_B \\ \dot{V}_i \end{bmatrix} = \begin{bmatrix} \dot{S}_E \\ \dot{S}_B \\ \dot{S}_i \end{bmatrix}$$

where subscript E, B and I respectively refer to internal, external and boundary system. In state estimation, leave out the inestimable external system, then the equation is transformed as:

(15) $$\begin{bmatrix} \dot{V}_B & 0 \\ 0 & \dot{V}_i \end{bmatrix} \begin{bmatrix} Y_{BB} + Y_{EQ} & Y_{BI} \\ Y_{IB} & Y_{II} \end{bmatrix} \begin{bmatrix} \dot{V}_B \\ \dot{V}_I \end{bmatrix} = \begin{bmatrix} \dot{S}_B + \Delta \dot{S}_B \\ \dot{S}_I \end{bmatrix}$$

where $\Delta \dot{S}_B = -\dot{V}_B Y_{BE} Y_{EE}^{-1} \dot{V}_E^{-1} S_B$, $Y_{EQ} = -Y_{BE} Y_{EE}^{-1} Y_{EB}$

In the contracted calculation network, set the number of equivalent nodes as n' , and the number of total measurement points related to the nodes as m' . It can be seen from the contraction process that $n' < n, m' < m$. According to the equivalent transformation process, set $v_t^{'}$ as the equivalent error vector, therefore $v_t^{'}$ still meets the normal distribution $N(0, \sigma^2)$.

Equivalent nodes have associated measurement, consequently the measurement Jacobian matrix *H* based on both sides of the nodes is full rank and in fine states. As for *H,* Rank(*H*)=2 N' -1, therefore the system is algebraically observable.

In consideration of topology, the initial network can be regarded as *G*=(*V*,*E*), which consists of *N* nodes and *b* branches. *V* is the set of nodes, and *E* is the set branches. A measurement subgraph *G'*=(*V'*,*E'*) is created from the measurement network, where *V'* is contained in *V*, and *E'* is contained in *E*.Regardless of the internal observability of equivalent nodes, from the external, in subgraph *G'* and graph *G* , *V'* is contained in *V*, and *E'* is contained in *E*, i.e. all points in *G* are contained in *G'*. Therefore, the equivalent system is observable according to the topology observability requirements analysis.

Distribution network contraction techniques based on CIM model is an effective way to solve low proportion of the measurement point, the large amount of actual computation and occupied storage space.Power system state estimation is to estimate the state variables in power system with redundant information. When the information is insufficient, existing measurement reconfiguration is to calculate the network state variable range, which is the observability analysis.

In the observability analysis with topological methods, it is common to add pseudo measurements in the lack-of-measurement points. The addition of pseudo measurements enlarge the observable range of the network but decrease the progress of state estimation.On the basis of specific distribution network observability analysis, the unobservable area, as a set of contracted equipments in Simplified Networks, calls the topological contraction function, where the boundary conditions are set according to the algorithm.

As for distribution network in which bus nodes can be measured, observability requirement is met with the contraction of the unobservable area and the simplification of the Virtual equivalent nodes.

4 The SE processing flow and test results

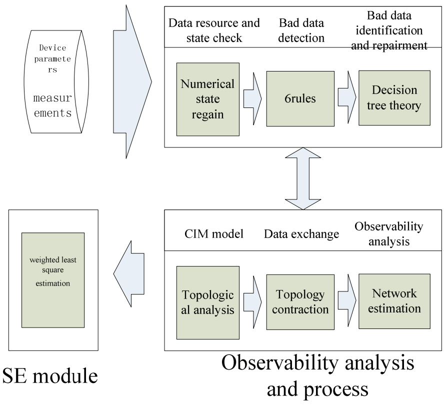


Fig 4 The overall process of quality-tag-based state estimation

Fig 4 illustrates the overall process of quality-tag-based state estimation. Multi-source measurements are integrated into IDP. Bad data detection and identification are based on 6 rules and the decision tree theory. Quality tags are initialed by decision tree and updated in data repairing module. And the modified measurements and their quality tags are use as input for SE. The quality tags are counted as weights in WLS.

Table 2 Comparison of iteration efficiency

	Traditional WLS		WLS based on quality tag	
Iteration number	2	3	2	3
Difference between the last two steps	2.178e-7	3.4854e-8	3.649e-7	1.846e-9
Assist criteria	1.898e-3	3.6967e-5	9.806e-5	2.178e-6

After the bad data processing, nodes without measurements are analyzed to be contracted. The contracted network formed by observability analysis and process module is able to meet the observability requirements, thus is estimable.

