Elsevier Editorial System(tm) for Electric Power Systems Research
Manuscript Draft

Manuscript Number:

Title: Robust WLS Estimator using Reweighting Techniques for Electric Energy Systems

Article Type: Research Paper

Keywords: Weighted least squares, power system state estimation, outlier detection, dependent Gaussian measurements.

Corresponding Author: Dr. Eduardo Caro, Ph.D.

Corresponding Author's Institution: Politechnical University of Madrid

First Author: Eduardo Caro, Ph.D.

Order of Authors: Eduardo Caro, Ph.D.; Roberto Minguez, Ph.D.; Antonio J. Conejo, Ph.D.

Abstract: The state estimator is a key tool in the operation of any real-world electric energy system. In this paper, a state estimator based on a weighted least squares model is proposed which is robust against outliers.
This algorithm presents two relevant features: robustness that is achieved by readjusting measurement weights,
and accuracy that is attained by considering measurement dependencies.
The proposed method is tested in a realistic system and the results are analyzed using Design of Experiments and ANOVA techniques.

# RESEARCH HIGHLIGHTS

A state estimator based on a weighted least squares model is proposed which is robust against outliers.

The proposed RWLS estimator takes into consideration measurement dependencies to improve accuracy and its weights are automatically readjusted to increase robustness.

A realistic 118-bus system is studied and the results are analyzed using Design of Experiments and ANOVA techniques.

Available online at www.sciencedirect.com

ScienceDirect

Electric Power Systems Research 00 (2012) 1–17

Electric Power
Systems Research

# Robust WLS Estimator using Reweighting Techniques for Electric Energy Systems

Eduardo Caro[1a], Roberto Mínguez[b], Antonio J. Conejo[c]

[a]*Universidad Politcnica de Madrid, Madrid, Spain*
[b]*Universidad de Cantabria, Cantabria, Spain*
[c]*Universidad de Castilla-La Mancha, Ciudad Real, Spain*

**Abstract**

The state estimator is a key tool in the operation of any real-world electric energy system. In this paper, a state estimator based on a weighted least squares model is proposed which is robust against outliers. This algorithm presents two relevant features: robustness that is achieved by readjusting measurement weights, and accuracy that is attained by considering measurement dependencies. The proposed method is tested in a realistic system and the results are analyzed using Design of Experiments and ANOVA techniques.



*Keywords:* Weighted least squares, power system state estimation, outlier detection, dependent Gaussian measurements.

## 1. Introduction

### *1.1. Motivation*

In any real-world electric energy system, it is required a Control Center to monitor and control the functioning of the network in real-time, ensuring operation security. To accomplish this task the Control Center needs to know accurately the actual state of the system (node voltages, power flows, etc.) at any time. These values are estimated by the State Estimator (SE).

The State Estimator is a mathematical algorithm which computes the most-likely state of the network, given a redundant set of measurements captured from the system. From the statistical point of view, the state estimation algorithm is a nonlinear multiple regression problem, whose parameters to be estimated are those which characterize the network state: node voltage magnitudes and angles.

[1]Corresponding Author:
E-mail address: Eduardo.Caro@upm.es, C/ José Gutiérrez Abascal, 2, 28006 Madrid (Spain). Phone: +34 913363149.

This estimated state is generally computed using the Maximum Likelihood Estimator, minimizing the weighted sum of the squared residuals (i.e., Weighted Least Squares approach). Once the most-likely state is obtained, the Control Center performs a "bad data detection and identification procedure" to detect and eliminate those measurements whose associated standardized errors are larger than a pre-established tolerance. The statistical tests commonly employed for these tasks are the $\chi^2$-test and the Largest Normalized Residual test, and are well established in the technical literature [1]. Once outliers have been removed, the nonlinear multiple regression problem is solved again, and the final state estimate is obtained.

If outliers are not properly detected or eliminated, the final estimate will be biased, and the Control Center will not have an accurate knowledge of the actual state of the system, leading eventually to an insecure operation of the network. For this reason, the detection and identification of bad measurements have a notorious relevance in the estimation process. In fact, an adequate and secure control is only achieved in the case that the SE procedure is robust enough to detect and eliminate the presence of corrupt measurements.

Traditionally, the "outlier elimination" problem is solved iteratively by detecting/removing suspected measurements and re-estimating the state disregarding the rejected data. These estimators are based on the weighted least squares, which shows a notable computational efficiency; however the lack of robustness deteriorates significantly their performance in the presence of bad measurements. Specifically, the presence of multiple conforming bad measurements in the measurement set may provoke a "masking effect": good measurements may be rejected whereas corrupted ones may not. This undesirable situation occurs when measurement dependencies are not properly modeled.

### *1.2. Aim*

The aim of this paper is to present a robust state estimator based on a weighted least squares regression, which carries out the estimation and the bad data detection/identification processes simultaneously by successively adjusting the weighting matrix and considering the effect of measurement dependencies. The obtained estimate does not require further bad measurement processing algorithms.

### *1.3. Literature Review*

The technical literature is rich in references concerning the state estimation problem, for instance, Schweppe and Wildes (1970) or [1]; and there is a significant number of references on outlier detection: [2, 3]; Cook and Weisberg (1982); [4]; Chatterjee and Hadi (1988); [5], [6], [7], . The previous works are focused mainly on the area of least squares linear regression. Other statistical models and estimation methods, such as reweighed techniques [8, 9, 10, 11], non-linear methods [12], heteroscedastic models [13], or some robust estimators [14, 15] have received comparatively less attention. Nevertheless, [16] report successful results from the application of the reweighted least deviances method developed by [10], to detect data related to hurricanes and typhoon on wave hindcast databases.

