Non-Technical Loss Identification by Using Data Analytics and Customer Smart Meters

Lívia M. R. Raggi, Fernanda C. L. Trindade, *Member*, *IEEE*, Vinícius C. Cunha, *Student Member*, *IEEE*, Walmir Freitas, *Member*, *IEEE*

***Abstract*—The smart meters installed at the customer premises are one of the main apparatuses promoting the modernization of the distribution systems. These devices collect a huge amount of data, demanding the development of analytic techniques to transform these data into useful information. In addition, by proposing new applications to data supplied by smart meters, more value is added to this equipment, allowing higher return on the associated investments. Utilities can have a business case by increasing their operational efficiency if smart meters are used, for instance, to identify non-technical losses, which represent an important cause of revenue losses. This paper presents a new data analytic technique for detection and location of non-technical losses caused by illegal connections of loads to distribution systems in the presence of smart meters. The data analytic technique relies on bad data analysis, similar to the ones used in state estimation methods, developed specifically for this application. A real 34-bus low voltage system is used to illustrate the main concepts of the proposed algorithm. Systematic tests are also conducted on a real 1,682-bus distribution system to evaluate the method performance considering electricity theft caused by medium and low voltage customers.**

***Index Terms*—Distribution management systems, non-technical losses, smart meters, bad data analysis.**

I. Introduction

TECHNOLOGICAL advances have motivated the monitoring of distribution systems by the development of an Advanced Metering Infrastructure (AMI). New features have been incorporated into the traditional energy meters, which characterize the so-called smart meters when associated with the possibility of two-way data communication. The large amount of available data resulting from this scenario can benefit several applications in the context of Advanced Distribution Management Systems (ADMS) [1].

Indeed, several applications based on the usage of data from smart meters have been proposed recently. Some examples are functionalities to: manage the life loss and failure risk of transformers [2]-[4], analyze harmonic levels at the customers [5], and locate faults in medium voltage (MV) networks [6]-[7]. These measurements also have the potential to better predict patterns on customers consumption [8], which indirectly improves the performance of data analytics techniques proposed to determine the distributed generation (DG) hosting capacity on low voltage (LV) networks [9], and maintain power quality within regulatory limits [10]-[11]. By developing proper applications for these data, the operational efficiency of distribution systems can be improved, so that more value is added to the smart meters, leading to a higher return on the investments in the AMI. One relevant issue that can be benefited from this scenario is the process of detection and location of non-technical losses (NTLs).

NTLs are a relevant cause of revenue loss for utilities and energy efficiency reduction in distribution systems. Most of the cases of NTLs result from meter tampering; illegal load connections; billing irregularities; malfunction of meters; and unpaid bills [12]. Most of these practices are associated with social and cultural issues or illegal activities. In Brazil, for instance, NTL associated costs are estimated in US$ 2.3 billion per year, corresponding to more than 27 TWh – approximately 8% of national residential and commercial energy consumption [13]. Even in developed countries, the level of NTL can be significant. The energy costs related to NTL considering only one major Canadian utility, for example, is approximately US$ 75 million a year, corresponding to 850 GWh [14]. Such high revenue losses and efficiency reductions in terms of energy motivate regulatory agencies and utilities to investigate and develop new methods of NTL detection and location.

Most of the traditional techniques of NTL detection and location are based on the past record of energy consumption, as in [15]-[19], which employ classification techniques for identification of abnormal customer load profile. Some of these previously proposed methods are based on the installation of sensors along the feeder, as in [20]. With the deployment of smart meters, other categories of detection methods and improved classification techniques have been investigated. In [21], a decision-making model based on a non-cooperative game is used to identify NTL, by comparing profiled usage with real-time usage obtained from smart meters. In [22], smart meter data are used to obtain information of customer's consumption behavior and a supervised learning technique is used to detect NTL. Smart meter bypass is modeled as a measurement bias in [23] and a privacy-preserving bias estimation is conducted based on Kalman filter. The Stochastic Petri Net formalism is used in [24] to locate NTL in microgrids. In [25], the energy consumption of distribution transformers is compared with the aggregated consumption recorded by smart meters to delimit areas suspected of NTL.

Despite the advances, these methods have not been able to properly solve the problem, as it can be certified by the high level of NTL still present in the systems. Therefore, in this work, we proposed a new approach derived from traditional methods assuming a scenario of wide-scale deployment of smart meters installed at the customers for billing, which is already reality in several utilities [26]. The proposed method detects and locates NTLs caused by illegal load connections in

This work was supported in part by CPFL Energia grant PD-0063-3048/2018, Sao Paulo Research Foundation (FAPESP) grants 2017/10476-3 and 2016/08645-9, and National Council for Scientific and Technological Development (CNPq) grants 304783/2016-1 and 432347/2018-6. The first author acknowledges the Brazilian Electricity Regulatory Agency (ANEEL) for the leave of absence to conduct part of this research. L. M. R. Raggi is with ANEEL, Brazil (e-mail: liviaraggi@aneel.gov.br). F. C. L. Trindade, V. Cunha and W. Freitas are with the Department of Systems and Energy, University of Campinas, Brazil (e-mails: fernanda@ieee.org; vcunha@ieee.org, walmir@ieee.org).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2020.2974132, IEEE Transactions on Power Delivery

MV and LV distribution systems with a high success rate. By using information obtained from smart meters, a data analytical technique, derived from a three-phase state estimator, is developed specifically for NTL detection and location. The approach uses the Weighted Least Squares (WLS) formulation and concepts of bad data analysis. A new index based on the measurement residual correlation is proposed to detect and locate NTLs. It is important to emphasize that the proposed method is designed to be performed offline and it does not intend to estimate the state of the distribution system per se, as traditional state estimation (SE) tools. Instead, the idea was developed as a specific application method for detection and location of NTLs.

This paper is organized as follows. Section II presents the NTL effects on load measurements and on the resulting measurement residuals. Section III presents the new method based on bad data analysis for NTL detection and proposes an index for NTL location. In Section IV, the algorithm performance is thoroughly evaluated in a real 34-bus LV system, and further validated in a real 1,682-bus MV/LV distribution system. The main conclusions are summarized in Section V.

II. Residual Behavior in the Presence of NTL

This section presents a discussion on measurement residual behavior in the presence of NTL. Fig. 1 schematically represents a generic customer, which, for the sake of generalization, may have a distributed generator (DG), with an illegal load connection, also known as irregular load, characterizing an NTL. It is also assumed that the customer meter can measure active and reactive power (P_2, Q_2) and voltage magnitude (V_2).

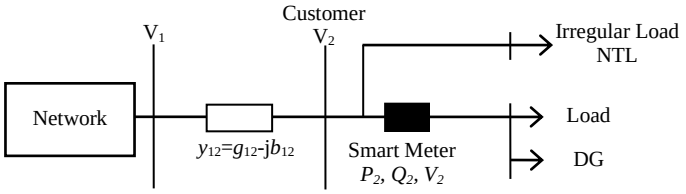


Fig. 1. NTL resulting from irregular load connection.

Because of the irregular load, P_2 and Q_2 measurement data do not correspond to the actual power consumption at the customer. However, information of voltage magnitude is consistent with reality as it includes the system voltage drop due to the total power consumption – from the irregular and regular loads. To put it simply, by bypassing the smart meter, one can tamper power measurements but cannot tamper voltage measurement. This fact is important as it allows identifying irregularities in meter reading as proposed in this paper.

