Pertanika J. Sci. & Technol. 25 (3): 825 - 834 (2017)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

The Control Chart Technique for the Detection of the Problem of Bad Data in State Estimation Power System

Zahid Khan[1]*, Radzuan B. Razali[1], Hanita Daud[1], Nursyarizal Mohd Nor[2] and Mahmud Fotuhi-Firuzabad[3]

[1]*Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS, 31750 UTP, Tronoh, Perak, Malaysia*

[2]*Department of Electrical Engineering, Universiti Teknologi PETRONAS, 31750 UTP, Tronoh, Perak, Malaysia*

[3]*Department of Electrical Engineering, Sharif University of Technology, 11155-1639 Tehran, Iran*

ABSTRACT

State estimation plays a vital role in the security analysis of a power system. The weighted least squares method is one of the conventional techniques used to estimate the unknown state vector of the power system. The existence of bad data can distort the reliability of the estimated state vector. A new algorithm based on the technique of quality control charts is developed in this paper for detection of bad data. The IEEE 6-bus power system data are utilised for the implementation of the proposed algorithm. The output of the study shows that this method is practically applicable for the separation of bad data in the problem of power system state estimation.

Keywords: Nonlinear estimation, weighted least squares method, bad data, Chi-square test, normalised residual test, Gauss-Newton algorithm

INTRODUCTION

The risk of blackouts in grids has increased the desperate need of monitoring the power system. State estimation is a fundamental method for online system monitoring, analysis and control functions. It is a technique of evaluating field data and developing a precise estimation of the system's parameters. The state variables estimated in this way are used to calculate other key functions of the power flow analysis. With the estimated power quantities, the operator in the control centre will be able to get information about the current status of the system and take necessary

Article history:
Received: 27 May 2016
Accepted: 14 November 2016

E-mail addresses:
zahidkhan_g02731@utp.edu.my, zahid.hazara@gmail.com (Zahid Khan),
radzuan_razali@petronas.com.my (Radzuan B. Razali),
hanita_daud@petronas.com.my (Hanita Daud),
nursyarizal_mnor@petronas.com.my (Nursyarizal Mohd Nor),
fotuhi@sharif.edu (Mahmud Fotuhi-Firuzabad)
*Corresponding Author

ISSN: 0128-7680

measurements in case of abnormal overloading and save the system from probably blackouts. It is fundamentally based for providing a reliable and accurate real-time data base, which in turn will be used by all other energy management system (EMS) functions.

The mathematical algorithm which is used by the state estimator is based on the weighted least squares (WLS) method. The WLS method works on the assumption that noisy measurements follow the Gaussian distribution. In the presence of noisy measurements only, the WLS provides reliable estimations for the state parameters. However, the final results of WLS in the existence of gross errors will be biased. Therefore, the detection of gross errors that cause bad data in SCADA measurements is also an important subject of interest in state estimation of a power system**.** The demand for electric energy is increasing because of the increased demand of consumers and their lifestyles. Thus, a power system must be able to generate enough energy to meet the required load for a region.

The revolutionary work of Schweppe (Schweppe & Rom, 1970; Schweppe & Wildes, 1970) has become the basis for using state estimation in the supervisory control unit of a power system. This model is built on measurements from snapshots that show the current status of the system (Monticelli, 2000). One of the key functions of state estimation is to detect bad data and eliminate them if possible (Mili, Van Cutsem, & Ribbens-Pavella, 1985). Sometimes, data are corrected by eliminating a bad value prior to estimation. An excellent discussion of these techniques can be found in Huang, Shih, Lee and Wang (2010).

The main objective of these techniques is saving iteration cost. Post estimation bad data techniques are based on residual analysis. The largest normalised residual is used to find the suspected measurement (Carvalho & Bretas, 2013). The normalised residual test does not always perform well (Khan, Razali, Daud, Nor, & Fotuhi-Firuzabad, 2015). The hypothesis-testing method is also used to find bad measurements in an iterative way (Mili & Van Cutsem, 1988). In the presence of bad data, many researchers also worked on the use of some robust methods like the least median squares and least trimmed squares estimator in calculating system state parameters, the details of which can be seen in several studies (Baldick, Clements, Pinjo-Dzigal, & Davis 1997; Kotiuga & Vidyasagar, 1982; Mili, Cheniae, & Rousseeuw 1994). A comparative study of these estimators in case of large and small bad measurements has been summarised in Habiballah and Irving (2000). There are a number of other robust estimators like the S-estimators, τ-estimators, and L-estimators that have reasonably high breakdown points that may be suitably applicable in the presence of bad values. The concept of outlier detection and leverage points within the context of robust regression is also a very important area. Introductory literature on these topics can be studied in statistical applications (Klebanov, Rachev, & Fabozzi 2009; Staudte & Sheather, 2011).

