ABSTRACT

Power system state estimation is a reliable tool used in Energy Management System (EMS) to identify the existence state of the system during its operating hours. The results of this estimation are values for unknown state parameters of the power system. The presence of systematic errors can alter the results of state estimation. The chi-square and normalized residual tests are the common post estimation procedures usually used for detection and identification of gross errors in the estimation algorithm. These tests are based on two separate test statistics and are not so powerful for detection of smaller magnitudes of gross errors. In this paper, an implementation of largest studentized residual (LSR) test is presented that combines both the results of chi-square and normalized test for detection and identification of bad data. Based on LSR test, a comprehensive strategy is developed for detection and identification of multiple gross errors which may exist simultaneously in the data. A six-bus power system data is used for the application of LSR test for detecting and identifying the gross errors in the processed measurements. The reporting results are presented showing that the method is most powerful and effective for practical implementation in conventional procedures of the state estimation problem.

Keywords: bad data identification, power system, LSR test.

INTRODUCTION

Motivation and Aim

The security of the power system from the major power outage and economical operations of the system highly depend on the correct information about the status of system during the operating hours. The control center of the power system is responsible for knowing the operations’ condition of the system which may change suddenly due to some assignable cause. Power system is much more dynamic as it is continuously supplying power to the consumers according to their load requirements. In such conditions, the corrective tools for supervising the system are necessary prior to any emergency state that the system may enter. The operator of the system in the control room has used these tools with the aim to operate the system in a secure manner. One of such measures is power system state estimation which is a mathematical algorithm built on the idea of real time modeling of the system’s current conditions. The information of the system is continuously extracted at different intervals and algorithm is executed subsequently by acquiring the information from the snapshots of system’s current conditions. This is an estimation procedure which is used to estimate the unknown values of the state vectors by utilizing the given information. The main goal associated with state estimation is to compute complex components of voltages at each bus of the power network. The addition purposes include detection of error in the model parameter and detection of bad data. The maximum likelihood method is generally implemented to estimate unknown state vector. This technique minimizes the weighted sum of square residuals. Once the estimated state is obtained, operator performs the detection of bad data for possible existence of gross errors. The state estimation technique crucially relies on the assumption that data measurements are not effected by gross errors but only the random errors are present. If this assumption does not hold, the resulting values of unknown parameters can be misleading. This invalidity in the results can then subsequently be wrongly guided to the system operator to take inaccurate decision, resulting in major problem in stability of the system. Therefore state estimation algorithm should be equipped with advanced mathematical methods to detect efficiency and identify the systematic errors in measurements. The statistical techniques like normalized residuals and chi-square tests are routinely practiced for detection and identification of bad data and well known in technical literature on state estimation problems. The application of chi-square test is only to ensure the presence of larger error as it is unable to locate the position of gross error whereas normalized residual test in connection with chi-square test possesses both identification and detection capabilities.

The main aim of this work is to employ a new statistical method that retains the preferable properties of bad data detection strategy. Our technique work is based on single test statistic and works well in presence of smaller amount of single gross errors. In this technique, both conventional statistical tests have been combined in well-established studentized residual test for the detection of outlier in measurements. It is also worth noting that the
serial compensation methodology is adopted rather than conventional serial elimination method for treatment of multiple gross errors.