However, no so many works address the power system WLS estimator using adjusted measurement weights. The pioneering work reported in [17] proposes a method for readjusting the measurement variances based on the residuals of previous estimations. Reference [18] develops this approach, improving the computational efficiency and ensuring mathematical convergence. [19] propose an iterative reweighted least-squares estimator that is based on Givens Rotations and improves the robustness against outliers.

In [20], the weights of the WLS estimator are artificially manipulated, leading to a more robust estimator with the properties of the weighted least absolute value approach. Recently, in [21], the WLS regression is addressed using estimated weights based on the measurement variances.

However, to the best of the authors' knowledge and in the framework of power system state estimation, no prior study has considered an iterative re-adjusting of a non-diagonal measurement variance-covariance matrix.

### *1.4. Contribution*

The contribution of this paper is to provide an iterative state estimator that (i) takes into consideration the measurement dependencies, (ii) is robust against multiple outliers, and (iii) is computationally efficient.

### *1.5. Paper Organization*

The rest of this paper is organized as follows. Section 2 develops and formulates the Reweighted Least Squares Estimator considering measurement dependencies. Section 3 applies the Design of Experiments and ANOVA procedures to the considered estimation problem. Section 4 provides and analyzes results from three realistic case studies. Finally, Section 5 provides some relevant conclusions.

## 2. Dependent State Estimation Model

Any state estimator can be formulated as a nonlinear multiple regression problem, where the unknown parameters are the node voltage magnitude and angle of every node, represented by $V_i$ and $\theta_i$, respectively. These two sets of variables form the state vector $\boldsymbol{x} = [\boldsymbol{V\theta}]^T$. There are $n$ state variables.

The unknown parameters are estimated using the information provided by observations $\{z_1, ..., z_m\}$. These observations are captured from the system using measuring devices, and are related with $\boldsymbol{x}$ by means of a multifunctional vector $\boldsymbol{h}(\boldsymbol{x})$. Depending on the measurement type, the functions $h_i(\boldsymbol{x})$ differ. The expression for this function for voltage measurements, active/reactive power flow, and active/reactive power injection measurements are provided below:

$$h_{V_i}(\boldsymbol{x}) = V_i \tag{1}$$

$$h_{P_i}(\boldsymbol{x}) = V_i \sum_{j\in\Omega} V_j \Big( G_{ij}\cos(\theta_i - \theta_j) + B_{ij}\sin(\theta_i - \theta_j) \Big) \tag{2}$$

$$h_{Q_i}(\boldsymbol{x}) = V_i \sum_{j\in\Omega} V_j \Big( G_{ij}\sin(\theta_i - \theta_j) - B_{ij}\cos(\theta_i - \theta_j) \Big) \tag{3}$$

$$h_{P_{ij}}(\boldsymbol{x}) = V_i V_j \left( G_{ij}\cos\left(\theta_i - \theta_j\right) + B_{ij}\sin\left(\theta_i - \theta_j\right) \right) - G_{ij} V_i^2 \tag{4}$$

$$h_{Q_{ij}}(\boldsymbol{x}) = V_i V_j \left( G_{ij}\sin\left(\theta_i - \theta_j\right) - B_{ij}\cos\left(\theta_i - \theta_j\right) \right) + V_i^2 \left( B_{ij} - b_{ij}^S/2 \right) \tag{5}$$

where parameters $G_{ij}$ and $B_{ij}$ are the real and imaginary parts of the node admittance matrix, $\Omega$ is the set of all nodes, and constant $b_{ij}^S$ is the shunt susceptance of line *i*–*j*.

Each observation $z_i$ is subject to a measurement error, i.e.

$$z_i^{\text{true}} = z_i + e_i \tag{6}$$

where $z^{\text{true}}$ is the true value for the observation, $z_i$ is the actual measurement, and $e_i$ is the measurement error. This error has been traditionally modeled as an independent unbiased Gaussian-distributed random variable.

The factual metering infrastructure within substations results in significant statistical correlations between measurements. Works [22] and [23] numerically show that these correlations are significant, and its consideration may improve the quality of the final estimate. Therefore, hereafter measurement errors are assumed to be dependent Gaussian-distributed unbiased random variables. The dependence structure is modeled by means of definite-positive non-diagonal variance-covariance matrix $\boldsymbol{C}_z$, which can be easily computed using the Point Estimate method [23].

### *2.1. State Estimation*

Given the previous assumptions, the estimation of the state variables are obtained by minimizing the weighted sum of squared errors of the multiple nonlinear regression model, leading to a nonlinear optimization problem:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad J = [\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})]^T \boldsymbol{C}_z^{-1} [\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})] \tag{7a}$$

subject to

$$\boldsymbol{c}(\boldsymbol{x}) = \mathbf{0} \qquad \boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0} \tag{7b}$$

where the scalar $J$ is the objective function and $\boldsymbol{c}(\boldsymbol{x})$ and $\boldsymbol{g}(\boldsymbol{x})$ are the equality and inequality constraints modeling zero-injections nodes and physical operating limits, respectively. Note that matrix $\boldsymbol{C}_z$ is not diagonal, since the considered model takes into account the statistical correlations between measurement errors, but definite-positive as it is a variance-covariance matrix.