Most of the aforementioned research work is addressed towards numerical efficiency. In this work, we have, therefore, proposed a technique for detection of bad data in a statistical point of view which is based on the structure of the data generating model and its modified assumptions.

The overall contribution of this work is to provide a new detection technique for bad data in state estimation environments. This method will be effective in time and cost savings. This approach will be new information for the power system engineers who in turn can apply the technique to many other related studies.

The rest of this article is structured as follows. Section 2 describes the mathematical formulation of the state estimation problem; Section 3 explains the procedure for the construction of the control chart for detection of bad data; Section 4 is related to the presentation of the proposed algorithm and its implementation on the IEEE 6-bus power system. Finally, a conclusion is drawn in Section 5.

MATHEMATICAL MODEL AND ESTIMATION METHOD

Noisy data have been collected from different measuring devices that were installed at different locations of a transmission system. The exact value of any physical measurement was not known. Therefore, the stochastic model that is used in power system state estimation was given by:

$$Z = HX + e \tag{1}$$

where $H = \frac{\partial h(X)}{\partial X}$ is the Jacobian matrix, $Z = [z_1, z_2, ...z_n]^T$ is the vector of measurements, $X = [x_1, x_2, ...x_k]^T$ is an unknown state vector consisting of complex components of voltages of all the buses except the slack bus, $e = [e_1, e_2 ...e_n]^T$ is a Gaussian error vector with the following assumptions:

$E(e) = 0$ (mean vector) and $E(ee^T) = \sigma^2 diag\{\sigma_1^2, \sigma_1^2 ...\sigma_n^2\}$ (covariance matrix).

Here, it is worth mentioning that weights are the known figures in the power system state estimation problem. Contrary to a conventional weights matrix we defined the weight for the i^{th} observation as:

$$w_i = \frac{1}{\sigma_i^2}.$$

Here, σ_i^2 is a measurement to reflect the accuracy of the i^{th} measuring instrument. These values of $\sigma_i^2, 1 \le i \le n$ are known numbers where σ^2 is either estimated by its maximum likelihood estimate in case of sample data or found directly from the data in case of a fully observable system. The objective function in matrix notation can be written as:

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

where W is a weight matrix.

The Gauss Newton algorithm is conventionally used in system state estimation for estimating the state of the power system (Monticelli, 2000).

In the Gauss Newton iterative scheme, the solution of X at rth iteration is as given below:

$$X^{r+1} = X^r - [G(X^r)]^{-1} g(X^r) \tag{3}$$

where $g\left(X^r\right) = -H^T\left(X^r\right)W\left[Z - h\left(X^r\right)\right]$ and $G\left(X^r\right) = \frac{\partial g\left(X^r\right)}{\partial X} = H^T\left(X^r\right)W\,H\left(X^r\right)$.

The most important sub-function of a state estimator is detection of bad data. Our objective function can be written as:

$$\varphi\left(X\right) = \frac{\sum_{i=1}^{n} w_i\left(z_i - h_i\left(\hat{X}\right)\right)^2}{n-k}$$

which follows the Chi-square distribution with $n-k$ degree of freedom. If the calculated value of the objective function at optimum value of the state vector X is greater than that of the set critical value, it is implied that at least one bad value is suspected in the measurements. The Chi-square test is only an indication of the existence of bad data. It does not indicate which value is the bad one. The other test that is used in state estimation of power is the normalised residual test (Carvalho & Bretas, 2013). Based on a sensitivity analysis we found a covariance matrix for residuals whose diagonal elements represented the corresponding variances of estimated residuals. Normalised residuals were then calculated as given below:

$$r_i^N = \frac{\left|r_i\right|}{\sqrt{\phi_{ii}}}\;;\quad r^N \sim N\left(0,1\right) \tag{4}$$

where ϕ_{ii} is the corresponding diagonal element of the variance covariance matrix of i^{th} residual. The largest normalised residual was in accordance with the value polluted by gross error. Thus, in this way, the value corresponding to the largest normalised residual was deleted from the data or corrected by its appropriate substitute.

