

## Distribution networks nontechnical power loss estimation: A hybrid data-driven physics model-based framework



Arturo S. Bretas (Conceptualization; Methodology; Software; Writing-Original draft preparation)<sup>a,\*</sup>, Aquiles Rossoni (Software; Visualization; Investigation; Writing-Reviewing and Editing)<sup>b</sup>, Rodrigo D. Trevizan (Software; Visualization; Investigation; Writing-Reviewing and Editing)<sup>c</sup>, Newton G. Bretas (Supervision; Visualization; Investigation; Writing-Reviewing and Editing)<sup>d</sup>

<sup>a</sup> Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, 32611-6200, USA

<sup>b</sup> School of Technology, Pontifical Catholic University of Rio Grande do Sul, Porto Alegre, RS, 90619-900, Brazil

<sup>c</sup> Energy Storage Technology & Systems Department, Sandia National Laboratories, Albuquerque, NM, 87123, USA

<sup>d</sup> Department of Electrical & Computer Engineering, University of São Paulo, São Carlos, SP, 13566-590, Brazil

### ARTICLE INFO

#### Keywords:

Distribution systems  
Nontechnical power loss  
Gross error analysis  
Synthetic measurements

### ABSTRACT

This paper presents a hybrid data-driven physics model-based framework for distribution networks nontechnical power loss estimation. Nontechnical power loss is defined as energy delivered to the consumers but not billed by the utility. These losses, unlike technical losses, are not inherent to the process of transportation of electricity. State-of-the-art solutions for nontechnical power loss estimation are either data-driven or physics model-based. However, due to the evolving nature of nontechnical power losses, data-driven solutions by themselves are not sufficient. Physics model-based analytical solutions, otherwise, which consider a quasi-static system model, rely solely on physics phenomena observation, however it is virtually impossible to model all grid dynamics. In this case, the nexus of data-driven physics model-based analytic models enable the solution of the problem. The hybrid framework is composed of three interdependent processes. First, an unbalanced load flow analysis is performed to obtain an initial estimate of the operating system state. Second, a data-driven method for consumer classification is applied. Third, synthetic measurements are created considering the measurement's innovation and  $n$ -tuple of critical measurements aiming to improve gross error analysis. Solution validation is made considering the IEEE 4-bus, 13-bus and 123-bus unbalance test feeders. Comparative test results highlight decreased nontechnical power loss estimation errors. Simplicity of implementation, with easy-to-obtain parameters, built on the classical weighted least squares state estimator, indicate potential aspects for real-life applications.

### 1. Introduction

Power system losses are defined as the difference between the electric energy injected into the system and the energy billed to consumers. Energy losses can be classified as technical and nontechnical. Technical losses (TL) are inherent to the process of transportation of electricity. Nontechnical losses (NTL), otherwise, correspond to energy delivered but not computed due to measurement errors, fraud or theft. Energy losses are most significant in distribution systems (DS), affecting the revenue of utilities, power quality and electricity rates [1].

NTL are currently mitigated through prevention, identification and correction of irregular consumption. State-of-the-art solutions detect

NTL through machine learning and anomaly detection techniques [2-4]. These solutions are referred in this work as consumer classification methods (CCM). CCM are data-driven based, such as support vector machines [5,6], extreme learning machines [7], fuzzy classification [8], rule induction models [9], optimum-path forest [10,11], game-theoretical inference-based models [12], statistical based methods [13,14], Markov chains [15], Petri Nets [16], unsupervised and supervised learning [17,18], wavelet-packet transform based [19] and A-star algorithms [20]. As unsupervised learning methods, it is also highlighted the use of unsupervised optimum-path forest [21], short-lived patterns [22] and a data-driven framework that combines methods as the maximum information coefficient, the fast search and the find of

\* Corresponding author.

E-mail address: [bretas.arturo@gmail.com](mailto:bretas.arturo@gmail.com) (A.S. Bretas).

Nomenclature	
<i>Acronyms</i>	
CCM	Consumer classification method
DS	Distribution system
FN	False negatives
FP	False positives
GE	Gross error
GMR	Global measurement redundancy
II	Innovation index
LF	Load flow
MV	Medium voltage
NRA	Normalized residual analysis-based method
NTL	Nontechnical losses
pdf	Probability density function
PR	Precision index
PSSE	Power system state estimation
RC	Recall index
RV	Reference value
SE	State estimation
TL	Technical losses
TP	True positives
WLS	Weighted least squares
<i>Variables and functions</i>	
$\mathbf{c}(\mathbf{x})$	Vector of equality constraints
$\mathbf{C}$	Jacobian matrix of $\mathbf{c}(\mathbf{x})$
$d$	Number of dimensions
$\mathbf{e}$	Vector of measurement errors
$\mathbf{E}$	Complex bus voltage vector
$e_{NTL}$	Percental estimation error of nontechnical losses
$e_{TL}$	Percental estimation error of technical losses
$\mathbf{FP}$	False positives
$\mathbf{FN}$	False negatives
$\mathbf{h}(\mathbf{x})$	Measurement function
$\mathbf{H}$	Jacobian matrix of $\mathbf{h}(\mathbf{x})$
$\mathbf{I}$	Complex bus current injection vector
$\mathbf{I}_{abc}^k$	$k$ -th bus current injections for phases a,b and c
$\mathbf{II}_k$	Innovation index of $k$ -th measurement
$J(\mathbf{x})$	Objective function of WLS
$\mathbf{K}$	Hat matrix of WLS state estimator
$n$	Number of buses in the system
$n_{rc}$	Number of consumers classified as regular
$n_{sc}$	Number of consumers classified as suspect
$NTL_f^k$	Estimate of nontechnical losses at bus $k$ and phase $f$
$P_{RV}$	Reference value of injected power
$P_f^k$	Estimate of real bus power injection in bus $k$ and phase $f$
$\bar{P}_f^k$	Measurement or pseudo measurement of bus power injection in bus $k$ and phase $f$
$P_f^{km}$	Real branch power flow from bus $k$ to bus $m$ in phase $f$
$S_f^k$	Apparent bus power injection in bus $k$ and phase $f$
$TL_f^{km}$	Technical losses on branch between buses $k$ and $m$ in phase $f$
$TP$	True positives
$\mathbf{W}$	Weight matrix
$\mathbf{x}$	Vector of states
$\mathbf{y}$	Feature vector of each consumer
$\mathbf{Y}_{bus}$	Complex bus admittance matrix
$\mathbf{z}$	Vector of measurements
$\boldsymbol{\lambda}$	Vector of Lagrange multipliers
$\boldsymbol{\mu}$	Sample distribution mean
$\nu$	Number of degrees of freedom of the $\chi^2$ distribution
$\sigma_i$	Standard deviation of $i$ -th measurement error
$\sigma_{up}$	Updated value of the standard deviation
$\Sigma$	Sample covariance matrix
$\Omega_k$	Set of buses connected to bus $k$

density peaks [23].

NTL can be further estimated through physics model-based solutions [24,25], as by power system state estimation (PSSE) solutions [13,14,26]. Considering such approach, NTL are estimated by the difference between measured and estimated power injections at each bus [14,26]. NTL can be further modeled as measurements with gross error (GE), which are defined as measurements with statistically large errors, and estimated during the PSSE process [27].

As Smart Grid technologies become more accessible, electro-mechanical meters are gradually being replaced by smart meters [20,28]. The increasing number of measurements in DS allows the precise estimation of system states while data provided by smart meters can be additional information to CCM. On the other hand, cyber-attacks, such as false data injections, can be launched on smart meters [29].

It is clear that while data-driven solutions are able to model temporal features of the grid, quasi-static physics based models are able to model spatial features [30]. PSSE and CCM are key complementary tools in smart grid operation and can compose a more robust NTL estimation framework. Physics model-based power system state estimation has been already used to ameliorate CCM [11,30].

Regarding hybrid data-driven physics model-based solutions for NTL estimation, recently strategies have been presented. In [13], first NTL are detected through physics model-based analytical tests, then, suspect consumers are identified by a CCM. In [14,26], authors' previous work, physics model-based analytical tests are used to detect, identify and estimate NTL. Since not all NTL are detected by the analytical tests, due to potential low measurement redundancy system areas, a subsequent CCM analysis is applied to improve NTL detection.

In previously presented hybrid solutions though, measurements are assumed to be installed in medium voltage (MV) buses. However, in real-life applications, smart meters are being allocated mainly on low voltage consumers [31]. Further, no model is presented to improve GE analysis in system areas with low measurement redundancy [14,26], which hinder real-life applications.