### 2.2. Reweighted Least Squares Formulation

The Weighted Least Squares model (7a) accounts for the heteroscedasticity of error measurements within the substation and their dependency. However, an alternative weighting matrix can be used accounting for i) heteroscedasticity, ii) dependency, iii) and degree of confidence related to each measurement.

In order to derive this weighting matrix let us consider vector $\boldsymbol{e} = \boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})$ and optimization problem (7). Using the Cholesky decomposition of matrix $\boldsymbol{C}_z$, the objective function in (7a) becomes

$$\begin{aligned} J &= [\boldsymbol{e}]^T \boldsymbol{C}_z^{-1} [\boldsymbol{e}] \\ &= \boldsymbol{e}^T \left(\boldsymbol{L}\boldsymbol{L}^T\right)^{-1} \boldsymbol{e} = \boldsymbol{e}^T \left(\boldsymbol{L}^{-1}\right)^T \boldsymbol{L}^{-1}\boldsymbol{e} \\ &= \left(\boldsymbol{L}^{-1}\boldsymbol{e}\right)^T \boldsymbol{L}^{-1}\boldsymbol{e} = \boldsymbol{u}^T\boldsymbol{u} \end{aligned} \tag{8}$$

where $\boldsymbol{u}$ a vector of standard independent normal random variables, with covariance matrix equal to the identity matrix $\boldsymbol{I}$, and $\boldsymbol{L}$ is the lower-triangular Cholesky factor of matrix $\boldsymbol{C}_z$.

The aim of most outlier detection methods is to determine whether or not a measurement should be considered as an outlier, without allowing for intermediate situations. In contrast, the method proposed in this paper, originally developed by [10], aims at empirically determining a diagonal matrix $\boldsymbol{W}$ to be included in model (8), i.e.

$$J_R = \boldsymbol{u}^T \boldsymbol{W} \boldsymbol{u} \tag{9}$$

where $w_{ii}$ is a weight for every observation ranging continuously from 0, for observations that are completely unreliable, up to 1, for observations that are completely reliable.

Considering (9), the objective function (8) becomes:

$$J_R = \boldsymbol{e}^T(\boldsymbol{L}^{-1})^T \boldsymbol{W} \boldsymbol{L}^{-1}\boldsymbol{e} = \boldsymbol{e}^T \boldsymbol{W}_R \boldsymbol{e}. \tag{10}$$

From (8) and (10), the following observations are in order:

- The measurement error vector $\boldsymbol{e}$ (dependent normal random variables) is transformed into a vector of independent standardized normal variables $\boldsymbol{u}$ [24].
- The objective function $J$ can be expressed as the sum of a set of squared independent standardized normal random variables.
- The objective function $J_R$ is computed as the *weighted* sum of squared independent standardized normal random variables. Each factor $u_i^2$ is multiplied by the weighting factor $w_{ii} \in [0, 1]$. If the $i$-th weighting factor is null ($w_{ii} = 0$), then the component $u_i^2$ is not considered in the objective function $J_R$. If, on the other hand, $w_{ii} = 1$, the component $u_i^2$ is fully considered in $J_R$.

The underlying idea of the RWLS method is to adjust empirically the weighting factors, based on the degree of confidence of each measurement. The coefficients for those measurements completely unreliable are adjusted to zero and, similarly, the weighting factors for those measurements completely reliable are adjusted to one.

The scheme of the iterative algorithm is the following:

- Step 0: Set $w_{ii} = 1;\ i = 1, \ldots, n$.
- Step 1: Compute an estimation using the objective function $J_R$.
- Step 2: New weights are computed using the last-fit residuals.
- Step 3: Repeat the two previous steps until convergence.

Several methods have been proposed to update the weights in step 2 (see [25], Chatterjee and Mächler (1997) or [10]). We use Tuckey's biweight:

$$w_{ii} = \begin{cases} \left[1 - \left(\frac{y_i}{6}\right)^2\right]^2 & \text{if} \quad |y_i| \leq 6 \\ 0 & \text{if} \quad |y_i| > 6 \end{cases} \tag{11}$$

where $y_i = \dfrac{u_i}{\sigma^*}$ is the standardized residual related to uncorrelated vector $\boldsymbol{u}$, and $\sigma^*$ is the scaled median absolute deviation estimator $\sigma^*$:

$$\sigma^* = \frac{\mathtt{median}_i |u_i|}{\Phi(3/4)} \approx \frac{\mathtt{median}_i |u_i|}{0.6745} \tag{12}$$

The selection of updating formula (11) is based on numerical simulations.

Thus, the RWLS problem formulation is:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad J = [\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})]^T \boldsymbol{W}_R [\boldsymbol{z} - \boldsymbol{h}(\boldsymbol{x})] \tag{13a}$$

subject to

$$\boldsymbol{c}(\boldsymbol{x}) = \boldsymbol{0} \qquad \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0} \tag{13b}$$

The algorithm of the RWLS estimator considering dependencies is:

