Analysis of Bad Data Processing Methodologies in Power System State Estimation

Sarbajit Basu, *Student Member IEEE*, Olga Lavrova, *Member IEEE*, Satish Ranade, *Member IEEE*
Electrical and Computer Engineering Department,
Klipsch School of Electrical and Computer Engineering
New Mexico State University, Las Cruces, NM 88003

***Abstract* — The reliability of results generated by state estimation are heavily impacted by measurement errors, topology errors, system transients and bad data injections. Detection and identification of such errors in a power system is important not only to monitor instrumentation performance but also to ensure system resilience. Despite their extensive application in power system state estimation research, the context and potency of WLS based bad data detection techniques is often overlooked. A better knowledge and implementation of such techniques will significantly improve the results by such work.**

This paper discusses the theoretical background and efficacy for the same. Results of simulation performed on a wide variety of systems to analyze the performance of each are presented.

2021 North American Power Symposium (NAPS) | 978-1-6654-2081-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/NAPS52732.2021.9654772

I. Introduction

The purpose of state estimation algorithms' is to generate accurate estimates of a given set of measurements using a well-defined mathematical model. The presence of any errors or bad data in the measurement set affects the estimations derived from them which, in turn, would have an adverse influence on the decisions based on the same.

State estimation algorithms should be resilient to a) measurement errors b) parameter errors, c) structural error and d) bad data. Multiple research groups across the world have been deploying the Weighted Least Squares (WLS) based state estimation for their varied applications including work on identifying generation and load islands, tie-line flow verification and so on. However, such an extensive bank of derivative work often leads to research that while being accurate, do not hold true to statistical basis.

This article discusses the WLS state estimation method and the relevant Chi-squares test and the Largest Normalized Residual (LNR) test for error and bad data detection and identification. The mathematical basis, the advantages, and the shortcomings of the same are discussed with the hope that the future work in this domain will be more precise and have greater applicability when deployed in real-time applications.

II. Problem Formulation

The purpose of a static state estimator, as defined by Handschin et. Al [5] can be expressed using Fig. 1. To impart more clarity to this definition, the following terms are defined:

System state: bus voltage and angle measurements

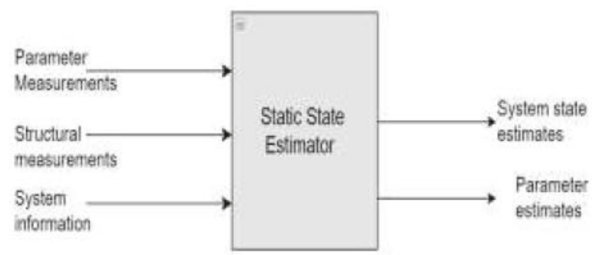

Fig 1: Function of Static State Estimator

Model structure: Telemetered switch and circuit breaker positions to determine network structure and meter location

Measurement error: Zero mean random vector error

Parameter error: Zero mean random vector pertaining to a small deviation of assumed values of a parameter measurement

Structure error: Error in system structure

Bad data: Large measurement error added to one or more measurements

Detection: Test to determine the presence of erroneous or bad data

Identification: Determination of erroneous or infected measurement

Redundancy: Ratio of number of measurements to the number of states

It can thus be declared that a properly designed state estimator should perform the following functions: a) excogitate structure b) estimate c) detect and d) identify. This article focuses on data collected at a single-time stamp with a focus on identifying that when is a measurement classified as erroneous or a bad data. While the paper does not explicitly work with the different kinds of error, it discusses the impact of the magnitude of an error or a bad data on different parameter measurements, and how the size of the system impacts the same.

III. Mathematical Model

A. The Model

Consider a linear model of the form:

$$y = Ax + \epsilon \qquad (1)$$

where

y = (n x 1) vector of linear, uncorrelated measurements with different measurements

A= (n x k) matrix of coefficients

X= (k x 1) vector of unknown parameters

ϵ = (n x 1) vector of 'error' random variables

The first and second order moment of ϵ are given as:

$$E(\epsilon) = 0$$
$$E(\epsilon\epsilon^t) = \sigma^2 R \qquad (2)$$

Where :

R = diagonal matrix of the model's measurement variances. The diagonal nature of the matrix signifies that the measurements are uncorrelated.

σ^2 is a scaling constant that impacts detection.