Although the measurement errors in P, Q, and V are assumed to be independent of each other, in a SE process, errors in these measurements result in strongly correlated residuals, as discussed in the sequence. The measurement residual vector $\mathbf{r}$ is defined as the difference between measurement vector $\mathbf{z}^{mea}$ and estimated measurement vector $\mathbf{z}^{est}$, as shown in (1).

$$\mathbf{r} = \mathbf{z}^{mea} - \mathbf{z}^{est} \tag{1}$$

From the NTL problem perspective, the residuals of active power (r_P), reactive power (r_Q) and voltage (r_V), calculated as in (2), are further investigated to address how they interact with each other.

$$\mathrm{r}_P = P^{mea} - P^{est}, \mathrm{r}_Q = Q^{mea} - Q^{est}, \mathrm{r}_V = V^{mea} - V^{est} \tag{2}$$

The residuals interaction can be mathematically and analytically analyzed by expanding the WLS formulation to a simple circuit, as the single-phase customer presented in Fig. 1. Firstly, the active and reactive power at bus 2 is calculated as:

$$P_2 - jQ_2 = \hat{V}_2^{*}(\hat{V}_1 - \hat{V}_2)y_{12} \tag{3}$$

Expanding (3) and neglecting the angle difference between the voltages of two consecutive buses, as this difference is small in distribution systems (*i.e.*, $\theta_{12} \approx 0$) [27], the estimated active and reactive power are shown in (4). The subscript ν refers to the iteration counter for the solution of the non-linear equations.

$$\begin{aligned} P_2^{\,est}{}_{\nu+1} &= -V_2^{\,est}{}_{\nu}(V_1^{mea} - V_2^{\,est}{}_{\nu})g_{12} \\ Q_2^{\,est}{}_{\nu+1} &= -V_2^{\,est}{}_{\nu}(V_1^{mea} - V_2^{\,est}{}_{\nu})b_{12} \end{aligned} \tag{4}$$

The sign convention adopted in (4) defines loads as negative power injections and, consequently, power generation as positive power injections. In the sequence, the classical WLS formulation is presented in (5).

$$\underbrace{\mathbf{H}^T\mathbf{W}\mathbf{H}}_{\boldsymbol{G}}\mathbf{x} = \mathbf{H}^T\mathbf{W}\underbrace{\left[\mathbf{z}^{\mathbf{mea}} - \mathbf{z}^{\mathbf{est}}\right]}_{\mathbf{r}} \tag{5}$$

where:

$$\mathbf{H} = \begin{bmatrix} \partial P_2/\partial V_2 \\ \partial Q_2/\partial V_2 \\ \partial V_2/\partial V_2 \end{bmatrix} = \begin{bmatrix} -V_1^{mea}g_{12} + 2V_2^{\,est}{}_{\nu}g_{12} \\ -V_1^{mea}b_{12} + 2V_2^{\,est}{}_{\nu}b_{12} \\ 1 \end{bmatrix}, \mathbf{z}^{\mathbf{mea}} = \begin{bmatrix} P_2^{mea} \\ Q_2^{mea} \\ V_2^{mea} \end{bmatrix}$$
$$\mathbf{W} = \begin{bmatrix} W_P & 0 & 0 \\ 0 & W_Q & 0 \\ 0 & 0 & W_V \end{bmatrix}, \mathbf{z}^{\mathbf{est}} = \begin{bmatrix} P_2^{\,est}{}_{\nu} \\ Q_2^{\,est}{}_{\nu} \\ V_2^{\,est}{}_{\nu} \end{bmatrix}, \mathbf{x} = \begin{bmatrix} V_2^{\,est}{}_{\nu+1} \end{bmatrix} \tag{6}$$

where W_P, W_Q, and W_V are the power and voltage weights. The result of the matrix operation in (5) is shown in (7). The non-linear equations (4) and (7) are solved iteratively until V_2^{est} presents a marginal variation compared to the previous iteration. Generalizing, the following information can be derived from this model:

- The measured active power P^{mea} is always higher than the real active power P^{real}, regardless the presence of DG, because of bypassed load, *i.e.*, the NTL. Thus, the estimated active power P^{est} will place within these two limits, P^{mea} and P^{real}, resulting in a positive r_P (as illustrated in Fig. 2);
- For most practical cases, the measured reactive power Q^{mea} is higher than the real reactive power Q^{real}, as the bypassed load, *i.e.*, the NTL, and the measured load typically have inductive power factors. Therefore, r_Q is also positive. Indeed, as long as the signal of the power factor of the measured load and the bypassed load are the same, *i.e.*, inductive or capacitive, r_Q is positive;
- The term that multiples the square brackets sum in (7) is positive when the system operates within regulatory limits. Therefore, the numerical convergence is obtained if and only if r_V presents opposite sign from square brackets sum in (7), that is, opposite sign from r_P and r_Q;

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2020.2974132, IEEE Transactions on Power Delivery

$$V_2^{est}{}_{\nu+1} = \frac{1}{\underset{positive\ non-zero}{A}} \left\{ \underbrace{(-V_1^{mea} + 2V_2^{est}{}_{\nu})}_{positive} \left[W_P g_{12} \overbrace{(P_2^{mea} - P_2^{est}{}_{\nu})}^{r_P} + W_Q b_{12} \overbrace{(Q_2^{mea} - Q_2^{est}{}_{\nu})}^{r_Q} \right] + W_V \overbrace{(V_2^{mea} - V_2^{est}{}_{\nu})}^{r_V} \right\} \quad (7)$$

$$A = W_P(-V_1^{mea} g_{12} + 2V_2^{est}{}_{\nu} g_{12})^2 + W_Q(-V_1^{mea} b_{12} + 2V_2^{est}{}_{\nu} b_{12})^2 + W_V$$

- Because of the previous points, the r_V has to be negative, what means that estimated V^{est} is higher than the measured V^{mea} and, consequently, the true V^{real};

The proportion of r_P, r_Q, and r_V required to converge the problem will be dictated by the power (W_P and W_Q) and voltage (W_V) weights. This means that the weights directly affect the converged values for the residuals. In other words, the influence of the NTL is not restricted to the estimated active or reactive power (P^{est} and Q^{est}) but also to the estimated voltage V^{est}. A detailed discussion about how to set the values of the weights and the corresponding impacts on the method performance are presented in Section IV.

The previous example can be generalized as illustrated in Fig. 2, assuming unity power factor for simplicity, for a customer k (a) demanding active power, *i.e.*, without DG or when the DG power is lower than the load consumption, and (b) injecting active power, with DG and when DG power is higher than the load consumption. Variable e represents measurement errors from the meter uncertainty or the NTL.

III. Proposed Method

One of the challenges to locate the NTL using the power residuals consists that the influence of the irregular load may be spread in others estimated parameters, as previously discussed. The solution presented in the literature for such problem involves analyzing the normalized residuals by the covariance matrix $\mathbf{\Omega}$ [28]. However, although such treatment is efficient to detect NTL, the location of these irregular loads, which is equally important, is still jeopardized by this residual behavior, resulting in wrong conclusions (*e.g.*, wrong bus pointed out with NTL). Therefore, to counteract this difficulty, this work proposes a method to locate NTL based on a new index, which is mathematically defined in the next subsections.

A. Bad Data Analysis

WLS state estimation assumes that measurement errors e_i have a Gaussian distribution with zero mean and standard deviation σ_i, *i.e.*, $e_i \sim N(0, \sigma_i)$. Bad data analysis is essentially based on residual properties for a certain probability distribution of the measurement error [28]. A method used to detect bad data in a measurement set is the normalized residual test. Measurement i residual is calculated as in (1) and normalized by using (8).

$$r_i^N = \frac{r_i}{\sqrt{\Omega_{ii}}} \quad (8)$$

where Ω_{ii} is the corresponding diagonal entry in the residual covariance matrix $\mathbf{\Omega}$ [28]. The normalized residuals are expected to have a standard Gaussian distribution with zero mean and unitary standard deviation, *i.e.*, $r_i^N \sim N(0,1)$.

