International Journal of Electrical and Computer Engineering (IJECE)
Vol.1, No.2, December 2011, pp. 85~92
ISSN: 2088-8708

Improving Bad Data Detection in State Estimation of Power Systems

M.Tarafdar Hagh[1], S. M. Mahaei[1], K. Zare[2]
1. Department of Electrical Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran
2. Department of Power Engineering, University of Tabriz, Tabriz, Iran
Correspond author: Tel: +989143114450, +984113300829, Fax: +984113300829
tarafdar@tabrizu.ac.ir

Abstract

In state estimation of power systems, it is possible that measurements include bad data, influencing on state estimations of power system. Several intelligent methods have been proposed to detect bad data which should be trained in various network situations but they are almost impractical because of abound situations of actual network. Some mathematical methods such as Chi-Square Distribution Test, Largest Normalized Residual Test and Hypotheses Testing Identification as the detectors of bad data have been presented, too. Sometimes these mathematical methods are not able to detect bad data. This paper proposes a method which can improve the detection of bad data in mentioned mathematical methods. Case studies have been done with different given errors on measurements of IEEE 14-bus system, and it was shown that this method is effective to improve the bad data detection.

***Keywords**: Bad Data; Chi-Square Distribution; Largest Normalized Residual; Hypotheses Testing Identification*

1. Introduction

In power systems due to the reasons including incorrect installation or servicing, disorders in communication system and auxiliary devices such as current transformers (CTs), potential transformers (PTs) or capacitive voltage transformers (CVTs), measured values can include errors. Measurement errors can affect the state estimations of power systems. Here, the question is that how can distinguish between the error and error-involved measurements?

So far many different methods have been proposed to detect bad data and a faulty measurement which can be classified in mathematical and intelligent methods. Among the mathematical methods such as the method presented in [1, 2] the Chi-Square Distribution Test (CSDT) is utilized to detect presence of bad data. This method can only detect existence of bad data, and cannot detect the faulty measurement. In [3-7], the Largest Normalized Residual Test (rNmax) is used to detect the faulty measurement; this method is one of the conventional methods for identification of bad data. In [8, 9], Hypotheses Testing Identification (HTI) is presented which is rNmax-based method. Among the intelligent methods, one can point to the fuzzy, neural, neuro-fuzzy and genetic methods which have been investigated in [10-13]. Intelligent methods can be divided into two parts: 1. Training-required methods which should be trained in various situation of network and because of abound situations of actual network they are almost impractical, 2. Methods do not require training, have very high computing time and they have not good effectiveness in practice. Consequently, in contrast to the intelligent methods, the mathematical methods are simple and have a relatively reasonable speed. However, mathematical methods fail to identify measurement error when the error percentage is low.

In this paper is proposed a method which increases the amount of error percentage untruly, so that detection of bad data is easier. This method will be tested using CSTD, rNmax and HTI techniques to improve the bad data detection. Obtained results of case studies on IEEE 14-bus system shows that the presented method was very effective in improving bad data detection.

2. State estimation

Consider the well known measurement model:

$$z = h(x) + e \tag{1}$$

Where:

z: measurement vector with size m
x: state vector with size n, with $m > n$
e: error vector
$h(x)$: vector with the non-linear functions relating measurements and states
The error vector e presents zero mean and covariance R_z.

Received Oct 6th, 2011; Revised Nov 20th, 2011; Accepted Nov 26th, 2011

The state vector x is obtained by minimizing the Weighted Least Square (WLS) index defined as:

$$J(x) = [z - h(x)]^T W [z - h(x)] \tag{2}$$

Where, the weighting factor W (diagonal $m \times m$ matrix) is usually the inverse of the covariance matrix of the measurements.

The condition for optimality is that the gradient of $J(x)$ vanished at optimal solution x, i.e:

$$H(x)^T [z - h(x)] = 0 \tag{3}$$

or,

$$x = G(x) H(x)^T W [z - h(x)] \tag{4}$$

Where, $G(x)^{-1} = H(x)^T WH(x)$.