PROPOSED METHOD FOR IDENTIFICATION OF BAD DATA

Our proposed method was based on the control chart methodology. The control charts are used for monitoring shifts either in the mean or standard deviation of an assumed distribution. A control chart is a useful statistical technique for monitoring the process/system in terms of statistical control. The term statistical control means the capability of the chart to figure out whether the system is operating under variation of chance causes only or not. In statistical terms, a process is said to be “out-of-control” if it is functioning in the presence of assignable causes. If bad measurements are not detected properly, the results of the state estimator will not be valid and in return the power system security analysis may be misguided. The question that came to mind as to which observation would be considered a bad one in context of state estimation of power. The value in which a gross error is higher than the expected accuracy of the meter would be considered bad.

In model form we can write it as:

$$z^{meas} = z^{true} + c\sigma$$

where c is the amount of bad data measured in terms of standard deviation, z^{meas} is the available value from the metering device and z^{True} is a value obtained from the measurement function $h(x)$.

Thus, the available meter measurement is the sum of the true value of z and the Gaussian random variable. So, the addition of an amount will shift the Gaussian curve to the right or left depending on the sign of the added term. The shape of the curve will remain the same i.e. a bad observation will not alter the scale of distribution. Hence, we put in the effort to monitor only the mean of the process. This method was based on post estimation because it used the residuals in its analysis. Here, the out-of-control process referred to the data that were used to find the state parameters of a power system that was not of good quality. Hence, the estimation based on bad quality data would not be reliable. A single realisation of n pairs of $\left(z_i, h_i(x)\right)$ can be represented as:

$$z_i = h_i(x) + e_i$$

where $h_i(x)$ has a specific known form that depends on the type of measurement and e_i is a normally distributed error term with a mean of zero and variance σ_i^2. Thus, the very observation that is used in the state estimation is a measurement of the dependent variable z at the non-linear measurement function $h(x)$. This relationship, in terms of statistical quality control, is called a profile. Thus, our proposed control chart was based on non-linear profiles monitoring through the residuals of a non-linear model given in Equation (1). The general model control chart for a quality characteristic that is described by the sample statistics states that P is given by (Montgomery, 2009):

$$\begin{aligned} UCL &= \mu_p + c\sigma_p \\ CL &= \mu_p \\ LCL &= \mu_p - c\sigma_p \end{aligned} \tag{5}$$

where μ_p is the mean and σ_p is the standard deviation of the statistic p. The model that is used in state estimation is just like the one used in multivariate weighted non-linear regression. Consider the following transformed form of Equation (1):

$$J(X) = \left[Z_t - h_t(X)\right]^T \left[Z_t - h_t(X)\right] \tag{6}$$

where $Z_t = W^{1/2} Z$, $h_t(X)$ $W^{/} h_t(X)$ and $W^{1/2}$ is the Cholesky factor matrix of W.

In this way, the solution is reduced to the estimation by the ordinary least squares (OLS) method. The OLS estimation of Equation (6) is thus given by:

$$\hat{X}_t = \left(H^T W H\right)^{-1} H^T W Z$$

The value of this estimation can be calculated using the Gauss Newton algorithm after linearising the objective function as given in Equation (6). Now let the following residuals vector from Equation (1):

$$r_w = Z - \hat{Z}_w \tag{7}$$

where $\hat{Z}_w = H\hat{X}_t$, which is an unbiased estimator of Z.
The Equation (7) implies that:

$$r_w = \left(I - H^*\right)e = Ke \tag{8}$$

where K is the sensitivity matrix of the residuals which is symmetric and idempotent.
By utilising the properties of Equation (8), the variance covariance matrix of the residuals can be found as:

$$vcor\left(r_w\right) = \sigma^2 KW^{-1} \tag{9}$$

Hence from Equation (8) and Equation (9), it yields that: $r_w \sim N\left(0, \sigma^2 S\right)$, where, $S = KW^{-1}$
Since each residual has its own variance, therefore, we defined the i^{th} studentised residual as:

$$r_{wi}^{\ S} = \frac{r_{wi}}{\hat{\sigma}\sqrt{s_{ii}}}, \; i = 1, 2, 3 \ldots n \tag{10}$$

where $\hat{\sigma}$ is independently computed at the optimum value of the state vector X.
Since each $r_{wi}^{\ s} \sim N\left(0,1\right)$, therefore, from Equation (5), the parameters of the control chart of the individual studentised residuals are given as:

$$UCL = 3, \; CL = 0, \; LCL = -3$$

where UCL and UCL signify the upper and lower control limits of the studentised residuals chart.

The value falling outside these limits was suspected to be a bad value. It needs correction or exclusion from the data and a re-estimation of the WLS on the corrected data. The residuals in Equation (9) are different from the conventional normalised residuals in Equation (3) because they also take into account the measurement of the closeness of the actual and estimated residuals into calculation.