**Literature review**

State estimation is performed in modern control centers of the power systems as one of the guarantee measures to compute network state conditions [1]. Literature on estimation of state parameters of the power system is well documented at both power transmission and distribution levels since the pioneering study of Schweppe et al. [1-4] and others [5]. As the presence of bad data cause biased results, therefore a lot of efforts are addressed in literature for elimination of these values both at pre and post estimation levels. A brief overview of pre estimation techniques can be seen in [6], [7] and [8]. The post estimation methods are residual based methods which can be categorized in two types of the methods, one that deals on robust estimation techniques [9-10], and others that take into account conventional weighted least square (WLS) algorithm of state estimation along with elimination strategy [11]-[12]. Although robust methods of estimation work well in presence of bad data but they are not so efficient. An efficient algorithm is another requirement for state estimation algorithm since power system is high dimensional non-linear problem. The WLS algorithm based, bad data detection techniques rely on statistical properties of the estimated residuals [13]. The chi-square test is used to detect the presence of bad data but is unable to compute the exact location of the bad value. The other widely used procedure is normalized residual test [14]. The value with largest normalized residual is suspected to be bad and corresponding value is removed from the measurements. All aforementioned methods are based on assumption of independent measurements from the measuring devices whereas state estimation and the largest normalized test has been developed in [15]-[16] by considering the dependent structure of measurements. The state estimation problem is similar to the multivariate nonlinear estimation problem [17]. The basic idea of bad data is originated from statistical outlier detection theory. Therefore number of statistical efficient tests can be implemented for the detection of gross errors in measurements. In statistical view point, we have therefore proposed a new detection and identification gross error method that is constructed on modification of assumptions of the state estimation model.

**Contribution**

The contribution of this work is twofold:

- Firstly, it provides new statistical procedure for handling the bad data in power system state estimation measurements. It is more efficient in sense that it combines both conventional methods in single test statistic for treatment of the gross errors.
- Secondly, serial compensation is adopted in this research rather than conventional serial elimination strategy for tackling the multiple gross errors where bad value are substituted by its estimates and are not deleted from the measurements.

**Paper organization**

The rest of this article is sequentially organized as follows: Section 2 describes the mathematical formation of bad values handling strategy; section 3 provides the procedure for construction of new test statistic for detection and identification of gross errors; section 4 explains methodology for tackling the multiple bad measurements; section 5 presents the application of proposed technique on six-bus data and results summary. Finally, section 6 provides the conclusions.

**Description of mathematical model and measurements system**

The exact values from the measuring devices in power system can never be known. Let us consider that the available measurements vector is represent by the vector $Z$ for $N$ bus system where $Z \in \mathbb{R}^{M \times 1}$ and system is over determined i.e. $M > (2N - 1)$. The vector $Z$ is therefore can be written as

$$Z = [z_1, z_2, ..., z_M]^T$$

Since these measurements are subject to random errors therefore the measurement vector $Z$ in model form can be written as:

$$Z = h(X) + \epsilon$$  \hspace{1cm}(1)$$

where

$$h(X) = [h_1(X), h_2(X), ..., h_M(X)]^T$$

is measurement function values of unknown state vector $X$.

The measurement function $h(X)$ can be a set of linear and non-linear equations for DC and AC network models respectively. As we are assuming the case of AC network model in our study therefore $h(X)$ considering $\pi$-model for each line, can be found as [5]:

- Real power and reactive power injection at bus $k$ are given by:

  $$P_k = V_k^2 \sum_{m=1}^{N} V_m \left( G_{km} \cos \theta_{km} + B_{km} \sin \theta_{km} \right)$$
  $$Q_k = V_k^2 \sum_{m=1}^{N} V_m \left( G_{km} \sin \theta_{km} - B_{km} \cos \theta_{km} \right)$$

- Real and reactive power flows from bus $k$ to $m$ are given by:

  $$P_{km} = V_k^2 G_{km} - V_k V_m G_{km} \cos \theta_{km} - V_k V_m h_{km} \sin \theta_{km}$$
  $$Q_{km} = -V_k^2 h_{km} + V_k V_m h_{km} \cos \theta_{km} - V_k V_m G_{km} \sin \theta_{km}$$
Flow magnitude of current in line from bus \( k \) to bus \( m \) is given by:

\[
I_{km} = \left( \frac{P_{km}^2 + Q_{km}^2}{V_k V_m} \right)^{1/2}.
\]

\[X = [x_1, x_2, ..., x_{2N-1}]^T\] is a vector of unknown values comprise of phase angles and voltage magnitudes for all bus nodes in the network.