In this work, a hybrid data-driven physics model-based framework for DS NTL estimation is presented. NTL are modeled as GE. NTL are detected, identified and estimated. The presented framework is composed of three interdependent processes. First, an unbalanced load flow analysis is performed to obtain an initial estimate of the system operating state. Second, a data-driven based method for consumer classification is applied. Third, synthetic measurements are created in low redundancy areas considering measurement's innovation and  $n$ -tuple of critical measurements.

The specific contributions of this work towards the state-of-the-art are:

- Identification of vulnerable NTL system areas based on concepts of measurement's innovation and  $n$ -tuple of critical measurements;
- Synthetic measurement model for GE analysis enhancement;
- Framework for NTL estimation in unbalanced distribution networks.

Previous work does not correlate system measurements with NTL estimation. In this work, through the concepts of measurement's innovation and  $n$ -tuple of critical measurements, synthetics measurements are created towards NTL estimation. According to the Central Limit Theorem, the sum of a large number of random variables that follow any distribution with bounded variance approximates a normal

distribution. Synthetic measurements increase model degrees of freedom, thus the chi-squared distribution also tends to a normal distribution. A framework for NTL estimation is presented considering unbalanced distribution network characteristics. The remainder of the paper is as follows. Section II presents the NTL estimation framework. Comparative case study is presented in section III. Section IV presents the conclusions of this work.

## 2. NTL estimation framework

Fig. 1 illustrates the data flow chart of the NTL estimation framework, named PSSE CCM. Fig. 2 illustrates one of the three framework's processes, the CCM-enhanced weighted least squares (WLS) PSSE.

Consider Fig. 1. Input data are system elements parameters, analog and status measurements. The CCM considers power injection measurements, provided by smart meters, to initially classify consumers as regular or suspect. This is just an initial guess, which estimates the ratio between the number of suspect and regular consumers in each bus. This estimate is used later in the process of creating synthetic measurements. Synthetic measurements are measurements created artificially, with the goal of to increase system measurement redundancy level. These are not to be mistaken with pseudo measurements, which are obtained from historical data [32]. An unbalanced load flow (LF) is then used to initially estimate the system state, which is then considered in a unbalanced WLS PSSE [27]. The goal is to have an initial system state estimation for the next process.

Consider Fig. 2. The CCM-enhanced WLS PSSE process runs a WLS PSSE twice. First, the WLS PSSE is performed with updated weights, estimated by the CCM, and without GE analysis. This first estimation aims to create synthetic measurements in all segments without real measurements, avoiding thus low values of  $H$  [33,34] and subset of critical measurements [35]. The second estimation, otherwise, considers real, synthetic and pseudo measurements and performs WLS PSSE with GE analysis [27]. NTL are modeled as GE. In this process, NTL are detected, identified and estimated through extended physics model-based analytical tests. The second WLS PSSE is performed though considering values of weights as a percentage of measurements [36], which are then used to compose the WLS weight matrix ( $\mathbf{W}$ ).

Fig. 3 presents an illustrative example of the framework's data flow, considering an illustrative 7-bus system.

In the step 1, smart meters power injection measurements, at buses 2 and 6, are considered by a CCM to initially classify consumers as regular or suspect, and compute updated standard deviations. Estimated standard deviations are used to compose the updated weight matrix  $\mathbf{W}$ , to be used in the first WLS PSSE. Also, in step 1, a LF analysis is performed to estimate the system state, the updated bus admittance matrix  $\mathbf{Y}$  and the pseudo measurements of power injection in the buses 2 and 6,  $PQ_2$  and  $PQ_6$  respectively. In step 2, considering the updated  $\mathbf{W}$ , a WLS PSSE is performed to obtain power flow synthetic measurements in MV segments without real measurements,  $PQ_{23}$ ,  $PQ_{34}$ ,  $PQ_{56}$  and  $PQ_{67}$ .

In step 2, GE analysis is not performed. In step 3, otherwise, a WLS PSSE considering now values of weights as a percentage of measurements is performed. In this step, GE analysis is performed, while considering synthetic measurements created previously. After, with estimated states, NTL, as well as TL are estimated. NTL estimation is performed in the MV system, given the lack of power flow measurements in the low voltage network.

Sections II.A, II.B and II.C describe the CCM, LF and the WLS PSSE, respectively. Synthetic measurements creation is presented in II.D.

### 2.1. CCM

The used CCM is a unsupervised statistics-based parametric anomaly detection method that considers a multivariate Gaussian distribution to model consumers load [14]. In this aspect, for each

consumer in a given class, the probability density function (pdf) is computed by (1).

$$f(\mathbf{y}, \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-\frac{1}{2}(\mathbf{y}-\mu)^T \Sigma^{-1}(\mathbf{y}-\mu)} \quad (1)$$

The number of dimensions ( $d$ ) in (1) is related to the number of measurements of each consumer considered to compute the pdf. The CCM's hypothesis is that, for the same type and class of consumer, most of them are regular and have a similar pattern of consumption. Therefore, the consumers which  $\mathbf{y}$  values are similar to  $\mu$  have a high pdf, and are considered as regular. It is also assumed that irregular consumers occur much less frequently than regular ones, and they have an anomalous consumption pattern, thus having a low pdf value. In this case, the inverse of the probability density function is an anomaly score [37], and the consumers with lower values are more likely to have an irregular consumption.

Usually, DS companies know with some precision the system NTL percentage related to injected power, without knowing where they occur [25]. Considering that the percentage of NTL corresponds to the percentage of irregular consumers, consumers with lowest pdf can be classified as suspect. Thus, in a DS with total NTL percentage of 5.6% and having 1000 consumers, 56 consumers with lowest pdf can be classified as suspect. Other approaches of establishing a pdf threshold value between suspect and regular consumers are discussed in [14]. Alternatively, other CCM, as those presented in the section I, can be considered in the presented framework.

Considering that inspections have high financial costs, one goal of NTL identification is having a high percentage of success through field evaluations. This performance is measured by the precision ( $PR$ ) index, which indicates the percentage of clients correctly classified as suspects. On the other hand, another goal is to detect the largest number of irregular consumers to reduce the amount of NTL. This performance is measured by the recall ( $RC$ ) index. The performance evaluation of a CCM is then given by  $PR$  and  $RC$  indexes, respectively computed by (2) and (3) [11].

$$PR = 100 \times TP / (TP + FP) \quad (2)$$

$$RC = 100 \times TP / (TP + FN). \quad (3)$$

In the previous equations,  $TP$  stands for true positives,  $FP$  are the false positives and  $FN$  means false negatives. The relationship between the number of consumers classified as suspect and as regular in each bus are considered in the weights of WLS PSSE for synthetic measurements creation, as presented in II.D. The information added by the CCM result is most important for the synthetic measurements creation, and enhances GE analysis.

This work does not contribute towards the state-of-art of CCM. Highlights that any CCM can be used in the framework. The framework considers the estimated number of suspect and regular consumers in each bus to update WLS PSSE weights. As a limitation, the presented

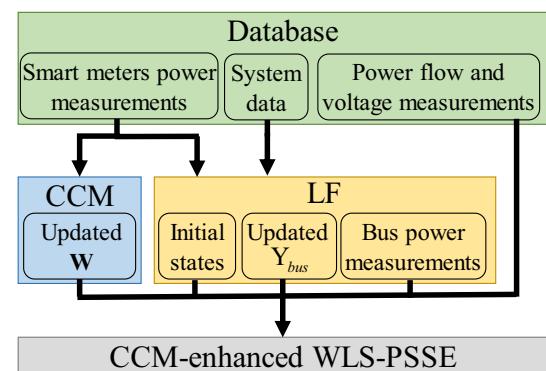


Fig. 1. PSSE CCM framework.

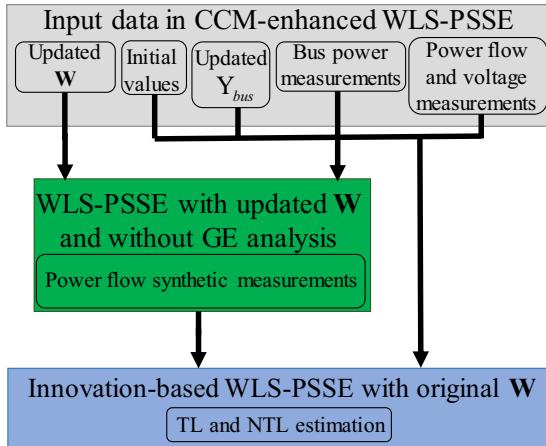


Fig. 2. CCM-enhanced WLS-PSSE.

