

A Physically Inspired Data-driven Model for Electricity Theft Detection with Smart Meter Data

Yuanqi Gao, *Student Member, IEEE*, Brandon Foggo, *Student Member, IEEE*, Nanpeng Yu, *Senior Member, IEEE*,

Abstract—Electricity theft is the third-largest form of theft in the United States. It not only leads to significant revenue losses, but also creates the risk of fires and fatal electrical shocks. In the past, utilities have fought electricity theft by sending field operation groups to conduct physical inspections of electrical equipment based on suspicious activity reported by the public. However, the recent rapid penetration of advanced metering infrastructure makes it possible to detect electricity theft by analyzing the information gathered from smart meters. In this paper, we develop a physically inspired data driven model to detect electricity theft with smart meter data. The main advantage of the proposed model is that it only leverages the electricity usage and voltage data from smart meters instead of unreliable parameter and topology information of the secondary network. Hence a speedy and widespread adoption of the proposed model is feasible. We show that a modified linear regression model accurately captures the physical relationship between electricity usage and voltage magnitude on the Kron-reduced distribution secondaries. Our results show that electricity theft on a distribution secondary will lead to negative and positive residuals from the regression for dishonest and honest customers respectively. The proposed model is validated with real-world smart meter data. The results show that the model is effective in identifying electricity theft cases.

Index Terms—Advanced metering infrastructure, data-driven analysis, electricity theft detection, smart meter

I. INTRODUCTION

Electricity theft refers to the practice of manipulating one's electricity data to reduce his or her electricity bill [1]. The practice raises electricity costs and overheats circuit devices to dangerous levels. It is also quite costly to electric utility companies. In the United States, utilities lose between 0.5% and 3.5% of their annual revenue to theft [2]. In some developing countries, the revenue loss from electricity theft is even larger [3] [4].

Electricity theft detection secures the revenue of utility companies around the world. But traditional detection methods rely on labor intensive inspections. These inspections can now be guided by newer detection methods, which instead rely on smart meter or other collected customer data [5]. Most, if not all, of these methods can be categorized into three groups based on the type of data that they use.

Methods in the first group assume that smart meter data is not available. Instead, they leverage ancillary information such as biannual electricity consumption and credit scores [5] [6] [7]. Such information can be used as features in supervised machine learning. Many supervised methods have been tested in literature. Examples include support vector

machines (SVM) [7], optimum-path forests [6], and artificial neural networks [8]. But these only work if verified cases of electricity theft are available to train on. If this is not the case, then unsupervised methods, which do not use electricity theft labels, must be used. Examples include fuzzy c-means clustering [5] and optimum-path forests clustering [9].

Methods in the second group assume that granular power consumption data is available. For example, reference [10] analyzed consumption profiles through a self-organizing map (SOM). Reference [11] proposed an entropy-based method to analyze the distribution of differenced consumption data. Reference [12] used an extreme learning machine to detect anomalies in electricity usage. Reference [13] combined a decision tree and an SVM to predict smart meter abnormalities. Reference [14] used a convolutional neural network trained on such data to perform detection.

Some studies in the second group assume the existence of a “central observer” [15] [16] [17] [18]. This observer measures the aggregated consumption from a group of customers. In particular, such central observers can be placed on the distribution transformers. Meter malfunction or tampering can thus be identified using linear regression.

Methods in the third group assume that network topology and parameter information are available. Under this assumption, state estimation based approaches become feasible. Early work on this direction [19] perform distribution system state estimation based on estimated load. The non-technical losses are then detected by comparing the results with the billed consumption. The approaches proposed in [20] [21] [22] first perform three phase state estimation procedures on the network. They then analyze variances [20] [22] or apply heuristic methods [21] to locate meter defects or tampering. Recently, these time-snapshot based methods have been improved by adopting phasor measurement unit (PMU) data [23]. Another method in this group formulates anomaly detection as an optimization problem [24]. The method finds a sparse power mismatch matrix whose non-zero elements correspond to the bypassed power from dishonest customers.

Perhaps, the most directly relevant work is [25], in which the authors proposed analyzing sample covariance matrices of smart meter measurement error statistics, voltage magnitude and active power data to detect electricity theft. Compared to [25], our proposed work does not need to make any assumption about the smart meter measurement error distributions. In addition, we provide a theoretical justification based on physical network model for using real power consumption and voltage magnitude measurement to detect electricity theft.

The existing literature on data-driven electricity theft detec-

tion has three limitations. First, it is not realistic to assume that the transformer power measurements, reliable topology documentation, and network parameter information are available to electric utilities. Network parameter information is typically known only up to the type of conductors. Good parameter estimation methods for single-phase models [26] and balanced three-phase models [27] do exist. But methods for estimating unbalanced three-phase network parameters are still in their infancy. Hence all techniques in the third group are usually infeasible. Furthermore, Pole mounted distribution transformers are generally not equipped with operational monitoring devices [28] in Europe and America. Thus the “central observer” techniques in the second group are infeasible as well. Second, residential customer loads are irregular and are dependent on many external factors [29]. Analyzing such profiles alone produces very limited interpretability and justification of the results. Worse yet, they might not distinguish between electricity theft and non-malicious customer activities. Many of these methods would detect the installation of a new electric device as theft. This diminishes the usability of methods in the second group. Finally, supervised approaches in the first and second groups need theft samples. But obtaining labeled datasets in this case is usually a hard task [9]. As a result, the number of labeled (inspected) customers is very small compared to the total number of customers.

In practice, data is limited to customer active power consumptions and voltage magnitudes. Network information is limited to the customer to transformer association map. We propose an unsupervised detection technique leveraging only these sources of information. Note that here ‘unsupervised’ means that electricity theft labels are not available for training purposes. Furthermore, the proposed method yields more interpretable results than consumption based methods. The unique contributions of this paper are listed as follows.

- To the best of our knowledge, this is the first method that detects electricity theft from customer active power and voltage magnitude measurements alone.
- This paper derives a linear model relating smart meter kWh measurements to line-line smart meter voltage magnitude measurements in distribution secondaries.
- This paper proves some exploitable behaviors expressed by the proposed linear model subject to normal and abnormal smart meter data.
- The proposed algorithm needs neither training samples for theft cases nor a complete network topology and parameter documentation; it requires only a customer to transformer association map.

The rest of the paper is organized as follows. Section II describes the proposed framework for electricity theft detection. Section III presents the modeling of distribution secondaries considering smart meter measurements. Section IV presents the technical methods for electricity theft detection. The experimental results are shown in Section V. Section VI provides the conclusion.

II. ELECTRICITY THEFT DETECTION FRAMEWORK

This section provides an overview of the framework for electricity theft detection. The framework consists of five

steps. Steps 1 through 4 are iteratively performed on a moving window of data. Step 5 combines the results from all the moving windows to get an anomaly score for each customer.

Step 1 first selects the initial training and testing data periods. It then removes outliers from the training dataset. We describe this outlier removal process in Subsection IV.A.

Step 2 fits a modified linear regression model with the training smart meter data. These models capture the relationship between smart meter voltage readings and electricity consumptions. The motivation for a modified linear model is the subject of Section III. The formal presentation of the model is the subject of Subsection IV.B.

Step 3 applies the trained models on the testing data and outputs an estimate of the electricity consumption for each customer. It then calculates the residual of the estimates to the true consumption. These residuals distinguish dishonest customers from honest ones. The former will have large negative residuals while the latter will have small positive residuals. We prove these observations in Subsection IV.C.

Step 4 produces an electricity theft anomaly score for each customer. These scores are obtained by running the electricity consumption residuals through post-processing techniques. The post-processing algorithm and the anomaly score are the subject of Subsection IV.D.

Step 5 ranks the customers according to their anomaly scores to help the system operator make further decisions.

The proposed framework assumes that each customer has a smart meter installed. It further assumes that each transformer serves a known list of customers. The proposed algorithm requires neither measurements from transformers nor information about network topology or parameters.

III. AN ANALYSIS OF DISTRIBUTION SECONDARIES

In this section, we linearize the physical relationship between smart meter voltage magnitudes and real power consumptions. A novel substitution method is then applied to handle nodes without measurements. No equations in this section are used for computation. Their purpose is to provide justifications of the algorithms presented in Section IV.

A. Linearization of Secondary Circuit Power Flow Equations

In North America, distribution secondaries serving residential customers typically have a 120/240V three-wire two-phase configuration. The two phases have voltages with an angle difference of 180 degrees. A sample distribution secondary with n_c customers is shown in Fig.1. To our knowledge, these secondaries do not have any existing linearized relationships. However, linearizations do exist for other power system configurations [30] [31].

To obtain this linearization, we adapt the approach of reference [30]. The idea is to find a tangent plane to the power flow manifold centered at a suitable point, which can be the modified flat voltage solution. We construct this solution as follows. First, assume that shunt admittances are zero. Denote the line to ground voltage phasor and current phasor at node m as \bar{u}_m and \bar{i}_m , and the real and reactive power injections as \bar{p}_m and \bar{q}_m . Then the modified flat voltage

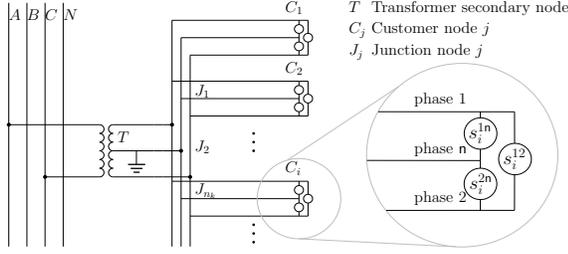


Figure 1: Triplex line secondary circuit

solution has, at each node m , the following set of values: $\bar{\mathbf{u}}_m = [1, 1 \cdot e^{-\pi j}]^T = [1, -1]^T$, $\bar{\mathbf{i}}_m = \mathbf{0}$, $\bar{\mathbf{p}}_m + j\bar{\mathbf{q}}_m = \mathbf{0}$.