\[e = [e_i, e_{i+1}, ..., e_{2N-1}]^T\] is the measurement error vector, having the following assumptions:

\[E(e_i) = 0, \quad \text{where } i = 1,2, ..., M\]

\[Cov(e) = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \]

Here \( \sigma^2_i, \quad i = 1, 2, ..., M \) is the variance of the error term which reflects the accuracy standard related with \( i \)th meter and \( \sigma^2 \) is common variance term. It is worth mentioning that the assumption considered above is contrary to the assumption used in conventional method. It was subsequently observed that formulation of the problem in this way does not alter the basic algorithm but improves its performance in treatment of bad data. The term \( \sigma^2 \) is rarely known and is estimated from its maximum likelihood estimate. The joint distribution of the error vector is multivariate normal with mean and covariance matrix as given above.

**Estimation method and objective function**

The method of maximum likelihood is generally used to find the value of unknown state vector. Since the measurements in vector \( Z \) are assumed to be independent, therefore the joint distribution can be written as:

\[g(z) = \prod_{i=1}^{M} g(z_i) \quad (2)\]

Hence the objective in state estimation is maximizing the log likelihood function of (2) in order to find most likely state vector \( X \) for the given observed data. Since under the normality assumption of error vector, maximum likelihood estimate is equivalent to estimates obtained by ordinary least squares method (OLS). Thus the objective is reduced to optimizing the following convex objective function:

\[
\phi(X) = \text{minimize} \left[ Z - h(X) \right]^T D \left[ Z - h(X) \right] \quad (3)
\]

where

\[
D = \begin{bmatrix} \frac{1}{\sigma^2_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma^2_u} \end{bmatrix}
\]

is weight matrix and known in power system state estimation problem.

Thus the problem is reduced to common procedure of weighted least squares (WLS). Due to nonlinear form of measurement functions an iterative procedure will be employed for the solution of most likely state vector \( X \). In general, the Gauss Newton algorithm is established routinely in state estimation problem for calculating the unknown state vector [13].

The value of \( X \) at \( i \)th iteration in the Gauss Newton method is obtained as:

\[
X^{i+1} = X^i - \left[ G(X^i) \right]^T g(X^i) \quad (4)
\]

where

\[
g(X^i) = -H^T(X^i) D \left[ Z - h(X^i) \right]
\]

\[
G(X^i) = \frac{\partial g(X^i)}{\partial X} = H^T(X^i) D H(X^i)
\]

where \( G(X) \) is symmetric, sparse and positive definite gain matrix.

**Gross errors processing**

The detection of the gross errors in measurements is the most significant sub part of state estimation algorithm. Our objective function in (3), can be rewritten as

\[
\phi(X) = \sum_{i=1}^{M} w_i \left( z_i - \bar{z}_i \right)^2 \quad (5)
\]

where \( \nu = M - 2N + 1 \)

which follows the \( \chi^2 \)-distribution with \( \nu \) degree of freedom. If the estimated value of (5) at state vector \( X \) is equal or greater than that of set critical point which indicates that at least one gross error is suspected, otherwise not. The \( \chi^2 \)-test is only the indication for the presence of bad measurements. It does not alarm which particular value is affected by gross error. Another statistical test which is carried out in the state estimator algorithm is the largest normalized residual (LNR) test.
using sensitivity study; variances of estimated residuals are obtained from corresponding diagonal elements of the covariance matrix for residuals. Normalized residuals are then found as:

$$r^n = \frac{|r_i|}{\sigma_i}$$  \(\sigma_i\) is the variance of \(i^{th}\) residual. Since each residual follows the standard normal distribution, therefore the gross error exist if \(|r^n| > 3\). Thus in this test, the value corresponding to largest normalized residual is deleted from the data.