B. Least Squares Estimation

The objective of least squares estimation is to minimize the function:

$$J(x) = (y - Ax)^T R^{-1}(y - Ax) \qquad (3)$$

Solution of Eq. (2) can be obtained using the orthogonality condition of the residual as:

$$A^t R^{-1}(y - Ax) = 0$$

The best estimate of parameter x, $\hat{x}$, is given by:

$$\hat{x} = (A^t R^{-1} A)^{-1}(A^t R^{-1} y) \qquad (4)$$

C. Expected values of $\hat{x}$ and $\hat{y}$

The expected value of $\hat{x}$, designated by $E(\hat{x})$ is given as:

$$\hat{x} = (A^t R^{-1} A)^{-1} A^t R^{-1}(Ax + \epsilon)$$

$$E(\hat{x}) = E[(A^t R^{-1} A)^{-1} A^t R^{-1}(Ax + \epsilon)]$$

$$E(\hat{x}) = E\ \{x + ((A^t R^{-1} A)^{-1} A^t R^{-1} \epsilon)\}$$

$$E(\hat{x}) = x + M{*}0 = x$$

Where

$$M = (A^t R^{-1} A)^{-1} A^t R^{-1}$$

The expected value of y, *E(y)*, can be calculated as:

$$E(y) = E[Ax + \epsilon] = y$$

The calculated value of the measured quantities is:

$$\hat{y} = A\hat{x}$$

Using equation (1), we can write the expected value of ($\hat{y}$), as:

$$E(\hat{y}) = A\ E(\hat{x}) = y \qquad (5)$$

Thus, an unbiased estimate of y can be generated.

IV. Chi-Squared Testing

A. Expected value of $J(\hat{x})$

The expected value of $J(\hat{x})$ can be expressed as:

$$J(\hat{x}) = (y - A\hat{x})^T R^{-1}(y - A\hat{x})$$

$$y - A\hat{x} = Ax + \epsilon - A(A^t R^{-1} A)^{-1} A^t R^{-1}(Ax + \epsilon)$$
$$= \epsilon\ \{I - A(A^t R^{-1} A)^{-1} A^t R^{-1}\} \qquad (6)$$

Putting Eq (6) in Eq. (3) and using the properties of idempotency, we get:

$$J(x) = \epsilon^t\ [R^{-1} - R^{-1} A(A^t R^{-1} A)\ AtR^{-1}]\ \epsilon \qquad (7)$$

Letting $B = R^{-1} - R^{-1} A(A^t R^{-1} A)\ AtR^{-1}$,

$$J(x) = \epsilon^t B \epsilon \qquad (8)$$

Taking the expected values of Eq. (8), and assuming all the ϵ are uncorrelated, we get:

$$E(J(\hat{x})) = o^2\ trace(RB)$$

Since, $RB = I - A(A^t R^{-1} A)\ AtR^{-1}$,

$$E(J(\hat{x})) = o^2\ (n - trace(RB)) \qquad (9)$$

Trace (RB) = trace (BR), so:

$$E(J(\hat{x})) = o^2\ trace\ (n - trace(I))$$

$$E[J(\hat{x})] = \sigma^2(n - k) \qquad (10)$$

Thus, an unbiased estimator can be constructed to estimate the value of σ^2. This estimator is considered unbiased.

B. Chi-squared property of $J(\hat{x})$

A chi-squared distribution is obtained when the sum of the squares of multiple random normally distributed variables is calculated.

For a set of m independent normally distributed random variables $X_1, X_2 \ldots\ldots X_m$,

$$Y = \sum_{i=1}^{m} {X_i}^2 \sim {\chi_m}^2 \tag{11}$$

Where m becomes the degrees of freedom of the chi-squared distribution.

If all the error variables ϵ, and hence the measurement residuals, follow a normal probability distribution, the transformation:

$$\frac{\epsilon - \hat{\epsilon}}{\sigma R}$$

Produces a unit normal variable, where $\hat{\epsilon}$ is the mean of the error variables.

Thus

$$\frac{J(\hat{x})}{\sigma^2} = \frac{1}{\sigma^2}[\epsilon^t R^{-1} \epsilon] \tag{12}$$

Is then chi-square.

And,

$$\frac{1}{\sigma^2} E[J(\hat{x})] = n - k = Z \tag{13}$$

Where Z becomes the degrees of freedom for the given system. By extension, therefore $\frac{J(\hat{x})}{\sigma^2}$: is chi- squared with (n-k) degrees of freedom.