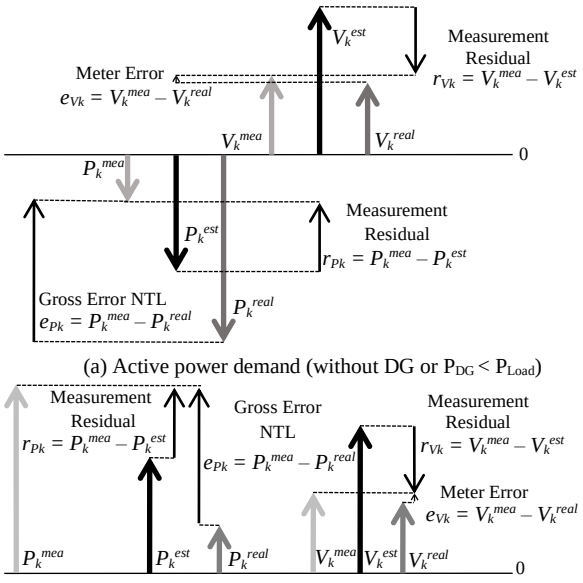


Fig. 2. Illustration of measurement residuals and errors in the presence of NTL at a customer k with (a) real power demand and (b) real power injection.

In the normalized residual test, the highest element in $\mathbf{r^N}$ is compared with a threshold δ to evaluate the existence of bad data. This threshold δ is defined as 3 times the respective standard deviation, as typically used in SE analysis [28]. Usually, once detected, the bad data is removed from measurement set. As removing measurement can result in an unobservable system, one can correct the erroneous measurement i by updating z_i as in (9) [28].

$$z_i = z_i^{bad} - e_i^{est} \quad (9)$$

where superscript *bad* denotes bad data and e_i^{est} corresponds to the estimated error.

B. Measurement Error Estimation

In this paper, the concept of Composed Measurement Error (CME) is used to estimate measurement errors [29]-[32]. According to this approach, the vector of measurement errors is composed of two components, which are orthogonal to each other, called detectable and undetectable components. The CME is given by the vector addition of these two components, and the residuals are the detectable error component. Once the erroneous measurement is detected and located, its adjustment can be done by using the corresponding Composed Normalized Error (CNE) multiplied by the respective measurement standard deviation σ_i, *i.e.*:

$$e_i^{est} = \sigma_i \cdot \frac{CME_i}{\sqrt{\Omega_{ii}}} = \sigma_i \cdot CNE_i \quad (10)$$

where:

$$CNE_i = \left[1 + \left(\frac{\sqrt{S_{ii}}}{\sqrt{1 - S_{ii}}} \right)^{-2} \right]^{1/2} \cdot r_i^N \quad (11)$$

S_{ii} is the i^{th} element of the main diagonal of the sensitivity matrix **S**, as described in Section III.C. More details about this approach can be obtained in [29]- [32].

C. Residual Vector Update

In traditional bad data treatment, after correcting an erroneous measurement, a new SE is performed to update the state variables and residual vector, allowing the identification of further bad data. However, the major interest in NTL detection and location is not to obtain the state of the system, but to identify suspicious customers. Therefore, a different approach is proposed to update the residual vector. This approach is based on the sensitivity matrix **S** [28], which reflects the sensitivity of measurement residuals **r** to the respective errors **e**, as in (12).

$$\mathbf{r} = \mathbf{S} \cdot \mathbf{e} \tag{12}$$

Element S_{mi} indicates the influence of measurement *i* error on the estimated value of measurement *m*, giving information about residual correlation. To update the residual vector in bad data treatment, firstly, the resulting variation of measurement error Δe_i is obtained by (13).

$$\Delta e_i = z_i - z_i^{bad} = -e_i^{est} \tag{13}$$

Then, the variation on the residual vector **r** can be calculated by means of (14).

$$\Delta \mathbf{r} = \mathbf{S}_i \cdot \Delta e_i \tag{14}$$

and

$$\mathbf{r}' = \mathbf{r} + \Delta \mathbf{r} \tag{15}$$

where $\mathbf{S}_i$ is the i^{th} column of **S** matrix and **r'** is the updated residual vector. Substantial load changes or the existence of bad data do not strongly influence **S** matrix if the system configuration is maintained. Consequently, under such conditions, **S** can be considered a constant matrix [33]. Based on this assumption, in this work, (13) to (15) are used to update the residual vector after a measurement correction instead of executing a new SE. Therefore, SE is performed only once to provide the initial residual vector and matrix **S**. However, it is important to highlight that the proposed index discussed in the next section could be calculated as a byproduct of a complete state estimation process.

D. Proposed Index Ψ

As discussed in Section II, NTL is characterized by positive residuals of active and/or reactive power and negative residuals of voltage magnitude. This behavior is used in this paper to propose a method to locate suspicious customers. Firstly, NTL detection is conducted by analyzing the normalized residuals of active and reactive power measurements, r^N_{Pk} and r^N_{Qk}, respectively. NTL is detected if r^N_{Pk} or r^N_{Qk} positive values exceed the specified threshold δ, indicating the presence of bad data. Then, suspect customer buses are selected by evaluating the index Ψ. This index is defined for each customer *k* whose normalized residuals, r^N_{Pk} or r^N_{Qk}, violate the specified threshold δ and is calculated by using (16), where *max*(*A*, *B*) is a function that calculates the maximum value between *A* and *B* and r^N_{Vk} is the normalized residual of voltage measurements.

$$\Psi_k = max(r^N_{P_k} - r^N_{V_k},\ r^N_{Q_k} - r^N_{V_k}) \tag{16}$$

When a three-phase system model is adopted, as in this work, index Ψ_k represents the highest residual composition of the phases of bus *k*. The bus with the highest index Ψ is initially selected as the suspect NTL location and the corresponding active or reactive power measurement is treated as bad data. Because measurement residuals are linear combination of measurement errors, not always the erroneous measurements present the highest residuals. For instance, power injection measurements from neighbor buses can present higher residuals than those from NTL associated buses. To overcome this issue, the proposed index Ψ takes advantage of voltage magnitude and power residuals correlation to highlight the affected bus (or buses), avoiding wrong conclusions that may occur when only power injection measurement residuals are assessed.

E. Proposed Algorithm

The following algorithm is proposed to detect and locate NTL in distribution networks:

Step 1: Calculate one solution of the WLS method and determine matrix **S**.

Step 2: Set the suspect irregular power for each bus *k* as zero, $Err_{NTLk} = 0$; the number of NTL suspect customers as zero, $N_{NTL} = 0$; and the iteration counter as 1, *iter* = 1.

Step 3: Calculate normalized residuals $\mathbf{r^N}$ and Composed Normalized Error **CNE**.

Step 4: Check if there is any normalized residual of active or reactive power measurements, r^N_{Pk} or r^N_{Qk}, higher than the specified threshold δ (which is set as three times the standard deviation, as typical in SE process). If "yes", proceed to *Step 5*. If "no", proceed to *Step 9*.

Step 5: For each bus *k* whose normalized residual r^N_{Pk} or r^N_{Qk} violates the threshold δ, calculate the index Ψ_k. The NTL suspect bus, k_{NTL}, is defined as the one with the highest value of index Ψ_k (17):

$$k_{NTL} \rightarrow \max(\Psi_k) \tag{17}$$

Step 6: If the current k_{NTL} has not been identified yet as a suspect bus, increment the number of suspect customers, $N_{NTL} = N_{NTL} + 1$. Otherwise, select the already identified k_{NTL}.

Step 7: Once the suspect bus k_{NTL} is selected, estimate the measurement error of the power injection by means of (10) and (11). Accumulate the estimated error in the corresponding complex variable Err_{NTLk} (18) as follows:

$$Err_{NTL_k}{}^{iter} = Err_{NTL_k}{}^{iter-1} + \begin{cases} e^{est}_{P_{k,ph}} & if\ \ bad\ data\ is\ in\ P^{mea}_{k,ph} \\ je^{est}_{Q_{k,ph}} & if\ \ bad\ data\ is\ in\ Q^{mea}_{k,ph} \end{cases} \tag{18}$$

where *ph* corresponds to *ph* phase of k_{NTL}-bus.