CCM does not consider distributed storage and generation elements, as prosumers, batteries and vehicle-to-grid.

## 2.2. Extended physics model-based analytical tests

### 2.2.1. System elements model

The system model is composed by segment and bus data. Segments are composed by lines, transformers and voltage regulators and can be modeled as a primitive admittance matrix. This matrix can be represented by the submatrices related to each bus and the interaction between them.

The primitive admittance matrix of segments relates the bus phase voltages ( $E$ ) with the lines currents ( $I$ ), as shown in (4)

$$\begin{bmatrix} I_{abc}^{km} \\ I_{abc}^{mk} \end{bmatrix} = Y_{prim\_km} \begin{bmatrix} E_{abc}^k \\ E_{abc}^m \end{bmatrix} = \begin{bmatrix} Y_{prim\_km}^{kk} & Y_{prim\_km}^{km} \\ Y_{prim\_km}^{mk} & Y_{prim\_km}^{mm} \end{bmatrix} \begin{bmatrix} E_{abc}^k \\ E_{abc}^m \end{bmatrix} \quad (4)$$

The values of the submatrices depend on the type and characteristics of each segment [38,39]. For a line, the submatrices are composed by the series impedance and shunt admittance matrices calculated by Carson equations and Kron reduction. For transformers and voltage regulators, the submatrices depend on transformers characteristics, as impedance of windings and connection type. The values of voltage regulator submatrices depend on tap position (*tap*) that affect the effective regulator ratio:  $a_R = 1 \pm 0.00625 \text{tap}$ . The variable sign depends of the regulator type. The value of *tap* is changed by a regulator control circuit that considers the voltages and currents measured and a set value for the line drop compensator. Bus data is composed by shunt admittance matrices ( $Y_{cap}^r$ ) that are connected in a single bus  $r$  and are related to capacitor banks. The admittance matrix of a capacitor bank depends on its nominal power, voltage and connection type.

Considering the segment and bus matrices, the system can be modeled as a bus admittance matrix that relates all bus phase voltages ( $E$ ) and bus injected currents ( $I$ ), given by (5). Diagonal submatrices are computed by (6), and the off diagonal submatrices by  $Y_{bus}^{kk} = Y_{prim\_km}^{kk}$ , where  $k$  and  $m$  are system buses and  $\Omega_k$  the set of buses connected to bus  $k$ .

$$I = Y_{bus} E \quad (5)$$

$$Y_{bus}^{kk} = Y_{cap}^k \sum_{m \in \Omega_k} Y_{prim\_km}^{kk} \quad (6)$$

### 2.2.2. Load flow model

Considering that we have a smart grid, active and reactive power of consumers are measured and can be modeled by a constant power model [38]. Considering a flat start, the vector of injected current in each bus can be computed by (7), where  $S_f^k$  and  $E_f^k$  are the complex

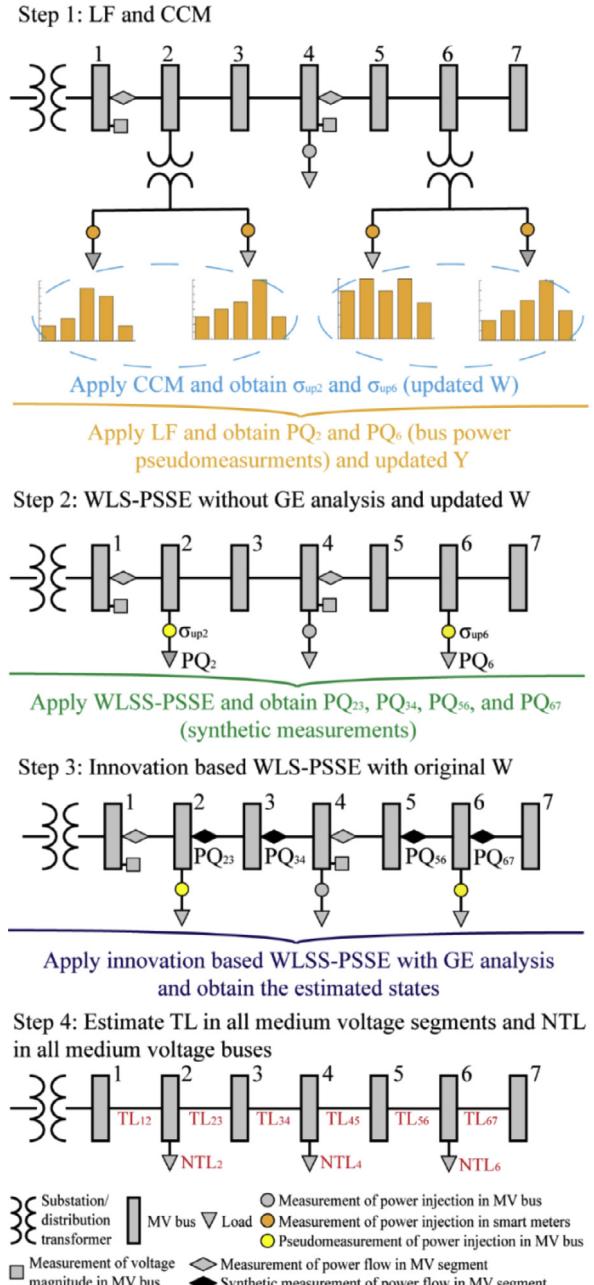


Fig. 3. Steps of PSSE CCM framework in an illustrative 7 bus system.

power and the voltage phasor in phase  $f$  of bus  $k$ . If the complex power represents a load, the value has a negative sign. If the power is measured between phases, the delta currents can be computed and transformed in injected currents [27].

$$I_{abc}^k = [(S_a^k/E_a^k)^* \quad (S_b^k/E_b^k)^* \quad (S_c^k/E_c^k)^*]^T \quad (7)$$

(5) can be rewritten as (8), where the subscripts *sw* and *ld* are related to substation bus and consumer buses, respectively. Given that the voltage at the substation bus and the calculated injected currents in consumers are known, the voltages at load buses can be estimated by (9).

$$\begin{bmatrix} I_{sw} \\ I_{ld} \end{bmatrix} = \begin{bmatrix} Y_{sw,sw} & Y_{sw,ld} \\ Y_{ld,sw} & Y_{ld,ld} \end{bmatrix} \begin{bmatrix} E_{sw} \\ E_{ld} \end{bmatrix} \quad (8)$$

$$E_{ld} = [Y_{ld,ld}]^{-1} [I_{ld} - Y_{ld,sw} E_{sw}] \quad (9)$$

Estimated voltages are then compared to their previous values. If the differences are lower than a defined tolerance, the method converged, otherwise, the injected currents and voltages are updated by (7) and (9). These updates are iteratively performed until convergence is achieved. In each iteration, the *tap* values of voltage regulators are updated and, as consequence,  $\mathbf{Y}_{bus}$  is updated. After convergence, the currents injected in both sides of the element are calculated by (4). The bus power injection in each phase is computed by  $S_f^k = (I_f^k)^*E_f^k$  and the branch power flow in each phase is calculated by  $S_f^{km} = (I_f^{km})^*E_f^k$ . The active ( $P_f^{km}$  and  $P_f^k$ ) and reactive ( $Q_f^{km}$  and  $Q_f^k$ ) power flows and injections can be obtained by separating real and imaginary parts. In a system model with both primary and secondary networks, the power flowing through a transformer connected in a bus  $k$  can be represented as the power injected in this bus  $k$ . In this way, LF process provides power injection measurements in MV buses for the WLS PSSE in addition to the initial states and updated  $\mathbf{Y}_{bus}$ . The power flow analysis is performed only once, before the WLS PSSE.

### 2.3. WLS PSSE

WLS PSSE is performed in the MV system considering real power flow and power injection measurements, provided by the load flow analysis, as described in section II-B. The power flow also provides the updated  $\mathbf{Y}_{bus}$  and the initial states for the WLS PSSE, improving the converge process.