C. Hypothesis testing of $J(\hat{x})$

Based on the normal distribution of the residuals, a statistical test based on hypothesis testing can be devised. The test is designed to test the validity of the null hypothesis H_0 against the alternative hypothesis H_1 such that the system is free of type-1 errors, i.e., H_0 is declared rejected when it was true.

If there is:

$$H_0 : \sigma^2 = 1 \tag{14}$$
$$H_1 : \sigma^2 > 1$$

$J(\hat{x})$ can be compared against ${\chi^2}_{n-k,\ b}$.

If

$$\frac{J(\hat{x})}{\sigma^2} \leq {\chi^2}_{n-k,\ b} \tag{15}$$

for a given level of confidence, b, on the set of measurements, then it can be said that the probability of a type-1 error is b.

Alternatively, the probability of H_0 being accepted is (1-b).

V. Residual Based Testing

A. Properties of measurement residual

The WLS estimate, $\hat{x}$, is based on:

$$J(\hat{x}) = (y - Ax)^T R^{-1} (y - Ax)$$

$\hat{x}$ is based on the optimality condition:

$$\frac{\partial J}{\partial x}\Big|_{x = \hat{x}} = H^T R^{-1}(y - Ax) = H^T R^{-1} r = 0 \tag{16}$$

where H = $\partial A / \partial x$ is the Jacobian matrix.

The residual vector $\hat{r}$

$$\hat{r} = y - \hat{y} = A(x) + \epsilon - A(\hat{x}) = \Omega_s \epsilon \tag{17}$$

with the residual sensitivity matrix Ω_s as:

$$\Omega_s = (I - H(H^T R^{-1} H)^{-1}) \epsilon$$

And the residual covariance matrix Ω_c as:

$$\Omega_c = R - H(H^T R^{-1} H)^{-1} H^T = \Omega_s R \tag{18}$$

The weighted residual r_w and normalized residual r_n are:

$$r_w = \sqrt{diag(R)}\, r\ , r_n = \sqrt{\mathrm{diag}(\Omega_c)} r$$

The normalized residual sensitivity matrix Ω_{sn} is given by:

$$\Omega_{sn} = \sqrt{[\mathrm{diag}(\Omega_c)]^{-1}} \tag{19}$$

It can be proven that:

$$|\Omega_{sn,\ jk}| > |\Omega_{sn,\ kk}| \tag{20}$$

Where $|\Omega_{sn,\ jk}|$ is the element in Ω_{sn} at the j^{th} column and k^{th} row and $|\Omega_{sn,\ kk}|$ is the element in Ω_{sn} at the k^{th} column and k^{th} row.

Thus, the value of the largest normalized residual would always correspond to the bad data point.

B. Hypothesis testing of r_n

The detection of errors and bad data is designed as a hypothesis testing problem given as:

H_0: no bad data or errors are present

H_1 : H_0 is not true

If

$$|r_{\hat{n}}, k| < \gamma \qquad (21)$$

H_0 is true and the threshold γ can be chosen. Otherwise H_0 is rejected or H_1 is accepted.

IV. Numerical Tests

Extensive simulations have been carried out to test the performance and reliability of all the testing methods discussed here. The objective was to identify the error detection and identification precision of each algorithm. Tests are performed on the 6,9,14,24, 39,57 and 118 bus system.

To ensure that the estimates generated were extremely accurate, high measurement redundancy was used. For any bus system, the number of states is always the sum of the number of bus voltage and angle measurements, except for the reference bus voltage. Work has been conducted [3] to reduce this dependency, but that is subject to presence or access to PMU measurements and requires further validation.

Error is added to one measurement at a time, of each type, for each bus case, as a multiplier of the corresponding measurement type's standard deviation. Larger error values are easier to detect those easier values and can often be removed by basic data pre-processing. For the sake of fairness, all parameter measurements of each type were subject to the analysis. This becomes even more relevant when working with reference bus measurements because all system parameter measurements are made based on that reference. Dependent parameters such as power injection and line flow also show considerable promise in this domain because this would highlight the impact of the independent measurements on the same, and the impact of error on the former for the same.

V. Analysis & Results

For best results, it becomes very important to design a state estimator with high accuracy as the generated estimates influence the whole error identification process. By carefully choosing the covariance matrix and proper Jacobian calculation, this can be easily achieved.