3. Measurement error detection using the Chi-Square Distribution Test

If C_i was an i^{th} random variable of a collection of independent random variables such as S, and having time probability distribution, another random variable U can be defined as follow:

$$U_k = \sum_{i=1}^{k} C_i^2 \tag{5}$$

Thus, U will have C_2 distribution with freedom degrees of k.

$$U_k \sim C_k^2 \tag{6}$$

Values of U_k for $Pr(U_k)=0.05$, $Pr(U_k)=0.01$ and $k=1$ to $k=30$ are calculated and given in the table. For $k> 30$, U will usually be normally distributed.

C_i for power systems is defined as follows:

$$C_i = \frac{Z_i - h_i(\hat{x})}{\sigma_i} \tag{7}$$

$$U = J(\hat{x}) \tag{8}$$

$$k = 2n - 1 - m \tag{9}$$

Hence, to identify the bad data following steps are pursued:

1. Selecting a given probability for square distribution $J(x)$ with freedom degrees of k, for example, 0.05
2. Obtaining the $J_c(x)$ with freedom degrees of k and with the probability of 0.05 from the Chi-square distribution table
3. If the obtained $J(x)$ from the estimation is greater than $J_c(x)$ $(J(x) > J_c(x))$ then there is bad data, otherwise there is no bad data.

The aforementioned method could only detect the presence of bad data and cannot detect the error-involved measurement.

4. Measurement error detection using the Largest Normalized Residual Test

Assuming that the state estimation has been finished and vector x_s obtained, therefore we have:

$$H = \frac{\partial h(\hat{x})}{\partial x}\Big|_{\hat{x}=\hat{x}_s} \quad (10)$$

By making linear (1) around the vector x_s it can be written:

$$z = Hx + e \quad (11)$$

The residual estimate vector can be written as:

$$r = Sz = S[Hx + e] \quad (12)$$

Where,

$$S = I - H\,G^{-1}H^{T}W \quad (13)$$

Once $SHx = 0$,
r is rewritten:

$$r = Se \quad (14)$$

For normalizing residual measurements is used measurement accuracy as follow:

$$r^{n} = (diag(R_r))^{0.5}\,r \quad (15)$$

Where, R_r is obtained from the following equation.

$$R_r = S\,R_z^{-1} \quad (16)$$

Accordingly, the measurement bad data procedure using r^{N}_{max} method will be as follows.

1. State Estimation
2. Forming the sensitivity matrix, S
3. Calculating the measurement residual vector, r
4. Obtaining the normalized residual measurement vector, r^n

   If $r_i^n > C$ (where C is a fixed value, for example 3), z_i has an bad data and if any of the above measurements have not the mentioned condition, algorithm is over,
5. Eliminating the error-involved measurement and return to Step 1

5. Measurement error detection using Hypotheses Testing Identification (HTI)

This method was first introduced in 1984 by L. Mili in [8] and is based on the hypothesis test; two hypotheses H_0 and H_1 are regarding as follows:

H_0: i^{th} measurement has no bad data.

H_1: i^{th} measurement has bad data.

Therefore, two types of errors may occur:

Error 1: i^{th} measurement has no bad data (H_0 is true) but it is identified bad data. (H_1 is rejected).

Error 2: i^{th} measurement has bad data (H_1 is true) but it is identified no-bad data. (H_1 is rejected).

If it is assumed that the probability distribution of the two mentioned errors are normalized as shown in Fig. 1, then mean value of error 1 is zero, and the mean value of error 2 is e_{si}, because in error 1 it is assumed that the i^{th} measurement has no bad data, but in error 2 the i^{th} measurement has bad data.

Accordingly, as shown in Fig. 1, the possibility of error 1 may be calculated using (17).

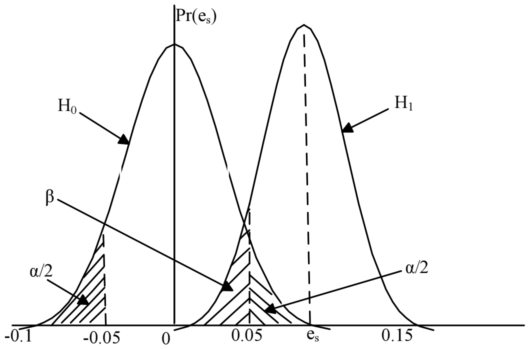

Figure 1. The porobability of existing errors 1 and 2

$$\alpha = P_r(H_0 \text{ has been rejected} \mid H_0 \text{ is true}) \\ = P_r(|\hat{e}_{si}| > \lambda_i) \quad (17)$$