1. *Initial non-dependent estimation.* An initial WLS estimation is performed to estimate the measurement variance-covariance matrix. Using the initial estimation obtained $\hat{\boldsymbol{x}}^{0)}$, matrix $\boldsymbol{C}_z^{0)}$ is computed via the Point-Estimate method in [23].
2. *Parameter initialization.* Weights $w_{ii}$ are set to one: $w_{ii}^{0)} = 1, \forall i \in \{1, m\}$.
   The iteration counter is set to 1, $\nu = 1$.
3. *Dependent state estimation*. The state estimation problem (13) is solved considering measurement dependencies and the reweighted matrix $\boldsymbol{W}_R^{\nu-1)}$. The obtained estimates are denoted as $\hat{\boldsymbol{x}}^{\nu)}$.
4. *Convergence checking.* Once the estimates $\hat{\boldsymbol{x}}^{\nu)}$ are available, if $||\hat{\boldsymbol{x}}^{\nu)} - \hat{\boldsymbol{x}}^{\nu-1)}|| > \varepsilon$ the estimation process continues in 5).
   Otherwise, a solution with a tolerance $\varepsilon$ is $\hat{\boldsymbol{x}}^{\nu)}$ and the algorithm concludes.
5. *Update weighting matrix $\boldsymbol{W}_R$.* Once the estimates of the state variables are available ($\hat{\boldsymbol{x}}^{\nu)}$), the weighting matrix ${\boldsymbol{W}_R}^{\nu)}$ is updated using (11).
   Set $\nu \longleftarrow \nu + 1$ and go to step 3).

Note that the computational efficiency of this algorithm can be improved by using $\hat{\boldsymbol{x}}^{\nu-1)}$ as the initial values of the estimation in step 3).

## 3. Design of Experiments

Section 2 above presents a novel algorithm to estimate the state of a power system in a robust manner. In this section, the statistical procedures "Design of Experiments" and "ANOVA" are briefly described as they are used (i) to analyze the performance of the proposed method, and (ii) to compare it with other existing methodologies.

The methods considered in this paper are listed below:

- WLS. The common Weighted Least Squares estimator is used as basic benchmark, using the $\chi^2$-test and the Largest Normalized Residual (LNR) test to detect and identify bad measurements, respectively. These algorithms are well-established in the technical literature [1]. The WLS results are the final estimates once the $\chi^2$-test and LNR test have been successfully passed.

- DWLS. The estimation, detection, and identification algorithms considering dependencies, proposed in [22] and [24], are employed to estimate the state and to detect bad measurements including the statistical correlation between measurements. Thus, the DWLS results correspond to the final estimate once the bad measurements have been removed.

### *3.1. Performance Assessment*

In order to rigorously determine which is the method with best performance, a design of experiments is carried out. This statistical procedure allows determining if the proposed algorithm is significantly better than the rest of the approaches with a pre-specified confidence level, and taking into consideration the dissimilarities among the considered measurement scenarios.

The performance of each method is assessed by means of the following metrics:

- Metric $\epsilon^V_{\text{MET},\omega}$, defined as the average absolute error of the voltage magnitude estimate for the $\omega$-th measurement scenario, considering the method MET, i.e.,

$$\epsilon^V_{\text{MET},\omega} = \frac{\sum_{i=1}^{n} |V^{\text{MET}}_{i,\omega} - V^{\text{true}}_{i,\omega}|}{n} \tag{14}$$

  Note that the previous metric is measured in p.u.

- Metric $\epsilon^\theta_{\text{MET},\omega}$, defined as the average absolute error of the voltage angle estimate for the $\omega$-th measurement scenario, considering the method MET, i.e.,

$$\epsilon^\theta_{\text{MET},\omega} = \frac{\sum_{i=2}^{n} |\theta^{\text{MET}}_{i,\omega} - \theta^{\text{true}}_{i,\omega}|}{n-1} \tag{15}$$

  The previous metric is measured in radians. Note that the considered reference angle is located at node 1, i.e., $\theta_1 = 0$ rad, for all the considered scenarios.

- Metric $\text{CPU}_{\omega}^{\text{MET}}$, defined as the required CPU time to obtain the final estimate considering the method MET for the $\omega$-th measurement scenario. Note that this metric is measured in seconds.

### *3.2. ANOVA Model*

The model employed in this design of experiments procedure comprises the factors "Method" and "Scenario". This model is described below,

$$y_{\text{MET},\omega} = \mu + \alpha_{\text{MET}} + \gamma_{\omega} + u_{\text{MET},\omega} \tag{16}$$

where $u_{\text{MET},\omega} \underset{iid}{\sim} N(0, \sigma^2)$ and:

$$\sum_{\text{MET}} \alpha_{\text{MET}} = 0 \quad ; \quad \sum_{\omega=1}^{n_{\omega}} \gamma_{\omega} = 0$$

where $\mu$ is the global effect, i.e., the average value of the considered metric $y_{\text{MET},\omega}$. Parameter $\alpha_{\text{MET}}$ is the main effect of the estimation method, and measures the increase/decrease of the average response of the factor "Method" (MET) with respect to the average level. Likewise, parameter $\gamma_{\omega}$ is the main effect of the block "Scenario" ($\omega$), and it measures the increase/decrease of the average response for all the methods with respect to the average level at the $\omega$-th measurement scenario. Finally, the random effect $u_{\text{MET},\omega}$ includes the effects of all other causes not modeled. Taking into consideration that the particularities of each measurement scenario may have influence on the method's performance, the effect "Scenario" is included in model (16).

The factors considered in this ANOVA analysis and the levels corresponding to each factor are provided in Table 1. Parameter $n_{\omega}$ stands for the number of measurement scenarios considered.

Table 1. ANOVA model: factors, blocks, and levels.

| | Names | Levels |
|---|---|---|
| Factors | Method (MET) | WLS<br>DWLS<br>RWLS |
| Block | Scenario ($\omega$) | 1,..., $n_{\omega}$ |

The background hypotheses of this model are: (i) normality, (ii) homocedasticity, and (iii) independence. To ensure this statistical properties, an appropriate diagnosis procedure is performed after the residual computation.