WLS PSSE is obtained through the solution of  $\mathbf{z} = \mathbf{h}(\mathbf{x}) + \mathbf{e}$ , where  $\mathbf{z}$  is an  $nm$  dimensional vector of measurements and  $\mathbf{h}(\mathbf{x})$  is an  $nm$  dimensional vector of equations that relate the  $ns$  dimensional vector of states  $\mathbf{x}$  with the measured quantities. The measurement errors are given by  $nm$  dimensional vector  $\mathbf{e}$ . States variables are the bus voltages magnitudes ( $V$ ) and angles ( $\theta$ ). The ratio between the number of measurements ( $nm$ ) and the number of states to estimate ( $ns$ ) is the global measurement redundancy:  $GMR = nm/ns$ . Power and voltage magnitude measurements are provided by substation and feeder equipment, by smart meters installed in MV consumers and by the result of LF in buses that feed low voltage consumers. The equations  $\mathbf{h}(\mathbf{x})$  are given by the states and the elements of  $\mathbf{Y}_{bus}$ , as described in [27].

In the WLS PSSE approach, the problem consists in solving (12), where the vector of constraints  $\mathbf{c}(\mathbf{x})$  is composed of measurement functions of zero power injections and  $\mathbf{W}$  is a diagonal weight matrix composed by the inverse of the squared values of measurement standard deviations ( $\sigma$ ):  $\mathbf{W} = \text{diag}([\sigma_1^{-2} \dots \sigma_{nm}^{-2}]^T)$ . Considering an initial value for  $\mathbf{x}$ , the states in each iteration  $v$  are updated by solving (13) and updating the states as  $\mathbf{x}^{v+1} = \mathbf{x}^v + \Delta\mathbf{x}^v$ .  $\mathbf{H}$  and  $\mathbf{C}$  are the Jacobian matrices of  $\mathbf{h}(\mathbf{x})$  and  $\mathbf{c}(\mathbf{x})$  and  $\lambda$  is the vector of Lagrange multipliers. The method converges when  $\Delta\mathbf{x}$  is lower than a defined tolerance.

$$\begin{aligned} \min \quad & J(\mathbf{x}) = [\mathbf{z} - \mathbf{h}(\mathbf{x})]^T \mathbf{W} [\mathbf{z} - \mathbf{h}(\mathbf{x})] \\ \text{s.t.} \quad & \mathbf{c}(\mathbf{x}) = \mathbf{0} \end{aligned} \quad (12)$$

$$\begin{bmatrix} \mathbf{H}^T \mathbf{W} \mathbf{H} & \mathbf{C} \\ \mathbf{C} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{x} \\ -\lambda \end{bmatrix} = \begin{bmatrix} \mathbf{H}^T \mathbf{W} [\mathbf{z} - \mathbf{h}(\mathbf{x})] \\ -\mathbf{c}(\mathbf{x}) \end{bmatrix} \quad (13)$$

After convergence, it is possible to perform GE analysis. A commonly performed error analysis is based on the chi-squared and the largest normalized residual tests [35], which are applied to detect and identify NTL in DS [13]. On the other hand, innovation-based analysis is a more robust approach to perform GE analysis [27,29,40]. The presented framework considers a modification of the innovation-based GE analysis described in [27]. First, characteristics of real-life DS not modeled in [27] are considered, such as single and two-phase lines, voltage regulators and zero power injection buses. Second, the use of LF as the initial state, the update of  $\mathbf{Y}_{bus}$  and the synthetic measurement generation are as well considered, which are not in [27].

After the WLS PSSE, TL in each phase  $f$  of segment  $km$  is calculated by (14). NTL in each phase  $f$  of bus  $k$  is computed by (15), where  $\bar{P}_f^k$

correspond to the measured value of active power injection. Total bus NTL can be obtained by the sum of NTL in each phase.

$$TL_f^{km} = P_f^{km} + P_f^m \quad (14)$$

$$NTL_f^k = P_f^k - \bar{P}_f^k \quad (15)$$

To assess the performance of the presented framework, TL and NTL percentage estimation errors are calculated by (16) and (17), respectively. Where  $RV$  subscript is reference values. Given that there are buses without NTL, to avoid a division by zero, the difference is divided by the reference value of injected power ( $P_{RV}$ ).

$$eTL = |(TL - TL_{RV})/TL_{RV}| \quad (16)$$

$$eNTL = |(NTL - NTL_{RV})/P_{RV}| \quad (17)$$

### 2.4. Synthetic measurements

WLS PSSE considers a normal distribution model of the measurement error. In view of the Central Limit Theorem [32], the sum of a large number of random variables that follow any distribution with bounded variance approximates a normal distribution. Thus, as the number of degrees of freedom  $v$  increases, the  $\chi^2$  distribution also tends to a normal distribution. If, however, the number of degrees of freedom of the measurement model is not large, the opposite would occur, and the hypothesis testing would not hold. The degrees of freedom of the measurement model is of course dependent on the resources available for the system operator to monitor accurately the power grid. The concepts of  $II$  and  $n$ -tuple of critical measurements are thus considered in this work for synthetic measurements generation. The core idea is to create, artificially, synthetic measurements on areas where the hypothesis test might fail. As described in [33,41], the  $II$  of a measurement  $k$  is obtained by (18), where  $\mathbf{K}$  the WLS-PSSE hat matrix:  $\mathbf{K} = \mathbf{H}[\mathbf{H}^T \mathbf{W} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{W}$ . The inverse of  $II$  is defined as the undetectability index and provides the distance of a measurement from the range space of the  $\mathbf{H}$ . In measurements with low  $II$ , the GE is less reflected in the residual and is masked in GE analysis. According to numerical simulation in transmissions systems [33], values of  $II$  lower than 0.75 can significantly mask GE.

$$II_k = \sqrt{1 - K_{kk}} / \sqrt{K_{kk}} \quad (18)$$

According to [35], an  $n$ -tuple of critical measurements correspond to a subset of  $n$  measurements that when eliminated from the measurement set, make the system unobservable. Numerical observability can be verified by the rank of  $\mathbf{H}$ . If the rank is lower than the  $ns$ , the system is unobservable. In a  $n$ -tuple of critical measurements,  $n-2$  GE are detectable and identifiable, however,  $n-1$  and  $n$  GE are detectable, but not identifiable by the largest normalized residual method. For example, let's consider a subset of the system measurements set containing four measurements. If the system became unobservable with the elimination of all four measurements, this subset is a 4-tuple of critical measurements. In this case, if one or two ( $n-2=2$ ) of these measurements have GE, the system is able to detect and identify them. With GE in three ( $n-1=3$ ) or four ( $n=4$ ) of these measurements, the system can detect the GE, but will not be able to correctly identify how many and which measurements contain GE. The lower the  $n$  value, the less measurements are detectable and identifiable. For  $n$  equal to one, which corresponds to a critical measurement, a GE in this measurement is not detected.

The detection of gross errors and estimation of losses based in the WLS PSSE with largest normalized residual analysis, as applied in [13], is named in this work as normalized residual analysis-based method (NRA).

To illustrate the concept of areas with low measurement redundancy, consider the modified balanced 4-bus IEEE test feeder (Fig. 4) [42]. Loads in buses 2, 3 and 4 are, respectively 600 +  $j250$

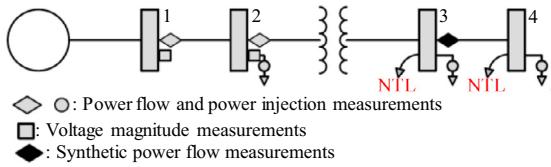


Fig. 4. 4-bus feeder: measurement framework and buses with NTL.

TABLE 1  
NTL and TL in balanced 4 bus feeder.

	Bus (NTL in kW)			Segment (TL in kW)		
	2	3	4	1-2	2-3	3-4
RV	0.0	90.0	63.0	9.6	10.1	28.9
NRA	0.0	0.0	148.7	9.6	10.1	33.1
PSSE CCM	0.0	90.0	63.0	9.6	10.1	28.9

TABLE 2  
Bus peak load and NTL in each phase (a, b and c) of unbalanced 4-bus system.

<i>f</i>	Bus 2: load kW	Bus 3: load (NTL) kvar	Bus 4: load (NTL) kvar
<i>a</i>	590	246	289 (86)
<i>b</i>	563	230	253
<i>c</i>	639	271	345 (111)

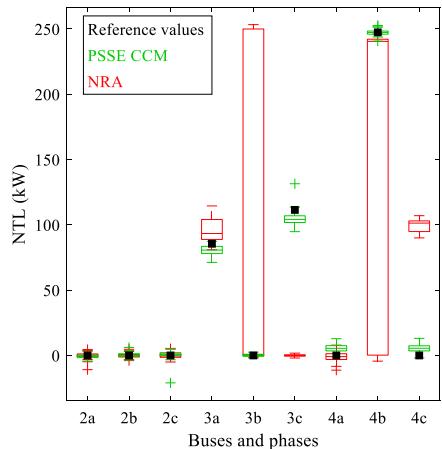


Fig. 5. NTL (kW) estimation in unbalanced 4 bus test system.