Where, λ_i is a constant value in the range of variance e_{si}, and e_{si} has the probability distribution as follows [14].

$$\hat{e}_{si} \sim N(0, \sigma_i^2 S_{ii}^{-1}) \quad (18)$$

By normalization of the probability distribution of e_{si}, α can be written as:

$$\alpha = P_r\left(\frac{|\hat{e}_{si}|}{\sigma_i\sqrt{S_{ii}^{-1}}} > \frac{\lambda_i}{\sigma_i\sqrt{S_{ii}^{-1}}} = N_{1-\frac{\alpha}{2}}\right) \quad (19)$$

In addition, the probability of error 2 can be as expressed in (18).

$$\beta = P_r(H_1 \text{ has been rejected} \mid H_1 \text{ is true}) \quad (20)$$

$$\beta = P_r(\hat{e}_{si} \leq \lambda_i) \quad (21)$$

In which, e_{si} has the possibility distribution as follows [14]:

$$\hat{e}_{si} \sim N(e_{si}, \sigma_i^2 (S_{ii}^{-1} - 1)) \quad (22)$$

By normalization of e_{si} we have:

$$\beta = P_r\left(\frac{\hat{e}_{si} - |e_{si}|}{\sigma_i\sqrt{S_{ii}^{-1} - 1}} \leq \frac{\lambda_i - |e_{si}|}{\sigma_i\sqrt{S_{ii}^{-1} - 1}} = N_\beta\right) \quad (23)$$

By comparing with N_β and $N_{(1-\alpha/2)}$ we can write:

$$\sigma_i N_\beta \sqrt{S_{ii}^{-1} - 1} = \sigma_i \sqrt{S_{ii}^{-1}}\, N_{1-\frac{\alpha}{2}} - |e_{si}| \quad (24)$$

Thus, bad data identification algorithm of measurement using the HTI method can be expressed as follows:

5.1. With fixed N_β

1. Selecting a value for N_β
2. State estimation and selecting suspected vector using r^n,
3. Calculating e_{si} and $N_{(1-\alpha/2)}$ for each member of the suspected vector

$$\hat{e}_{si} = S_{ii}^{-1} r_{si} \tag{25}$$

$$N_{(1-\frac{\alpha}{2})i} = \frac{|e_{si}| + \sigma_i N_\beta \sqrt{S_{ii}^{-1} - 1}}{\sigma_i \sqrt{S_{ii}^{-1}}} \tag{26}$$

$$\lambda_i = d_i \sqrt{S_{ii}^{-1}}\, N_{(1-\frac{\alpha}{2})i} \tag{27}$$

4. Selecting measurements such that $|e_{si}| > \lambda_i$.
5. Selected measurements in step 4 are chosen as new suspected vector and steps 3 and 4 are repeated again, if the vector in step 1 is identical with step 4, therefore the algorithm is over and suspected vector is known as the bad data measurements.

5.2. With fixed $N_{(1-\alpha/2)}$

1. State estimation and selecting fixed value for $N_{(1-\alpha/2)}$
2. Selecting suspected vector using r^n
3. Calculating e_{si} and λ_i for each member of the suspected vector using (22) and (24),
4. Selecting measurements such that $|e_{si}| > \lambda_i$
5. Selected measurements in step 4 are chosen as new suspected vector and steps 3 and 4 are repeated again, if the vector in step 1 is equal to step 4, therefore the algorithm is over and suspected vector is known as the bad data measurements.

6. Improving detection of measurement error

In a power system, it is assumed that the there would be two measurements *i* and *j*, so that:

$$Z_i >> Z_j \tag{28}$$

Using the (12) for the two measurements we have,

$$\begin{bmatrix} r_i \\ r_j \end{bmatrix} = \begin{bmatrix} S_{ii} & S_{ij} \\ S_{ji} & S_{jj} \end{bmatrix} \begin{bmatrix} Z_i \\ Z_j \end{bmatrix} \tag{29}$$

$$\begin{aligned} r_i &= S_{ii} Z_i + S_{ij} Z_j \\ r_j &= S_{ji} Z_i + S_{jj} Z_j \end{aligned} \tag{30}$$