This novel method was employed in SE of the test lines. Table 2 compares the results between the traditional WLS and the novel method. Figures in Table 2 show that within the same iteration step, the novel method has higher accuracy. The novel method is convergent with 1.8459e-9, and the traditional one is with 2.17811e-7. Results in Table 3 demonstrate the calculation results of both methods. It is obviously shown through the estimation results of node 8,9,10 and 11 that the SE results from the proposed novel method are more compatibles.

Table 3 State estimation result

Node Number	Node Voltage		Node Phase Angle		Active Power		Reactive Power	
	Traditional WLS	WLS considered quality tag	Traditional WLS	WLS considered quality tag	Traditional WLS	WLS considered quality tag	Traditional WLS	WLS considered quality tag
1	1	0.989655	0	0	0.0036	0.00247577	0.0016	0.00339743
2	0.9974	0.985316	-0.2018	-0.110124	0	-0.00205306	0	-0.003526
3	0.9939	0.984969	-0.1669	-0.123491	0.0003	0.000292524	0.0002	7.80E-05
4	0.9906	0.984326	-0.1383	-0.136848	0.0001	0.00141947	0.0001	0.000365213
5	0.9831	0.979589	-0.076	-0.166476	0.0002	-0.000643804	0.0001	-0.00014974
6	0.9786	0.977546	-0.0405	-0.184306	0.0001	-7.68E-05	0.0001	-1.78E-05
7	0.9748	0.975817	-0.0108	-0.200211	0.0001	0.00204933	0.0001	0.000524463
8	0.9658	0.964809	0.0456	-0.238234	0	-0.00524348	0	0.00466842
9	0.9658	0.9658	0.0456	0	0	0.00280393	0	-0.00527805
10	0.9658	0.9658	0.0456	0	0	0	0	0
11	0.9632	0.963954	0.0623	-0.249601	0	0.00161719	0	0.000416251
12	0.9564	0.957783	0.1053	-0.277046	0.0004	-0.000884639	-0.0003	-0.000163297
13	0.9526	0.955249	0.1664	-0.289496	0.0002	0.00143022	0.0002	0.000392037

14	0.9519	0.954197	0.1667	−0.291064	0	−0.000328107	0	−8.61E−05
15	0.9518	0.954189	0.1667	−0.291465	0.0002	−1.93E−05	0.0001	8.72E−06
16	0.9462	0.946364	0.2732	−0.307467	0	−0.00290929	0	0.0015067
17	0.9462	0.9462	0.2733	0	0	0.000461666	0	−0.00211973
18	0.9417	0.946279	0.3511	−0.309602	0	0.00106878	0	0.000278667
19	0.936	0.945475	0.4635	−0.313044	0	0.0126492	0	0.0105663
20	0.936	0.936	0.4639	0	0	−0.0110501	0	−0.0101688
21	0.9289	0.931448	0.6109	−0.332639	0.0002	−0.0016987	0.0001	−0.000399146
22	0.9278	0.930936	0.6354	−0.334766	0.0004	0.000131591	0.0002	4.94E−05
23	0.9265	0.929374	0.6693	−0.336559	0	−0.000440372	0	−0.00010607
24	0.9262	0.929314	0.6785	−0.336897	0.0003	−6.72E−05	0.0003	−1.03E−05
25	0.9407	0.943091	0.3636	−0.320355	0.0004	−0.000514336	0.0002	−2.71E−05
26	0.9403	0.942909	0.3725	−0.315662	0.0003	−0.000125657	0.0002	−9.49E−05

5 Conclusions

This paper carried out a pre-process of the measurements for distribution network SE based on the current data acquisition systems and the multi-data source situation. The proposed work minimized the effects of the bad data on the SE program, and overcome the disadvantage of evenly distributed bad data within all the data. This method combined the nature of the measurements in the distribution networks into the theory of decision tree, and accelerated the bad data detection speed and improved the accuracy. Therefore the risks of the misjudging bad data are reduced. In the data repairing module, this work improved the quality of the input of the SE by repairing the bad data according to the redundant measurement data, together with the concept of quality tag.

The weighed least square SE based on quality tag was employed in the process of the SE of the testing lines. The results show that this method improves the reliability of the measurements and reduces the effect of bad data, which is promising and practical for distribution automation.

This paper is supported by the National High-tech R&D Program of China (863 Program with issue number: 2012AA050212) and Program for New Century Excellent Talents in University (No.NCET-07-0547).