TABLE 3  
NTL and TL percental estimation error in unbalanced 4-bus feeder.

	Average error: mean ( $\sigma$ ) eNTL (%)		Maximum error: mean ( $\sigma$ ) eNTL (%)	
	eNTL (%)	eTL (%)	eNTL (%)	eTL (%)
NRA	10.3 (13.3)	9.4 (4.8)	100.0 (30.6)	53.1 (14.3)
PSSE CCM	0.7 (0.5)	0.6 (0.3)	5.4 (1.2)	4.3 (0.9)

kVA,  $300 + j150$  kVA and  $900 + j400$  kVA. The standard deviation considered for the substation measurements is 0.1% and for all other measurements 1% is adopted. The same setup is considered in the case studies.

Disregarding synthetic measurements, the GMR is equal to 1.71 and due to the lack of a power flow measurement between buses 3 and 4, the power injections in these buses ( $P_3$ ,  $P_4$ ,  $Q_3$  and  $Q_4$ ) form a 4-tuple of critical measurements. In this case, the NRA can detect and identify only two of these measurements, if they have GE. Thus, the GE detection test is prone to fail in this area. Further, the values of  $II$  for measurements  $P_3$  and  $P_4$  are respectively 0.28 and 1.53, indicating that GE in  $P_3$  will be difficult to detect by NRA.

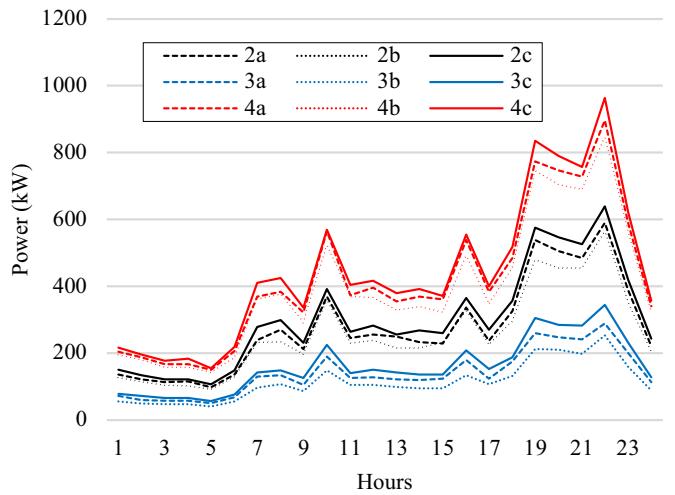


Fig. 6. 24-hour RV of injected power in each phase of 4-bus test system.

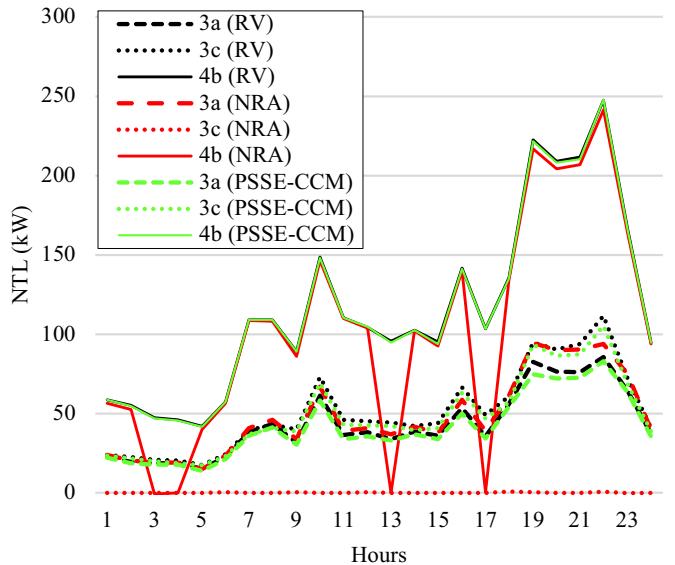


Fig. 7. 24-hour NTL estimation in 4-bus test system – phases with NTL.

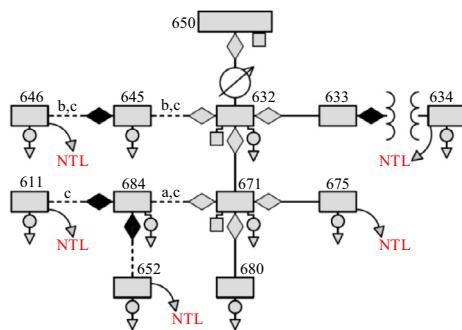


Fig. 8. 13-bus test feeder: measurement framework and buses with NTL

For example, considering that buses 3 and 4 have NTL with values of  $90 + j45$  kVA and  $60 + j48$  kVA, respectively. This results in GE in the measurements  $P_3$ ,  $P_4$ ,  $Q_3$  and  $Q_4$ . If it is considered that, for illustrative purposes, no measurement noise is present, NRA can identify GE in  $P_4$  and  $Q_3$ , resulting in the TL and NTL estimation presented in Table 1. Only the active power values are presented. The estimation errors of TL are related to the incorrect estimation of NTL. In this case, the

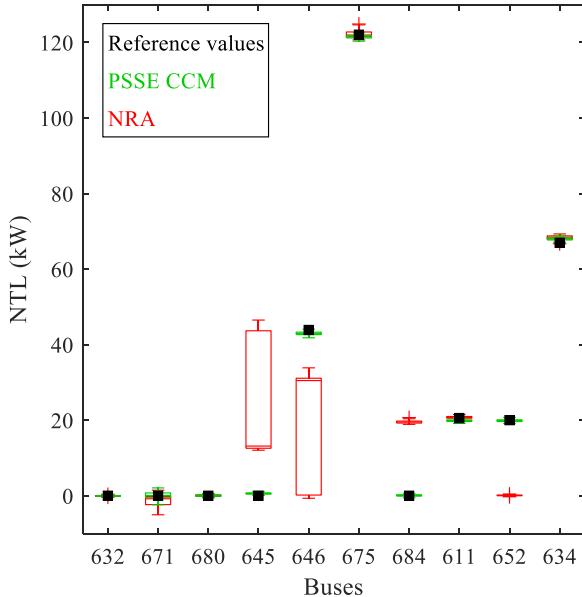


Fig. 9. NTL (sum of kW in all phases) estimation in unbalanced 13 bus system.

TABLE 4

NTL and TL percental estimation error in unbalanced 13-bus feeder.

	Average error: mean ( $\sigma$ )		Maximum error: mean ( $\sigma$ )	
	eNTL (%)	eTL (%)	eNTL (%)	eTL (%)
NRA	7.5 (0.7)	3.8 (0.2)	26.8 (1.4)	23.4 (0.6)
PSSE CCM	0.3 (0.1)	0.5 (0.1)	2.8 (0.5)	4.4 (0.7)

TABLE 5

123-bus feeder: segments and buses with measurements and buses with NTL.

Power flow measurements	150-149; 13-18; 18-35; 54-57; 67-72; 67-97; 13-52; 23-25; 44-47; 60-62; 76-86; 101-105
Voltage measurements	150; 13; 18; 54; 67; 13; 23; 44; 60; 76; 101
NTL	4; 20; 30; 151; 43; 59; 52; 62; 73; 90; 98; 73; 90

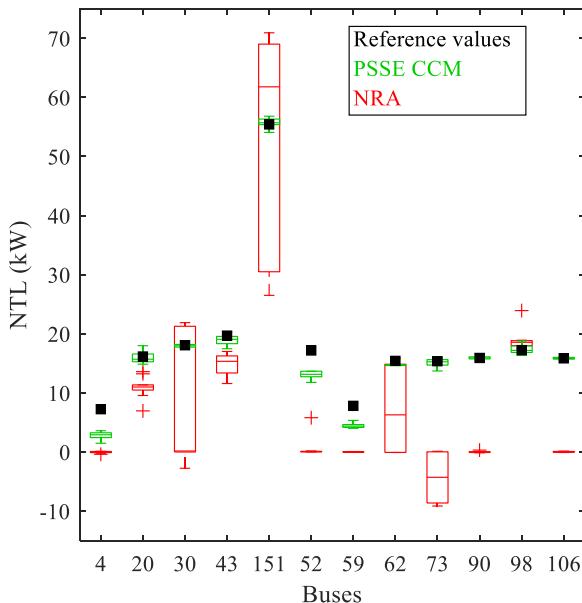


Fig. 10. NTL estimation in unbalanced 123 bus system – buses with NTL.