A number of problems associated with these tests in statistical view point. The LNR test in conjunction with chi-square test is unable to detect and identify the smaller amount of gross errors in measurements. From statistical perspective, these tests do not have greatest capabilities for treating the gross errors in measurements i.e. not so powerful. Therefore, for achieving best performance, the LSR test for detecting and identifying the gross errors in process measurements is proposed in Section-III

### Proposed LSR test for detection and identifying the gross errors

Alike the LNR and chi-square tests, the proposed method also utilizes the statistical properties of calculated residuals. The LNR test has been modified in such way that it takes into account the mean square error (MSE) of the given observations in its calculation. In this way, considering the addition factor of common variance resulting out its greatest capabilities for detecting and identifying the bad data. The idea of the gross error detection is originated from statistically outlier detection methods. Under the given assumptions, inherent variability in measurements follows the normal distribution with certain mean and variance. Therefore most of standardized residual falls within confidence interval except the erroneous measurements. In power system, observations are gathered from different meters and the value is called bad or erroneous if more than expected error of its measuring device is included in it. Mathematically erroneous measurement can be written as:

$$z = \bar{z} + k\sigma + \varepsilon$$  \(\bar{z}\) is observed measurement from the measuring device, \(\bar{z}\) is true measurement and \(k\sigma\) is amount of gross error in terms of standard deviation. Therefore (1), for the grossly data can be written as:

$$Z = h(X) + h\mu + \varepsilon$$  (8)

Where \(b = k\sigma\) and \(\mu\) is a vector with zeros entries except unity at \(i^{th}\) position.

The structure of (8) shows that presence of bad data in \(i^{th}\) measurement does not affect scale of the distribution but resulting the shift in mean of the observation. Since formulation of state estimation problem is similar to nonlinear WLS problem in regression analysis, therefore results of nonlinear regression can be considered for the treatment of gross errors.

The equivalent transform form follow from of (3) can be derived as:

$$\phi(X) = \text{minimize} \left[ Z_i - h_i(X) \right]^T \left[ Z_i - h_i(X) \right]$$  (9)

Where

$$Z_i = \sqrt{D} \bar{Z}\text{ and } h_i(X) = \sqrt{D} h(X).$$

The desired value of state vector from (9), can be obtained by least squares (LS) method and is given by:

$$\hat{X} = \left( H_i^T H_i \right)^{-1} H_i^T Z_i$$

Because of nonlinearity of objective function as given in (9), the Gauss Newton algorithm is routinely applied for the solution of this estimate.

From (1), now let us define the residuals as:

$$\hat{e} = Z - \hat{Z}$$  (10)

where \(\hat{Z} = H\hat{X}\), and \(E[\hat{Z}] = Z\) (unbiased property).

The (10), implies that:

$$\hat{e} = (I - H(H^T D H)^{-1} H^T) e = S e$$  (11)

Where \(S\) is called sensitivity matrix which is idempotent and symmetric.

Thus the variance covariance matrix from (11) can be derived as:

$$\text{Cov} (\hat{e}) = \sigma^2 SD$$  (12)

Hence it yields from (11) and (12)
\[ e_i \sim N(0, \sigma^2 K), \text{ where } K = SD \]

The \(^i\)th studentized residual is defined as:

\[ \hat{e}_i = \frac{e_i}{\hat{\sigma} \sqrt{k_{ii}}}, \quad i = 1, 2, 3 \ldots M \]

where \( k_{ii} \) is \(^i\)th diagonal element of variance covariance matrix given in (12) and \( \hat{\sigma}^2 \) is maximum likelihood estimate of common variance \( \sigma^2 \).

After finding the optimum value of state vector, say \( \hat{x} \), the value of \( \hat{\sigma} \) is computed independently as follow:

\[ \hat{\sigma} = \sqrt{\frac{\sum w_i (z - h_i(x))^2}{\nu}}, \quad \text{where } \nu = M - (2N - 1). \]

Hence the ratio given in (12), has student t-distribution with \( \nu \) degree of freedom. Since power system state estimation is a high dimensional problem, therefore each studentized residuals in (12), has approximately \( N(0,1) \) for larger set of observations.