Denote v_m^p as the deviation of the line-to-ground voltage from the flat voltage solution \bar{v}_m^p at node m and phase p . Let \mathbf{v}^1 and \mathbf{v}^2 be the reindexed voltage vectors:

$$[\mathbf{v}^{1T}, \mathbf{v}^{2T}] = [v_1^1, v_2^1, \dots, v_n^1, v_1^2, v_2^2, \dots, v_n^2]$$

where n is the number of non-zero injection nodes in a secondary circuit. The voltage angles θ^1, θ^2 , real power injections $\mathbf{p}^1, \mathbf{p}^2$, and reactive power injections $\mathbf{q}^1, \mathbf{q}^2$ are defined in a similar manner. The linearization around the modified flat voltage solution yields:

$$\begin{bmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \\ \mathbf{q}^1 \\ \mathbf{q}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{G}^{11} & -\mathbf{G}^{12} & -\mathbf{B}^{11} & \mathbf{B}^{12} \\ -\mathbf{G}^{21} & \mathbf{G}^{22} & \mathbf{B}^{21} & -\mathbf{B}^{22} \\ -\mathbf{B}^{11} & \mathbf{B}^{12} & -\mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{B}^{21} & -\mathbf{B}^{22} & \mathbf{G}^{21} & -\mathbf{G}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{v}^1 \\ \mathbf{v}^2 \\ \theta^1 \\ \theta^2 \end{bmatrix} \quad (1)$$

where \mathbf{G}^{ij} and \mathbf{B}^{ij} are the real and imaginary blocks of the reindexed nodal admittance matrix \mathbf{Y}^r :

$$\mathbf{Y}^r = \begin{bmatrix} \mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{G}^{21} & \mathbf{G}^{22} \end{bmatrix} + j \begin{bmatrix} \mathbf{B}^{11} & \mathbf{B}^{12} \\ \mathbf{B}^{21} & \mathbf{B}^{22} \end{bmatrix} \quad (2)$$

which is a permutation of the well-known bus admittance matrix \mathbf{Y} . Explicitly, \mathbf{Y}^r is obtained by taking every odd indexed row and column of \mathbf{Y} and relocating them to the bottom-most and right-most positions respectively. In the following, we refer to (1) as $\mathbf{y}^r = \mathbf{L}^r \mathbf{x}^r$. The derivation of (1) is provided in Appendix A.

In (1), $\mathbf{p}^1, \mathbf{p}^2, \mathbf{q}^1$ and \mathbf{q}^2 are single-phase net injections. But electric loads at the customers' site can be single-phase or two-phase as shown in Fig.1. We can get the single-phase net injections from the electric loads by using (3):

$$\begin{bmatrix} s_i^1 \\ s_i^2 \end{bmatrix} = \begin{bmatrix} \frac{u_i^{1n}}{u_i^{1n} + u_i^{2n}} & 0 & 1 \\ -\frac{u_i^{2n}}{u_i^{1n} + u_i^{2n}} & 1 & 0 \end{bmatrix} \begin{bmatrix} s_i^{12} \\ s_i^{2n} \\ s_i^{1n} \end{bmatrix} \quad (3)$$

The derivation of (3) is in Appendix B. This can be simplified near the flat voltage operating condition to:

$$\begin{bmatrix} p_i^1 \\ p_i^2 \end{bmatrix} + j \begin{bmatrix} q_i^1 \\ q_i^2 \end{bmatrix} = \begin{bmatrix} s_i^1 \\ s_i^2 \end{bmatrix} \approx \begin{bmatrix} \frac{1}{2} & 0 & 1 \\ \frac{1}{2} & 1 & 0 \end{bmatrix} \begin{bmatrix} s_i^{12} \\ s_i^{2n} \\ s_i^{1n} \end{bmatrix} \quad (4)$$

B. Conversion to Smart Meter Measurements

In practice, smart meters read line-line voltage magnitudes $|u_i^1 - u_i^2|$ and the sum of single-phase powers $p_i^1 + p_i^2$. But (1)

relates single-phase net injections to line-ground voltages. We need to change (1) such that it relates the former quantities.

To do this, we first assume that the following approximation holds (u_i^1, u_i^2 are line-to-ground voltage phasors, θ_i^1, θ_i^2 are line-to-ground voltage angles, of node i phase 1 and 2):

$$\begin{aligned} |u_i^1 - u_i^2| &= |(v_i^1 + 120)\cos(\theta_i^1) + j(v_i^1 + 120)\sin(\theta_i^1) \\ &\quad - (v_i^2 + 120)\cos(\theta_i^2) - j(v_i^2 + 120)\sin(\theta_i^2)| \\ &\approx |v_i^1 + v_i^2 + 240| \end{aligned} \quad (5)$$

This is valid when all θ_i^1, θ_i^2 are near those of the modified flat voltage solution. Thus the measured voltage magnitude can be approximately written as $|u_i^1 - u_i^2| \approx v_i^1 + v_i^2 + 240$.

Next, we introduce two new vectors \mathbf{x}^s and \mathbf{y}^s :

$$\begin{aligned} \mathbf{x}^{sT} &= [(\mathbf{v}^1 + \mathbf{v}^2)^T, (\mathbf{v}^1 - \mathbf{v}^2)^T, \theta^{1T}, \theta^{2T}] = [\mathbf{v}^{sT}, \theta^{sT}] \\ \mathbf{y}^{sT} &= [(\mathbf{p}^1 + \mathbf{p}^2)^T, (\mathbf{p}^1 - \mathbf{p}^2)^T, \mathbf{q}^{1T}, \mathbf{q}^{2T}] = [\mathbf{p}^{sT}, \mathbf{q}^{sT}] \end{aligned}$$

which are related to \mathbf{x}^r and \mathbf{y}^r via $\mathbf{M} = \text{diag}(\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{I} & -\mathbf{I} \end{bmatrix}, \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix})$: $\mathbf{x}^r = \mathbf{M}^{-1} \mathbf{x}^s$ and $\mathbf{y}^r = \mathbf{M}^{-1} \mathbf{y}^s$. Substituting these relationships into (1) yields $\mathbf{y}^s = \mathbf{M} \mathbf{L}^r \mathbf{M}^{-1} \mathbf{x}^s = \mathbf{L}^s \mathbf{x}^s$, or:

$$\begin{bmatrix} \mathbf{p}^s \\ \mathbf{q}^s \end{bmatrix} = \begin{bmatrix} \mathbf{L}_{11}^s & \mathbf{L}_{12}^s \\ \mathbf{L}_{21}^s & \mathbf{L}_{22}^s \end{bmatrix} \begin{bmatrix} \mathbf{v}^s \\ \theta^s \end{bmatrix} \quad (6)$$

Further, we remove the dependency on θ^1 and θ^2 to obtain:

$$\mathbf{p}^s = (\mathbf{L}_{11}^s - \mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{L}_{21}^s) \mathbf{v}^s + \mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{q}^s \quad (7)$$

Where $\mathbf{L}_{22}^{s\dagger}$ is the pseudoinverse of \mathbf{L}_{22}^s . We prove (7) in Appendix C.

Now, if each node has a constant lagging power factor over the analysis window, we can write $\begin{bmatrix} \mathbf{q}^1 \\ \mathbf{q}^2 \end{bmatrix} = \mathbf{D} \begin{bmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \end{bmatrix}$ where \mathbf{D} is a positive definite diagonal matrix. Then we have $\mathbf{q}^s = \mathbf{D} \mathbf{M}_u^{-1} \mathbf{p}^s$ where \mathbf{M}_u is the upper left block of \mathbf{M} . We can then simplify (7) to

$$\mathbf{p}^s = (\mathbf{I} - \mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{D} \mathbf{M}_u^{-1})^{-1} (\mathbf{L}_{11}^s - \mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{L}_{21}^s) \mathbf{v}^s \quad (8)$$

We argue that $(\mathbf{I} - \mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{D} \mathbf{M}_u^{-1})$ is nonsingular in practice in Appendix C. The first n_c equations of (8) are

$$\underbrace{\mathbf{p}^1 + \mathbf{p}^2}_{-p} = \mathbf{L}_{pv}^{(+)} \underbrace{(\mathbf{v}^1 + \mathbf{v}^2)}_{v-1 \cdot 240} + \mathbf{L}_{pv}^{(-)} (\mathbf{v}^1 - \mathbf{v}^2) \quad (9)$$

where p and v are smart meter net power consumption and voltage measurements respectively. (9) shows that the real power measurements depend on the observed voltage sums and the unobserved voltage differences. This latter term is negligible. Accounting for measurement errors with a noise term ϵ , we thus have

$$p = \mathbf{L}_{pv} (\mathbf{v} - \mathbf{1} \cdot 240) + \epsilon = \mathbf{L}_{pv} \mathbf{v} + \epsilon \quad (10)$$

where $\mathbf{L}_{pv} = -\mathbf{L}_{pv}^{(+)}$ whose nullspace contains $\mathbf{1}$.

C. A Remedy for Not Having Transformer Data

We can partition (10) with respect to transformer node and customer nodes:

$$\begin{bmatrix} p_T \\ p_C \end{bmatrix} = \begin{bmatrix} l_{TT} & l_{TC} \\ l_{CT} & l_{CC} \end{bmatrix} \begin{bmatrix} v_T \\ v_C \end{bmatrix} + \begin{bmatrix} \epsilon_T \\ \epsilon_C \end{bmatrix} \quad (11)$$

In most cases, transformers do not have smart meters installed on them. Thus there are no measurements for v_T and p_T in (11). We can remedy this by using conservation of energy to write $p_T \approx -\mathbf{1}^T \mathbf{p}_C$. This relationship is exact when there are no losses. We can then eliminate v_T from (11) and replace the remaining p_T term to obtain:

$$\mathbf{p}_C = -\mathbf{l}_{CT} \mathbf{l}_{TT}^{-1} \mathbf{1}^T \mathbf{p}_C + (\mathbf{L}_{CC} - \mathbf{l}_{CT} \mathbf{l}_{TT}^{-1} \mathbf{l}_{TC}) \mathbf{v}_C + \epsilon'_C \quad (12)$$

Where $\epsilon'_C = \epsilon_C - \mathbf{l}_{CT} \mathbf{l}_{TT}^{-1} \epsilon_T$. (12) motivates the use of $\mathbf{1}^T \mathbf{p}_C$ as a covariate in model estimation. We will use it in Subsection IV.B where we develop the modified linear model.

IV. TECHNICAL METHODS

This section details each step of the proposed framework outlined in Section II. To avoid tedious notation, all quantities correspond to *one* rolling window.

A. Data Preprocessing

Most real-world smart meter datasets contain missing values and outliers. The time stamps with missing values and/or power outages are discarded from the analysis. The amount of discarded data is usually negligible. Outliers are much more frequent and can hurt our regression models [32]. We discuss the problem of outliers in voltage data in this subsection.

A properly trained model would be sensitive to voltage measurement errors because the voltages vary around flat condition by a very small amount. In this paper, the time stamps where voltage error is large will be removed. The method is as follows. First, we train a regression model that is robust to outliers on the training dataset. We then apply the model to each customer $i \in \{1, 2, \dots, n_c\}$ and search for *training* time stamps $\mathcal{T}_i^{\text{out}} = \{t_{i,1}^{\text{out}}, \dots, t_{i,i_o}^{\text{out}}\}$ with large residuals. For notational clarity, all references to quantities involving robust regression will carry the superscript *rb*. Then:

$$t \in \mathcal{T}_i^{\text{out}} \quad \text{if} \quad \begin{aligned} & (\tilde{y}_i^{rb}(t) / \text{var}(\tilde{y}_i^{rb}))^2 > F_{\chi_1^2}^{-1}(0.999) \quad \text{and} \\ & (\mathbf{v}(t) - \bar{\mathbf{v}})^T \Sigma_{\mathbf{v}}^{-1} (\mathbf{v}(t) - \bar{\mathbf{v}}) > F_{\chi_{n_c}^2}^{-1}(0.999) \end{aligned} \quad (13)$$

where $\tilde{y}_i^{rb}(t) = y_i(t) - \hat{y}_i^{rb}(t)$ is the estimation residual for customer i at time t , $\text{var}(\tilde{y}_i^{rb}) = \sum_t (\tilde{y}_i^{rb}(t) - \bar{\tilde{y}}_i^{rb})^2 / (T - 1)$ is its empirical variance. $\bar{\mathbf{v}}$ and $\Sigma_{\mathbf{v}}$ are the sample mean vector and covariance matrix of voltage measurements. $F_{\chi_1^2}^{-1}$ is the inverse of chi-square CDF with one degree of freedom.