With the mentioned approximation (28), (29) and (30) will be simplified as follows:

$$\begin{aligned} r_i &= S_{ii} Z_i \\ r_j &= S_{ji} Z_i \end{aligned} \tag{31}$$

According to the (12) to (14), Z_i and Z_j can be replaced by e_i and e_j, respectively.

$$\begin{aligned} r_i &= S_{ii} e_i \\ r_j &= S_{ji} e_i \end{aligned} \tag{32}$$

The above equations show that the j^{th} measurement error has roughly no impact on the value of i^{th} measurement residual. In other words, if the j^{th} measurement has error ($e_j \neq 0$, $e_i=0$) then based on (31) and (32), j^{th} measurement is detected with no error. Also if the i^{th} measurement has error, j^{th} measurement is also detected error-involved.

To solve this problem, in this paper proposed the value of each measurement covariance, σ^2, would be divided into the per unit value of that measurement.

$$
\begin{aligned}
W_{ii} &= \sigma_i^{-2} / |Z_i| \\
W_{jj} &= \sigma_j^{-2} / |Z_j|
\end{aligned}
\tag{33}
$$

Changes in the matrix W according to the (33), influence on matrix S according to (13) so that the values of S_{ii} and S_{jj} will be balanced. Hence, $S_{jj}^{new} >> S_{jj}^{old}$. Therefore even if regarding (28), Z_j cannot be omitted from (30). Further, if Z_j includes error, r_j will be more affected than r_i and vice versa.

7. Case studies

Case studies were carried out on IEEE 14-bus system, and the measurement bad data was tested using CSTD, r^N_{max} and HTI. As shown in Fig. 2, this system has several power flow, injection power and voltage measurements.