Since the aim of this study is to find the most accurate estimation method and to check if it is significantly different from the other methods, the following tests are performed:

$$\begin{cases} H_0 : \alpha_{\text{MET}} = 0, \forall \text{MET} \\ H_1 : \exists \text{MET} \mid \alpha_{\text{MET}} \neq 0 \end{cases} \quad \begin{cases} H_0 : \gamma_\omega = 0, \forall \omega = 1, ..., n_\omega \\ H_1 : \exists \omega \mid \gamma_\omega \neq 0 \end{cases}$$

The null hypothesis for the first test corresponds to the no statistically-significant influence of the method on the average performance. The alternative hypothesis establishes that it exists at least one method performing different from the average. The second test is analogous but related to the factor "Scenario".

To perform the above two statistical hypothesis testing, the ANOVA table is computed and analyzed. Table 2 provides the general structure for this table, particularized for the problem under consideration. The computation of the elements of the table is well-established in the technical literature [26].

Table 2. ANOVA table structure

| Source | Squared Sum | Deg. of freedom | Mean-sq | $F$-stat |
|---|---|---|---|---|
| Method | $SS_{\text{M}}$ | $3-1$ | $\hat{s}_\alpha^2$ | $\hat{s}_\alpha^2/\hat{s}_R^2$ |
| Scenario | $SS_{\text{D}}$ | $n_\omega - 1$ | $\hat{s}_\gamma^2$ | $\hat{s}_\gamma^2/\hat{s}_R^2$ |
| Residual | $SS_{\text{error}}$ | $(n_\omega - 1)(3-1)$ | | |
| Total | $SS_{\text{total}}$ | $3n_\omega - 1$ | | |

In case that the $F$-statistic for the factor "Method" on Table 2 is higher than the critical value for the $F$-distribution with two degrees of freedom and a given confidence level $(1-\alpha)$, then the related null hypothesis is rejected and it is concluded that the response variable is significatively affected by the factor "Method". Then, the average values for each method with the confidence intervals are plotted to determine which is the method with best performance.

## 4. Case Study

In this section, three case studies are analyzed to check the estimation accuracy and computational efficiency of the proposed state estimator.

The network under study is the 118-bus IEEE system[2]. To obtain statistically sound conclusions: (i) a set of one hundred randomly-generated measurement scenarios is considered, and (ii) an ANOVA procedure is performed to analyze the obtained results.

Each scenario involves:

**(i)** a random active/reactive power consumption level,

[2]Power Systems Test Case Archive. Available at: `http://www.ee.washington.edu/research/pstca/`

**(ii)** random locations of voltage and active/reactive power meters (ensuring observability of the whole system),

**(iii)** a random redundancy level, and

**(iv)** Gaussian-distributed random errors in all measurements, (standard deviations of 0.01 pu and 0.02 pu for voltage and power measurements, respectively).

The computational analyses have been performed using a Windows-based personal computer with a 64-bits eight-core i7 processor at 1.73 GHz and 8 Gb of RAM.

### *4.1. First Case Study: No Gross Errors*

In this case, the measurement vector $\boldsymbol{z}$ is free of gross errors. For each measurement scenario, the estimates for methods WLS, DWLS, and RWLS are computed and an ANOVA analysis is performed.

Table 3 provides the ANOVA analysis for metrics $\epsilon^V_{\text{MET}}$, $\epsilon^\theta_{\text{MET}}$, and $\text{CPU}^{\text{MET}}$, and Fig. 1 depicts the average value for these metrics and the confidence intervals for a 95% confidence level. The last column of Table 3 provides the p-value. The obtained value for this parameter allows deciding whether or not the null hypothesis $H_0$ should be rejected. If the p-value is smaller than 0.05, then the corresponding null hypothesis is rejected. Otherwise, there is not enough statistical evidence to reject $H_0$.

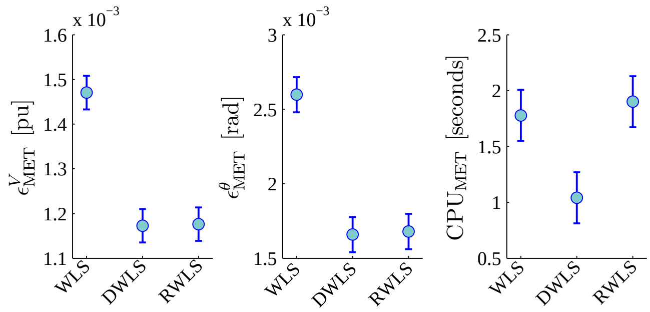


Figure 1. Results for the case study with no gross errors: performance comparison.

From Table 3 and Fig. 1, the following observations are in order:

- From Table 3, note that the three p-values corresponding to the factor “Method” denote that there are statistically significant differences regarding the performance of the considered methods, for numerical accuracy and computational efficiency. Thus, the confidence interval plots (Fig. 1) are studied to determine which method provides the best performance.

- Regarding estimation accuracy (left and center subplots), the WLS estimator is the least accurate, whereas the DWLS and RWLS procedures provide the most accurate results for a confidence level of 95%.

Table 3. ANOVA table for the first case study.