Step 8: Determine the error variation using (13) and update the residual vector using (14) and (15). Do *iter* = *iter* + 1 and return to *Step 3*.

Step 9: If $N_{NTL} = 0$, there is no suspect bus, finalize the process. Otherwise, go to *Step 10*.

Step 10: For each bus *k*, determine Err_{NTLk}(%) using (19).

$$Err_{NTL_k}(\%) = \frac{\left|Err_{NTL_k}\right|}{\sum_{N_{NTL}} \left|Err_{NTL_k}\right|} \cdot 100\% \tag{19}$$

where $|Err_{NTLk}|$ is the absolute value of Err_{NTLk}.

Step 11: Rank vector Err_{NTL}(%) in descending order for all buses. Finalize the process.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2020.2974132, IEEE Transactions on Power Delivery

As one can observe, the iterative process is required to (a) identify multiple NTL, as the algorithm detects and locates NTL one by one, and (b) to update the suspect irregular power for each bus, as the suspect irregular power identified for a bus in a single iteration may not represent the total NTL. In the last case, the next iteration will indicate an increment in the suspect irregular power to the evaluated bus, which will tend to the true NTL value. Therefore, the suspect irregular active and reactive power for each bus k_{NTL} is accumulated in the complex variable $\boldsymbol{Err_{NTL}}$ along the iterative process. Err_{NTLk}(%), which is a real number as defined in (19), is used at the end of the procedure to rank suspect buses in descending order, indicating the most likely NTL locations. The algorithm result is a list of suspect buses ranked by Err_{NTLk}(%).

IV. Case Studies

Several case studies are presented to assess the efficacy of the proposed method for NTL detection and location. Firstly, a study is performed in a real 34-bus LV distribution system in order to illustrate the algorithm concepts. Then, a sensitivity study is conducted in a real 1,682-bus distribution feeder, whose data is available in [34], considering the presence of NTL in LV and MV customers, as well as multiple NTL occurrences. For all simulated cases, it is considered that at each customer bus k the active and reactive power injection, P_k^{mea} and Q_k^{mea}, and voltage magnitude, V_k^{mea}, are supplied by smart meters. In addition, at each no-load bus k the active and reactive power injection, P_k^{mea} and Q_k^{mea}, are assumed as virtual measurements, which are set equal to zero as they are associated with buses with no injection.

The per-unit values of the standard deviation σ used in the weight matrix of the WLS method are shown in Table I, where the base power is 1 MVA and the base voltage is the rated voltage. To test the method, measurements are synthetically generated by adding random noise to power flow results. These noises follow a Gaussian distribution and are limited to meter accuracy, defined as 0.2% for voltage magnitude measurements and 1.0% for active and reactive power measurements.

In traditional state estimation procedures, the standard deviations used in the WLS weight matrix reflects the accuracy of the measurement devices. In this work, the standard deviation assigned to V_k^{mea} is calculated from meter accuracy, by dividing the maximum expected error (in pu) by 3, as in [35] (considering V_k^{mea} = 1 pu). However, the standard deviations of P_k^{mea} and Q_k^{mea} are chosen so that a defined range of NTL magnitude can be detected (which corresponds to a few kWs), and measurement noises are not identified as gross errors. Therefore, to detect NTL above 1 kW ($1{\cdot}10^{-3}$ pu) in LV systems, one can assume that this value is the maximum expected error of P_k^{mea}, which corresponds to a standard deviation of approximately $3{\cdot}10^{-4}$. The same logic is applied to power measurements in MV systems. Also, as can be observed from Table I, the same standard deviation (or weight) is assigned to similar magnitude measurements, for instance, the value $\sigma = 3{\cdot}10^{-4}$ pu is assigned to all P_k^{mea} located in LV systems. This improves the method performance, as the same importance is given to similar measurements. The small values of σ assigned to P_k^{mea} and Q_k^{mea} in LV systems is because P_k^{mea} and Q_k^{mea} and their corresponding noises, in pu, are smaller in these systems than in the MV feeder and the HV/MV substation.

TABLE I
STANDARD DEVIATIONS ASSIGNED TO MEASUREMENTS IN THE WEIGHT MATRIX

Data source	Data type	σ (pu)	
		MV system (13.8 kV)	LV system (0.22 kV)
Smart Meter	P_k^{mea},Q_k^{mea}	$1{\cdot}10^{-3}$	$3{\cdot}10^{-4}$
	V_k^{mea}	$6.7{\cdot}10^{-4}$	$6.7{\cdot}10^{-4}$
Distribution HV/MV Substation	P_k^{mea},Q_k^{mea}	$5{\cdot}10^{-2}$	-
	V_k^{mea}	$6.7{\cdot}10^{-4}$	-
Virtual Measurement	P_k^{mea},Q_k^{mea}	$1{\cdot}10^{-4}$	$1{\cdot}10^{-5}$

A. Case Study 1 – Method analysis (34-bus system)

The following example illustrates the method results for a real LV distribution system with NTL. This system is supplied by a 112.5 kVA, 13.8/0.22 kV transformer and composed of 34 buses (as shown in Fig. 3). System data is available in [34]. It is simulated a 3-kW NTL at bus 23, which is associated to an irregular load connected between phases *a* and *b*. The bad data, consequently, is present in active power measurement at bus 23, P_{23}^{mea}, at phases *a* and *b*. This example will be used to illustrate in details the algorithm defined in Section III.E.

Fig. 4 and Fig. 5 present the normalized residuals $\mathbf{r^N}_P$ and $\mathbf{r^N}_V$ for the first iteration obtained in Step 3. In these figures, only buses where the corresponding phases have P_k^{mea}, Q_k^{mea} and V_k^{mea} measurements are discriminated on axis *x*. Notice, in Fig. 4, that four buses (6, 11, 17, and 23) have $\mathbf{r^N}_P$ higher than 3, which means that, according to Step 4, an NTL is detected. After detecting an NTL, the next step (Step 5) corresponds to locate the bus with the illegal load, which is done by evaluating the index Ψ. The highest index Ψ refers to the bus 23 on phase b, indicating this bus as the suspect one. The next steps refer to the estimation of the NTL, *i.e.*, Err_{NTL23}, (Step 7) and update of the residual vector (Step 8), eliminating the influence of the identified NTL. This ends the first iteration of the algorithm.

As the system may have multiple NTLs or the identified NTL may only represent part of the power from the true NTL, new iterations must be conducted as long as Step 4 returns a positive outcome for the detection. In this example, another iteration to update the NTL of bus 23 (*i.e.*, Err_{NTL23} from 1.4 kVA to 2.1 kVA) is enough to converge the algorithm, as summarized in Table II. The first column refers to algorithm iteration. The other columns present r^N_{Pk}, r^N_{Qk}, r^N_{Vk}, the iteration suspect bus, k_{NTL}, the corresponding phase, *ph*, and Ψ_k for the customer with maximum index. The next column presents the accumulated $|Err_{NTLk}|$. Finally, the last column indicates the primary suspect bus based on Err_{NTLk}(%), obtained in Step 10. In this case, only one bus is correctly selected as suspect.