IMPLEMENTATION OF RESIDUAL BASED CONTROL CHART

We developed a programme in MATLAB for estimating the model and finding the residuals as explained in the previous section. The algorithm of our technique can be summarised in the following steps:

Step 1: Initialise the values by taking $r = 0$, flat start of state vector X^o and convergence quantity ε.

Step 2: Calculate the measurement function $h\left(X^r\right)$, Jacobian function and gain matrix $G\left(X^r\right)$ as given in Equation (2).

Step 3: Solve Equation (2), for new values of state variable X and check the solution convergence by comparing $\left|\Delta X^r\right| \le \varepsilon$; if the solution converges, extract residuals from the model; otherwise, go to step 2, updating $r = r + 1$.

Step 4: Construct residual-based control chart. If the process is in control, display the output for X; otherwise, display the control chart for the out-of-control process and go to step 3.

Step 5: Update the measurements vector by deleting the corresponding bad value from the data in accordance with the largest studentised residual and go to step 2.

The IEEE 6-bus standard test network was selected for the implementation of the results obtained in Section 3. The data on the used test network and system's topology can be found in Wood and Wollenberg (2012). The main features of this system are shown in Table 1.

Table 1
Main Features of IEEE 6-Bus Test Network

Types of Measurement	**STD**
Voltage measurement	2 kV
Real and reactive power injections	3 MV
Real and reactive power flows	3MVR

For the implementation of WLS, we assumed that each observation in both systems was weighted by the accuracy measurement of its corresponding measuring meter. These weights were defined in terms of the standard deviation (STD) of the measurements as given in Table 2.

Table 2
Weights of Different Measurements

Features	IEEE 6-bus test network
Total nodes	6
Total branches	11
Total measurements	62
Total parameters to be estimated	11

These values of weights, like the other measurements, were also converted in per unit system for their usage in the elaborated algorithm in Section 3 by dividing them by their base case values of 230 kV, 100 MV and 100 MVR. The termination criteria as stated in the algorithm was set at 10^{-4} for the convergence of the values of the vector. The actual values were assumed from the results of load-flow analysis whereas the noisy measurements were generated by perturbing the values with normal distribution random generation having a mean parameter equal to zero and variance corresponding to defined weights for each measurement as shown in Table 2. For the inclusion of a single bad value in the measurements, a value of the real power injection at bus 1 of the IEEE 6-bus system, was corrupted intentionally by adding 10 times its per unit standard deviation. The algorithm was converged at the third iteration and the results obtained as shown in Table 3.

Table 3
State Estimation Results with and Without Bad Data

$\hat{X}$ without bad data	$\hat{}$ with bad data
1.0449	1.0477
1.0478	1.0473
1.0685	1.0677
0.9864	0.9863
1.0021	0.9792
-3.7282	1.0013
-4.2486	-4.3158
-3.7282	-4.9734
-4.2883	-4.8452
-5.3444	-5.9877
-5.8367	-6.5629
( ) =52.1745	$\varphi(X)$=88.8270

The computed value of the Chi-square statistic showed that bad data were suspected when actually our data had a bad value. However, we did not know which value was actually bad and how bad it was. In order to answer questions on the detection and identification of bad value, we constructed the studentised residual-based control chart with the parameters as listed in Section 3. The control charts, which were the output of our proposed algorithm, are shown in Figure 1 and Figure 2.