After the sets $\{\mathcal{T}_i\}_{i=1}^{n_c}$ are found, we remove any time instances that are a member of *two or more* of these sets. That is, if $t_p \in \mathcal{T}_i^{\text{out}} \cap \mathcal{T}_j^{\text{out}}, i \neq j$, then the measurements $p_i(t_p), v_i(t_p)$ are discarded for all $i \in \{1, 2, \dots, n_c\}$. The reasoning behind this final rule is as follows. If there is a voltage outlier, then at least two of the customers' regression residuals will be severely affected. This fact has been confirmed experimentally and can be understood from (10).

Robust regression methods such as least median of squares (LMS) [33]; M-estimator [34]; and random sample consensus (RANSAC) [35] can be used. In this work, we use RANSAC for its simplicity and efficiency.

B. Modified Linear Model

The ideas outlined in Subsection III.A-III.C are combined to produce the following *modified linear model (MLM)*:

$$\begin{aligned} y_i(t) &= [\mathbf{x}(t)^T \quad \sum_{j=1}^{n_c} y_j(t)] \begin{bmatrix} \beta_i^x \\ \beta_i^y \end{bmatrix} + \epsilon'_i(t) \\ &= \mathcal{X}(t)^T \mathcal{B}_i + \epsilon'_i(t) \end{aligned} \quad (14)$$

where $\mathbf{x}(t) = [v_1(t), v_2(t), \dots, v_{n_c}(t)]^T$: vector of voltage readings at time t , $y_i(t) = p_i(t)$: kWh readings of customer i at time t , and $\epsilon'_i(t)$ accounts for measurement noise and unobserved dependencies as in (12).

The parameter vectors $\{\mathcal{B}_i\}_{i=1}^{n_c}$ will be estimated by using ordinary least squares (OLS) on the training data $(\mathcal{X}^D, \mathbf{y}^D)$ [36], which is a portion of the rolling window. This estimate is achieved by solving the normal equations:

$$(\mathcal{X}^{D^T} \mathcal{X}^D) [\mathcal{B}_1, \dots, \mathcal{B}_{n_c}] = \mathcal{X}^{D^T} [\mathbf{y}_1^D, \dots, \mathbf{y}_{n_c}^D] \quad (15)$$

Variations of OLS such as total least squares (TLS) [37] [38] can be used instead, but these do not exhibit the properties described in the next subsection.

The fitted model is then used to predict kWh consumption values for the testing data $(\mathcal{X}, \mathbf{y})$ within the rolling window:

$$[\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_{n_c}] = \mathcal{X} [\mathcal{B}_1, \dots, \mathcal{B}_{n_c}] \quad (16)$$

The LHS are used to calculate the residual time series $\tilde{\mathbf{y}}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i$, which are used to perform electricity theft detection.

In this work, the length of each rolling window is chosen to be 67 days. The first 60 days and the last 7 days form the training and testing data, respectively. Each rolling window is 1 day ahead of the one preceding it.

C. Properties of Residuals Under Theft

The residuals of the MLM change when there is an energy thief. Denote $\mathbf{y}_i^{(e)}$ as the kWh meter data for customer i when one of the customers in the same secondary is a thief. Let the original symbol \mathbf{y}_i denote the kWh meter data of the same customer had there been no theft activities. Suppose without loss of generality that customer i is the thief, then

$$\mathbf{y}_i^{(e)} = \mathbf{y}_i - \mathbf{y}_i^s; \quad \mathbf{y}_j^{(e)} = \mathbf{y}_j \quad \forall j \neq i$$

where the non-negative vector \mathbf{y}_i^s denotes the difference between the imagined kWh measurement and the actual one. Let $\tilde{\mathbf{y}}_i^{(e)}$ and $\tilde{\mathbf{y}}_i$ denote the out-of-sample residual time series for the energy thief. Then the following results hold

Lemma 1.

$$\tilde{\mathbf{y}}_i^{(e)} - \tilde{\mathbf{y}}_i = - \sum_{j \neq i} \beta_j^y \mathbf{y}_j^s \quad (17)$$

Lemma 2.

$$\sum_j \tilde{\mathbf{y}}_j^{(e)} = \sum_j \tilde{\mathbf{y}}_j = \mathbf{0} \quad (18)$$

Lemma 3. For any $\delta > 0$, there exists a training data window length $T > 0$ such that for each j

$$\mathbb{P}(\beta_j^y \geq -\delta) > 1 - \delta \quad (19)$$

Lemma.1 and Lemma.3 combine to show that a thief's residuals will become negative once he or she begins to steal power. Lemma.2 shows that the residuals of the other customers will raise in order to balance their sum. These conclusions are useful to the design of the postprocessing method discussed in Subsection IV.D. The proofs of Lemma.1, Lemma.2 and Lemma.3 are given in Appendix D.

D. Energy Theft Detection

We define an anomaly score in terms of the residuals \tilde{y}_i for each customer. Customers with high anomaly scores are likely to be thieves or have malfunctioning smart meters.

Before defining the anomaly score, we post-process the residuals in two steps. The first step removes outliers. This step is analogous to the preprocessing stage except here, we substitute its residual value by that of the nearest future non-outlier. The second step sets all positive residuals to zero. This rule comes from experimentation and the lemmas of the previous subsection. We denote the resulting residual time series after the two steps as \tilde{y}'_i .

Until now, we have ignored the subscript for the rolling windows. It is necessary to introduce it here. We use the symbol $f = 1, 2, \dots$ to index the rolling windows. The anomaly score for each customer i and each rolling window f is defined as $d_i(f) = w_i(f) \|\tilde{y}'_i(f)\|_2$ where $w_i(f) = \sqrt{|t^{\mathcal{D}}(f)|} / \|\tilde{y}^{\mathcal{D},i}(f)\|_2$ is a weighting coefficient. Energy thefts are identified by ranking $d_i(f)$ for all i and all f . The higher $d_i(f)$ is, the higher priority of investigation customer i should have. This ranking method is simplified to ranking $\max_f d_i(f)$ for all i when theft time is unimportant.

V. EXPERIMENTAL RESULTS

This section evaluates the performance of the proposed method on a real dataset with synthetic electricity theft cases. In Subsection V.A, we describe the dataset in detail. In Subsection V.B, we test the performance of the modified linear model without energy theft. In Subsections V.C-V.E, we demonstrate the impact of energy theft on out-of-sample residuals and anomaly scores. In Subsection V.F, we compare the performance of our proposed anomaly detection method with comparable methods.

A. Experimental Data Description

1) *Real-world Smart Meter Data*: The smart meter dataset comes from a 12 KV distribution feeder in Southern California Edison (SCE)'s service territory. The schematic of the testing distribution feeder is shown in Fig.2. Measurements were taken from August 1, 2015 to Feb 1, 2016, including the customers' hourly average voltage magnitudes and electricity consumption. A majority of the customers on the distribution feeder are residential customers. The transformer to customer association information is also provided by SCE. 190 such transformers were selected for the experimental study. This accounts for 980 residential customers.

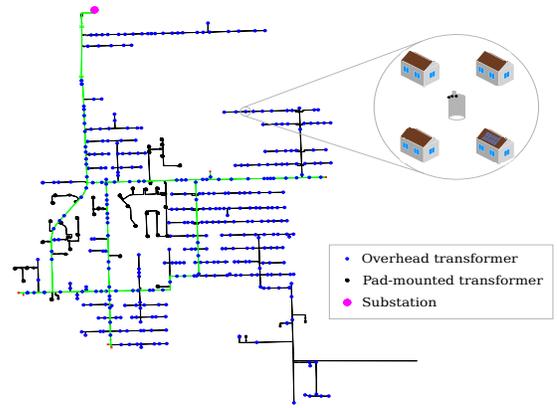


Figure 2: Schematic of the test distribution feeder.

2) *Synthetic Electricity Theft Data*: Similar to other literature, we synthesize electricity theft data [17] [39] [40]. The *attack invariant* principle [41] is followed during the data synthesis process: $\sum_{t \in \mathcal{T}_e} p_k^{(e)}(t) < \sum_{t \in \mathcal{T}_e} p_k(t)$ where the k th customer is stealing power during time period \mathcal{T}_e . $p_k(t)$ denotes the actual electric power consumed by the k th customer. $p_k^{(e)}(t)$ is the electricity consumption of the k th customer recorded by the electric utility.

The amount of electricity theft from the k th customer during hour t , $p_k^s(t)$, is defined as $p_k^s(t) = p_k(t) - p_k^{(e)}(t)$ where $0 \leq p_k^s(t)$. Within the attack invariant principle, four electricity theft cases are simulated.

- Case 1: 100% of electricity theft for n hours: $p_k^s(t) = p_k(t)$
- Case 2: A constant amount of electricity theft: $p_k^s(t) = \alpha_{c2}$
- Case 3: A uniformly distributed electricity theft: $p_k^s(t) \sim \mathcal{U}(0, \alpha_{c3})$
- Case 4: A constant percentage of electricity theft: $p_k^s(t) = \alpha_{c4} p_k(t)$

In this paper, we assume the time period when electricity theft occurs is a consecutive subset of all time stamps of our dataset, that is, $\mathcal{T}_e(t_1^{(e)}, t_2^{(e)}) = \{t : t_1^{(e)} \leq t < t_2^{(e)}\}$. Data synthesis is performed within $\mathcal{T}_e(t_1^{(e)}, t_2^{(e)})$.

The synthetic electricity theft case for the k th customer is created as follows. If customer k does not have DERs, then $p_k(t) - p_k^{(e)}(t) \mapsto p_k^{(e)}(t)$ and $\max(p_k^{(e)}(t), 0) \mapsto p_k^{(e)}(t)$. The synthesized electricity consumption of a customer without DERs should be higher than zero. If customer k does have DERs, then the floor for net electricity consumption recording should be the electricity delivered back to the grid.