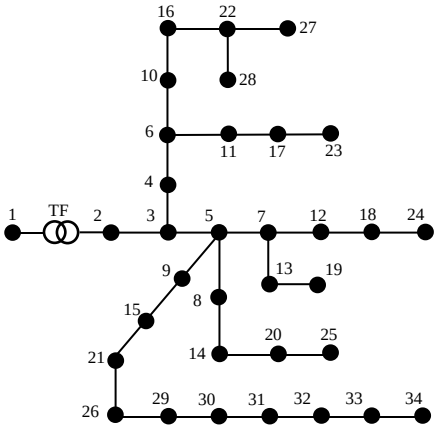
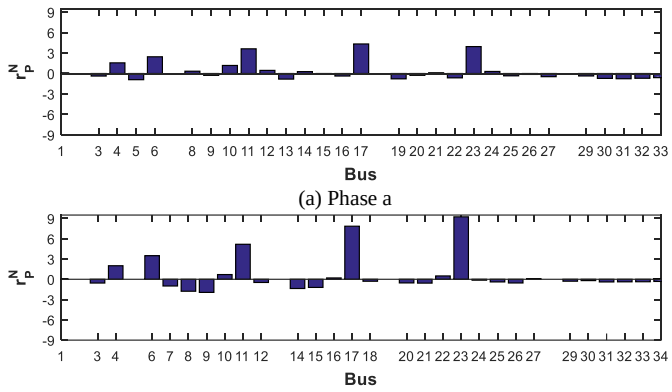


Fig. 3. Real 34-bus LV network.

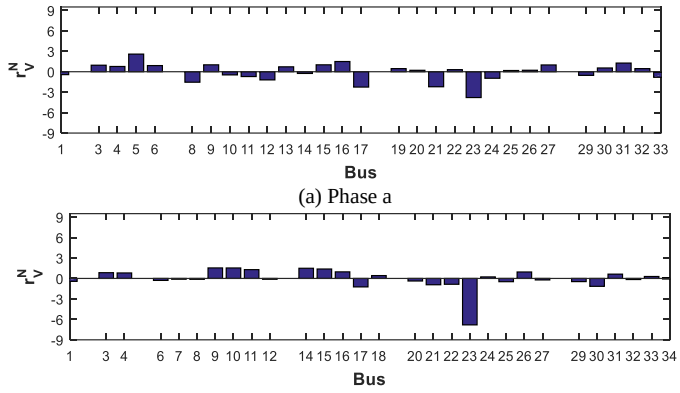
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2020.2974132, IEEE Transactions on Power Delivery



(a) Phase a

(b) Phase b

Fig. 4. Normalized residuals of active power measurements.



(a) Phase a

(b) Phase b

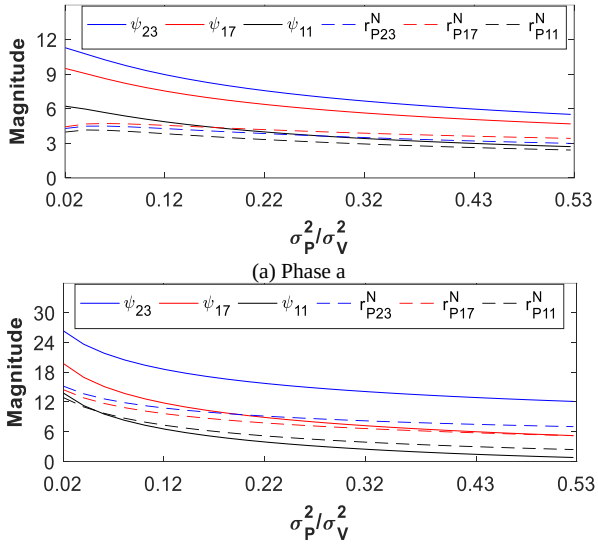
Fig. 5. Normalized residuals of voltage magnitude measurements.

TABLE II
METHOD RESULTS FOR A 3-KW NTL AT BUS 23

iter	r^N_{Pk}	r^N_{Qk}	r^N_{Vk}	k_{NTL}	*ph*	Ψ_k	$\lvert Err_{NTLk}\rvert$ (kVA)	Primary suspect
1	9.2	-	-6.8	23	*b*	16.0	1.4	23
2	4.0	-	-2.3	23	*a*	6.3	2.1	23

In Fig. 6, in order to assess how robust is the index Ψ for NTL location, this index is compared to the conventional index r^N for several ratios of power (active and reactive) and voltage variances (σ_P^2/σ_V^2). It is possible to note that the residual margin, *i.e.*, the difference between the first and the second highest residuals, for index Ψ ($\Psi_{23} - \Psi_{17}$) is higher than for the r^N index ($r^N_{P23} - r^N_{P17}$). This indicates that the proposed index is more robust for NTL selection. For example, for phase *a* (Fig. 6 (a)), index Ψ is the only index that correctly selects bus 23 as suspicious, as r^N selects bus 17, although both the indexes correctly locate the NTL for phase *b* (Fig. 6 (b)). Moreover, considering the several standard deviations tested, it can be observed that the proposed index is practically not affected by the increase of σ_P and σ_Q, as the residual margin remains almost constant for all evaluated cases. Therefore, the usage of Ψ for NTL location is potentially more suitable, as this index is not strongly affected by the weight matrix.

Table III presents the results when an illegal load of 4 kW is connected to bus 23. In this case, buses 23 and 17 are indicated as suspects, with $|Err_{NTL23}|$ = 2.7 kVA (82% of the estimated NTL) and $|Err_{NTL17}|$ = 0.6 kVA (18%). Despite the indication of bus 17 as an additional irregular load, the first customer in the list of suspect buses is the correct one and the difference from the second suspect is significant, as Err_{NTL}(%) is 82% for bus 23 and 18% for bus 17. In addition, the method reinforces the NTL region as bus 17 is a direct neighbor of bus 23.



(a) Phase a

(b) Phase b

Fig. 6. Comparison of proposed (Ψ) and traditional (r^N) index to locate NTL on customers in a LV distribution network.

TABLE III
METHOD RESULTS FOR A 4-KW NTL AT BUS 23

iter	r^N_{Pk}	r^N_{Qk}	r^N_{Vk}	k_{NTL}	*ph*	Ψ_k	$\lvert Err_{NTLk}\rvert$ (kVA)	Primary suspect
1	12.2	-	-8.5	23	*b*	20.7	1.9	23
2	5.5	-	-3.1	23	*a*	8.6	2.7	23
3	3.6	-	-1.8	17	*c*	5.4	0.6	23

B. Case Study 2 – Sensitivity analysis for LV illegal loads (1,682-bus system)

The second case study analyzes the algorithm performance applied to a real 13.8-kV feeder, which supplies 55 LV systems and 64 MV customers, totalizing 1,682 buses. Feeder data are available in [34]. A sensitivity analysis of the method performance is conducted by varying the NTL magnitude from 1 kW to 10 kW considering illegal loads connected at the LV systems. In these cases, illegal loads with unity power factor and 0.84 lagging power factor are analyzed to investigate the effect of reactive power deviation. All the possibilities of presence of illegal LV loads are analyzed, *i.e.*, the occurrence of NTL is simulated at each load bus of all 55 LV networks (~1,000 occurrences), with the illegal load connected between two phases. The tests are conducted for each load bus separately.

This study aims to evaluate the algorithm performance considering the following issues:

a) Number of cases in which NTL is detected, *i.e.*, at least one bus has r^N_{Pk} or r^N_{Qk} higher than the threshold δ.
b) Number of cases in which NTL bus is selected among the suspect buses, *i.e.*, the bus with the illegal load has Err_{NTLk}(%) > 0.
c) Number of cases in which NTL bus is selected as the primary suspect, *i.e.*, the NTL bus has the maximum Err_{NTLk}(%) value.
d) Number of cases in which the NTL bus or a first neighbor bus is selected with the maximum Err_{NTLk}(%) value.
e) Number of buses selected as suspects of NTL (N_{NTL}) for each case, *i.e.*, the number of buses with Err_{NTLk}(%) > 0.