Metric $\epsilon^{V}_{\text{MET}}$

| Source | Squared Sum | DoF | Mean-sq | F | p-value |
|---|---|---|---|---|---|
| Method | $5.9 \cdot 10^{-6}$ | 2 | $2.9 \cdot 10^{-6}$ | 81.0 | $< 10^{-25}$ |
| Scenario | $4.1 \cdot 10^{-5}$ | 99 | $4.2 \cdot 10^{-7}$ | 11.6 | $< 10^{-46}$ |
| Residual | $7.1 \cdot 10^{-6}$ | 198 | $3.6 \cdot 10^{-8}$ | | |
| Total | $5.4 \cdot 10^{-5}$ | 299 | | | |

Metric $\epsilon^{\theta}_{\text{MET}}$

| Source | Squared Sum | DoF | Mean-sq | F | p-value |
|---|---|---|---|---|---|
| Method | $5.7 \cdot 10^{-5}$ | 2 | $2.8 \cdot 10^{-5}$ | 91.0 | $< 10^{-25}$ |
| Scenario | $6.5 \cdot 10^{-5}$ | 99 | $6.5 \cdot 10^{-7}$ | 1.8 | $1.8 \cdot 10^{-4}$ |
| Residual | $7.1 \cdot 10^{-5}$ | 198 | $3.5 \cdot 10^{-7}$ | | |
| Total | $1.9 \cdot 10^{-4}$ | 299 | | | |

Metric $\text{CPU}_{\text{MET}}$

| Source | Squared Sum | DoF | Mean-sq | F | p-value |
|---|---|---|---|---|---|
| Method | 43.2 | 2 | 21.6 | 16.1 | $< 10^{-6}$ |
| Scenario | 183.9 | 99 | 1.8 | 1.3 | 0.028 |
| Residual | 265.8 | 198 | 1.3 | | |
| Total | 493.0 | 299 | | | |

- There is no significant difference between the accuracy provided by the DWLS and RWLS methods.
- Regarding the required CPU time, the DWLS method is the most efficient. Note that the efficiency provided by the WLS and RWLS algorithms are statistically similar.

### *4.2. Second Case Study: Three Gross Errors*

In this case, three randomly-chosen substations are affected by a set of correlated gross errors. Then, the measurement vector $\boldsymbol{z}$ contains three sets of multiple bad data.

Again, Table 4 provides the ANOVA analysis for metrics $\epsilon^{V}_{\text{MET}}$, $\epsilon^{\theta}_{\text{MET}}$, and $\text{CPU}^{\text{MET}}$, and Fig. 2 depicts the average value for these metrics and the confidence intervals for a 95% confidence level.

From Table 4 and Fig. 2, the following observations are in order:

- From Table 4, note that the three p-values corresponding to the factor "Method" denote that there are statistically significant differences regarding the performance of the considered methods, for numerical accuracy and

Table 4. ANOVA table structure

Metric $\epsilon^{V}_{\text{MET}}$

| Source | Squared Sum | DoF | Mean-sq | F | p-value |
|---|---|---|---|---|---|
| Method | $1.5 \cdot 10^{-5}$ | 2 | $7.4 \cdot 10^{-6}$ | 81.0 | $< 10^{-27}$ |
| Scenario | $4.7 \cdot 10^{-5}$ | 99 | $4.7 \cdot 10^{-7}$ | 11.6 | $< 10^{-24}$ |
| Residual | $1.7 \cdot 10^{-5}$ | 198 | $8.5 \cdot 10^{-8}$ | | |
| Total | $7.8 \cdot 10^{-5}$ | 299 | | | |

Metric $\epsilon^{\theta}_{\text{MET}}$

| Source | Squared Sum | DoF | Mean-sq | F | p-value |
|---|---|---|---|---|---|
| Method | $1.1 \cdot 10^{-4}$ | 2 | $5.2 \cdot 10^{-5}$ | 70.1 | $< 10^{-23}$ |
| Scenario | $1.3 \cdot 10^{-4}$ | 99 | $1.4 \cdot 10^{-6}$ | 1.8 | $2.1 \cdot 10^{-4}$ |
| Residual | $1.8 \cdot 10^{-4}$ | 198 | $7.5 \cdot 10^{-7}$ | | |
| Total | $3.9 \cdot 10^{-4}$ | 299 | | | |

Metric $\text{CPU}_{\text{MET}}$

| Source | Squared Sum | DoF | Mean-sq | F | p-value |
|---|---|---|---|---|---|
| Method | 341.2 | 2 | 170.7 | 69.9 | $< 10^{-22}$ |
| Scenario | 390.9 | 99 | 3.9 | 1.6 | 0.0023 |
| Residual | 483.8 | 198 | 2.4 | | |
| Total | 1214 | 299 | | | |

computational efficiency. Thus, the confidence interval plots (Fig. 2) are studied to determine which method provides the best performance.

- Regarding estimation accuracy (left and center subplots), the WLS estimator is the less accurate, whereas the DWLS and RWLS procedures provide the most accurate results for a confidence level of 95%.
- There is no significant difference between the accuracy provided by the DWLS and RWLS methods.
- Regarding the required CPU time, the RWLS estimator is the most efficient. The efficiency provided by the WLS and DWLS algorithms are statistically similar.

### *4.3. Third Case Study: Six Gross Errors*

In this case, six substations are randomly chosen, and a set of multiple gross errors is located in each substation. Thus, the measurement vector is corrupted by six sets of multiple bad data.

Table 5 provides the ANOVA analysis for metrics $\epsilon^{V}_{\text{MET}}$, $\epsilon^{\theta}_{\text{MET}}$, and $\text{CPU}^{\text{MET}}$, and Fig. 3 depicts the average value for these metrics and the confidence intervals for a 95% confidence level.