B. Performance of the Modified Linear Model

Consider a distribution secondary circuit consisting of 4 residential customers as highlighted in Fig.2. The rolling window under study is set up as follows. The training dataset $t^{\mathcal{D}}$ starts at hour 1 and ends at hour 1440 from 60 days. The testing dataset $t^{\mathcal{D}^a}$ includes 168 consecutive hours from 7 days following the training dataset.

We first show that the proposed MLM accurately estimates the electricity consumption of a given customer. This customer's true consumption, estimated consumption, and residuals are depicted in Fig.3. We plot this data for the first 100

hours of the in-sample and out-of-sample periods. The average

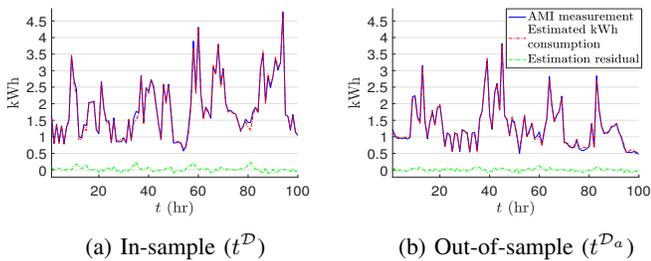


Figure 3: Electricity consumption residuals from the MLM for one customer.

electric load consumed by this customer is 1.6 kWh. The mean of the estimation residual is -0.01 kWh and its standard deviation is 0.1 kWh. Both of these are small compared to the customer's average load. This result shows that the MLM estimates the electricity consumption of a customer quite well.

We next apply this analysis to every customer on the feeder over 59 rolling windows. The first window is the same as above. Each other window is 24 hours ahead of the one preceding it. The box plot of the in-sample residual sample standard deviation, out-of-sample residual sample mean and standard deviation for all customers are shown in Fig.4. The statistics of the example customer shown in Fig.3 are highlighted by the yellow dashed lines. Most customers' have a small residual

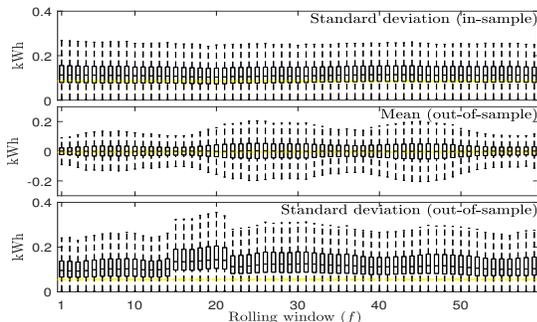


Figure 4: Residual statistics of all customers on the distribution feeder.

mean and standard deviation. Hence the proposed MLM is accurate in estimating the consumption of most customers on the feeder. Yet, some customers do have relatively large residuals. These are likely due to errors in the customer to transformer mapping and noisy smart meter data.

Finally, we compare the performance of the MLM with three nonlinear regression models. A Feed-forward Neural Network (FNN), a Radial Basis Function Network (RBF) [42], and a Support Vector Regression (SVR) model [43]. The inputs and outputs of the nonlinear models are the same as MLM. The number of hidden units in the FNN is one plus the number of inputs, the number of neurons in the RBF is 200, and the kernel for the SVR is a degree 2 polynomial. Five equally spaced rolling windows between the 1st and 59th - shown in Fig.4 - were selected to perform the regression analysis. All other experimental setups are identical as in Fig.4. The

results are reported in Table.I. Each cell of Table.I shows the $\mu \pm 2\sigma$ of the corresponding performance measure. All values are in kWh and have been rounded to 2 decimal places. Table.I shows that all regression models perform similarly over

Table I: Comparison of regression models

	std (in-sample)	mean (out-of-sample)	std (out-of-sample)
MLM	0.12 ± 0.09	0.00 ± 0.09	0.13 ± 0.10
FNN	0.11 ± 0.08	0.00 ± 0.09	0.13 ± 0.10
RBF	0.12 ± 0.07	0.00 ± 0.08	0.17 ± 0.13
SVR	0.12 ± 0.09	0.01 ± 0.08	0.13 ± 0.10

a wide range of customers. But we value interpretability and simplicity over sophistication and complexity. For this reason, our experimental studies shall focus on the proposed modified linear model.

C. Properties of the Anomaly Score

This subsection studies the proposed energy theft detection scheme of section IV. The goal of the following experiments is twofold. First, it confirms that ranking the maximum anomaly score $\max_f d_i(f)$ for all i is a good way to detect energy thieves. Second, we show that $\max_f d_i(f)$ generally occurs during a rolling window with nice properties. The properties in question are cleanliness of the training dataset and theft strength of the testing dataset.

For illustrative purposes, we consider the following experiment. First, we give synthetic theft data to the customer depicted in Fig.3. This synthesized data follows case 3 with parameter $\alpha_{c3} = 1.8$ kWh. We then increase $|\mathcal{T}_e \cap t^{\mathcal{D}}|/|t^{\mathcal{D}}|$ and $|\mathcal{T}_e \cap t^{\mathcal{D}_a}|/|t^{\mathcal{D}_a}|$ from 0 to 1 and average the results from 10 such simulations. The resulting anomaly scores for the first window are shown in Fig.5. Fig.5 shows that the anomaly

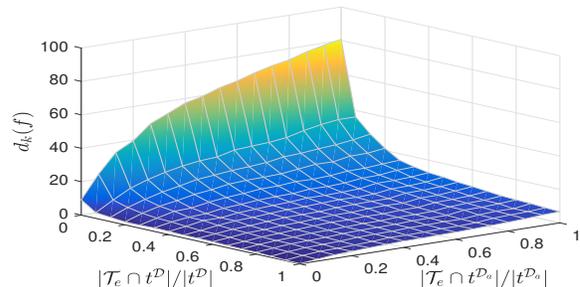


Figure 5: Anomaly scores for the example customer

score increases with the amount of \mathcal{T}_e contained in the testing dataset. However, it decreases with the amount of \mathcal{T}_e contained in the training dataset. A maximum occurs when $|\mathcal{T}_e \cap t^{\mathcal{D}}|/|t^{\mathcal{D}}|$ is 0 and $|\mathcal{T}_e \cap t^{\mathcal{D}_a}|/|t^{\mathcal{D}_a}|$ is 1. We extend these properties of the anomaly score to all rolling windows, all synthetic theft cases, and all customers. To do this, we consider 5 different theft intervals \mathcal{T}_e , which begin at 20%, 30%, 40%, 50%, and 60% of the way through the dataset and end at the last sample. The 4 synthetic theft cases were considered with parameters given by $\alpha_{c2} = 1$ kWh; $\alpha_{c3} = 1.8$ kWh; $\alpha_{c4} = 0.5$. In total, we have 20 different synthesized datasets for each customer.

Table II: Maximum anomaly score $\max_f d_i(f)$ and ranking percentile averaged over all customers

\mathcal{T}_e	20%	30%	40%	50%	60%
Case 1	16.2 (55)	30.7 (10)	79.3 (1)	88.8 (1)	94.0 (1)
2	16.6 (53)	26.9 (16)	67.4 (1)	70.2 (1)	69.8 (1)
3	17.5 (47)	27.1 (16)	61.3 (1)	64.3 (1)	64.7 (1)
4	17.0 (50)	25.2 (19)	43.4 (3)	46.8 (2)	49.6 (2)

 Table III: Difference $|t^{\mathcal{D}_a} \cap \mathcal{T}_e|/|t^{\mathcal{D}_a}| - |t^{\mathcal{D}} \cap \mathcal{T}_e| \cdot |t^{\mathcal{D}}|$

\mathcal{T}_e	20%	30%	40%	50%	60%
Case 1	0.44	0.77	0.92	0.93	0.93
2	0.37	0.77	0.93	0.92	0.92
3	0.35	0.76	0.92	0.92	0.92
4	0.25	0.71	0.88	0.89	0.87

The 59 rolling windows described in Subsection V.B were simulated $980 \cdot 20$ times. That is, for each rolling window, there is a simulation for each customer in each of the 20 theft modes. For each simulation, denote k as the index of that simulation's thief. We report the value and ranking percentile of $\max_f d_k(f)$ among $\max_f d_i(f)$ of all other customers i in Table.II. The numbers in the parenthesis are the ranking percentile of the anomaly score in that cell.

In the first two columns of Table II, the anomaly periods \mathcal{T}_e intersect the training dataset. As a result, the maximum anomaly scores $\max_f d_i(f)$ are indistinguishable from the anomaly scores of non-thieves. However, these scores increase as \mathcal{T}_e takes up smaller portions of the training dataset as shown in the last three columns of Table II. Thus the rolling window approach is useful when there is no theft for the first part of the analysis. The exact amount of time necessary for this part of the analysis depends on the length of the training window.

The highest anomaly scores correspond to the rolling window which has a maximum amount of clean data in training set and a minimum amount of clean data in testing set. This intuition is confirmed in Table.III. Each cell is the difference $|t^{\mathcal{D}_a} \cap \mathcal{T}_e|/|t^{\mathcal{D}_a}| - |t^{\mathcal{D}} \cap \mathcal{T}_e| \cdot |t^{\mathcal{D}}|$ averaged over all customers.

The detection abilities of these anomaly scores need a window that is both clean and strong in theft. Such a window will exist so long as the thief does not steal power throughout the entire analysis. Table.II shows that this window will be recognizable because the anomaly score of the thief will increase substantially during this window. We may idealize this window as one with $|t^{\mathcal{D}} \cap \mathcal{T}_e| \cdot |t^{\mathcal{D}}| = 0$ and $|t^{\mathcal{D}_a} \cap \mathcal{T}_e| \cdot |t^{\mathcal{D}_a}| \approx 1$. Since this particular rolling window is of crucial importance, we study it in detail in Subsections V.D and V.E.

D. The Impact of Energy Theft on Out-of-sample Residuals

This subsection focuses on the behavior of out-of-sample residuals when the training set is clean. It emphasizes a difference in behavior between theft and non-theft scenarios.

We first synthesize smart meter data for customer k under synthetic case 3. We assume that the electricity theft activities occur from hour $t_1^{(e)} = 25$ to hour $t_2^{(e)} = 168$ in the out-of-sample period. The amount of electricity theft is assumed to follow a uniform distribution with $p_k^s(t) \sim \mathcal{U}(0, 1.8)$ (kWh).