The results obtained for the cases of single illegal LV loads are summarized in Table IV and Table V for power factors equal to 1.00 and 0.84 inductive, respectively. It is important to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2020.2974132, IEEE Transactions on Power Delivery

highlight again that all the possibilities of illegal loads in the LV system are investigated, *i.e.*, ~1,000 occurrences. Each column presents the results associated with items a) to e), showing the successful percentage for the simulated cases for items a) to d) and the 50th and 90th percentiles in the simulated cases for item e). Based on this tables, it can be verified that the method has a good performance when the NTL is 2 kW or higher. In addition, one can also observe that the higher the NTL magnitude, the more accurate the proposed method. On the other hand, its selectivity is less effective, as it indicates a higher number of buses as possible location of NTL (higher N_{NTL}). For example, for a 5-kW NTL with unity power factor, in 94.6% of the cases, the correct bus is indicated as the most likely NTL location and in 97.8% of the cases, the correct bus or a first neighbor bus is indicated, so that the correct region is located. In addition, from column (e), in 90% of the cases, it is indicated up to 3 suspect buses and in 50% 2 suspect buses are selected.

By comparing the results from Table IV and Table V, one can see that the presence of reactive power component in the illegal load improves the NTL detection and location, as it increases the NTL apparent power. However, the method selectivity is slightly reduced for some cases, as the number of suspect buses increases (higher N_{NTL} values).

It is important to note that the detection effectiveness for a range of NTL magnitudes depends on the standard deviation σ assigned to the measurements in the weight matrix. As verified on the previous example, the method fails for most of 1-kW NTL cases. In order to detect this NTL magnitude, the standard deviation assigned to the active and reactive power measurements should be smaller. Nevertheless, the consequence of decreasing this parameter is the reduction in method selectivity (higher N_{NTL} values). In this context, a sensitivity analysis is conducted to evaluate the effect of the standard deviation σ assigned to active and reactive power measurements. Fig. 7 and Fig. 8 present the method performance varying σ values from $1 \cdot 10^{-4}$ pu to $5 \cdot 10^{-4}$ pu. The same previous simulations were conducted varying the NTL magnitude from 1 to 10 kW for a lagging power factor equal to 0.84. Fig. 7 presents the percentage of the simulated cases where the NTL is detected, for each tested value of NTL. As one can see, the smaller the standard deviation σ, the higher the NTL detection rate. However, for a 3-kW NTL and higher, the method performance is roughly the same for all values of tested standard deviation σ. Fig. 8 presents the number of detected illegal loads considering the 90th percentile, *i.e.*, the N_{NTL} 90th percentile, for the simulated cases according to NTL magnitude. This figure shows that, in general, the smaller the standard deviation σ, the higher the number of buses indicated as NTL suspects.

C. Case Study 3 – Sensitivity analysis for MV illegal loads (1,682-bus system)

In this section, it is analyzed the presence of illegal three-phase loads at customers connected directly to MV systems, such as large commercial or industrial installations. The NTL magnitude is varied from 20 kW to 200 kW, with a lagging power factor of 0.84. For this study, the customer is connected by a dedicated transformer to each MV bus – except for the MV buses corresponding to the primary bus of the distribution transformers. The data of the dedicated transformer added to the simulations are provided in Table VI. The tests are conducted for each load bus separately, however, considering all the possibilities of illegal load connections (56 occurrences).

TABLE IV
METHOD RESULTS FOR LV NTL WITH UNITY POWER FACTOR

NTL (kW)	Successful Cases (%)				e) N_{NTL} (percentile)	
	a)	b)	c)	d)	50th	90th
1	24.8	19.6	19.6	23.9	1	1
2	86.9	71.8	70.7	81.5	1	1
3	97.3	88.3	85.1	92.8	1	2
4	99.3	95.3	91.0	96.9	1	2
5	99.3	97.7	94.6	97.8	2	3
6	99.5	98.6	95.9	98.8	2	3
7	99.5	98.9	96.6	98.9	2	3
8	99.5	99.1	97.0	99.3	2	3
9	99.5	99.1	97.7	99.3	2	4
10	99.5	99.1	98.4	99.5	2	4

a) NTL is detected; b) NTL bus is among the suspect buses; c) NTL bus is indicated with the maximum Err_{NTLk}(%) value; d) NTL bus or a first neighbor bus is indicated with the maximum Err_{NTLk}(%) value; e) Number of buses indicated as suspects of NTL.

TABLE V
METHOD RESULTS FOR LV NTL WITH 0.84 LAGGING POWER FACTOR

NTL (kW)	Successful Cases (%)				e) N_{NTL} (percentile)	
	a)	b)	c)	d)	50th	90th
1	39.4	32.9	32.2	37.2	1	1
2	95.5	86.5	80.9	89.9	1	2
3	99.1	97.1	90.1	96.6	2	3
4	99.3	98.6	91.6	96.4	2	3
5	99.8	99.1	93.6	97.7	2	4
6	99.8	99.1	95.6	98.4	3	4
7	99.8	99.1	96.1	98.9	3	4
8	99.8	99.1	97.2	99.1	3	4
9	99.8	99.0	97.6	99.3	3	5
10	99.8	99.3	98.6	99.6	3	5

a) NTL is detected; b) NTL bus is among the suspect buses; c) NTL bus is indicated with the maximum Err_{NTLk}(%) value; d) NTL bus or a first neighbor bus is indicated with the maximum Err_{NTLk}(%) value; e) Number of buses indicated as suspects of NTL.

The results for these simulations are presented in Table VII. As one can see, the proposed method can locate NTL magnitudes higher than or equal to 23 kW, with successful detection and location around 100%. Again, the selectivity of the method slightly decreases with the increase in NTL magnitude. This is observed for the values of the 90th percentile of the N_{NTL}, which increases to 2 when the NTL magnitude is 200 kW. Nevertheless, in the case of illegal loads connected at MV level, the corresponding number of buses indicated as suspects of NTL is lower than those obtained for LV customers, indicating a better selectivity. One reason for this observation is the fact that the illegal load is connected to the MV feeder by a dedicated transformer, so that bad data from measurements associated with this bus have minor effects on measurements in other parts of the system (*i.e.*, weak interaction).

D. Case Study 4 – Multiple illegal loads (1,682-bus system)

The objective of this example is to assess the method performance for cases with multiple illegal loads. In this case study, 7 illegal loads are simultaneously connected to 3 out of 55 LV systems of the 1,682-bus distribution feeder. Table VIII shows the buses and magnitude of each NTL, and the corresponding LV systems. The NTLs are considered with 0.84 inductive power factor.

Table IX presents the method results, showing the buses suspected of NTL in descending order according to Err_{NTL}(%). As one can observe, the NTL buses are properly indicated with

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TPWRD.2020.2974132, IEEE Transactions on Power Delivery

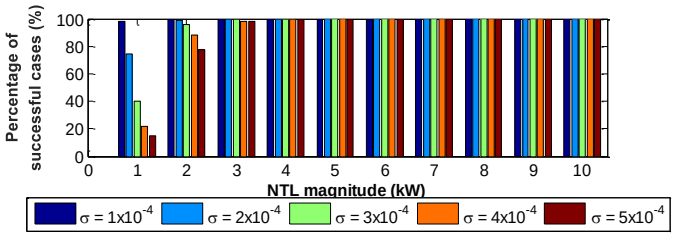


Fig. 7. NTL detection rate considering illegal load with 0.84 lagging power factor.

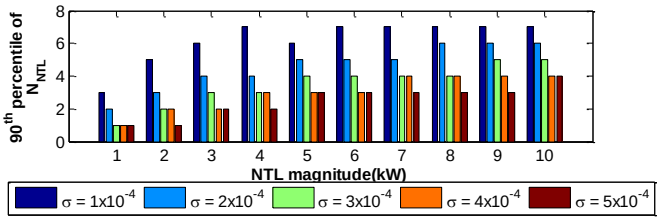


Fig. 8. N_{NTL} 90th percentile for simulated cases considering illegal load with 0.84 lagging power factor.