Therefore the measurement is considered as bad data value if \(|\hat{e}_i| > \tau\) where \( \tau \) is the set critical point and is appropriately taken as 3 since 99.8% observations from any specific meter \( i \), are believed to have fallen within the range \([-3\hat{\sigma}, 3\hat{\sigma}]\). Once the bad measurement \( z \) is confirmed, it can be substituted by its estimate. The estimated value can be calculated by from (11) and is given by:

\[ z_i^{est} = z_i - \frac{e_i}{s_i} \]

Where \( e_i \) is corresponding residual of bad measurement and \( s_i \) is the \(^i\)th diagonal element of sensitivity matrix. Moreover it can be observed that standardized residuals in (12) are different from the conventionally used residuals as given in (6). These modified residuals consider the MSE which provides additional information in the form of the closeness of values to its estimates.

**Strategy for multiple bad data**

The described technique in previous section pertains to handle the single bad value in measurements. However, its power like the other conventional tests is decreased dramatically in presence of multiple bad values.

For the treatment of multiple non-interacting bad measurements, a sequential application of the LSR test is proposed, such that bad measurement is compensated for error but not deleted. In this scheme, the bad value is treated one at a time i.e. if a bad value is confirmed, the corresponding measurement is corrected by its estimate and the algorithm is repeated until no further bad values are found. The flowchart of this scheme is shown in Figure-1.

It should be noted that the successful application of above described scheme depends upon the association among the multiple bad values. This scheme performs well in presence of weakly correlation existing among the bad measurements i.e. the detection of a bad value is not related to existence of other bad value. If this is not the case, then contrary observations may be resulted however the scheme achieves best results when numerical efficiency is considered another important requirement which is always true in power system state estimation.

![Figure-1](image-url)  
*Figure-1. The methodology of LSR test for tackling the multiple bad measurements.*

**Application of proposed LSR test to six-bus power system data**

In this section, we have applied our proposed residuals based test statistic to six-bus power network. The network topology and data on six-bus power system can
be seen in [18]. The detail on main properties of 6-bus test network is summarized in Table-1.

- **Table-1.** The characteristics of six-bus test network.

<table>
<thead>
<tr>
<th>System characteristics</th>
<th>Total number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of buses</td>
<td>6</td>
</tr>
<tr>
<td>Number of power lines</td>
<td>11</td>
</tr>
<tr>
<td>Number of observations</td>
<td>62</td>
</tr>
<tr>
<td>Number of parameters to be estimated</td>
<td>11</td>
</tr>
</tbody>
</table>

The unknown vector $X$ is comprised of phase angles and voltage magnitudes of all buses. The bus one is assumed to be a referenced bus so the defined phase angle for this is zero. The process data on bus voltages, real and reactive power injections, real and reactive power flows, are converted to per unit (pu) system measurements by dividing them $\frac{230kV}{230kV}, \frac{100MV}{100MV}$ and $\frac{100MVA}{100MVA}$ respectively. The available data from the source [18], are true values i.e. not perturbed by random errors. For the utilization of the weighted least squares (WLS) algorithm, weights are assumed to each measurement according to degree of accuracy of its corresponding measuring device in terms of standard deviation (Std). These weights, like the other measurements, are converted in per unit system and are shown in Table-2.

- **Table-2.** The assignment of weights to different measurements.

<table>
<thead>
<tr>
<th>Measurement Types</th>
<th>Std in pu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage measurements</td>
<td>$8.7 \times 10^{-3}kV$</td>
</tr>
<tr>
<td>Power injection measurements</td>
<td>$3 \times 10^{-3}MV$</td>
</tr>
<tr>
<td>Power flows measurements</td>
<td>$3 \times 10^{-3}MV$</td>
</tr>
</tbody>
</table>

A program in MATLAB has been written for estimating the state vector $X$ and computing the residuals as described in section 3. Generally, the algorithm of our residual base statistical test for detecting and locating the position of single error can be described in steps as follows:

- **Step I** Provide flat start values to state vector $X$ at iteration number $i = 0$, and chose the convergence number $\varepsilon$.