Table IV: Anomaly Scores

d_k	Ranking	$\sum_i d_i/N$	$\text{PR}(d_i, 95)$	$\max_{i \neq k} d_i$
79.4	1 (0.1%)	8.1	14.6	42.2

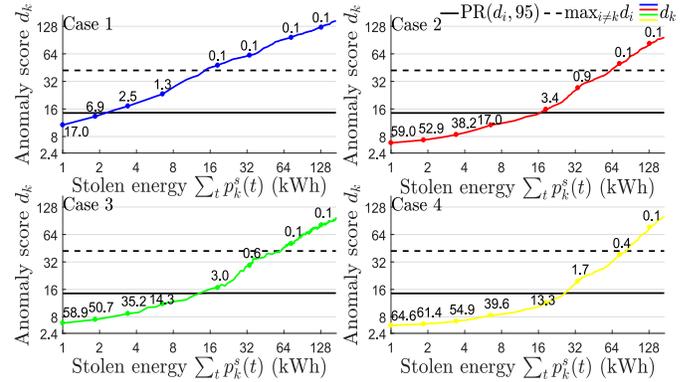


Figure 6: Anomaly scores versus amount of stolen electricity

The MLM is applied for all 4 customers in the same secondary. The out-of-sample residuals for the 4 customers are shown in Fig.7b. The figure represents the residuals of customer k by the solid green line. The other 3 customers' residuals are represented by blue dash lines. The out-of-sample residuals obtained from the original data (without electricity theft) are shown in Fig.7a for comparison purposes. As shown in Fig.7b,

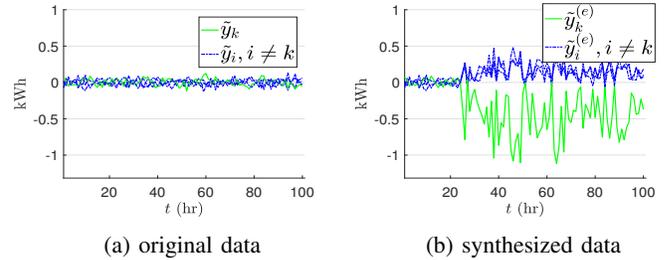


Figure 7: Out-of-sample residuals.

customer k has negative residuals while the honest customers have positive residuals. The sum of them at any given hour is zero as stated in Lemma.2. For all customers, the regression coefficients $\sum_{\ell \neq j} \beta_{\ell}^y$ are positive. In this case, $\sum_{\ell \neq k} \beta_{\ell}^y$ took on the values of 0.74, 0.77, 0.67 and 0.81 for $j = 1, 2, 3$ and 4. Hence, the residuals of the dishonest customer k will always be negative according to Lemma.1. These results show that the residual plots of all customers on the same secondary are helpful in detecting electricity theft.

E. The Impact of Energy Theft on Anomaly Scores

Next, we will show that electricity theft can be easily detected by anomaly scores in a wide variety of cases. We further show that the anomaly scores increase with the amount of stolen electricity.

We first calculate the anomaly scores for all customers on the distribution feeder under the experiment detailed in Subsection V.D. The anomaly score of customer k and the

summary statistics of all customers' anomaly scores are reported in Table IV. As shown in the table, the k th customer has an anomaly score of 79.4. This is the highest among all 980 customers in the distribution feeder. The second highest anomaly score of any customer is 42.2 which is much lower than that of customer k . The average and 95th percentile of all customers' anomaly scores $\text{PR}(d_i, 95)$ are 8.1 and 14.6. Both of these are much lower than that of customer k . In this case, the anomaly score easily detects the electricity theft activity.

Synthetic electricity theft datasets for customer k are then created for each of the synthetic cases. The parameters used for the cases are as follows. In case 1, the theft activity starts from hour 1 in the out-of-sample period. Multiple datasets for this case are then created by increasing the theft ending hour in the out-of-sample period. We create a dataset for each ending hour from hour 2 to hour 168. In cases 2-4, the theft activity starts and ends with hours $t_1^{(e)} = 0.2|t^{\mathcal{D}_a}|$ and $t_2^{(e)} = |t^{\mathcal{D}_a}|$. Multiple datasets for this case are then created by increasing the parameters α_{c2} , α_{c3} , and α_{c4} . The parameters are varied such that the total amount of stolen electricity ranges from 1 kWh to 128 kWh. Again the electricity theft activities are assumed to occur during the out-of-sample period.

We then calculate the residual and anomaly score for each constructed dataset. The anomaly score of customer k and the summary statistics of all customers' anomaly scores are depicted in Fig.6. The colored solid curve in each subplot represents the anomaly scores of customer k . The numbers along the curve show the ranking percentile of customer k 's anomaly score. The solid black line represents the 95th percentile of all customers' anomaly scores. The dashed black line represents the average anomaly score of all customers. Both axes are on a logarithmic scale.

The figure shows that the anomaly score of customer k increases monotonically with the amount of stolen electricity. In all cases, customer k 's anomaly score will surpass the 95th percentile of all customers if it steals more than 32 kWh. This averages to 0.19 kW of power. A stronger result holds for cases 1-3. In these cases, customer k 's anomaly score will be the absolute largest of all customers if it steals more than 0.38 kW of power.

To further prove the validity of the proposed framework, we extend this analysis to all customers. That is, the previous case study is repeated 980 times. Each new set of cases sets a new customer as the thief.

The anomaly scores of the customers who are stealing electricity are binned and reported in Fig.8. The x- and y-axes of the figure represent the amount of stolen electricity and the anomaly scores. The z-axis represents the number of customers who have an anomaly score which falls into a particular bin. The color of the each bar indicates the ranking (in percentage) of anomaly score of the customers in that bin against all honest customers. A darker color represents a higher ranking. Each row of bars add up to the total number of customers in the distribution feeder. When the amount of stolen electricity increases, the distribution of dishonest anomaly scores shifts to the right. The ranking of the anomaly scores also increases.

Finally, the figure can be used to predict the probabilities of detection. For example, if a customer steals more than 0.38

kW of power, then it has a 97 percent chance of surpassing the 95th percentile of all customers. It further has a 57 percent chance that its anomaly score will be the highest among all customers. These results show that framework is effective in catching even small amounts of theft.

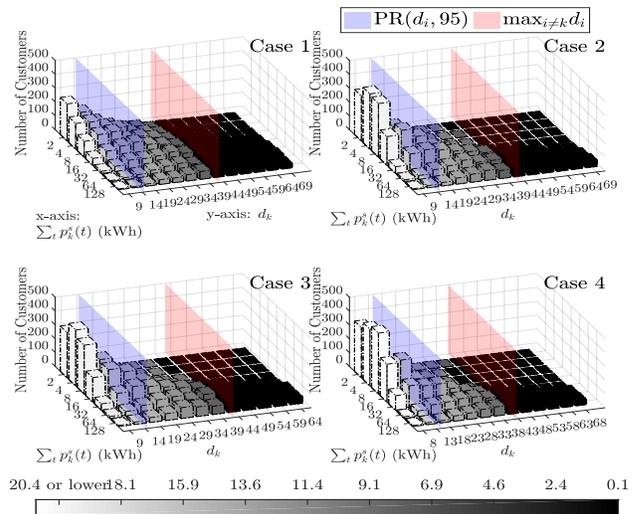


Figure 8: Numerical evaluation for all customers on the distribution feeder

F. Comparison with Existing Techniques

We compare the performance of the proposed anomaly detection method with the Fuzzy C-means (FCM) based method [5], the Self-Organizing Maps (SOM) based method [10], and Random Matrix Theory (RMT) based method [25]. We excluded the comparisons with state estimation based methods, supervised machine learning based methods, and the ‘‘central observer’’ based methods such as [18] because they represent solutions to different classes of problems. The original methods described in [5], [10], and [25] need to be modified slightly to match our experimental data and our performance measure.

For the FCM method. First, missing values were imputed by the average values of the two nearest time stamps of the same customer. Next, the time series data was dimension-reduced via the t-distributed Stochastic Neighbor Embedding (t-SNE) algorithm [44]. This was performed separately on the consumption and voltage time series data. The results were concatenated to form a feature vector. Second, the definition of the anomaly score in [5, Fig.3] is modified to be $d_k = \|\mathbf{u}_{0k} - \mathbf{u}_k \mathbf{P}^*\|_2$ where \mathbf{P}^* is a column permutation matrix such that $\mathbf{P}^* = \text{argmin} \|\mathbf{U}_0 - \mathbf{U}\mathbf{P}\|$. For the SOM method. First, the kWh and voltage time series of each customer were converted to kWh and voltage daily profiles. Missing values and outage values were treated similarly to the FCM method above. The step of comparing with the contracted power demand was removed. We further defined the minimum quantization error [45] as the anomaly score. For the RMT method. First, each distribution transformer secondary is considered as a region. Next, a window of 67 days of hourly

two-phase voltage magnitude and active power measurements are collected and undergo the same preprocessing procedure as described in Section IV.A. The active power measurement noise is assumed to be zero mean normal with a standard deviation of 0.02 times the range of active power.

We setup experiments as follows. The four different anomaly cases discussed in Subsection V.A will be simulated with the parameters $\mathcal{T}_e, \alpha_{c2}, \alpha_{c3}$ and α_{c4} being varied such that the total amount of bypassed electricity ranges from 2 kWh to 128 kWh. We performed the experiments using the same set of customers and training/testing dataset as depicted in Fig.8. The results are shown in Table.V. Each cell is the anomaly score ranking for anomalous customers with respect to normal customers averaged for all selection of anomalous customers and expressed in percentage. Table.V shows that the

Table V: Performance comparison with [5], [10], and [25]

$\sum_t p_k^s(t)$ (kWh)	2	4	8	16	32	64	128
Case 1: disconnection of meters model							
FCM	49.65	50.30	48.94	47.75	42.75	38.64	34.34
SOM	47.07	44.69	42.19	38.18	30.90	21.12	16.37
RMT	49.22	46.10	40.06	30.42	20.70	15.03	10.95
MLM	13.29	7.07	4.20	2.30	1.29	0.85	0.73
Case 2: constant bypassing model							
FCM	50.49	49.49	49.48	48.40	45.10	40.59	36.94
SOM	49.86	49.51	48.38	44.59	34.21	22.33	18.97
RMT	51.23	51.09	50.87	50.60	48.92	41.35	26.99
MLM	40.54	32.23	19.95	8.55	2.73	1.21	0.91
Case 3: random uniform bypassing model							
FCM	50.11	49.51	48.93	48.35	44.96	41.02	38.52
SOM	49.83	49.43	47.93	43.38	33.46	24.55	21.07
RMT	50.95	51.17	50.74	49.73	45.04	34.88	22.79
MLM	40.15	31.00	17.96	7.10	2.39	1.29	0.99
Case 4: constant percentage bypassing model							
FCM	50.32	50.33	49.60	48.89	46.27	41.55	36.09
SOM	50.33	50.53	50.71	50.08	45.17	28.78	19.36
RMT	51.27	51.19	51.33	50.90	47.62	33.29	16.20
MLM	44.30	38.93	28.65	15.07	5.15	1.43	0.83

proposed method beats the modified existing techniques in all cases. For all four methods, the rankings of the anomaly scores decrease in response to increasing level of anomaly. However, only the proposed method consistently ranks the anomalous customers at the top.