TABLE VI
DEDICATED TRANSFORMER DATA – MV LOADS

S (kVA)	V_1 (kV)	V_2 (kV)	R (%)	X (%)	Connection
500	13.8	0.4	1.3	3	Δ - Y_g

TABLE VII
METHOD RESULTS FOR MV NTL WITH 0.84 LAGGING POWER FACTOR

NTL (kW)	Successful Cases (%)				e) N_{NTL} (percentile)	
	a)	b)	c)	d)	50th	90th
20	0	0	0	0	0	0
21	0	0	0	0	0	0
22	39.6	39.6	39.6	39.6	1	1
23	96.9	96.9	96.9	96.9	1	1
24/25/50/100/150	100	100	100	100	1	1
200	100	100	100	100	1	2

a) NTL is detected; b) NTL bus is among the suspect buses; c) NTL bus is indicated with the maximum Err_{NTLk}(%) value; d) NTL bus or a first neighbor bus is indicated with the maximum Err_{NTLk}(%) value; e) Number of buses indicated as suspects of NTL.

the highest Err_{NTL}(%) (ranking position 1 to 7). Except for bus 8012, this rank respects the descending order of NTL magnitude, *i.e.*, it orders from the highest NTL bus to the lowest NTL bus, which represents a considerable advantage when the utility schedules an inspection. Other 2 buses (ranking positions 8 to 9) from secondary systems 2 and 47 are also indicated, so that N_{NTL} is equal to 9. Although these buses are wrongly selected, they are indicated as the lowest NTL probability and, in addition, the method correctly indicates only the affected LV systems, reducing the required inspection area.

The method efficacy in identifying multiple cases of NTL is due to the consideration of voltage magnitude residuals to compute the index Ψ. The same simulation of multiple NTL was conducted using the r^N index and the results are shown in Table X, where, the index Err_{NTL} is calculated by using r^N. As one can observe, one bus without NTL is indicated with the highest Err_{NTL}(%). Another problem faced is that the ranking is not as accurate as when using the index Ψ to order the NTLs from the highest to the lowest. Moreover, N_{NTL} is equal to 12, indicating the lower selectivity of normalized residual as the index for NTL location. It is important clarify that the direct usage of r^N would lead to an even worse performance.

Additionally, 100 other scenarios are simulated, considering different NTL magnitudes in each one of the 7 NTL buses indicated in Table VIII. In each scenario, random values of

TABLE VIII
MULTIPLE NTL LOCATIONS

LV system number	LV system buses	NTL buses (Power – kW)
2	8002-8025	8010 (9), 8011 (8), 8012 (7)
45	13002-13028	13016 (4), 13024 (3)
47	20002-20034	20013 (6), 20019 (5)

TABLE IX
METHOD RESULTS FOR MULTIPLE NTL – BUSES RANKED BY Err_{NTL}(%)

Ranking Position	k_{NTL}	Err_{NTL} (%)	Ranking Position	k_{NTL}	Err_{NTL} (%)
1	**8010**	27.4	**6**	**13016**	8.7
2	**8011**	22.0	**7**	**13024**	7.1
3	**20013**	11.8	8	20012	1.1
4	**20019**	10.9	9	8016	1.1
5	**8012**	9.9	-	-	-

TABLE X
RESULTS FOR MULTIPLE NTL USING R^N INDEX– BUSES RANKED BY Err_{NTL}(%)

Ranking Position	k_{NTL}	Err_{NTL} (%)	Ranking Position	k_{NTL}	Err_{NTL} (%)
1	8004	28.8	**7**	**13024**	5.3
2	**8010**	16.3	**8**	**8012**	4.9
3	**20013**	12.3	9	8017	1.7
4	**20019**	11.7	10	8014	1.0
5	**8011**	9.4	11	8016	0.5
6	**13016**	7.7	12	20007	0.4

NTL magnitude are allocated to each NTL bus (the NTL magnitudes can assume values from 3 kW to 7 kW, with 0.84 inductive power factor).

When the proposed index Ψ is used to identify NTL buses, the following results are observed:

- In 92 simulations (92% of the cases), all 7 NTL buses are within the ranking position 1 to 7.
- In 100 simulations (100% of the cases) all 7 NTL buses are within the ranking position 1 to 8.
- In 90 simulations (90% of the cases) the number of buses selected as suspects of NTL (N_{NTL}) is smaller than or equal to 10 (despite having only 7 NTL buses).

Otherwise, when the index r^N is used, the following results are observed:

- In 14 simulations (only 14% of the cases), all seven NTL buses are within the ranking position 1 to 7.
- In 53 simulations (only 53% of the cases) all seven NTL buses are within the ranking position 1 to 8.
- In 90 simulations (90 % of the cases) the number of buses selected as suspects of NTL (N_{NTL}) is smaller than or equal to 14 (despite having only 7 NTL buses).

The mentioned results reinforce the outperformance of the proposed index Ψ in NTL location, demonstrating its better selectivity.

V. CONCLUSIONS

Customer smart meter is an apparatus with potential to benefit the management of distribution systems, provided that new techniques are developed to transform the huge amount of data supplied by these devices into useful information. In this context, this paper proposes a new data analytic technique derived from bad data analysis for NTL detection and location caused by illegal connections. The NTL detection is initially conducted by using the normalized residuals of active and reactive power measurements r^N_P and r^N_Q. However, as these residuals do not have information from voltage, which cannot be tamper by meter bypassing, a new index Ψ is proposed to

select (locate) the suspicious customers. This new index combines the normalized residuals of active and reactive power measurements, r^N_P and r^N_Q, with the normalized residuals of voltage magnitude r^N_V, improving the method efficiency and robustness considerably, as this new index is more accurate and less affected by the weight matrix. Finally, the suspicious customers are ranked by using the estimated accumulated error Err_{NTLk}. Based on the results obtained by using a real MV/LV 1,682-bus distribution feeder, the method can detect and locate NTL as small as 2 kW for LV illegal loads and 23 kW for MV illegal loads. The proposed method benefits from the scenario of wide-scale deployment of smart meters, without the need for further investments in infrastructure, aggregating value to these new devices.

VI. References

[1] EPRI, "Integrating Smart Distributed Energy Resources with Distribution Management Systems," Technical Results, 2012.

[2] M. Dong, A. B. Nassif, B. Li, "A data-driven residential transformer overloading risk assessment method", *IEEE Trans. Power Delivery*, vol. 34, no. 1, pp 387-396, Feb. 2019.

[3] Y. Lo, S. Huang, C. Lu, "Transformational benefits of AMI data in transformer load modeling and management", *IEEE Trans. Power Delivery*, vol. 29, no. 2, pp. 742-750, April 2014.

[4] D. Martin, J. Marks, T. Saha, O. Krause, N. Mahmoudi, "Investigation into modeling Australian power transformer failure and retirement statistics", *IEEE Trans. Power Delivery*, vol. 33, no. 4, pp. 2011-2019, Aug. 2018.

[5] S. Elphick, V. Gosbell, V. Smith, S. Perera, P. Ciufo, G. Drury, "Methods for harmonic analysis and reporting in future grid applications", *IEEE Trans. Power Delivery*, vol. 32, no. 2, pp. 989-995, April 2017.

[6] F. C. L. Trindade, W. Freitas, and J. C. M. Vieira, "Fault location in distribution systems based on smart feeder meters," *IEEE Trans. Power Delivery*, vol. 29, no. 1, pp. 251–260, Feb. 2014.

[7] K. Sun, Q. Chen, Z. Gao, "An automatic faulted line section location method for electric power distribution systems based on multisource information", *IEEE Trans. Power Delivery*, vol. 31, no. 4, pp. 1542-1551, Aug. 2016.