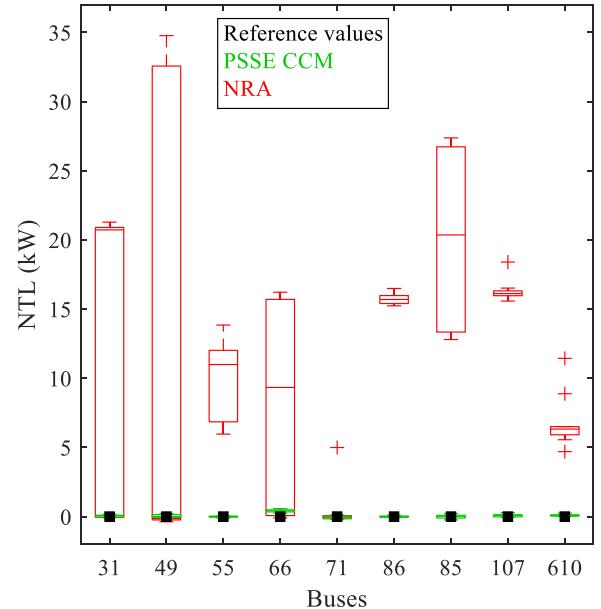


Fig. 11. NTL estimation in unbalanced 123 bus system – buses without NTL.

TABLE 6

NTL and TL percentual estimation error in unbalanced 123 bus feeder.

	Average error: mean ( $\sigma$ )		Maximum error: mean ( $\sigma$ )	
	eNTL (%)	eTL (%)	eNTL (%)	eTL (%)
NRA	7.9 (0.8)	9.7 (1.0)	100.2 (9.6)	161.6 (33.0)
PSSE CCM	0.8 (0.1)	1.4 (0.1)	25 (3.1)	43.0 (4.5)

estimated power flows do not significantly change with the wrong estimation of NTL, therefore, the TL estimation difference from RV are less expressive than those for NTL. In the following tables and figures, the term PSSE CCM is used for the presented framework.

By adding a synthetic power flow measurement from bus 3 to bus 4, the subset formed by the synthetic measurement and the injected power measurements of bus 3 and 4 form a 6-tuple of critical measurements ( $P_3$ ,  $P_4$ ,  $P_{34}$ ,  $Q_3$ ,  $Q_4$  and  $Q_{34}$ ), which is easily verified by inspecting matrix  $H$ . The addition of the synthetic measurement increases the measurement redundancy in this area. Being a 6-tuple of critical measurements, it is now possible to detect and identify GE in four of these measurements. Additionally, the values of  $\Pi$  for measurements related to  $P_3$  and  $P_4$  increased to 0.55 and 1.77. The GMR is artificially increased from 1.71 to 2.

In the presented framework, synthetic measurements are created by performing first a WLS PSSE with weights updated with the CCM result. Authors previous work [14,26] demonstrate that this approach improves the estimation accuracy in DS with NTL. However, in the presented framework, this approach is considered to obtain synthetic measurements values and not to obtain the estimation results as in [14,26]. To obtain a relationship between CCM result and standard deviations that define the weights, first suppose that the NTL values in all buses are known. The  $\sigma$  of each measurement reflects the expected accuracy of the corresponding meter. Given that 99.7% of measurement data are within  $\pm 3\sigma$  of the mean, considering a normal pdf, the maximum acceptable error of a meter can be approximated by  $3\sigma$ . A power injection measurement with NTL will have then the measurement error plus the value of NTL, therefore, the maximum error is equal to  $3\sigma$  added to the NTL value. Thus, an updated value of standard deviation ( $\sigma_{up}$ ) can be considered as (19).

$$\sigma_{up} = (NTL + 3\sigma)/3 \quad (19)$$

In buses without NTL,  $\sigma_{up}$  will be equal to  $\sigma$ , whereas buses with

large NTL values,  $\sigma_{up}$  will be approximately  $NTL/3$ . The higher the NTL in a bus, the lesser accurate measurements are considered in this bus. The updated weight is then computed by  $W_{up} = \sigma_{up}^{-2}$ . This model can be applied for both active and reactive power injections. As a limitation, this weighting approach is defined for distribution systems without distributed storage and generation elements. A change in this model is necessary for MV buses that contain these elements. However, the presented framework remains unchanged.

Considering (19) as updated standard deviation of all power injection measurements, the weight matrix is updated and WLS PSSE is performed without gross error analysis, as illustrated in Fig. 2. The obtained values of active and reactive power flow from bus 3 to bus 4 are then added to the measurement set as synthetic measurements. Considering this new set of measurements, the WLS PSSE is performed again with innovation-based GE analysis [27], considering the original values of standard deviations, as illustrated in Fig. 2.

Considering the modified balanced 4-bus IEEE test feeder, applying the PSSE CCM, all measurements with GE are detected and identified, while the TL and NTL are correctly estimated, as presented in Table 1.

However, NTL values in system buses are previously unknown. To solve this problem, first a CCM should be applied to classify all consumers, obtaining the number of consumers classified as regulars (*nrc*) and suspect (*nsc*) in each phase of each bus. Consider that the mean power consumption ( $Pm_f^k$ ) for a same group of consumers is similar, and that the consumption of suspect consumers it is not computed. Then, the mean consumption can be approximated by (20) and the NTL by (21). For example, for a bus with power injection of 90 kW for a given instant having 360 consumers classified as regular and 40 as suspect, the estimated NTL is 10 kW. The approximation of NTL obtained for each bus by (21), can be used in (19), and considered in the first WLS PSSE to obtain the synthetic measurements. For reactive power, the approach is the same. Note that this is not the final estimation of NTL, just a rough initial guess to update the weight matrix.

$$Pm_f^k \approx \bar{P}_f^k / nrc \quad (20)$$

$$NTL_f^k \approx Pm_f^k nsc \approx \bar{P}_f^k (nsc / nrc) \quad (21)$$

Even in smart DS scenarios with large measurement redundancy, it is unusual to have real-time measurements in all segments. Considering that the lack of power flow measurements is related to power injection measurements with low *II* values and with *n*-tuple of critical measurements, the PSSE CCM considers the values of power flow estimated by the first WLS PSSE in all segments without real-time measurements as synthetic measurements for the second WLS PSSE.

### 3. Case Study

Power measurements were created based on the most typical consumption range of Brazilian consumers, 101-200 kWh per month, and a typical daily load curve for this range of consumption, presented in [43]. Considering a log-normal distribution, several types of consumer patterns were created. The aggregated consumption corresponds to the peak demand values (10 p.m.), which was set equal to IEEE test feeders loads. NTL was created by multiplying consumer load curves by a reducing factor, randomly set between 0 and 0.7. Since the IEEE test feeders considered do not present the low voltage network, bus power injections in each bus were obtained by adding the measured power of the consumers of each spot load. In this case, delta loads were transformed to wye phase power injections in LF to provide the initial values of voltages and the updated bus admittance matrix. In IEEE 13-bus and 123-bus feeders there are voltages regulators, zero injection power buses and two-phase and single-phase lines. The CCM was applied considering the consumer mean power as data input and that the percentage of total NTL is known with some error ( $\pm 1\%$ ). The estimation is presented only for NTL, given that TL are direct related to the NTL estimation. Estimation errors are presented for both TL and NTL.

Simulations were performed in a personal computer Intel Core i7-6500U CPU @ 2.50 GHz with 8.00 GB of installed memory (RAM) using MATLAB software. The computation time to run the PSSE CCM in the 123-bus test system was less than 1 min.

#### 3.1. Unbalanced 4-bus test feeder

Consider the IEEE 4-bus test system (Fig. 4), with changes only in the loads. The load and NTL (total of 8.3%) for the peak demand are presented in Table 2. They correspond to a total of 11450 consumers, 1497 of which with NTL. The CCM obtained the values of 82.4% and 53.7% for *PR* and *RC* indexes.

Before adding synthetic measurements, the *II* of bus 3 active power injection measurements are 0.27, 0.25 and 0.30, and of bus 4 are 1.56, 1.54 and 1.53, phases *a*, *b* and *c*, respectively. Applying the NRA in 100 simulations with a gaussian noise of zero mean and standard deviation of 0.33% of measured value in all measurements, the results of NTL and TL estimation are presented in Fig. 5 and Table 3.