VI. CONCLUSION

This paper developed a physically inspired data-driven algorithm for electricity theft detection. The proposed algorithm leverages an approximate linear relationship between the power consumption and voltage data of customers on the same secondary. The proposed MLM produces accurate estimates of the electricity consumption for the majority of the customers. The MLM model is able to detect inconsistencies among smart meter measurements of a group of customers from the same distribution secondary thereby identifying electricity thefts. An evaluation of the proposed electricity theft detection algorithm was then performed with real-world smart meter data and synthesized electricity theft cases. The evaluation

results show that the proposed anomaly score developed in this paper is effective in identifying electricity theft cases even when the amount of stolen electricity is small. The method was compared with existing unsupervised electricity theft detection techniques. The comparison results show that the proposed method is more effective in identifying the electricity thefts.

APPENDIX A

LINEARIZATION OF DISTRIBUTION SECONDARY POWER FLOW EQUATIONS

We wish to approximate the nonlinear power flow equation as a linear one:

$$\mathcal{F}(\mathbf{v}, \boldsymbol{\theta}, \mathbf{p}, \mathbf{q}) = \mathbf{0} \quad \rightarrow \quad \mathbf{F}_{\bar{\mathbf{X}}}[\mathbf{v}^\top, \boldsymbol{\theta}^\top, \mathbf{p}^\top, \mathbf{q}^\top]^\top = \mathbf{0}$$

where $\mathbf{v} = [\mathbf{v}^{1\top}, \mathbf{v}^{2\top}]^\top$ (same token for $\boldsymbol{\theta}, \mathbf{p}, \mathbf{q}$); $\mathbf{F}_{\bar{\mathbf{X}}}$ is the Jacobian matrix of \mathcal{F} evaluated at some operating point $\bar{\mathbf{X}} = [\bar{\mathbf{v}} \quad \bar{\boldsymbol{\theta}} \quad \bar{\mathbf{p}} \quad \bar{\mathbf{q}}]^\top$. This point must itself be a solution to the power flow equation $\mathcal{F}(\bar{\mathbf{X}}) = \mathbf{0}$. When $\bar{\mathbf{X}}$ is fixed, this Jacobian is given by [30]

$$\mathbf{F}_{\bar{\mathbf{X}}} = [(\langle \text{diag}(\mathbf{Y}\mathbf{u})^* \rangle + \langle \text{diag}(\mathbf{u}) \rangle \mathbf{N}_{2n} \langle \mathbf{Y} \rangle) R(\mathbf{u}) \quad -\mathbf{I}] \quad (20)$$

where \mathbf{Y} is the bus admittance matrix, \mathbf{u} is the vector of complex bus voltages, $\mathbf{N}_{2n} = \begin{bmatrix} \mathbf{I}_{2n} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_{2n} \end{bmatrix}$, and

$$R(\mathbf{u}) = \begin{bmatrix} \text{diag}(\cos(\boldsymbol{\theta})) & -\text{diag}(\mathbf{v} \sin(\boldsymbol{\theta})) \\ \text{diag}(\sin(\boldsymbol{\theta})) & \text{diag}(\mathbf{v} \cos(\boldsymbol{\theta})) \end{bmatrix}$$

$$\langle \mathbf{A} \rangle = \begin{bmatrix} \text{Re}\{\mathbf{A}\} & -\text{Im}\{\mathbf{A}\} \\ \text{Im}\{\mathbf{A}\} & \text{Re}\{\mathbf{A}\} \end{bmatrix}$$

Recall that our modified flat voltage solution is given by $\bar{\mathbf{u}} = [\mathbf{1}_n, -\mathbf{1}_n]^\top$, $\bar{\mathbf{p}} + j\bar{\mathbf{q}} = \mathbf{0}$. Assuming that no shunt resistances are present, this is a solution to the power flow manifold with zero branch currents. Thus $\text{diag}(\mathbf{Y}\mathbf{u}) = \mathbf{0}$.

Furthermore, $\text{diag}(\mathbf{u}) = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_n \end{bmatrix} \triangleq \mathbf{N}_n$, so $\langle \text{diag}(\mathbf{u}) \rangle =$

$R(\mathbf{u}) = \begin{bmatrix} \mathbf{N}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_n \end{bmatrix}$. Thus the left hand block matrix of (20) reduces to

$$\begin{bmatrix} \mathbf{N}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_n \end{bmatrix} \begin{bmatrix} \mathbf{G}^r & -\mathbf{B}^r \\ -\mathbf{B}^r & -\mathbf{G}^r \end{bmatrix} \begin{bmatrix} \mathbf{N}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{N}_n \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{N}_n \mathbf{G}^r \mathbf{N}_n & -\mathbf{N}_n \mathbf{B}^r \mathbf{N}_n \\ -\mathbf{N}_n \mathbf{B}^r \mathbf{N}_n & -\mathbf{N}_n \mathbf{G}^r \mathbf{N}_n \end{bmatrix}$$

where \mathbf{G}^r and \mathbf{B}^r are the real and imaginary components of \mathbf{Y}^r . In this final expression, each product $\mathbf{N}_n \mathbf{A} \mathbf{N}_n$ negates the off diagonal blocks of \mathbf{A} , yielding the desired linearization

$$\begin{bmatrix} \mathbf{p}^1 \\ \mathbf{p}^2 \\ \mathbf{q}^1 \\ \mathbf{q}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{G}^{11} & -\mathbf{G}^{12} & -\mathbf{B}^{11} & \mathbf{B}^{12} \\ -\mathbf{G}^{21} & \mathbf{G}^{22} & \mathbf{B}^{21} & -\mathbf{B}^{22} \\ -\mathbf{B}^{11} & \mathbf{B}^{12} & -\mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{B}^{21} & -\mathbf{B}^{22} & \mathbf{G}^{21} & -\mathbf{G}^{22} \end{bmatrix} \begin{bmatrix} \mathbf{v}^1 \\ \mathbf{v}^2 \\ \boldsymbol{\theta}^1 \\ \boldsymbol{\theta}^2 \end{bmatrix} \quad (21)$$

APPENDIX B

CONVERSION FROM LOADS TO NET INJECTIONS

(2) is derived as the follows. First define the reference direction of voltages and currents as shown in Fig.9. u variables refer to voltages, i variables refer to currents, and s variables refer to VA power consumptions.

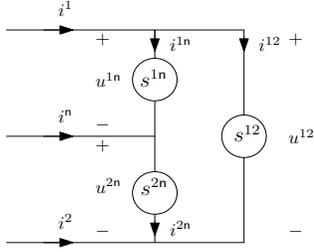


Figure 9: A triplex line load with reference direction defined

We then have:

$$\begin{aligned} s^1 &= u^{1n} i^{1*} = u^{1n} (i^{1n*} + i^{12*}) = s^{1n} + u^{1n} \frac{s^{12}}{u^{1n} + u^{2n}} \\ s^2 &= u^{2n} i^{2*} = u^{2n} (-i^{2n*} - i^{12*}) = s^{2n} - u^{2n} \frac{s^{12}}{u^{1n} + u^{2n}} \end{aligned} \quad (22)$$

APPENDIX C

THE ELIMINATION OF DEPENDENCIES ON VOLTAGE ANGLES AND REACTIVE POWERS

We first show that the pseudoinverse of \mathbf{L}_{22}^s can eliminate the voltage angle dependence in our model. We first rearrange the model equations to

$$\mathbf{p}^s = \mathbf{L}_{11}^s \mathbf{v}^s + \mathbf{L}_{12}^s \boldsymbol{\theta}^s \quad (23)$$

$$\mathbf{L}_{22}^s \boldsymbol{\theta}^s = \mathbf{q}^s - \mathbf{L}_{21}^s \mathbf{v}^s \quad (24)$$

Since (24) is enforced by our model, its right hand side is in $\text{Range}(\mathbf{L}_{22}^s)$. Let $\boldsymbol{\theta}^*$ denote its least norm solution. Then any other solution can be written as $\boldsymbol{\theta}^s = \boldsymbol{\theta}^* + \boldsymbol{\theta}_n$ where $\boldsymbol{\theta}_n$ is in $\text{Null}(\mathbf{L}_{22}^s)$.

Now,

$$\mathbf{L}_{22}^s = \begin{bmatrix} -\mathbf{G}^{11} & \mathbf{G}^{12} \\ \mathbf{G}^{21} & -\mathbf{G}^{22} \end{bmatrix} \quad (25)$$

and each of the blocks \mathbf{G}^{ij} is a Laplacian matrix having nullspace $\mathbf{1}$. In practice the mutual conductances are much smaller than the self conductances [46]. Hence the entries in \mathbf{G}^{12} and \mathbf{G}^{21} are much smaller than that in \mathbf{G}^{11} and \mathbf{G}^{22} . Then the overall matrix \mathbf{L}_{22}^s has nullspace

$$\text{Null}(\mathbf{L}_{22}^s) = \text{Span} \left(\begin{bmatrix} \mathbf{1}_n \\ \mathbf{0}_n \end{bmatrix}, \begin{bmatrix} \mathbf{0}_n \\ \mathbf{1}_n \end{bmatrix} \right)$$

But

$$\mathbf{L}_{12}^s = \begin{bmatrix} -\mathbf{B}^{11} + \mathbf{B}^{21} & \mathbf{B}^{12} - \mathbf{B}^{22} \\ -\mathbf{B}^{11} - \mathbf{B}^{21} & \mathbf{B}^{12} + \mathbf{B}^{22} \end{bmatrix} \quad (26)$$

where each block \mathbf{B}^{ij} is again a Laplacian matrix. Thus the nullspace of this matrix contains the above spanning vectors, so $\text{Null}(\mathbf{L}_{22}^s) \subseteq \text{Null}(\mathbf{L}_{12}^s)$. It follows that $\mathbf{L}_{12}^s \boldsymbol{\theta}^s = \mathbf{L}_{12}^s (\boldsymbol{\theta}^* + \boldsymbol{\theta}_n) = \mathbf{L}_{12}^s \boldsymbol{\theta}^*$. We can then write the above system as

$$\mathbf{p}^s = \mathbf{L}_{11}^s \mathbf{v}^s + \mathbf{L}_{12}^s \boldsymbol{\theta}^* \quad (27)$$

$$\mathbf{L}_{22}^s \boldsymbol{\theta}^* = \mathbf{q}^s - \mathbf{L}_{21}^s \mathbf{v}^s \quad (28)$$

where $\boldsymbol{\theta}^* = \mathbf{L}_{22}^{s\dagger} (\mathbf{q}^s - \mathbf{L}_{21}^s \mathbf{v}^s)$ because it is the least norm solution. Substituting this into (27) yields the desired result

$$\mathbf{p}^s = (\mathbf{I}_{11}^s - \mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{L}_{21}^s) \mathbf{v}^s + \mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{q}^s \quad (29)$$

We conclude this appendix by showing that $(\mathbf{I} - \mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{D} \mathbf{M}_u^{-1})$ is nonsingular. We do this by showing that 1 is *not* an eigenvalue of $\mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{D} \mathbf{M}_u^{-1}$.