[8] T. Xu, H. Chiang, G. Liu, C. Tan, "Hierarchical K-means method for clustering large-scale advanced metering infrastructure data", *IEEE Trans. Power Delivery*, vol. 32, no. 2, pp. 609-616, April 2017.

[9] R. Torquato, D. Salles, C. O. Pereira, P. C. M. Meira, W. Freitas, "A comprehensive assessment of PV hosting capacity on low-voltage distribution systems", *IEEE Trans. Power Delivery*, vol. 33, no. 2, pp. 1002-1012, April 2018.

[10] S. Elphick, P. Ciufo, G. Drury, V. Smith, S. Perera, V. Gosbell, "Large scale proactive power-quality monitoring: An example from Australia", *IEEE Trans. Power Delivery*, vol. 32, no. 2, pp. 881-889, April 2017.

[11] H. Pezeshki, A. Arefi, G. Ledwich, P. Wolfs, "Probabilistic voltage management using OLTC and dSTATCOM in distribution networks", *IEEE Trans. Power Delivery*, vol. 33, no. 2, pp. 570-580, April 2018.

[12] T. B. Smith, "Electricity theft: A comparative analysis," *Energy Policy*, vol. 32, pp. 2067-2076, 2004.

[13] Brazilian Electricity Regulatory Agency (ANEEL). [Online]. Available: http://www.aneel.gov.br/aplicacoes/noticias/Output_Noticias.cfm?Identidade=4160&id_area=90.

[14] BC Hydro. Smart Metering & Infrastructure Program Business Case. [Online]. Available: https://www.bchydro.com/content/dam/BCHydro/customer-portal/documents/projects/smart-metering/smi-program-business-case.pdf.

[15] L. T. Faria, J. D. Melo, A. Padilha-Feltrin, "Spatial-temporal estimation for nontechnical losses", *IEEE Trans. Power Delivery*, vol. 31, no. 1, pp. 362-369, Feb. 2016.

[16] J. Nagi, K. Yap, S. K. Tiong, S. Ahmed, M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines", *IEEE Trans. Power Delivery*, vol. 25, no. 2, pp. 1162-1171, April 2010.

[17] E. W. S. Angelos, O. R. Saavedra, O. A. C. Cortés, A. N. De Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems", *IEEE Trans. Power Delivery*, vol. 26, no. 4, pp. 2436-2442, Oct. 2011.

[18] C. C. O. Ramos, A. N. De Souza, A. X. Falcão, J. P. Papa, "New insights on nontechnical losses characterization through evolutionary-based feature selection", *IEEE Trans. Power Delivery*, vol. 27, no. 1, pp. 140-146, Jan. 2012.

[19] C. Muniz, M. Vellasco, R. Tanscheit, K. Figueiredo, "A neuro-fuzzy system for fraud detection in electricity distribution," *Joint Int. Fuzzy Syst. Assoc. World Congr. Eur. Soc. Fuzzy Logic Technol. Conf.* , pp. 1096-1101, Jul. 2009.

[20] L. G. O. Silva, A. A. P. Silva, A. T. Almeida-Filho, "Allocation of power-quality monitors using the P-Median to identify nontechnical losses", *IEEE Trans. Power Delivery*, vol. 31, no. 5, pp. 2242-2249, Oct. 2016.

[21] C. H. Lin, S. J. Chen, C. L. Kuo, J. L. Chen, "Non-cooperative game model applied to an advanced metering infrastructure for non-technical loss screening in micro-distribution systems", *IEEE Trans. Smart Grid*, vol. 5, n. 5, pp. 2468-2469, Sep. 2014.

[22] M. M. Buzau, J. T. Aguilera, P. C. Romero, A. G. Expósito, "Detection of non-technical losses using smart meter data and supervised learning", *IEEE Trans. Smart Grid*, vol. 10, n. 3, pp. 2661-2670, May 2019.

[23] S. A. Salinas, P. Li, "Privacy-preserving energy theft detection in microgrids: a state estimation approach," *IEEE Trans. Power Systems*, vol. 31, n. 2, pp. 883-894, Mar. 2016.

[24] M. Tariq, H. V. Poor, "Electricity theft detection and localization in grid-tied microgrids", *IEEE Trans. Smart Grid*, vol. 9, n. 3, pp.1920-1929, May 2018.

[25] P. Jokar, N. Arianpoo, V. C. M. Leung, "Electricity theft detection in AMI using customers' consumption patterns," *IEEE Trans. Smart Grid*, vol. 7, n. 1, pp. 216-226, Jan. 2016.

[26] J. Zheng, D. W. Gao, L. Lin, "Smart meters in smart grid: an overview," in *Proc. 2013 IEEE Green Technologies Conf.*, pp. 57-64, April 2013.

[27] N. Jenkins, R. Allan, P. Crossley, D. Kirschen, and G. Strbac, Embedded Gen. IEE Power and Energy Series 31, 2000.

[28] A. Abur, A. G. Expósito, *Power System State Estimation: Theory and Implementation*, New York, Marcel & Dekker, 2004.

[29] N. G. Bretas, J. B. A. London, L. F. C Alberto, R. A. S. Benedito, "Geometrical approach on masked gross errors for power systems state estimation", *in Proc. 2009 IEEE Power & Energy Society General Meeting (PES'09)*, pp. 1-7, Jul. 2009.

[30] N. G. Bretas, J. B. A. London, L. F. C. Alberto, R. A. S. Benedito, "Geometrical approaches for gross errors analysis in power systems state estimation," *in Proc. 2009 IEEE Power Tech*, pp. 1-7, Jun. 2009.

[31] N. G. Bretas, S. A. Piereti, A. S. Bretas, A. C. Martins, "A geometrical view for multiple gross errors detection, identification, and correction in power system state estimation," *IEEE Trans. Power Systems*, vol. 28, n. 3, pp. 2128-2135, Aug. 2013.

[32] N. G. Bretas, A. S. Bretas, and A. C. P. Martins, "Convergence Property of the Measurement Gross Error Correction in Power System State Estimation, Using Geometrical Background," IEEE Trans. on Power Systems, vol. 28, n. 4, pp. 3729-3736, Nov. 2013.

[33] X. Nian-de, W. Shi-ying, Y. Er-keng, "A new approach for detection and identification of multiple bad data in power system state estimation," *IEEE Trans. Power Apparatus and Systems*, vol. PAS-101, n. 2, pp. 454-462, Feb. 1982.

[34] Test feeder data. [Online]. Available: www.dsee.fee.unicamp.br/~fernanda/testfeeder_ARD18.7z.

[35] A. P. Grilo, P. Gao, W. Xu, M. C. de Almeida, "Load monitoring using distributed voltage sensors and current estimation algorithms", *IEEE Trans. on Smart Grid*, vol. 5, n. 4, pp. 1920-1928, Jul. 2014.

VII. Biographies

Lívia M. R. Raggi received the Ph.D. degree in electrical engineering from the University of Campinas, Brazil, in 2018. Currently she is a senior specialist of the Brazilian Electricity Regulatory Agency. Her research interests include the analysis of distribution systems and smart meter applications.

Fernanda C. L. Trindade (S'09, M'13) received the Ph.D. degree in electrical engineering from the University of Campinas, Brazil, in 2013, where, currently, she is an Assistant Professor. Her research interests are power system automation, monitoring and protection.

Vinícius C. Cunha (S'16) received the M.Sc. degree in electrical engineering from the University of Campinas, Brazil, in 2017, where, currently, he is a Ph.D. student. His research interests are analysis of distribution systems and low-carbon technologies.

Walmir Freitas (S'96-M'02) received the Ph.D. degree in electrical engineering from the University of Campinas, Brazil, in 2001, where currently he is a Professor. His research interests are analysis of distribution systems, distributed generation and power quality. Dr. Freitas is an Editor of the IEEE Trans. on Power Delivery and IEEE Power Engineering Letters.