The NRA incorrectly identifies and estimates the power injection of phase *c* of bus 4, instead of bus 3, in all simulations, and the power injection of phase *b* of bus 3, instead of bus 4, in some simulations. Applying the PSSE CCM framework, power flow synthetic measurements are created from bus 3 to bus 4, active and reactive in all phases. The *II* of active power injection measurements of bus 3 increase to 0.52, 0.48 and 0.57, phase *a*, *b* and *c*, respectively. All measurements with NTL are detected and identified, and NTL average estimation errors are lower than 1%. The results of PSSE CCM are also shown in Fig. 5 and Table 3. Table 3 shows a reduction in estimation error when using PSSE CCM. By analyzed Fig. 5, it is also highlighted that NRA identifies and attributes NTL to buses without NTL.

Considering that the previous results were only obtained for the peak demand, a 24-hour simulation was performed to analyze the framework's performance under time-varying load conditions. Note that the CCM is performance once, therefore, the *nrc* and *nsc* remain the same for all simulation hours, on the other hand, synthetic measurements are created for each hour. The Fig. 6 shows the 24-hour reference values of injected power in each phase of the 4-bus system and Fig. 7 presents the NTL reference values (RV) and NTL estimations by PSSE-CCM and NRA. It is only presented the phases and buses with NTL. The results show that PSSE-CCM correctly estimates the NTL in all phases at all hours. The NRA approach fails to estimate NTL in phase *c* of bus 3 (dotted line) at all hours. NRA identifies NTL at some hours for phase *b* of bus 4 (continuous line), and at all hours for phase *a* of bus 3 (dashed line). When both methods identify the NTL, it is possible to graphically analyze that the PSSE-CCM presents lower estimation errors than NRA (black lines closer to the green lines than the red lines). The results of Fig. 7 are in accordance with Fig. 5, demonstrating that the NRA is not able to identify NTL in phase *c* of bus 3 and is not robust for NTL identification in phase *b* of bus 4. The PSSE-CCM remains robust in NTL estimation under load variation. These results were expected given that the synthetic measurements of power flow avoid small sets of critical measurements and increase the *II*. Given that the 24-hour simulation provided the same conclusions of the previous analysis, this simulation was not performed for other test systems.

#### 3.2. Unbalanced 13-bus test feeder

The tests in IEEE 13-bus feeder aims to simulate a smart grid measurement condition with a GMR of 1.92, similar as those considered in [13]. The feeder measurement set and buses with NTL (total of 8.2%) are shown in Fig. 8. The total number of consumers is 7014, and the number with NTL is 852. The CCM obtained the values of 83.4% and 54.2% for *PR* and *RC* indexes.

Synthetic measurements were created in the segments without flow measurements, increasing the GMR from 1.92 to 2.15, and improving the local measurement redundancy in the neighborhood of buses 611,

646 and 652. The local measurement redundancy for buses 675 and 634 is less critical, since there is a flow measurement in the direction of 675 and, in the case of 634, although there is no flow measurement, 634 is preceded by a bus with zero power injection, 633, treated as an equality constraint. It is expected that NTL estimation by NRA perform better for buses 634 and 675, when considering other buses with NTL. The NTL estimation results are presented in Fig. 9. The estimation errors are in presented in Table 4.

NRA fails to identify NTL in bus 646 and 611, wrongly indicating NTL at buses 645 and 684, but correctly estimating the NTL of buses 652, 634 and 675. PSSE CCM correctly identifies all buses with NTL, with values of errors significantly smaller than the NRA. It is important to note that the PSSE CCM also does not incorrectly identifies NTL in buses where it does not exist.

### 3.3. Unbalanced 123-bus test feeder

The IEEE 123-bus test feeder was considered with a GMR of 1.72. Buses with power flow and voltage magnitude measurements and buses with NTL (total of 5.6%) are described in Table 5. Power injection measurements are provided by the LF in all load buses [42]. The total number of consumers is 8052, and 685 consumers with NTL. The CCM obtained the values of 82.0% and 53.3% for PR and RC indexes.

Synthetic measurements were created in all segments without flow measurements, increasing the GMR from 1.72 to 2.53. Considering the NRA, NTL were correctly identified in buses 20, 43, 151, 98, 30, 62, while NTL were not identified in buses 4, 52, 59, 73, 90 and 106, and NTL were wrongly attributed to buses 31, 49, 55, 66, 71, 86, 85, 107 and 610. The PSSE CCM identify all buses with NTL in all simulations. Estimations results are presented in Figs. 10 and 11.

Estimation errors are presented in Table 6. Estimation errors increased in this test system due to the low initial GMR, however, PSSE CCM continues to present significantly better results than the NRA. Additionally, the NRA attributed NTL in buses without NTL, while PSSE CCM identified only buses with NTL. The larger estimation errors in PSSE CCM are related to buses with lower NTL and the TL estimation in segments connected to these buses.

For all test systems, the vulnerable areas of NTL non-identification were identified through the low values of  $II$  and the formation of sets of  $n$ -tuples of critical measurements. These subsets are composed by bus injection measurements connected by segments without flow measurements, as presented in section II.D. The addition of synthetic flow measurements, as proposed in this work, increased the local measurement redundancy of these vulnerable areas, avoiding low subset of critical measurements and increasing the  $II$  of these measurements. The presented model for creating synthetic measurements for the PSSE with information provided by the CCM resulted in a GE analysis improvement. This is highlighted by the results for all study cases, which demonstrate that the GE were correctly estimated by the PSSE-CCM, while NRA performed less well, even assigning NTL in buses without NTL. Similar results were also obtained in 24-hour simulation under load variation. Overall, the presented framework was able to estimate NTL in unbalanced DS with lower estimation errors than NRA.

## 4. Conclusions

In this work a hybrid data-driven physics model-based framework for NTL estimation is presented. Test results show that system segments without power flow measurements tends to form power injection measurements with low  $II$  and  $n$ -tuple of critical measurements, making it most difficult to identify NTL. To increase the local measurement redundancy, the presented framework artificially creates power flow synthetic measurements in all segments without real-time measurements. Synthetic measurements are created by a WLS PSSE weighted by the result of a CCM. Additionally, a LF is considered to obtain power injection measurements in MV buses, the initial states and the updated

bus admittance matrix for WLS PSSE. Solution validation is made in three IEEE test feeders with realistic DS characteristics. Comparative test results highlight decreased average NTL estimation errors. With some modifications in the framework processes, it is possible to consider distributed storage and generation elements. Additionally, NRA incorrectly attributed NTL in buses without NTL for all test systems, which did not occur for the presented framework. This is most important, since consumer inspections are guided by loss estimation results. Finally, the presented framework is easy-to-implement, built on the classical WLS PSSE, without hard-to-derive parameters, which highlight potential aspects for real-life applications.

## Funding

This work was supported by NSF grant ECCS 1809739.