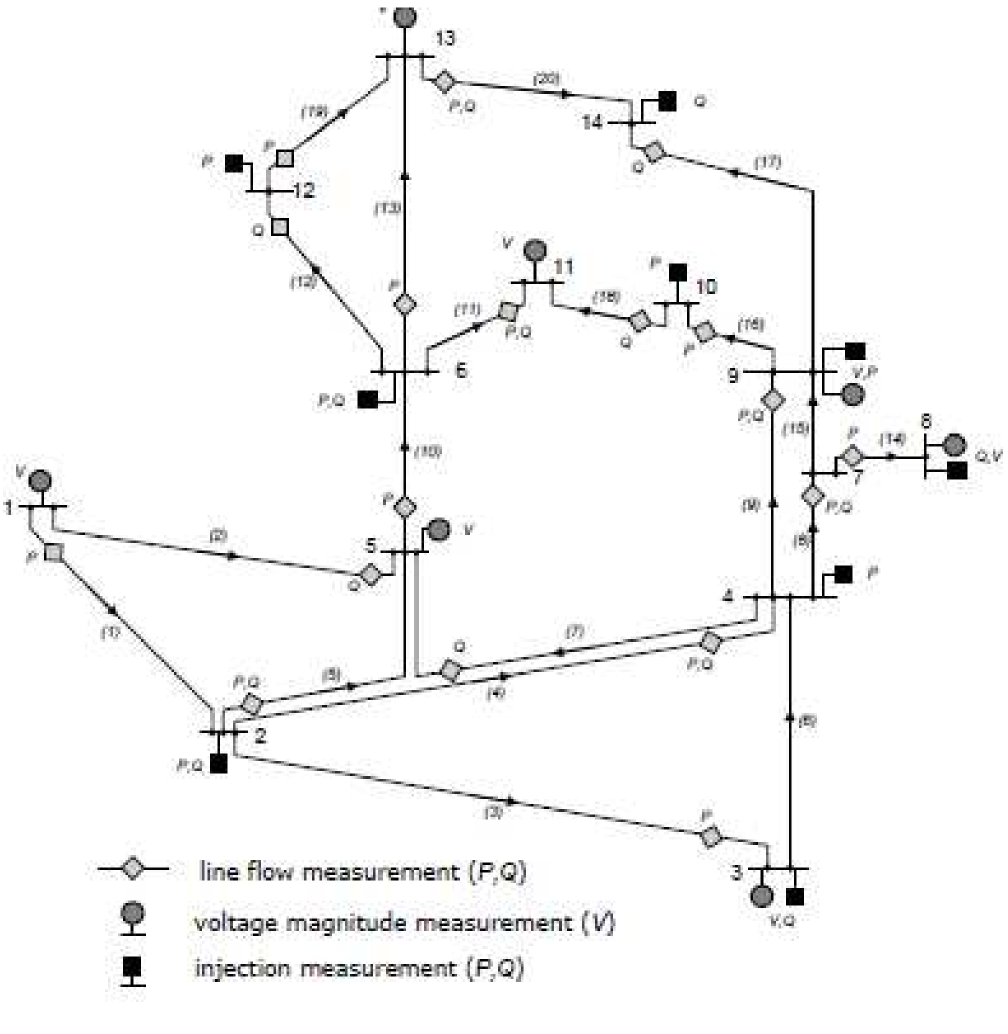

Figure 2. IEEE 14-bus system

At first step, each of measurements are considered individually with 25%-error (actual value × 0.75 = value of error-involved measurement) and the state estimation is done for each situation. The numbers of times that the CSTD, r^N_{max}, and HTI identify the bad data are given in Table 1. For the other tests, the error value is regarded 100% and 200% (0=measurement value with 100% error and the actual value × -1 = value of measurement with

200% error) and for any of the situation, state estimation are performed. The number of times that the CSTD, r^N_{max}, and HTI are able to detect the bad data in 100% and 200% error are given in Table 1, too.

Table 1. Bda data detection by CSDT, r^N_{max} and HTI

Percentage of error	CSDT	r^N_{max}	HTI	Average length of final suspected vector
25%	9	7	5	7
100%	16	12	5	24.2
200%	20	13	20	8.85

However, according to the section V and Table 1 it is shown that HTI is not able to exactly identify the bad data and only can reduce the suspected vector, unless that the final suspected vector is just error-involved measurement. For example, in Table 1, the mean value of final suspected vector for 75%, 100% and 200% errors were given, and as it can be seen, among 42 measurements, HTI is just could detect those measurements as final suspected measurements which it is possible the error-involved measurements are in same suspected measurements.

In the next step, the mentioned study is done by changing (33). Results will be obtained according to the Table 2.

Table 2. Proposed bda data detection by CSDT, r^N_{max} and HTI

Percentage of error	CSDT	r^N_{max}	HTI	Average length of final suspected vector
25%	15	7	10	15.6
100%	35	13	18	16.38
200%	34	15	34	17.64

Comparing Tables 1 and 2, show that the proposed method can improve the bad data detection by CSTD, r^N_{max} and HTI.

8. Conclusion

In state estimation of power systems, measurements may have bad data which affects correct state estimation as well as measurements values. It is therefore required to detect these error-involved measurements and put aside them from measurements collection. So far several methods have been presented to identify these bad data, which are classified into mathematical and intelligent methods. Intelligent methods either need network training for various situations not practical due to the existing many situations or require more calculations time, therefore mathematical methods major Chi-Square Distribution Test as the error presence detector, Largest Normalized Residual Test and Hypotheses Testing Identification as the bad data detectors of measurement are used. However, these tests are not able especially when the measurement error is low to detect the bad data. In this paper, a method was presented which was able to improve the detection of bad data. Various case studies with different error percentages were carried out on IEEE 14-bus system. The results showed that the proposed method is able to increase ability of Chi-square Distribution, Largest Normalized Residual and Hypothesis Testing Identification.

Acknowledgements

This paper is supported by Islamic Azad University- Ahar Branch for MSc tesis of Mr. Mahaei entiteled “Detection of Bad Data and Identification Topology in Power System State Estimation”

References

[1] Monticelli, A. Garcia, “Reliable bad data processing for real-Time state Estimation”, *IEEE Transactions on Power Apparatus and Systems*, Vol. PAS-102, No. 5, May 1983, pp 1126-1139.

[2] K. L. Lo, P. S. Ong, R. D. Mcoll, A. M. Moffatt, J. L. Sulley, “Development of a Static State Estimation Part I: Estimation and Bad data Suppression”, *IEEE Transactions on Power Apparatus and Systems*, Vol. 102, No. 8, August 1983, pp 2486-2491.