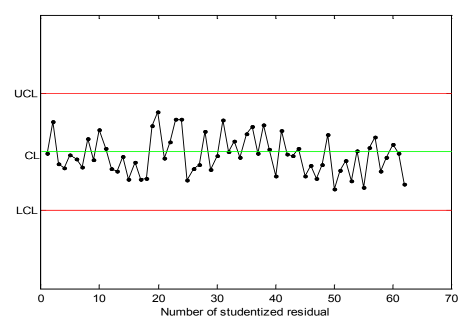

Figure 1. The snapshot of in-control data for the IEEE 6-bus system

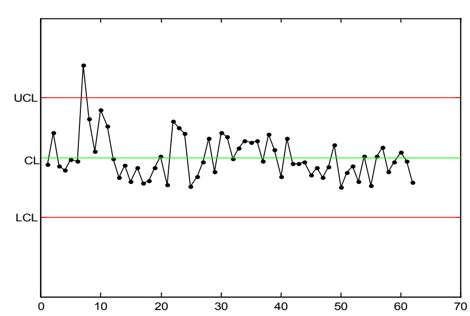

Figure 2. The snapshot of out-of-control data for the IEEE 6-bus system

The structure of the points in the control chart as given in Figure 2 showed that the quality of data was out of control at the seventh studentised residual. Hence, the corresponding value of the residual was suspected to be bad and it needed to be deleted from the data. Our proposed method has a visualisation feature that may help the operator in the separation of good and bad values. This method is better in the sense of having both properties of detection and identification of bad values whereas the Chi-square test has only the feature of sounding the alarm at the presence of bad data. Although the normalised residual test also had both properties, there was a risk of deletion of those values that were actually not bad whereas our proposed method had a clear benchmark limits to decide bad values. Our test was a modified form of the normalised residual test because it included mean square error (MSE) in its calculation. As the value of bad data increased, the MSE also increased in the denominator of the proposed studentised residual, which kept the ratio within the defined limits.

CONCLUSION

State Estimation has become a significant part of the controlling, monitoring and planning of modern power systems. Therefore, to rely on the results of a state estimator, it is mandatory to study those factors that can alter the results of the state estimator. One of these main factors is the existence of bad data. In view of its importance, a new algorithm has been proposed in this work for sounding the alarm in case of the existence of bad values in measurements. The technique has additional features of visualisation and clear threshold limits to discriminate between good and bad values. The algorithm on one hand is a new technique for power engineers for separation of bad data and is compatible with the conventional method of WLS in state estimation of power, but on the other, it is an extension of applications of the control chart in statistical process control (SPC) for monitoring non-linear profiles.

REFERENCES

Baldick, R., Clements, K. A., Pinjo-Dzigal, Z., & Davis, P. W. (1997). Implementing non-quadratic objective functions for state estimation and bad data rejection. *IEEE Transactions on Power Systems*, *12*(1), 376–382.

Carvalho, B., & Bretas, N. (2013). Analysis of the largest normalized residual test robustness for measurements gross error processing in the WLS state estimator. *Journal of Systemics, Cybernetics and Informatics, 11*(7), 1–6.

Habiballah, I. O., & Irving, M. R. (2000). A comparative study of three LS-based power system state estimators for bad data identification. *Electric Machines and Power Systems*, *28*(2), 105–114.

Huang, C. H., Shih, K. K., Lee, C. H., & Wang, Y. J. (2010). Application of Kalman filter to bad data detection in power system. In V. Kordic (Ed.), *Kalman filter* (pp. 127–144). Croatia: InTech.

Khan, Z., Razali, R. B., Daud, H., Nor, N. M., & Fotuhi-Firuzabad, M. (2015). The largest studentized residual test for bad data identification in state estimation of a power system. *ARPN Journal of Engineering and Applied Sciences, 10*(21), 10184–10191. Retrieved from http://www.arpnjournals.org/jeas/research_papers/rp_2015/jeas_1115_3046.pdf

Klebanov, L. B., Rachev, S. T., & Fabozzi, F. J. (2009). *Robust and non-robust models in statistics*. United States of America, USA: Nova Science Publishers.

Kotiuga, W. W., & Vidyasagar, M. (1982). Bad data rejection properties of weighted least absolute value techniques applied to static state estimation. *IEEE Transactions on Power Apparatus and Systems, PAS-101*(4), 844–853.

Mili, L., Cheniae, M. G., & Rousseeuw, P. J. (1994). Robust state estimation of electric power systems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, *41*(5), 349–358.

Mili, L., & Van Cutsem, T. (1988). Implementation of the hypothesis testing identification in power system state estimation. *IEEE Transactions on Power Systems, 3*(3), 887–893.

Mili, L., Van Cutsem, T., & Ribbens-Pavella, M. (1985). Bad data identification methods in power system state estimation: A comparative study. *IEEE Transactions on Power Apparatus and Systems, PAS-104*(11), 3037–3049.

Montgomery, D. C. (2009). *Introduction to statistical quality control* (7th Ed.). New York, NY: John Wiley & Sons.

Monticelli, A. (2000). Electric power system state estimation. *Proceedings of the IEEE, 88*(2), 262–282.

Schweppe, F. C., & Rom, D. B. (1970). Power system static-state estimation, Part II: Approximate model. *IEEE Transactions on Power Apparatus and Systems, PAS-89*(1), 125–130.

Schweppe, F. C., & Wildes, J. (1970). Power system static-state estimation, Part I: Exact model. *IEEE Transactions on Power Apparatus and Systems, PAS-89*(1), 120–125.

Staudte, R. G., & Sheather, S. J. (2011). *Robust estimation and testing* (Vol. 918). New York, NY: John Wiley & Sons.

Wood, A. J., & Wollenberg, B. F. (2012). *Power generation and control*. New York, NY: John Wiley & Sons.