## References

- [1] J. Agüero, Improving the efficiency of power distribution systems through technical and non-technical losses reduction, Proc. IEEE PES T&D (2012) 1–8.
- [2] J.R. Viegas, P.R. Esteves, R. Melício, V.M.F. Mendes, S.M. Vieira, Solutions for detection of non-technical losses in the electricity grid: A review, Renewable Sustainable Energy Rev. 80 (Dec. 2018) 1256–1268.
- [3] G.M. Messinis, N.D. Hatziairgiou, Review of non-technical loss detection methods, Electr. Power Syst. Res., Lausanne 158 (Jan. 2018) 250–266.
- [4] R. Jiang, R. Lu, Y. Wang, J. Luo, C. Shen, X.S. Shen, Energy-theft detection issues for advanced metering infrastructure in smart grid, Tsinghua Sci. Technol. 19 (2) (April 2014) 105–120.
- [5] J. Nagi, S. Tap, S.K. Tiong, S.K. Ahmed, M. Mohamad, Nontechnical loss detection for metered customers in power utility using support vector machines, IEEE Trans. Power Del 24 (4) (Oct. 2009) 2117–2124.
- [6] G.M. Messinis, A.E. Rigas, N.D. Hatziairgiou, A hybrid method for non-technical loss detection in smart distribution grids, IEEE Trans. Smart Grid 10 (6) (Nov. 2019) 6080–6091.
- [7] A.H. Nizar, Z.Y. Dong, Y. Wang, Power utility nontechnical loss analysis with extreme learning machine method, IEEE Trans. Power Syst 23 (3) (Aug. 2008) 946–955.
- [8] S. Chen, T. Zhan, C. Huang, J. Chen, C. Lin, Nontechnical loss and outage detection using fractional-order self-synchronization error-based fuzzy petri nets in micro-distribution systems, IEEE Trans. Smart Grid 6 (1) (Jan. 2015) 411–420.
- [9] C. Leon, F. Biscarri, I. Monedero, J.I. Guerrero, J. Biscarri, R. Millan, Variability and trend-based generalized rule induction model to NTL detection in power companies, IEEE Trans. Power Syst 26 (4) (Nov. 2011) 1798–1807.
- [10] C.C.O. Ramos, A.N. de Souza, A.X. Falcao, J.P. Papa, New insights on nontechnical losses characterization through evolutionary-based feature selection, IEEE Trans. Power Del 27 (1) (Jan. 2012) 140–146.
- [11] R.D. Trevizan, A.S. Bretas, A. Rossoni, Nontechnical losses detection: a discrete cosine transform and optimum-path forest based approach, Proc. IEEE PES NAPS, 2015, pp. 1–6.
- [12] T. Zhan, S. Chen, C. Kao, C. Kuo, J. Chen, C. Lin, Non-technical loss and power blackout detection under advanced metering infrastructure using a cooperative game based inference mechanism, IET Gener., Trans. Distrib. 10 (4) (2016) 873–882 10.3.
- [13] S. Huang, Y. Lo, C. Lu, Non-technical loss detection using state estimation and analysis of variance, IEEE Trans. Power Syst. 28 (3) (Aug. 2013) 2959–2966.
- [14] A. Rossoni, S.H. Braunstein, R.D. Trevizan, A.S. Bretas, N.G. Bretas, Smart distribution power losses estimation: A hybrid state estimation approach, Proc. IEEE/ PES GM (2016) 1–5.
- [15] L.T. Faria, J.D. Melo, A. Padilha-Feltrin, Spatial-temporal estimation for non-technical losses, IEEE Trans. Power Del 31 (1) (Feb. 2016) 362–369.
- [16] M. Tariq, H.V. Poor, Electricity theft detection and location in grid-tied microgrids, IEEE Trans. Smart Grid 9 (3) (May 2018) 1920–1929.
- [17] G.M. Messinis, N.D. Hatziairgiou, Unsupervised classification for non-technical loss detection, Proceedings of the 2018 Power Systems Computation Conference, 2018.
- [18] M.M. Buzau, J. Tejedor-Aguilera, P. Cruz-Romero, A. Gomez-Exposito, Detection of non-technical losses using smart meter data and supervised learning, IEEE Trans. Smart Grid 10 (3) (May 2019) 2661–2670.
- [19] N.F. Avila, G. Figueiroa, C.C. Chu, NTL detection in electric distribution systems using the maximal overlap discrete wavelet-packet transform and random under-sampling boosting, IEEE Trans. Power Syst 33 (6) (Nov. 2018) 7171–7180.
- [20] J.B. Leite, J.R.S. Mantovani, Detecting and locating non-technical losses in modern distribution networks, IEEE Trans. Smart Grid 9 (2) (March 2018) 1023–1032.
- [21] L.A.P. Junior, C.C.O. Ramos, D. Rodrigues, D.R. Pereira, A.N. De Souza, K.A. Pontara da Costa, J.P. Papa, Unsupervised non-technical losses identification through optimum-path forest, Electric Power Syst. Res. 140 (2016) 413–423.
- [22] M. Zanetti, E. Jamhour, M. Pellenz, M. Penna, V. Zambenedetti, I. Chueiri, A tunable fraud detection system for advanced metering infrastructure using short-lived patterns, IEEE Trans. Smart Grid 10 (1) (2019) 830–840.
- [23] K. Zheng, Q. Chen, Y. Wang, C. Kang, Q. Xia, A novel combined data-driven

approach for electricity theft detection, *IEEE Trans. Ind. Inform.* 15 (3) (2019) 1809–1819.

[24] E.A.C. Aranha Neto, J. Coelho, Probabilistic methodology for technical and non-technical losses estimation in distribution system, *Electr. Power Syst. Res.* 97 (1) (Apr. 2013) 93–99.

[25] M.E. Oliveira, A. Padilha-Feltrin, A top-down approach for distribution loss evaluation, *IEEE Trans. Power Del* 24 (4) (Oct. 2009) 2117–2124.

[26] A. Rossoni, S.H. Braunstein, R.D. Trevizan, A.S. Bretas, N.G. Bretas, Contribution to Distribution Systems Technical and Nontechnical Losses Estimation using WLS State Estimator, *Proc. IEEE/PES GM* (2017) 1–5.

[27] A.S. Bretas, N.G. Bretas, S.H. Braunstein, A. Rossoni, R.D. Trevizan, Multiple gross errors detection, identification and correction in three-phase distribution systems WLS state estimation: A per-phase measurement error approach, *Electr. Power Syst. Res.* 151 (1) (Oct. 2017) 174–185.

[28] S.A. Salinas, P. Li, Privacy-preserving energy theft detection in microgrids: a state estimation approach, *IEEE Trans. Power Syst* 31 (2) (Mar. 2016) 883–894.

[29] A.S. Bretas, N.G. Bretas, B.B. Carvalho, E. Baeyens, P.P. Khargonekar, Smart grids cyber-physical security as a malicious data attack: An innovation approach, *Electr. Power Syst. Res.* 149 (2017) 210–219.

[30] C. Ruben, S. Dhulipala, K. Nagaraj, A. Starke, A.S. Bretas, A. Zare, J. McNair, Hybrid data-driven physics model-based framework for enhanced cyber-physical smart grid security", (accepted for publication), *IET Smart Grids* (2020).

[31] P. Jokar, N. Arianpoo, V.C.M. Leung, Electricity theft detection in ami using customers' consumption patterns, *IEEE Trans. Smart Grid* 7 (1) (Jan. 2016) 216–226.

[32] A. Monticelli, State Estimation in Electric Power Systems: A Generalized Approach, 1st ed, (1999) vol. 1, Springer.

[33] N.G. Bretas, A.S. Bretas, S.A. Piereti, Innovation concept for measurement gross error detection and identification in power system state estimation, *IET Gener., Transm. Distrib.* 5 (6) (June 2011) 603–608.

[34] N.G. Bretas, A.S. Bretas, A.C. Martins, Convergence property of the measurement gross error correction in power system state estimation, using the geometrical background, *IEEE Trans. Power Syst.* 28 (4) (Nov 2013) 3729–3736.

[35] A. Abur, A. Gómez-Expósito, *Power System State Estimation: Theory and Implementation*, 1st ed, Marcel Dekker, New York, 2004 vol. 1.

[36] N.G. Bretas, A.S. Bretas, The Extension of the gauss approach for the solution of an overdetermined set of algebraic non linear equations, *IEEE Trans. Circuits Syst., II, Exp. Briefs* 65 (9) (Sept. 2018) 1269–1273.

[37] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection: a survey, *ACM Comput. Surv.* 41 (1) (2009) 1–72 Set.

[38] W.H. Kersting, *Distribution System Modeling and Analysis*, 3th ed, CRC Press, Boca Raton, 2012, pp. 1–450 vol. 1.

[39] M. Bazrafshan, N. Gatsis, Comprehensive modeling of three-phase distribution systems via the bus admittance matrix, *IEEE Trans. Power Syst.* 33 (2) (Mar. 2018) 2015–2029.

[40] A.S. Bretas, N.G. Bretas, B.B. Carvalho, Further contributions to smart grids cyber-physical security as a malicious data attack: proof and properties of the parameter error spreading out to the measurements and a relaxed correction model, *Int. J. Electr. Power Energy Syst.* 104 (2019) 43–51.

[41] N.G. Bretas, S.A. Piereti, A.S. Bretas, A.C.P. Martins, A Geometrical View for multiple gross errors detection, identification, and correction in Power System State Estimation, *IEEE Trans. Power Syst* 28 (3) (Jan 2013) 2128–2135.

[42] IEEE/PES Distribution Systems Analysis Subcommittee. (2018, Aug.). Radial Test Feeders. [Online]. Available: <http://ewh.ieee.org/soc/pes/dsacom/testfeeders.html>.

[43] J.A. Jardini, C.M.V. Tahan, M.R. Gouvea, S.U. Ahn, F.M. Figueiredo, Daily load profiles for residential, commercial and industrial low voltage consumers, *IEEE Trans. Power Del* 15 (1) (Jan. 2000) 375–380.