[3] B. M. Zhanet, S. Y. Wars, N. D. Xiang, “A Linear Recursive Bad Data Identification Method with Real Time Application for Power System State Estimation”, *IEEE Transactions on Power Systems*, Vol. 7, No, 3, August 1992, pp 1378-1385.

[4] G. N. Korres, G. C. Contaxis, “A Reduced Model for Bad Data Processing in State Estimation”, *IEEE Transactions on Power Systems*, Vo1.6, No. 2, May 1991, pp 550-557.
[5] B. M. Zang, K. L. LO, “A Recursive Measurement Identification Method for Bad Data Analysis in Power System State Estimation”, *IEEE Transactions on Power Systems*, Vo1.6, No.1, February 1991, pp 191-198.
[6] A. Monticelli, F. Wu, M. Yen, “Multiple Bad Data Identification for State Estimation by Combinatorial Optimization”, *IEEE Transactions on Power Delivery*, Vol. 1, No. 3, July 1986, pp. 361-369.
[7] A. Monticelli, Felix F. Wu, Maosong Yen, “Multiple Bad Data Identification for State Estimation by Combinatorial Optimization”, *IEEE Transactions on Power Delivery*, Vol. 1, No. 3, July 1986, pp 361-369.
[8] L. Mili, Th. Van Cutsem, M. Ribbens-Pavella, “Hypothesis Testing Identification: A New Method for Bad Data Analysis in Power System State Estimation”, *IEEE Transaction on Power Apparatus and Systems*, Vol. 103, No. 11, November 1984, pp 3229-3259.
[9] F. Zhuang, R. Balasubramanian, “Bad Data Processing in Power System State Estimation by Direct Data Detection and Hypothesis Tests”, *IEEE Transactions on Power Systems*, Vol. 2, No. 2, May 1987, pp 321-327.
[10] S. Naka, T. Genji, T. Yura and Y.Fukuyama, “A hybrid particle swarm optimization for distribution state estimation,” *IEEE Transactions Power System*, vol. Pas 18, Nov 2002, pp 55-57.
[11] D.M.V. Kumar, S. C. Srivastava, S. Shah and S. Mathur, “Topology Processing and Static State Estimation Using Artificial Neural Networks” *Proc. Inst. Elect. Eng. C*, vol 148, Jan, 1996, pp 51-60.
[12] M. Shahidehpour and D Labudda, “ A Fazy Molti-objective Approach to Power System State Estimation”, *Fourth International Symposium on Expert System Application to Power System*, 1993, pp218-223.
[13] D Chauhan and D Singh and J P Pandey, “Topology Identification, Bad Data Processing, and State Estimation Using Fuzzy Pattern Matching”, *IEEE Transaction Power System*, vol. Pas 20, Aug 2005, pp 1570-1579.
[14] A.Abur and A.Gomez Exposito, “Power System State estimation: Theory and Implementation”, New York, Marcel Dekker, 2004.

Bibliography of authors

Mehrdad Tarafdar Hagh (S'98-M'06-M'2001) received his B. Sc. and M. Sc. both with first honor in 1988 and 1992, respectively and Ph. D. in 2000, all in power engineering from University of Tabriz, Iran. He has been with the faculty of electrical and computer engineering, University of Tabriz since 2000, where he is currently an Associate Professor. He has published more than 140 papers in power system and power electronics related topics. His interest topics include power system operation, FACTS and power quality.

Seyyed Mehdi Mahaei was born in Tabriz, Iran, in 1984. He received the B.Sc. degree (with First Class Honors) in electrical engineering from University of Applied Science and Technology, Tabriz, Iran, in 2007, and M.Sc. degree from Islamic Azad University, Ahar, Iran, in 2011. Also he is project supervisor in Azerbaijan Regional Electric Company, Tabriz, Iran.
His research interests include power system, optimization and operation.

Kazem Zare received his B. Sc. and M. Sc. in 2000 and 2003, respectively from University of Tabriz, Iran and Ph. D. in 2009 from Tarbiat Modares University, Tehran, Iran all in power engineering. He has been with the faculty of electrical and computer engineering, University of Tabriz since 2009. He has published many papers in power system related topics. His interest topics include reliability, reconstruction and distribution systems.