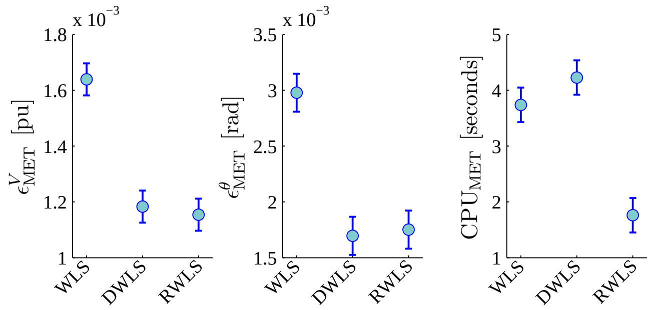


Figure 2. Results for the case study with three gross errors: performance comparison.

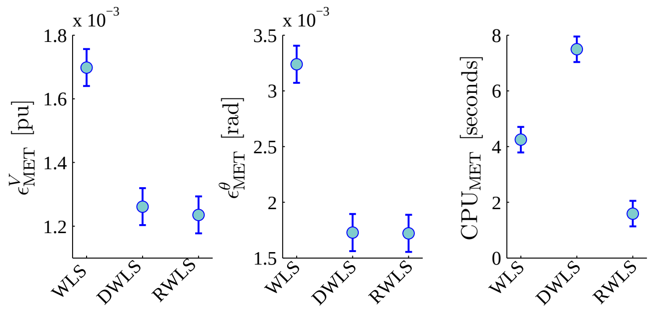


Figure 3. Results for the case study with six gross errors: performance comparison.

From Table 5 and Fig. 3, the following observations are in order:

- Results from Table 5 allow withdrawing the same conclusions as in the previous section: there are significant differences among methods' accuracy and efficiency.
- Again, the WLS estimator is the least accurate, whereas the DWLS and RWLS procedures provide the most accurate results, without statistically-significant differences between these two approaches.
- Regarding the required CPU time, the RWLS estimator is the most efficient, and the DWLS algorithm is the less efficient one.
- The required CPU time for the RWLS estimator is approximately 77% and 59% smaller than the CPU times required by the DWLS and WLS methods, respectively.

### 4.4. *Results Comparison*

Analyzing jointly the results obtained for the three case studies, the following general conclusions can be withdrawn:

1. The estimation accuracy of the proposed RWLS algorithm is significantly better than the one provided by the traditional WLS procedure, for a 95% confidence level and considering scenarios with zero, three, and six sets

Table 5. ANOVA table structure

Metric $\epsilon^{V}_{\text{MET}}$

| Source | Squared Sum | DoF | Mean-sq | F | p-value |
|---|---|---|---|---|---|
| Method | $1.35 \cdot 10^{-5}$ | 2 | $6.7 \cdot 10^{-6}$ | 78.0 | $< 10^{-24}$ |
| Scenario | $5.25 \cdot 10^{-5}$ | 99 | $5.3 \cdot 10^{-7}$ | 6.13 | $< 10^{-26}$ |
| Residual | $1.7 \cdot 10^{-5}$ | 198 | $8.6 \cdot 10^{-8}$ | | |
| Total | $8.3 \cdot 10^{-5}$ | 299 | | | |

Metric $\epsilon^{\theta}_{\text{MET}}$

| Source | Squared Sum | DoF | Mean-sq | F | p-value |
|---|---|---|---|---|---|
| Method | $1.5 \cdot 10^{-4}$ | 2 | $7.6 \cdot 10^{-6}$ | 107.1 | $< 10^{-31}$ |
| Scenario | $1.3 \cdot 10^{-4}$ | 99 | $1.3 \cdot 10^{-7}$ | 1.9 | $< 10^{-4}$ |
| Residual | $1.4 \cdot 10^{-4}$ | 198 | $7.1 \cdot 10^{-8}$ | | |
| Total | $4.2 \cdot 10^{-4}$ | 299 | | | |

Metric $\text{CPU}_{\text{MET}}$

| Source | Squared Sum | DoF | Mean-sq | F | p-value |
|---|---|---|---|---|---|
| Method | 1747 | 2 | 1032 | 873.5 | $< 10^{-41}$ |
| Scenario | 911 | 99 | 6.95 | 9.2 | 0.001 |
| Residual | 1070 | 198 | 5.4 | | |
| Total | 37291 | 299 | | | |

of multiple gross errors.

2. For a 95% confidence level, there is no significant difference between the estimation accuracy degree of methods RWLS and DWLS. That is, the obtained estimate for both methods have the same accuracy.
3. The computational efficiency of the RWLS remains unaltered with independence of the number of multiple gross errors corrupting the measurement set. This consideration can be graphically illustrated by plotting the average required CPU time for each method as a function of the number of multiple bad measurement sets (see Fig. 4).
4. Note that the most accurate estimates does not correspond to the WLS results. This is so because WLS estimation do not consider measurement dependencies.

## 5. Conclusion

The method proposed in this paper is statistically proven to be superior to alternative algorithms reported in the technical literature, especially if the number of gross errors or outliers in the measurement set is large.