Different measurements were considered for each bus system, and it was found that, for a single erroneous parameter measurement, the threshold would enable detection of bad data greater than a standard deviation multiplier of 7.5. This result was consistent for all the bus cases.

As shown in Figure 2(a), each voltage measurement, taken one at a time, follows a similar pattern of impacting the values of the chi-squared statistic, when the error added to the same is varied. Active power measurements, as shown in figure 2(b), while following a similar pattern, show significant differences in the amount of influence based on them being a PV bus or a PQ bus or slack bus.

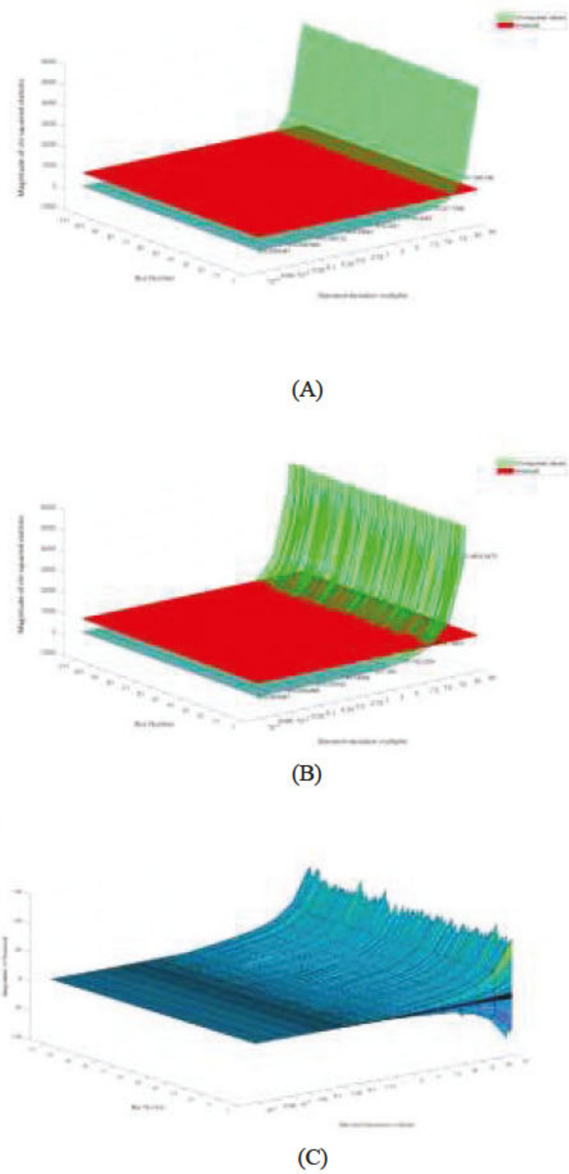

Fig. 2. For 118 bus case and one measurement at a time, a) Chi square statistic variation with variation of error signal added to each voltage measurement b) Chi-square statistic variation with variation of error signal to each active power injection measurement c) Variation of residue corresponding to each active power injection measurement when the error signal added to each is varied.

The residual test, when considered for voltage and angle measurements, revealed that that the corresponding injection and flow measurement residuals are impacted when the former is subject to an error. For a standard deviation multiplier greater than or equal to 0.1, the residual corresponding to the erroneous measurement, was significantly higher than the others, having at least a difference of a magnitude of 10. For multiplier greater than 5, the corresponding residue becomes significantly larger. The dependent residuals increase too, but still maintain a notable difference from the impacted measurement. For multipliers less than 0.1, the impact on the residue is very small, and in certain cases, the residue is greater on the dependent injection measurement than on the actual measurement, thereby causing wrong erroneous measurement identification.

When line flow and injection measurements were considered as show in figure 2(C), both the active and reactive power measurements, in each case, were considered together as the same sensor and telemetry device would be under question. Since they don't influence other measurements, the impact of error/bad data from them on the system only due to them is minimal. For standard deviation multiplier less than 0.1, it points to the inflicted measurement. The rest of the findings were like that of voltage and angle measurements.