REFERENCES

[1]. Yu Erkeng. 1985, *Power System State Estimation,* Water Power Press, Beijing, China.

[2]. F. F. Wu, 1990, "Power System State Estimation: A Survey", *Electrical Power & Energy Systems*， Vol. 12， No. 2， April.

[3]. Celik M K, Abur Ali, 1992, "A robust WLAV state estimation using transformations ". *IEEE Transactions on Power Systems*, vol. 7(1)

[4]. Liu Jian, Ma Li, Wei Li, Dong Haipeng, 2004, "Analysis of Large Scale Distribution Networks with Reduced Calculation", *Power System Technology*, vol. 28(8).

[5]. Cai Zhongqin, Guo Zhizhong, 2000, "Newton Load Flow for Radial Distribution Network Based on Upstream Labeling Technique", *Proceedings of the CSEE*，vol. 20(6), 13-16.

[6]. Sun Hongbin, Zhang Boming, Xiang Niande, 1999, "Study on Convergence of Back/Forward Sweep Distribution Power Flow", *Proceedings of the CSEE*, vol. 19(7), 26-29.

[7]. Lu Jiangang , Wen Bojian, Jiang Caiyun, Gu Quan, 2002, "Research of Multi-source Data and Its Quality Control in Energy Management System", *Automation of Electric Power Systems,* vol.26(23):73-75

[8]. Ye Feng, He Hua, Gu Quan, Zhang Gaofeng, 2006, "Bad Data Identification and Correction for Load Forecasting in Energy Management System", *Automation of Electric Power Systems*, vol.30(15):85-88.

[9]. Liu Dong, Ding Zhenhua, Teng Letian, 2004, "Key Technology and Its Development in The Practicability of Distribution Automation Systems", *Automation of Electric Power Systems*, vol. 28(7), 16-19

[10]. George F． Luger, 2004, *Artificial Intelligence: Structures and Strategies for Complex Problem Solving, Fifth Edition* .China Machine Press, Beijing, China.

[11]. Liu Jian, et al,1998, *Analysis and Optimization of Large Scale Distribution Networks*, Electrical Power Press, Beijing, China.

[12]. Li Haiying, Li Yuzeng, Zhao Shaohua. An Energy Acquisition Model for a Distribution Company With Distributed Generation and Interruptible Load Options[J]. Proceedings of the CSEE, 2008, 28(10):88-93. (in Chinese)

[13]. Bagen W, Billinton R. Incorporating well-being considerations in generating systems using energy storage. IEEE Transactions on Energy Conversion 2005; 20(1): 225–230.

[14]. Billinton R, Bai G. Generating capacity adequacy associated with wind energy. IEEE Transactions on Energy Conversion 2004; 19(3): 641–646.

[15]. Wangdee W, Billinton R. Considering load-carrying capability and wind speed correlation of WECS in generation adequacy assessment. IEEE Transactions on Energy Conversion 2006; 21(3): 734–741.

[16]. Billinton R, Wangdee W. Reliability-based transmission reinforcement planning associated with large-scale wind farms. IEEE Transactions on Power Systems 2007; 22(1): 34–41.

[17]. TradeWind Project. Integrating wind: developing Europe's power market for the large-scale integration of wind power, EWEA, Feb. 2009.

[18]. Liao Huaiqing, Liu Dong, Huang Yuhui. A Study on Compatibility of Smart Grid Based on Large scale Energy Storage System[J]. Automation of Electric Power Systems, 2010,34(2):15-19. (in Chinese)

[19]. Huang Yuhui,Liu Dong,Liao Huaiqing. Network Reconfiguration Solution Space Optimization Considering Grid Characteristics[J]. Automation of Electric Power Systems (Accepted).

Authors

Huang Yuhui: corresponding author, born in Weihai, China, in 1983. He received his Bachelor degree from North Electric University, China in 2001 and Master degree from Shanghai Jiao Tong University, China in 2007. Now he is a Doctor Candidate in Electrical Engineering of Shanghai Jiao tong University, China. His fields of interest are Automation of power system.E-mail:hyh8307@sjtu.edu.cn

LIU Dong: born in October 1968, Ph.D.，Researcher, doctor supervisor, be honored with Ministry of Education New Century Excellent Talents. He is now a professor researcher in the School of Electronic, Information and Electrical Engineering at Shanghai Jiao Tong University. His fields of interest are Smart Grid, Automation of Electric Network Management & Computer Information Process, Distribution Automation & Distribution Management System.E-mail: dongliu@sjtu.edu.cn