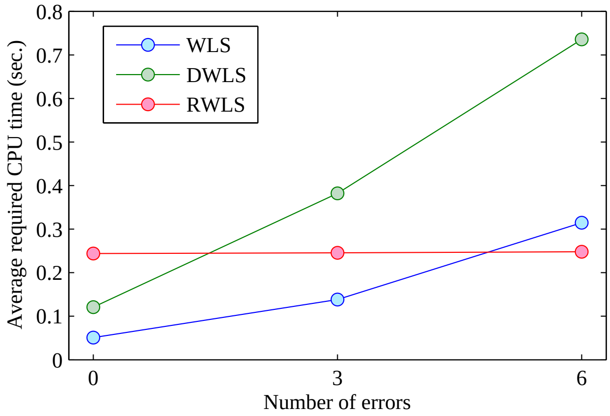


Figure 4. Required CPU time comparison.

The proposed RWLS estimator takes into consideration measurement dependencies to improve accuracy and its weights are automatically readjusted to increase robustness. In summary, it provides accurate estimates, is robust against outliers, and is computationally efficient.

The method's performance is tested using a realistic case study with single and multiple gross errors, and considering a high number of scenarios. These results are compared in detail using ANOVA techniques, which allows proving the outperformance of the proposed method from a statistical point of view.

Since most of the state estimation algorithms used in practice are based on WLS techniques, such algorithms can be easily adapted or modified to include RWLS features.

## References

[1] A. Abur, A. Gómez-Expósito, Power System State Estimation. Theory and Implementations, Marcel Dekker, New York, 2004.
[2] D. M. Hawkins, Identification of Outliers, Chapman and Hall, London, 1980.
[3] D. A. Belsley, E. Kuh, R. E. Welsch, Regression Diagnostics: Identifying Influential Data and Sources of Multicollinearity, John Wiley & Sons, New York, 1980.
[4] A. C. Atkinson, Plots, Transformations, and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis, Oxford: Clarendon Press, New York, 1985.
[5] V. Barnett, T. Lewis, Outliers in Statistical Data, 3rd Edition, John Wiley & Sons, New York, 1994.
[6] Confirmation of multiple outliers in generalized linear and nonlinear regressions, Computational Statistics & Data Analysis 25 (1) (1997) 55–65.
[7] J. W. Wisnowski, D. C. Montgomery, J. R. Simpson, A comparative analysis of multiple outlier detection procedures in the linear regression model, Computational Statistics & Data Analysis 36 (3) (2001) 351–382.
[8] A. Luceño, Estimation of missing values in possibly partially nonstationary vector time series, Biometrika 84 (1997) 495–499.
[9] A. Luceño, Detecting possibly non-consecutive outliers in industrial time series, Journal of the Royal Statistical Society. Series B (Statistical Methodology) 60 (1998) 295–310.
[10] A. Luceño, Multiple outliers detection through reweighted least deviances, Comput. Stat. Data Anal. 26 (1998) 313–326.
[11] N. Billor, S. Chatterjee, A. S. Hadi, Iteratively re-weighted least squares method for outlier detection in linear regression, Bulletin of the International Statistical Institute 1 (2001) 470–472.
[12] E. Castillo, A. S. Hadi, A. J. Conejo, A. Fernández-Canteli, A general method for local sensitivity analysis with application to regression models and other optimization problems, Technometrics 46 (4) (2004) 430–445.
[13] T.-C. Cheng, Robust diagnostics for the heteroscedastic regression model, Comput. Stat. Data Anal. 55 (2011) 1845–1866.
[14] P. J. Rousseeuw, A. M. Leroy, Robust Regression and Outlier Detection, John Wiley & Sons, New York, 1987.
[15] P. J. Rousseeuw, K. Van Driessen, A fast algorithm for the minimum covariance determinant estimator, Technometrics 41 (1999) 212–223.

[16] R. Mínguez, B. G. Reguero, A. Luceño, F. J. Méndez, Regression models for outlier identification (hurricanes and typhoons) in wave hindcast databases, Journal of Atmospheric and Oceanic Technology 29 (2012) 267–285.
[17] L. Guangyi, Y. Erkeng, Y. Song, Novel algorithms to estimate and adaptively update measurement error variance using power system state estimation results, Electric Power Systems Research 47 (1) (1998) 57–64.
[18] S. Zhong, A. Abur, Auto tuning of measurement weights in WLS state estimation, IEEE Transactions on Power Systems 19 (4) (2004) 2006 – 2013.
[19] R. Pires, A. Simoes Costa, L. Mili, Iteratively reweighted least-squares state estimation through givens rotations, IEEE Transactions on Power Systems 14 (4) (1999) 1499–1507.
[20] R. Jabr, B. Pal, Iteratively reweighted least-squares implementation of the wlav state-estimation method, IEE Proceedings Generation, Transmission and Distribution 151 (1) (2004) 103–108.
[21] D. Wiens, Designs for weighted least squares regression, with estimated weights, Statistics and Computing (2012) 1–11.
[22] E. Caro, A. J. Conejo, R. Mínguez, Power system state estimation considering measurement dependencies, IEEE Transactions on Power Systems 24 (4) (2009) 1875–1885.
[23] E. Caro, A. J. Conejo, J. Morales, R. Mínguez, Calculation of measurement correlations using point estimate, IEEE Transactions on Power Delivery 25 (4) (2010) 2095–2103.
[24] E. Caro, A. J. Conejo, R. Mínguez, M. Zima, G. Andersson, Multiple bad data identification considering measurement dependencies, IEEE Transactions on Power Systems 26 (4) (2011) 1953–1961.
[25] P. J. Huber, Robust Statistics, New York. John Wiley, 1981.
[26] H. Scheff, The Analysis of Variance, Wiley, New York, 1959.