First, it is easy to show that $\mathbf{L}_{22}^r = \mathbf{L}_{22}^s$, and $\mathbf{L}_{12}^s = \mathbf{M}_u \mathbf{L}_{12}^r$, so

$$\mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{D} \mathbf{M}_u^{-1} = \mathbf{M}_u \mathbf{L}_{12}^r \mathbf{L}_{22}^{r\dagger} \mathbf{D} \mathbf{M}_u^{-1} \quad (30)$$

Thus, if 1 is an eigenvalue of $\mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{D} \mathbf{M}_u^{-1}$, then 1 is an eigenvalue of $\mathbf{L}_{12}^r \mathbf{L}_{22}^{r\dagger} \mathbf{D}$. Then there exists a vector \mathbf{p} such that

$$\mathbf{p} = \mathbf{L}_{12}^r \mathbf{L}_{22}^{r\dagger} \mathbf{D} \mathbf{p} \quad (31)$$

Then there exists a vector $\boldsymbol{\theta}_x \in \text{Range}(\mathbf{L}_{22}^{r\dagger})$ such that

$$\mathbf{L}_{22}^r \boldsymbol{\theta}_x = \mathbf{Q} \mathbf{D} \mathbf{p} \quad (32)$$

$$\mathbf{L}_{12}^r \boldsymbol{\theta}_x = \mathbf{p} \quad (33)$$

where \mathbf{Q} is the orthogonal projector onto the range of \mathbf{L}_{22}^r . Then, since $\boldsymbol{\theta}_x \perp \text{Null}(\mathbf{L}_{22}^r)$ we have

$$(\mathbf{L}_{22}^r - (\mathbf{I} - \mathbf{N}) \mathbf{D} \mathbf{L}_{12}^r) \boldsymbol{\theta}_x = \mathbf{0} \quad (34)$$

$$\mathbf{N} \boldsymbol{\theta}_x = \mathbf{0} \quad (35)$$

Where \mathbf{N} is the orthogonal projector onto the nullspace of \mathbf{L}_{22}^r and is given by

$$\mathbf{N} = \frac{1}{n} \begin{bmatrix} \mathbf{1} \mathbf{1}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \mathbf{1}^T \end{bmatrix} \quad (36)$$

Thus, for a solution to exist, the following augmented matrix cannot have full column rank ($= 2n_c$):

$$\begin{bmatrix} \mathbf{L}_{22}^r - (\mathbf{I} - \mathbf{N}) \mathbf{D} \mathbf{L}_{12}^r \\ \mathbf{N} \end{bmatrix} \quad (37)$$

But clearly $(\mathbf{I} - \mathbf{N}) \mathbf{D} \mathbf{L}_{12}^r$ has the same range and nullspace as \mathbf{L}_{12}^r . The sum of the first n_c rows is therefore zero. The same holds for the last n_c rows. Thus we can perform row operations to show that this has the same rank as the matrix with the n_c^{th} and $2n_c^{\text{th}}$ rows removed. The upper matrix also has the property that the sum of the first n_c columns is zero and the sum of the last n_c columns is zero. Thus column operations show that our matrix has the same rank as

$$\begin{bmatrix} [\mathbf{L}_{22}^r - (\mathbf{I} - \mathbf{N}) \mathbf{D} \mathbf{L}_{12}^r]_{red} & \mathbf{0} & \mathbf{0} \\ \mathbf{1} \mathbf{1}^T & \mathbf{0} & \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \mathbf{1}^T & \mathbf{0} & \mathbf{1} \end{bmatrix} \quad (38)$$

where the *red* subscript indicates that the n_c^{th} row, n_c^{th} column, $2n_c^{\text{th}}$ row, and $2n_c^{\text{th}}$ column have been removed.

Now, the lower right hand block of this matrix indicates two pivots, so a solution can only exist if $[\mathbf{L}_{22}^r - (\mathbf{I} - \mathbf{N}) \mathbf{D} \mathbf{L}_{12}^r]_{red}$ does not have full rank ($= 2(n_c - 1)$). Since this matrix only removes rows and columns from its constituents, we can write it as $[\mathbf{L}_{22}^r]_{red} - [(\mathbf{I} - \mathbf{N}) \mathbf{D} \mathbf{L}_{12}^r]_{red}$ where the constituent matrices now have full rank.

This matrix subtraction is unlikely to have less than full rank for two reasons. First, it is an extremely precise requirement on the relationship between the network parameters and the power factors. It is precise in the sense that the set of all invertible matrices sum to a singular matrix has Lebesgue measure zero. Second, the matrix $[\mathbf{L}_{22}^r]_{red}$ contains conductance values and the matrix $[(\mathbf{I} - \mathbf{N}) \mathbf{D} \mathbf{L}_{12}^r]_{red}$ contains

transformed susceptance values. Since susceptance values are typically much larger than conductance values, it follows that the rows of the matrix subtraction will be primarily dominated by $[(\mathbf{I} - \mathbf{N})\mathbf{DL}_{12}^{\dagger}]_{red}$ which has full rank. Thus a real network is unlikely to have this difference be singular or even close to singular. Therefore in realworld cases, $(\mathbf{I} - \mathbf{L}_{12}^s \mathbf{L}_{22}^{s\dagger} \mathbf{DM}_u^{-1})$ will be invertible.

APPENDIX D

PROOF OF LEMMA.1 LEMMA.2 AND LEMMA.3

A. Lemma.2

Proof. Suppose without loss of generality that customer i is the electricity thief. Suppose that our training window lasts T time instances. Then at any time t :

$$\begin{aligned} & \sum_j (\tilde{y}(t)_j^e - \tilde{y}(t)_j) \\ &= \sum_j ((y(t)_j^e - y(t)_j) - (\mathcal{X}(t)^e - \mathcal{X}(t))\beta_j^y) \\ &= (y(t)_i^e - y(t)_i) - \sum_j \beta_j^y \sum_k (y(t)_k^e - y(t)_k) \\ &= (y(t)_i^e - y(t)_i)(1 - \sum_j \beta_j^y) \end{aligned} \quad (39)$$

because $(y(t)_j^e - y(t)_j)$ is nonzero at index i only and \mathcal{X}^e and \mathcal{X} differ only in their last component.

Now, due to the use of ordinary least squares, β_j is the pseudoinverse of the matrix $[\mathcal{X}_v \ \mathbf{y}_\Sigma]$ applied to \mathbf{y}_j^D . Here, \mathcal{X}_v is a T by $n_c + 1$ matrix of in sample voltage measurements, \mathbf{y}_Σ is a T dimensional vector of in sample power sums, and \mathbf{y}_j^D is a T dimensional vector of in sample power measurements. We can write the pseudoinverse in block form [47] to obtain

$$\begin{bmatrix} \beta_j^x \\ \beta_j^y \end{bmatrix} = \begin{bmatrix} (\mathcal{X}_v^T Q_y \mathcal{X}_v)^{-1} \mathcal{X}_v^T Q_y \\ (\mathbf{y}_\Sigma^T Q_x \mathbf{y}_\Sigma)^{-1} \mathbf{y}_\Sigma^T Q_x \end{bmatrix} \mathbf{y}_j^D \quad (40)$$

where Q_y and Q_x are the residual projection matrices

$$\begin{aligned} Q_y &= \mathbf{I} - \mathbf{y}_\Sigma (\mathbf{y}_\Sigma^T \mathbf{y}_\Sigma)^{-1} \mathbf{y}_\Sigma^T \\ Q_x &= \mathbf{I} - \mathcal{X}_v (\mathcal{X}_v^T \mathcal{X}_v)^{-1} \mathcal{X}_v^T \end{aligned}$$

Now, $\sum_j \mathbf{y}_j^D = \mathbf{y}_\Sigma$, so summing (40) over j yields

$$\sum_j \begin{bmatrix} \beta_j^x \\ \beta_j^y \end{bmatrix} = \begin{bmatrix} (\mathcal{X}_v^T Q_y \mathcal{X}_v)^{-1} \mathcal{X}_v^T Q_y \mathbf{y}_\Sigma \\ (\mathbf{y}_\Sigma^T Q_x \mathbf{y}_\Sigma)^{-1} \mathbf{y}_\Sigma^T Q_x \mathbf{y}_\Sigma \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} \quad (41)$$

where the $\mathbf{0}$ comes from the residual of the projection of \mathbf{y}_Σ onto itself. (39) and (41) show that $\sum_j (\tilde{y}(t)_j^e - \tilde{y}(t)_j) = 0$. Vectorizing over time yields the left hand equality of the lemma. Finally, since

$$\hat{\mathbf{y}}_j^D = [\mathcal{X}_v \ \mathbf{y}_\Sigma] \begin{bmatrix} \beta_j^x \\ \beta_j^y \end{bmatrix} \quad (42)$$

we have

$$\sum_j \hat{\mathbf{y}}_j^D = \mathcal{X}_v \sum_j \beta_j^x + \mathbf{y}_\Sigma \sum_j \beta_j^y = \mathbf{y}_\Sigma = \sum_j \mathbf{y}_j^D \quad (43)$$

Subtracting the leftmost term from the rightmost term yields the right hand equality of the lemma for in sample data. A similar argument shows that the above equation also holds for out of sample data. \square

B. Lemma.1

Proof. Repeating the derivation of (39), but omitting index i from the sum yields

$$\sum_{j \neq i} (\tilde{y}(t)_j^e - \tilde{y}(t)_j) = (y(t)_i^e - y(t)_i)(1 - \beta_i^y) \quad (44)$$

The remaining terms present in the right hand side of (39) but absent in (44) are $-(y(t)_i^e - y(t)_i) \sum_{j \neq i} \beta_j^y$. Therefore it must be the case that $\tilde{y}(t)_i^e - \tilde{y}(t)_i = -(y(t)_i^e - y(t)_i) \sum_{j \neq i} \beta_j^y$. Vectorizing over time yields the result

$$\tilde{\mathbf{y}}_i^{(e)} - \tilde{\mathbf{y}}_i = - \sum_{j \neq i} \beta_j \mathbf{y}_i^s \quad (45)$$

\square

C. Lemma.3

Proof. Consider the true (unestimated) model

$$p_j(t) = \mathbf{x}(t)^T \mathbf{r}_j + c_j y_\Sigma(t) \quad (46)$$

Consider further the hypothetical scenario where *only* the transformer voltage deviates from its flat value. In this scenario, we have for customer j

$$p_j(t) = c_j y_\Sigma(t) \quad (47)$$

so $c_j = p_j(t) / \sum_{i=1} p_i(t)$ is the portion of the total power injection contributed by customer j in this scenario. But if all voltages are flat except the transformer voltage, then the power injections must all have the same sign, so $c_j \geq 0$.