- **Step II** Compute $h(X^i), g(X^i)$ and $G(X^i)$ quantities according to (4) and get the new set of values for state vector $X$.

- **Step III** Find the convergence by computing $|X^{i+1} - X^i| \leq \varepsilon$, if it holds, calculate the residuals according to (12) and go to **Step IV** otherwise go to step II by taking $i = i + 1$.

- **Step IV** Perform the significance test for each residual by computing $|\varepsilon_i| > 3$ and indicate the measurement corresponding to significant residual.

The termination strategy as described in the algorithm is chosen to be $10^{-3}$ as convergence criterion for values of unknown vector $X$. The stream of noisy measurements is generated by adding random number from the normal distribution with zero mean and variance equal to corresponding weight as shown in Table-2. A grossly measurement is intentionally generated then by adding in it some multiple of its standard deviation. For checking the performance of our test, a bad value is generated by adding 10 times standard deviation in the measurement $P_1$ (real power injection at bus one). A larger multiple of standard deviation has been taken here due to smaller error variance. The results of our LSR test and LNR test without and with single grossly measurement are shown in Figure-2 and Figure-3, respectively.

- **Figure-2.** Residual plot of both tests without bad value.

- **Figure-3.** Residuals plot of both tests with single bad value.

The overlapping pattern of the residual series is indicated in Figure-1, showing equivalence between both strategies. Hence, in the absence of any bad value, both residuals behave almost similar and showing none of the significant residual. In the presence of single bad measurement $P_1$, our proposed residual test is showing significance value exactly at the position of grossly measurement whereas LNR test beside the erroneous residual also showing significant results for the other residuals. In this way, the bad data algorithm based on conventional LNR test has higher risk to exclude other values which are not actual bad values.

Our residual test statistic combines the information provided by chi-square test and LNR test in a single test statistic and has a greater power to detect smaller amount of gross error, since the performance of...
both conventional tests is being checked together with our LSR test which is analytically impossible. Here the joint testing of conventional tests means that once the chi-square will detect (because of its detection property only) the presence of gross error then LNR test will subsequently identify which measurement is bad. The test will be the powerful if it has higher probability to correctly confirm the bad value when actually bad data exist in measurements. Therefore; we have carried out a simulation study to validate the performance of LSR test in terms of its power. This power comparison is done under the assumption of single bad value among the measurements since none of these tests perform best in occurrence of multiple bad values which will be observed later. In this simulation experiment, the power of the test under the null hypothesis of no gross error in \( i \)th measurement is defined as:

\[
\text{Power} = \frac{\text{The count of bad data correctly identify in } i \text{th measurement}}{\text{The count of bad data simulated}}
\]  

(15)

To see the performance of both strategies in presence of smaller to larger gross error, the power is estimated in following steps:

**Step I.** A bad value is created intentionally in random chosen value \( P_{15} \) according to (7) at different values of \( k \) where \( k = 1, 2, \ldots, 10 \).

**Step II.** At each chosen value of \( k \), data is simulated 10000 times according to model (8) and solved by state estimation algorithm.

**Step III.** The power given in (15) is calculated i.e., how many times a given strategy rejects the null hypothesis of no gross error in \( P_{15} \), the result of this power comparison is shown in Figure 4.

**CONCLUSIONS**

The detection and identification tests of grossly measurements in power state estimation problem have been investigated by many researchers in terms of numerical efficiency but insufficient research is available on statistical rigor. By using the statistical properties of residuals, a new test has been presented in this paper. The test is the result of the modified assumptions of the model common used in power system state estimation. The developed test is actually the modification of LNR test such that it incorporates the information provided by chi-square test in its calculation. Based on the proposed test, a modified bad data algorithm strategy has been described. In this work, it has been shown that the proposed scheme has greatest power for treating the bad measurements as compared to joint implementation of conventional tests. Additionally, to save system from observability problem and information loss, we applied the given technique in sequential way by replacing the grossly measurements one at a time by their estimates rather than their deletion from the measurements.

**REFERENCES**