VI. Conclusions

This article discusses the statistical basis behind the WLS based state estimation and the classically used techniques for detection and identification of measurement error and bad data. The tests have been subject to extensive simulation studies and thus implemented on a wide number of bus systems with varying number of parameter measurements. To ensure estimate accuracy, high fidelity was enforced on the models. The Chi-square test does not detect the bad data point thereby limiting its application. The LNR test holds superior to the former because it can detect the presence of bad data as well as identify the corresponding measurement number. Based on the definitions of errors and bad data explained in this paper, it was found that the chi-square test and LNR test can identify the presence of all bad data, but it failed to identify any errors. However, most of the measurements in question are in per unit values, and a minor digression in such measurements can lead to more gross errors.

A better understanding of these techniques would help current researchers in the domain of state estimation and thereby generate more exciting results, subject to computational complexity, level of trust on measurement devices and the expanse of the test network.

Acknowledgement

This work was partially supported by the NSF Grants \#OIA-1757207 (NM EPSCoR), HRD-1345232, HRD-1914635 and funding from Sandia National Laboratories Campus Executive (CE) LDRD Supplemental Project 20-0656, and award number GR0006770, and funding from the Electric Utility Management Program (EUMP) at the New Mexico State University.

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The views expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

References

[1] A. Abur and A. Gomez-Exposito, "Power System State Estimation: Theory and Implementation", book, Marcel Dekker, 2004.

[2] J. F. Dopazo, O. A. Klitin and A. M. Sasson, "State estimation for power systems: detection and identification of gross measurement errors", Paper X-2, PICA Conference Minneapolis, 1913

[3] George N. Korres, Nikolaos M. Manousakis, 'State estimation and bad data processing for systems including PMU and SCADA measurements', Electric Power Systems Research, Volume 81, Issue 7, 2011, Pages 1514-1524,

[4] H. M. Merrill and F. C. Schweppe, "Bad data suppression in power system static-state estimation", IEEE Trans., PAS-90, pp. 2718-2725, 1971

[5] E. Handschin, F. C. Schweppe, J. Kohlas, and A. Fiechter, "Bad Data Analysis for Power System Static State Estimation", IEEE Transactions on Power Apparatus and Systems, vol. PAS-94, Mar/Apr 1975.

[6] A. Monticelli and 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

[7] A. Monticelli, "Electric Power System State Estimation", Proceedings of the IEEE, vol. 88, no 2, February 2000.

[8] F. Aboytes and B. J. Cory, " Identification of Measurement, Parameter and Configuration Errors in Static State Estimation ", Proceedings of the IEEE PICA Conference, 1975.

[9] F. C. Schweppe, J. Wildes and D. P. Rom, "Power System Static State Estimation", IEEE Transactions on Power Apparatus and Systems, vol. PAS-89, pp. 120-135, Jan 1970.

[10] E. W. S. Ângelos and E. N. Asada, "Bad data processing for real-time equivalent networks," 2015 IEEE Power & Energy Society General Meeting, 2015, pp. 1-5, doi: 10.1109/PESGM.2015.7286458.

[11] N. G. Bretas, S. A. Piereti, A. S. Bretas and A. C. P. Martins, "A Geometrical View for Multiple Gross Errors Detection, Identification, and Correction in Power System State Estimation," in IEEE Transactions on Power Systems, vol. 28, no. 3, pp. 2128-2135, Aug. 2013, doi: 10.1109/TPWRS.2012.2234768.

[12] A. C. Aitken, "On Least Squares and Linear Combinations of Observations", Proc. Royal Society of Edinburg, 1935, vol. 35, pp. 42-48.

[13] I. Idehen and T. J. Overbye, "PMU Time Error Detection Using Second-Order Phase Angle Derivative Measurements," 2019 IEEE Texas Power and Energy Conference (TPEC), College Station, TX, USA, 2019, pp. 1- 6. doi: 10.1109/TPEC.2019.866213

[14] I. N. Kolosok, E. S. Korkina and A. E. Mahnitko, "Detection of systematic errors in PMU measurements by the power system state estimation methods," 2015 56th International Scientific Conference on Power and Electrical Engineering of Riga Technical University (RTUCON), Riga, 2015, pp. 1-4

[15] M. Göl and A. Abur, "A modified Chi-square's test for improved bad data detection," 2015 IEEE Eindhoven PowerTech, 2015, pp.1-5, doi: 10.1109/PTC.2015.7232283.

[16] [Quiroz, J. E., Reno, M. J., Lavrova, O., & Byrne, R. H. (2017, April). Communication requirements for hierarchical control of volt-Var function for steady-state voltage. In 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT) (pp. 1-5). IEEE.