Now β_j^y is an estimator of c_j . We will repeat its equation here:

$$\beta_j^y = (\mathbf{y}_\Sigma^T Q_x \mathbf{y}_\Sigma)^{-1} \mathbf{y}_\Sigma^T Q_x \mathbf{y}_j^D \quad (48)$$

This estimator is *biased*. This is because the term \mathbf{y}_Σ is confounded by the sum of all noise terms for each individual dependent variable. Thus this estimator suffers from the *Classical Errors in Variable Problem* [48]. But since this lemma only relies on the sign of β_j^y , this does not pose much of a problem. We still have

$$\text{plim } \beta_j^y = \lambda c_j, \quad 0 < \lambda < 1 \quad (49)$$

Then there are two cases. If $c_j > 0$, then $\mathbb{P}(\beta_j^y < 0) \leq \mathbb{P}(|\beta_j^y - \lambda c_j| \geq \lambda c_j) \rightarrow 0$ as the training window length goes to infinity. Thus for any $\delta > 0$, there exists a window length T_1^j such that $\mathbb{P}(\beta_j^y < 0) < \frac{\delta}{2}$. If, however, $c_j = 0$, then there exists a window length T_2^j such that $\mathbb{P}(|\beta_j^y| \geq \delta) < \frac{\delta}{2}$. Let $T^j = \max(T_1^j, T_2^j)$. Then for window length T^j , $\mathbb{P}(\beta_j < -\delta) \leq 2(\frac{\delta}{2}) = \delta$. Letting $T = \max_j \{T^j\}$ completes the proof. \square

REFERENCES

- [1] S. McLaughlin, D. Podkuiko, and P. McDaniel, *Energy Theft in the Advanced Metering Infrastructure*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 176–187.
- [2] T. B. Smith, “Electricity theft: a comparative analysis,” *Energy Policy*, vol. 32, no. 18, pp. 2067–2076, 2004.
- [3] S. S. R. Depuru, L. Wang, and V. Devabhaktuni, “Support vector machine based data classification for detection of electricity theft,” in *Power Systems Conference and Exposition (PSCE), 2011 IEEE/PES*. IEEE, 2011, pp. 1–8.

- [4] T. Ahmad, H. Chen, J. Wang, and Y. Guo, "Review of various modeling techniques for the detection of electricity theft in smart grid environment," *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2916–2933, 2018.
- [5] E. W. S. Angelos, O. R. Saavedra, O. A. C. Cortés, and A. N. de Souza, "Detection and identification of abnormalities in customer consumptions in power distribution systems," *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2436–2442, Oct 2011.
- [6] C. C. O. Ramos, A. N. de Sousa, J. P. Papa, and A. X. Falcao, "A new approach for nontechnical losses detection based on optimum-path forest," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 181–189, Feb 2011.
- [7] J. Nagi, K. S. Yap, S. K. Tiong, S. K. Ahmed, and M. Mohamad, "Nontechnical loss detection for metered customers in power utility using support vector machines," *IEEE Transactions on Power Delivery*, vol. 25, no. 2, pp. 1162–1171, April 2010.
- [8] B. Costa, B. L. A. Alberto, A. M. Portela, M. W, and E. O. Eler, "Fraud detection in electric power distribution networks using an ann-based knowledge-discovery process," vol. 4, pp. 17–23, 11 2013.
- [9] L. A. P. Júnior, C. C. O. Ramos, D. Rodrigues, D. R. Pereira, A. N. de Souza, K. A. P. da Costa, and J. P. Papa, "Unsupervised nontechnical losses identification through optimum-path forest," *Electric Power Systems Research*, vol. 140, pp. 413–423, 2016.
- [10] J. E. Cabral, J. O. P. Pinto, and A. M. A. C. Pinto, "Fraud detection system for high and low voltage electricity consumers based on data mining," in *2009 IEEE Power Energy Society General Meeting*, July 2009, pp. 1–5.
- [11] S. K. Singh, R. Bose, and A. Joshi, "Entropy-based electricity theft detection in ami network," *IET Cyber-Physical Systems: Theory Applications*, vol. 3, no. 2, pp. 99–105, 2018.
- [12] A. H. Nizar, Z. Y. Dong, and Y. Wang, "Power utility nontechnical loss analysis with extreme learning machine method," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 946–955, Aug 2008.
- [13] A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, "Decision tree and svm-based data analytics for theft detection in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 3, pp. 1005–1016, June 2016.
- [14] Z. Zheng, Y. Yang, X. Niu, H. N. Dai, and Y. Zhou, "Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 4, pp. 1606–1615, April 2018.
- [15] C. J. Bandim, J. E. R. Alves, A. V. Pinto, F. C. Souza, M. R. B. Loureiro, C. A. Magalhaes, and F. Galvez-Durand, "Identification of energy theft and tampered meters using a central observer meter: a mathematical approach," in *2003 IEEE PES Transmission and Distribution Conference and Exposition (IEEE Cat. No.03CH37495)*, vol. 1, Sept 2003, pp. 163–168 Vol.1.
- [16] S.-C. Yip, K. Wong, W.-P. Hew, M.-T. Gan, R. C.-W. Phan, and S.-W. Tan, "Detection of energy theft and defective smart meters in smart grids using linear regression," *International Journal of Electrical Power & Energy Systems*, vol. 91, pp. 230–240, 2017.
- [17] P. Jokar, N. Arianpoo, and V. C. M. Leung, "Electricity theft detection in AMI using customers consumption patterns," *IEEE Transactions on Smart Grid*, vol. 7, no. 1, pp. 216–226, Jan 2016.
- [18] M. Tariq and H. V. Poor, "Electricity theft detection and localization in grid-tied microgrids," *IEEE Transactions on Smart Grid*, vol. 9, no. 3, pp. 1920–1929, May 2018.
- [19] R. V. Cruz, C. V. Quintero, and F. Perez, "Detecting non-technical losses in radial distribution system transformation point through the real time state estimation method," in *2006 IEEE/PES Transmission Distribution Conference and Exposition: Latin America*, Aug 2006, pp. 1–5.
- [20] S. C. Huang, Y. L. Lo, and C. N. Lu, "Non-technical loss detection using state estimation and analysis of variance," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 2959–2966, Aug 2013.
- [21] W. Luan, G. Wang, Y. Yu, J. Lin, W. Zhang, and Q. Liu, "Energy theft detection via integrated distribution state estimation based on ami and scada measurements," in *2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, Nov 2015, pp. 751–756.
- [22] S. Salinas, C. Luo, W. Liao, and P. Li, "State estimation for energy theft detection in microgrids," in *9th International Conference on Communications and Networking in China*, Aug 2014, pp. 96–101.
- [23] C. Carquex and C. Rosenberg, "Multi-timescale electricity theft detection and localization in distribution systems based on state estimation and pmu measurements," in *Proceedings of the Ninth International Conference on Future Energy Systems*, ser. e-Energy '18. New York, NY, USA: ACM, 2018, pp. 282–290.
- [24] D. Drzajic, "Energy theft detection using compressive sensing methods," Semester thesis, ETH Zürich, 2015.
- [25] F. Xiao and Q. Ai, "Electricity theft detection in smart grid using random matrix theory," *IET Generation, Transmission Distribution*, vol. 12, no. 2, pp. 371–378, 2018.
- [26] S. Park, D. Deka, and M. Chertkov, "Exact topology and parameter estimation in distribution grids with minimal observability," *CoRR*, vol. abs/1710.10727, 2017. [Online]. Available: <http://arxiv.org/abs/1710.10727>
- [27] B. Das, "Estimation of parameters of a three-phase distribution feeder," *IEEE Transactions on Power Delivery*, vol. 26, no. 4, pp. 2267–2276, Oct 2011.
- [28] Z. Zumar, "Last Mile Asset Monitoring: Low Cost Rapid Deployment Asset Monitoring," Master's thesis, Portland State University, 2014.
- [29] A. Albert and R. Rajagopal, "Smart meter driven segmentation: What your consumption says about you," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4019–4030, Nov 2013.
- [30] S. Bolognani and F. Dörfler, "Fast power system analysis via implicit linearization of the power flow manifold," in *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on*. IEEE, 2015, pp. 402–409.
- [31] D. Deka, S. Backhaus, and M. Chertkov, "Structure learning and statistical estimation in distribution networks-part I," *arXiv preprint arXiv:1501.04131*, 2015.
- [32] P. J. Rousseeuw and A. M. Leroy, *Robust regression and outlier detection*. John wiley & sons, 2005, vol. 589.
- [33] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 871–880, 1984.
- [34] J. Fox and S. Weisberg, *An R Companion to Applied Regression*. SAGE Publications, 2011.
- [35] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.
- [36] C. Rao, A. Fieger, C. Heumann, H. Toutenburg, T. Nittner, and S. Scheid, *Linear Models: Least Squares and Alternatives*, ser. Springer Series in Statistics. Springer New York, 2006.
- [37] I. Markovsky and S. V. Huffel, "Overview of total least-squares methods," *Signal Processing*, vol. 87, no. 10, pp. 2283–2302, 2007, special Section: Total Least Squares and Errors-in-Variables Modeling.
- [38] S. Rhode, K. Usevich, I. Markovsky, and F. Gauterin, "A recursive restricted total least-squares algorithm," *IEEE Transactions on Signal Processing*, vol. 62, no. 21, pp. 5652–5662, Nov 2014.
- [39] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 7, pp. 1319–1330, July 2013.
- [40] M. Zanetti, E. Jamhour, M. Pellenz, M. Penna, V. Zambenedetti, and I. Chueiri, "A tunable fraud detection system for advanced metering infrastructure using short-lived patterns," *IEEE Transactions on Smart Grid*, vol. PP, no. 99, pp. 1–1, 2017.
- [41] D. Mashima and A. A. Cárdenas, "Evaluating electricity theft detectors in smart grid networks," in *International Workshop on Recent Advances in Intrusion Detection*. Springer, 2012, pp. 210–229.
- [42] C. M. Bishop, *Neural networks for pattern recognition*. Oxford university press, 1995.
- [43] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004.
- [44] W. Wang and N. Yu, "Advanced metering infrastructure data driven phase identification in smart grid," 2017.
- [45] J. Tian, M. H. Azarian, and M. Pecht, "Anomaly detection using self-organizing maps-based k-nearest neighbor algorithm," in *Proceedings of the European Conference of the Prognostics and Health Management Society*. Citeseer, 2014.
- [46] W. H. Kersting, *Distribution system modeling and analysis, Third edition*. CRC press, 2012.
- [47] J. K. Baksalary and O. M. Baksalary, "Particular formulae for the moore–penrose inverse of a columnwise partitioned matrix," *Linear algebra and its applications*, vol. 421, no. 1, pp. 16–23, 2007.
- [48] S. Pischke, "Lecture notes on measurement error," *Lecture Notes on Measurement Error*, 2007.